Design and Implementation of a
Research Data Management System:
The CRC/TR32 Project Database (TR32DB)

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Constanze Curdt

aus Göttingen

Köln 2014

Abstract

Research data management (RDM) includes all processes and measures which ensure that research data are well-organised, documented, preserved, stored, backed up, accessible, available, and re-usable. Corresponding RDM systems or repositories form the technical framework to support the collection, accurate documentation, storage, back-up, sharing, and provision of research data, which are created in a specific environment, like a research group or institution. The required measures for the implementation of a RDM system vary according to the discipline or purpose of data (re-)use. In the context of RDM, the documentation of research data is an essential duty. This has to be conducted by accurate, standardized, and interoperable metadata to ensure the interpretability, understandability, shareability, and long-lasting usability of the data. RDM is achieving an increasing importance, as digital information increases. New technologies enable to create more digital data, also automatically. Consequently, the volume of digital data, including *big data* and *small data*, will approximately double every two years in size. With regard to e-science, this increase of data was entitled and predicted as the *data deluge*. Furthermore, the paradigm change in science has led to *data intensive science*. Particularly scientific data that were financed by public funding are significantly demanded to be archived, documented, provided or even open accessible by different policy makers, funding agencies, journals and other institutions. RDM can prevent the loss of data, otherwise around 80-90 % of the generated research data disappear and are not available for re-use or further studies. This will lead to empty archives or RDM systems. The reasons for this course are well known and are of a technical, socio-cultural, and ethical nature, like missing user participation and data sharing knowledge, as well as lack of time or resources. In addition, the fear of exploitation and missing or limited reward for publishing and sharing data has an important role.

This thesis presents an approach in handling research data of the collaborative, multidisciplinary, long-term DFG-funded research project Collaborative Research Centre/Transregio 32 (CRC/TR32) "Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation". In this context, a RDM system, the so-called CRC/TR32 project database (TR32DB), was designed and implemented. The TR32DB considers the demands of the project participants (e.g. heterogeneous data from different disciplines with various file sizes) and the requirements of the DFG, as well as general challenges in RDM. For this purpose, a RDM system was established that comprises a well-described self-designed metadata schema, a file-based data storage, a well-elaborated database of metadata, and a corresponding user-friendly web interface. The whole system is developed in close cooperation with the local Regional Computing Centre of the University of Cologne (RRZK), where it is also hosted.

The documentation of the research data with accurate metadata is of key importance. For this purpose, an own specific TR32DB Metadata Schema was designed, consisting of multi-level metadata properties. This is distinguished in general and data type specific (e.g. data, publication, report) properties and is developed according to the project background, demands of the various data types, as well as recent associated metadata standards and principles. Consequently, it is interoperable to recent metadata standards, such as the Dublin Core, the DataCite Metadata Schema, as well as core elements of the ISO19115:2003 Metadata Standard and INSPIRE Directive. Furthermore, the schema supports optional, mandatory, and automatically generated metadata properties, as well as it provides predefined, obligatory and self-established controlled vocabulary lists. The integrated mapping to the DataCite Metadata Schema facilitates the simple application of a Digital Object Identifier (DOI) for a dataset.

The file-based data storage is organized in a folder system, corresponding to the structure of the CRC/TR32 and additionally distinguishes between several data types (e.g. data, publication, report). It is embedded in the Andrew File System hosted by the RRZK. The file system is capable to store and backup all data, is highly

scalable, supports location independence, and enables easy administration by Access Control Lists. In addition, the relational database management system MySQL stores the metadata according to the previous mentioned TR32DB Metadata Schema as well as further necessary administrative data.

A user-friendly web-based graphical user interface enables the access to the TR32DB system. The web-interface provides metadata input, search, and download of data, as well as the visualization of important geodata is handled by an internal WebGIS. This web-interface, as well as the entire RDM system, is self-developed and adjusted to the specific demands.

Overall, the TR32DB system is developed according to the needs and requirements of the CRC/TR32 scientists, fits the demands of the DFG, and considers general problems and challenges of RDM as well. With regard to changing demands of the CRC/TR32 and technologic advances, the system is and will be consequently further developed. The established TR32DB approach was already successfully applied to another interdisciplinary research project. Thus, this approach is transferable and generally capable to archive all data, generated by the CRC/TR32, with accurately, interoperable metadata to ensure the re-use of the data, beyond the end of the project.

Zusammenfassung

Forschungsdatenmanagement beinhaltet alle notwendigen Abläufe und Maßnahmen, die sicherstellen, daß Forschungsdaten gut organisiert, dokumentiert, erhalten, gesichert, erreichbar, verfügbar und wiedernutzbar sind. Entsprechende Systeme und Repositorien für das Forschungsdatenmanagement bilden diesbezüglich den technischen Rahmen, um Daten von ForscherInnen und Institutionen entsprechend zu sammeln, zu dokumentieren, zu speichern, zu sichern und zu verteilen. Die notwendigen Implementierungsmaßnahmen variieren zwischen den einzelnen Disziplinen und dem jeweiligen Zweck. Eine wesentliche Aufgabe im Forschungsdatenmanagement ist die Dokumentation von Forschungsdaten. Dies wird über genaue, standardisierte und interoperable Metadaten erreicht, die eine Interpretation, Verteilung, ein Verständnis und eine lange Nachnutzung der Daten ermöglichen. Die Wichtigkeit von Forschungsdatenmanagement steigt mit der zunehmenden Masse von digitalen Informationen. Neue Technologien erfassen immer mehr Daten, zunehmend auch automatisch. Entsprechend verdoppelt sich das Volumen von *small data* und *big data* ungefähr alle zwei Jahre. Im Hinblick auf *e-science* wird dieses Wachstum als *data deluge* („Datensintflut") bezeichnet und ist vorhergesagt worden. Darüber hinaus findet ein Paradigmenwechsel zur datenintensiven Wissenschaft statt. Vor allem Forschungsdaten, die durch öffentliche Mittel finanziert worden sind, sollen ausgehend von Entscheidungsträgern, Geldgebern, Zeitschriften und anderen Institutionen archiviert, dokumentiert und zugänglich gemacht werden. Forschungsdatenmanagement verhindert außerdem den Verlust von Forschungsdaten, die andernfalls zu 80-90% verloren gingen oder nicht mehr für eine Nachnutzung zur Verfügung stünden. Dieses Problem kann zu leeren Datenarchiven und Systemen führen. Die Gründe dafür sind das Resultat bekannter technischer, sozio-kultureller und ethischer Probleme, wie beispielsweise fehlende Mitwirkung und fehlendes Wissen über Datenweiternutzung sowie mangelnde Zeit und Ressourcen. Zusätzlich besteht Angst vor Ausnutzung und mangelnder Honorierung für das Publizieren und Teilen von Forschungsdaten.

Diese Arbeit stellt einen Ansatz zum Umgang mit Forschungsdaten im Rahmen eines kollaborativen, multi-disziplinären, langfristigen Forschungsprojektes vor, das im Rahmen des von der DFG geförderten Sonderforschungsbereiches/Transregio 32 (SFB/TR32) „Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation" anzusiedeln ist. Ein Forschungsdatenmanagementsystem, die sogenannte SFB/TR32 Projektdatenbank (TR32DB), wurde geplant und implementiert. Die TR32DB berücksichtigt die Anforderungen der ProjektteilnehmerInnen (z.B. heterogene Datenbestände mit unterschiedlichen Dateigrößen) und der DFG, sowie alle generellen Herausforderungen des Forschungsdatenmanagements. Für diesen Zweck wurde ein System entwickelt, das aus einem voll dokumentierten, selbst entwickelten Metadaten Schema, einem dateibasierten Datenspeicher, einer gut ausgearbeiteten Datenbank und einer nutzerfreundlichen Weboberfläche besteht. Das gesamte System ist in enger Zusammenarbeit mit dem Regionalen Rechenzentrum der Universität zu Köln (RRZK) entwickelt worden und wird hier auch bereitgestellt.

Die Dokumentation von Forschungsdaten mit genauen Metadaten ist von zentraler Bedeutung. Für diesen Zweck wurde ein spezifisches TR32DB Metadaten Schema entwickelt, das aus mehrstufigen Bestandteilen von Metadaten besteht. Es wird zwischen generellen und datentypspezifischen Eigenschaften (z.B. Daten, Publikationen und Berichte) unterschieden. Das Schema ist entsprechend dem Projekthintergrund, den Anforderungen der verschiedenen Datentypen sowie den aktuellsten Metadatenstandards und -prinzipien entwickelt worden. Entsprechend ist dieses Schema interoperabel zu dem Dublin Core Metadaten Standard, dem DataCite Metadaten Schema und den Kernelementen des ISO Metadatenstandards 19115:2003 sowie der INSPIRE-Richtlinie. Weiterhin unterstützt das Schema optionale, verpflichtende und automatische Metadatenelemente und stellt vordefinierte, obligatorische und selbsterstellte kontrollierte Begriffslisten

bereit. Die integrierte Zuordnung zum DataCite Metadaten Schema unterstützt die einfache Beantragung eines *Digital Object Identifier (DOI)* für einen Datensatz.

Der dateibasierte Datenspeicher ist in einer Ordnerstruktur entsprechend der Struktur des SFB/TR32 angelegt, und es wird zwischen den einzelnen Datentypen (z.B. Daten, Publikationen, Präsentationen und Berichte) unterschieden. Diese Struktur ist in das vom RRZK bereitgestellte Andrew Dateisystem eingebettet. Das Dateisystem ermöglicht es, alle Daten zu speichern und zu sichern, ist gut skalierbar, unterstützt Ortsunabhängigkeit und erlaubt eine Administration über Zugriffslisten. Die Metadaten des TR32DB Metadaten Schemas sowie weitere administrative Daten werden in einem relationalen Datenbankverwaltungssystem MySQL gespeichert.

Eine nutzerfreundliche, web-basierte Benutzeroberfläche ermöglicht den Zugang zum TR32DB System. Die Weboberfläche erlaubt die Eingabe von Metadaten, eine Suche und das Herunterladen von gespeicherten Forschungsdaten des Projektes sowie die Visualisierung wesentlicher Geodaten durch ein internes WebGIS. Die Weboberfläche ist, wie das gesamte System, selbst entwickelt und an die spezifischen Anforderungen angepasst.

Insgesamt ist das SFB/TR32 Forschungsdatenmanagementsystem entsprechend den Ansprüchen und Bedürfnissen der WissenschaftlerInnen des SFB/TR32 entwickelt worden und entspricht den Anforderungen der DFG sowie den generellen Erfordernissen und Herausforderungen derartiger Systeme. Unter Beachtung der sich ändernden Ansprüche des SFB/TR32 und technologischen Verbesserungen wird das System konsequent weiterentwickelt. Der etablierte Ansatz der TR32DB ist erfolgreich auf ein anderes interdisziplinäres Forschungsprojekt angewendet worden. Somit ist dieser Ansatz übertragbar und generell in der Lage, alle Daten, die vom SFB/TR32 erzeugt werden, mit passenden, interoperablen Metadaten zu archivieren, um eine Wiederverwendung dieser Daten über das Ende des Projektes hinaus zu ermöglichen.

Acknowledgements

Table of Contents

Figures

Tables

Abbreviations

A	Automatic
ACL	Access Control List
AFS	Andrew-File-System
ANSI	American National Standards Institute
API	Application Programming Interface
ASP	Active Server Pages
BIBO	Bibliographic Ontology
C	Conditional
CMS	Content Management System
CRC	Collaborative Research Centre
CRC/TR32	Collaborative Research Centre/Transregio 32
DB	Database
DBMS	Database Management System
DC	Dublin Core
DCC	Digital Curation Centre
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Initiative
DDC	Dewey Decimal Classification
DDI	Data Documentation Initiative
DFG	German Research Foundation
DMP	Data Management Plan
DOI	Digital Object Identifier
DOP	Digital Orthophoto
DWD	German Weather Forecast
ER	Entity Relation
FGDB	File Geodatabase
GEMET	General Multilingual Environmental Thesaurus
GESIS	German Leibniz Institute for the Social Sciences
GFZ	German Research Centre for Geosciences
GIS	Geographic Information System
HPC	High Performance Computing
HTML	Hypertext Markup Language
ID	Identifier
IDF	International DOI Foundation
IETF	Internet Engineering Task Force
INF	Information - Project
INSPIRE	Infrastructure for Spatial Information in the European Community
IP	Internet Protocol
IR	Implementation Rule
ISO	International Organization for Standardization

JISC	Joint Information Systems Committee
JS	JavaScript
M	Mandatory
MD	Metadata
MVC	Model View Controller
NAS	Network Attached Storage
NASA	National Aeronautics and Space Administration
NISO	National Information Standards Organization
O	Optional
OB	Obligation
OCC	Occurrence
OGC	Open Geospatial Consortium
Opt	Optional
OWL	Web Ontology Language
PDF	Portable Document Format
PHP	Hypertext Preprocessor
RDF	Resource Description Framework
RDM	Research Data Management
Req	Required
REST	Representational State Transfer
RFC	Request for Comments
RRZK	Regional Computing Centre of the University of Cologne
RU	Research Unit
SAN	Storage Area Network
SDI	Spatial Data Infrastructure
SFTP	Secure File Transfer Protocol (SSH File Transfer Protocol)
SOA	Service-oriented Architecture
SOAP	Simple Object Access Protocol
SOFS	Scale out File Services
SP	Sub-Project
SQL	Structured Query Language
SVA	Soil-Vegetation-Atmosphere
TCP	Transmission Control Protocol
TIB	German National Library of Science and Technology
TR32DB	CRC/Transregio 32 Database
UML	Unified Modelling Language
URI	Uniform resource identifier
URL	Uniform resource locator
URN	Uniform resource name
USGS	United States Geological Survey
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1 Introduction and background

1.1 Introduction

Over the past decade, the creation of digital information and data has increased in many fields (HILBERT & LOPEZ, 2011), as the digital revolution or digital age and new technologies enable us to create more data, automatically and digitally. In April 2014, the seventh *Digital Universe* study (EMC, 2014) was published by the EMC Corporation and analysed by the International Data Corporation (IDC). This study gives a forecast on the growth of data volume in the digital universe from 2013 to 2020. In this time period, the digital data is predicted to grow by a factor of 10 (from 4.4 zettabytes to 44 zettabytes). Consequently, the volume of the digital data will approximately double in size every two years.

This digital information is created in every possible field of the private and public sector as well as in cultural life and science. For instance, in the private and public sector, Amazon (WIRED, 2014), eBay (ITNEWS, 2013) or the LIBRARY OF CONGRESS (2014) have to handle a huge amount of data every day with an increasing trend. In addition, the growth of users in the context of common social-media applications, like Facebook (DE WITT et al., 2012), Twitter or YouTube (WIKIBON BLOG, 2012), has led to a strong growth in digital information as well. In this context, the development of hardware, such as automatic or wireless sensors in agriculture (DETER, 2013; PAJARES, 2011), and software improvements also enable the faster generation of data. For instance, in physical sciences, huge amounts of data are automatically created each day by the Pan-STARRS telescope (Panoramic Survey Telescope and Rapid Response System) located on Maui (Hawaii, USA) the Large Hadron Collider (LHC) of the European Organization for Nuclear Research (CERN) or the Sloan Digital Sky Survey (SDSS) telescope at New Mexico's (USA) Apache Point Observatory (DE WITT et al., 2012). Likewise, global climate data, including model results, satellite imagery, and in situ measurements, dramatically increase in complexity and volume (OVERPECK et al., 2011). In the context of the fast growing amount of data volumes, the usage of the term *big data* has become very common in recent years and has been discussed in various fields (FREYTAG, 2014; LYNCH, 2008) and special issues (ERCIM NEWS, 2012; NATURE, 2008; POPULAR SCIENCE, 2011).

With regard to science, the increase of data, for instance, due to developments of new technologies, was predicted a few years ago and entitled the *data deluge* (HEY & TREFETHEN, 2003). Moreover, the change of science to a collaborative, linked digital research environment, known as *e-research* or *e-science* (HEY & TREFETHEN, 2002), has led to the increase of data. Additionally, science is performing the paradigm change to the *data intensive science* (MCNALLY et al., 2012), also known as the fourth paradigm (HEY et al., 2009). In this context, both *big data* and *small data* have to be considered (KLUMP & BERTELMANN, 2013; ONSRUD & CAMPBELL, 2007). Recently, particularly data that were generated in the area of e-science and financed by public funding are significantly demanded, for instance, by different policy makers (DFG, 2014b; EUROPEAN COMMISSION, 2007a; HOCHSCHULREKTORENKONFERENZ, 2014; NATIONAL SCIENCE FOUNDATION, 2011b), journals (AGU, 2013; PLOS, 2014) or other institutions in the society (e.g. public transport, energy industry, disaster protection). For example, Neelie Kroes, Vice-President of the European Commission, has outlined that "Data is the new gold […] just as oil was likened to black gold, data takes on a new importance and value in the digital age." (Kroes, 2011: 2). Additionally, the journal POPULAR SCIENCE published a special issue in 2011 entitled "Data is power" (POPULAR SCIENCE, 2011), as well as Nature's special issue headed "Data - What data?" (NATURE, 2009) or the special issue of Science titled "Dealing with Data" (SCIENCE, 2011). All highlight the importance of data. Thus, the access to open research data is encouraged by many initiatives and organisations, such as the THE ROYAL SOCIETY (2012), the EUROPEAN COMMISSION (2012b) or the White House (HOLDREN, 2013).

Scientific data is demanded by society and policy makers for several reasons. These include the re-use of the data for analysis or simulation of recent events (e.g. simulation of climate change or traffic flows, manage transportation, coping with diseases). Data is requested with the purpose to repeat and reproduce research results, to prevent manipulation of data as well as to protect the scientists from false accusation. Consequently, safeguarding, maximal availability (HANSON et al., 2011) as well as sharing (BORGMAN, 2012) of data are very important. Research data management (RDM) and the establishment of corresponding RDM infrastructures support the management and sharing of data. As stated by Dr. Stefan Winkler-Nees (DFG) in 2011 in the ODE report "Without the infrastructure that helps scientists manage their data in a convenient and efficient way, no culture of data sharing will evolve." (ODE PROJECT, 2011:20). In addition, RDM can prevent the loss of scientific data. Recent studies in ecology have shown that the availability of the raw data declines dramatically as the research articles are ageing. In measurable terms, after 20 years about 80 % of the underlying data is no longer accessible (GIBNEY & VAN NOORDEN, 2013). This also corresponds with the statement of Dr. Stefan Winkler-Nees (DFG) pointing out that from all collected research data (e.g. including data created in preliminary studies, minutes of field campaigns) around 80-90 % disappear and will consequently not be available for re-use and further studies (KRAMER, 2014). Furthermore, the funding organizations' are aware of the importance of RDM due to the commitment to provide so-called *Data Management Plans (DMP)* with the application of grants (DFG, 2014b; NATIONAL SCIENCE FOUNDATION, 2011b; NERC, 2014a; WELLCOME TRUST, 2014).

RDM is the key to progress of science (AILAMAKI et al., 2010). It is particular important in interdisciplinary, long-term research projects, as highlighted by MÜCKSCHEL et al. (2006), as well as strongly demanded by KLAR & ENKE (2013) and the DFG (2013d). In the context of RDM, the documentation of the research data with accurate, standardized, interoperable metadata is a significant issue. For the description of the data, appropriate metadata have to be chosen to enable the reusability of the data as well as to ensure the understanding of the data (VAN NOORDEN, 2013). However, it is stated that "Good metadata is expensive" (LEVINE, 2014: 136). The creation of the metadata requires expertise and resources. Inconsistent and incomplete metadata implicate that the described data is useless and with minor quality (LEVINE, 2014). In general, RDM solves the problems of organizing, exchanging, searching, data mining, documentation, accessing, storage and re-use of research data (CORTI et al., 2014). With regard to technical solutions to set up RDM infrastructures, several software components are available (e.g. CKAN, 2014b; DSPACE, 2014a; EPRINTS, 2014a). As stated by OSSWALD et al. (2012), these available standard solutions are not suitable as a standard solution for all disciplines, institutions, and research projects.

RDM also has problems. While considering the prediction of a data deluge (ECONOMIST, 2010), RDM also faces problems of empty data archives (NELSON, 2009). These empty archives are a consequence of missing user participation and data sharing knowledge. In addition, further factors that constrain researchers from publishing and sharing their data can be summarized in challenges of technical, socio-cultural, and ethical nature as identified by HARTTER et al. (2013). These include the lack of time and resources, the lack of experience or expertise in RDM, loss of data, lack of standards as well as ignorance about available data archives to store the data and their usage. Furthermore, the fear of exploitation and competitive factors as well as missing or limited reward for publishing and sharing of data have an effect (GRIFFITHS, 2009; HARTTER et al., 2013; NELSON, 2009; TENOPIR et al., 2011). In addition, RDM barriers are outlined by SALLANS & LAKE (2014) from the researcher's perspective, which include key factors like lack of training or ignorance of best practices in RDM as well as insufficient support and guidance. Thus, restrictions, for example in terms of data policies, might be necessary as well as user-friendly systems and features, such as the distribution of Digital Object Identifiers (DOI) for an accurate citation of datasets. Furthermore, a need to train and guide researchers in RDM principles, techniques and the actual usage of specific RDM systems is essential (JONES, 2013).

The recent relevance and the importance of RDM becomes apparent, for example, by the number of conferences related to this topic as well as recently published books (e.g. PRYOR et al., 2013; RAY, 2014;

CORTI et al., 2014), special issues of journals (e.g. THE ECONOMIST, 2010; LIBRARY CONNECT, 2013; NATURE, 2009; SCIENCE, 2011) or articles in magazines (e.g. KRAMER, 2014).

The current discussion about RDM and the announcement of new funding projects related to the set-up of RDM infrastructures point out that it is a contemporary issue, which is not yet solved. For example, the Joint Information Systems Committee (JISC) has launched the digital infrastructure research data management programme (JISC, 2014a) in the UK which covers the areas of research data management, research information management, and support of infrastructure for repositories and curation. In addition, the German Research Foundation (DFG) has announced funding programs with regard to Information Infrastructures for Research Data (DFG, 2013a) and Virtual Research Environments (DFG, 2013b). In this context, research projects like the RADIESCHEN ("Rahmenbedingungen einer disziplin-übergreifenden Forschungsdaten-Infrastruktur"; funding period: 05/2011-04/2013) project (RADIESCHEN, 2013) or the RADAR ("Research Data Repositorium"; funding period: 09/2013-08/2015) project (RADAR, 2014) are funded by the DFG to identify framework requirements as wells as to support the establishment of RDM in Germany.

Consequently, the present thesis will present an approach to handle research data in a collaborative, multidisciplinary research environment by designing and implementing a research data management infrastructure. This approach also aims at solving the outlined challenges and problems of RDM. In addition, the RDM is designed according to the requirements of the research funding organisation, demands of the research project, as well as according to recent standards and principles.

1.2   Aim of the thesis

Generally, this doctoral thesis is related to geoinformatics, also known as spatial information sciences or spatial informatics, a scientific discipline that focuses on the development and implementation of methods and concepts of informatics on spatial problems and questions (BILL, 2010). Thus, geoinformatics connect informatics and geosciences (DE LANGE, 2013), and is acknowledged as a part of applied geography (DICKMANN & ZEHNER, 2001; EHLERS & SCHIEWE, 2010). According to DE LANGE (2013), one part of geoinformatics investigates the creation and handling of databases and metadatabases in a spatial context. Likewise, standardization, interoperability, data organization and management, as well as database systems are issues of this scientific area (DE LANGE, 2013), which are equally all used in this contribution.

The thesis is embedded within the context of the interdisciplinary, long-term research project Collaborative Research Centre/Transregio 32 (CRC/TR32) "Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation" funded by the German Research Foundation (DFG) since 2007. The CRC/TR32 involves research groups from the Universities of Aachen, Bonn, Cologne, and the Research Centre Jülich. The scientists involved work in the field of meteorology, geophysics, soil and plant sciences, hydrology, mathematics, geography as well as geoinformatics. As a result of their research, the scientists create a lot of heterogeneous research data in various spatial and temporal scales as well as in various file sizes (few kilobytes to gigabyte per file). These data are generated by conducting field measurement, laboratory analysis, and are an output of spatial modelling approaches (e.g. KORRES et al., 2013) or spatial analysis (e.g. WALDHOFF et al., 2012; HOFFMEISTER et al., 2013) as well. Almost all of these data are captured and modelled in the spatial domain and consequently have a spatial reference. Finally, the scientists publish their research results in journal papers or give a presentation at conferences.

Within the framework of this study, embedded in the CRC/TR32 project section *Information Infrastructure (INF-*project*)*, the aim was to design and implement a comprehensive, user-friendly project database for the CRC/TR32 (TR32DB). This research data management (RDM) system is intended to ensure the exchange as well as a secure and sustain long-term storage and back-up of all data, created by the scientists of the CRC/TR32. In order to achieve this objective, the description and documentation of all data with accurate

metadata according to recent standards and principles is essential. For this purpose, the main aim was to design and implement a comprehensive metadata schema which covers all requirements in terms of different data types and interoperability to recent metadata standards and principles. Furthermore, to support the accurate citation of all data, stored in the TR32DB, the application of persistent identifiers, such as a Digital Object Identifier (DOI), was requested. Moreover, the visualisation of purchased and generated data of the scientists with a spatial reference was desired.

For the design and implementation of the TR32DB, several specific demands and challenges have to be considered. These include general challenges and problems of RDM, as already indicated in section 1.1. Furthermore, requirements resulting from the interdisciplinary research focus of the CRC/TR32, for example, in term of data heterogeneity (e.g. modelled or measured data, publications, presentations) or file size need to be regarded. In addition, requirements of the DFG have to be fulfilled, such as the cooperation with a computing centre or library with regard to the development of the RDM system or the request to accurately document the data with metadata (DFG, 2013d). In addition, the sustainable, long-term storage, and backup of the data beyond the end of the project is required. For example, the *Proposals of Safeguarding Good Scientific Practice* demand the long-term storage of data for ten years after their creation (DFG, 2013e). Moreover, technological developments, expandability, and specifications have to be considered.

In summary, the main aims with regard to the design and implementation of the TR32DB are:

- Establishment of a RDM system in the environment of a computing centre or library which supports sustainable long-term storage, backup, and exchange of all heterogeneous project data
- Provision of a user-friendly graphical user web-interface to access the data
- Development of a metadata schema for an accurate, interoperable documentation of all data according to project demands and current metadata schemas and standards
- Provision of accurate citation of the data with a persistent identifier
- Visualisation of purchased and generated data of CRC/TR32 scientists with a spatial reference
- Compliance with demands of the CRC/TR32 scientists and DFG
- Consideration of general problems and challenges in RDM

Consequently, this thesis outlines the realisation of the above-mentioned aims. Therefore, a general introduction to RDM is provided, followed by the used methods to design and implement the TR32DB. Finally, all components and features of the TR32DB will be presented and discussed.

1.3   Outline of the thesis

In order to achieve the stated aims of this thesis, outlined in the previous section, chapter 2 gives an initial introduction to research data management (RDM) principles. This includes a general, introducing overview about RDM (2.1), an introduction to metadata (2.2), and the description of technical components used to set up a RDM system (2.3). Finally, an overview is presented about available software solutions for RDM (2.4) as well as sample projects and institutions (2.5) which successfully operate RDM.

Subsequently, chapter 3 describes the specific data, methods and components used within the design of the TR32DB architecture. After a description of the project background of the CRC/TR32 (3.1), the overall architecture of the TR32DB will be characterized (3.2). Section 3.3 depicts the first component of the TR32DB, the design and documentation process of the TR32DB Metadata Schema. The setting and structure of the TR32DB data storage in the Regional Computing Centre of the University of Cologne (RRZK), as a second component of the TR32DB, is outlined in section 3.4. The associated database structure, for example containing the descriptive metadata of the data, is the third component of the TR32DB and described in section 3.5. Finally, section 3.6 provides an overview of the methods used to establish the TR32DB web-interface, which is the fourth component of the TR32DB.

The overall functionalities of the TR32DB are presented in chapter 0. At first, the requirements of the TR32DB are depicted (4.1). Consecutively, the entire structure of the TR32DB is described (4.2), followed by the single components. The TR32DB metadata schema is presented in section 4.3, followed by the storage and organisation of the data and corresponding metadata (4.4). Finally, the functionalities and features of the web-interface are demonstrated (4.5).

In chapter 5, the outlined methods and results of the thesis will be discussed. At first, an evaluation of the TR32DB systems is presented (5.1). Subsequently, the TR32DB is compared to other RDM systems (5.2). Finally, the transferability of the TR32DB approach to another research project is depicted (5.3).

Chapter 6 provides a brief conclusion and future perspectives.

Supplementary material and information is provided in the appendix of this thesis. Appendix A comprises the results of a survey. Furthermore, complementary information about the web-interface (appendix B), the database structure (appendix C), and the documentation of the TR32DB Metadata Schema (appendix D) are available in the appendix.

Note: All specific features, specific properties, commands, file paths mainly related to the TR32DB are displayed in another font type (e.g. `titleType`). In addition, source code is presented in grey shaded areas.

2 Introducing research data management principles

This chapter gives an overview of the basic principles and components in research data management (RDM). As an introduction, an overview will be given of the research life cycle in a research project which also involves the management of research data that were created during the project duration. RDM supports the organization of research data, including the handling of the data through the entire research cycle from the entry stage, through to the dissemination stage, and finally to the archiving stage (WHYTE & TEDDS, 2011).

A research life cycle model defines and illustrates the overall complex process of research, usually from the beginning to the end of a certain research project. This model identifies explicit stages, components as well as activities and responsibilities, usually associated with research data management and curation (CARLSON, 2014). Different kinds of research life cycle models have been developed by research projects, organizations, and communities. In the higher education institutional domain, models are developed for example by the ERASMUS UNIVERSITEIT ROTTERDAM (2014), the UNIVERSITY OF CENTRAL FLORIDA LIBRARIES (2014), and the UNIVERSITY OF LONGHBOROUGHT LIBRARY (2014). Some research life cycle models are extended with an integrated data life cycle model, such as the models developed by the UNIVERSITY OF VIRGINIA LIBRARY (2014), the Joint Information Systems Committee (JISC, 2014b), or the model presented by RICE (2007). For example, Figure 2-1 presents the research life cycle model for a certain research project, developed by the OREGON STATE UNIVERSITY LIBRARIES (2014). This life cycle is also extended with a project data life cycle and based upon the models of the UNIVERSITY OF VIRGINIA LIBRARY (2014) and the INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (2012). The research life cycle presents several stages, represented as boxes and data management activities. With regard to the transition from one stage to another, the responsibility can change. This might cause data loss, lack of documentation or failure in stage moving (OREGON STATE UNIVERSITY LIBRARIES, 2014).
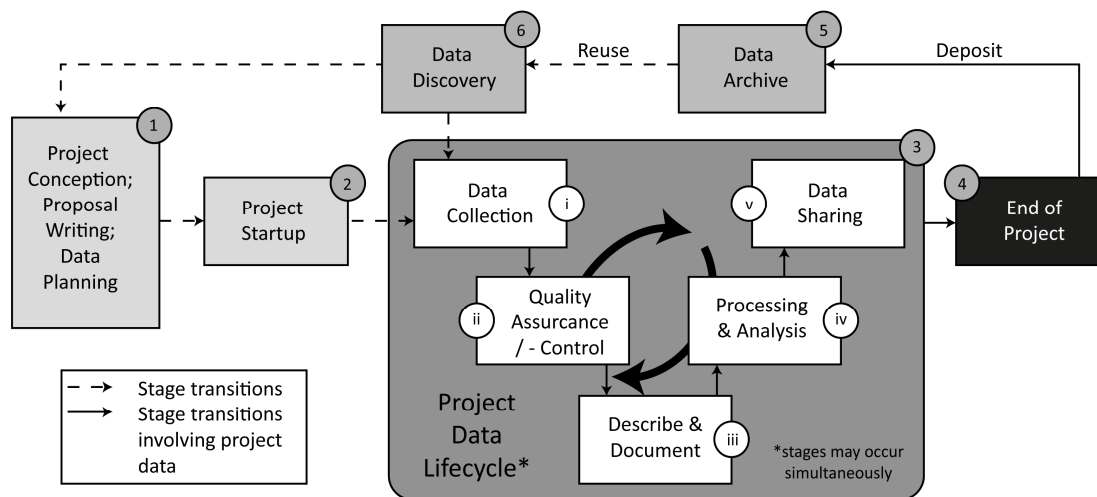
Figure 2-1: The research life cycle model for a particular research project developed by and modified after the OREGON STATE UNIVERSITY LIBRARIES (2014)

The overall research life cycle (Figure 2-1) is distinguished in three parts: (i) project planning (1-2), (ii) project data lifecycle (3), and (iii) finally wrap-up of the project (4-6). All three parts are related to data.

> **Project planning** This part involves the application of the project (e.g. including review of current data, recent studies, consideration of data costs, sharing, management, storage and archiving). In general, a Data Management Plan (DMP) should be established, often required by the project funder. Finally, the project will start and move to the project data lifecycle part.

Project data lifecycle This part describes different stages of project data. It includes the data collection, re-use of available data, file organisation, quality assurance, and quality control. In addition, it is important to describe and document the collected data with regard to the context (e.g. documentation of objectives, collection methods and instruments, temporal and geographic coverage) and data-level (e.g. description of records, values, variables, missing values, and code). Processing and analysis of the data has also to be documented (e.g. file versions). Finally, the data can be shared with the project members in determined data file formats and through publications or presentations the data can be provided in a descriptive form.

Project wrap-up Finally, the data can be deposited to a data repository or archive and final reports can be provided to the funder. Data discovery and re-use is possible. (OREGON STATE UNIVERSITY LIBRARIES, 2014)

According to HIGGINS (2011), the output of the academic research activities, respectively of the research life cycle, are the data created or collected during the study and the published papers resulting from the analysis of the data. Both need to be managed, most likely by an RDM system. Furthermore, HIGGINS (2011) states that until recently, data was only treated as a by-product used in the publication process, which is nowadays changing. Accordingly, AILAMAKI et al. (2010) complement that the key of scientific progress is the management of the huge amount of research data, created and collected during the research lifecycle. Finally, GILCHRIST (1986) asserts that an accurate, efficient management of research data in a research project is the "nuts and bolts" (GILCHRIST 1986: 251) in the research process respective in a research life cycle.

Thus, the following sections will describe several parts and complementary components of the previously described research life cycle and the role of RDM. Hence, the first section 2.1 will provide a general introduction to RDM. Section 2.2 gives an introduction to metadata which are needed and used to accurately document research data. Common metadata standards and schemas are described. Basic technical components used to implement a RDM system, a data repository or an archive to store research data are presented in section 2.3. Finally, available RDM software solutions (section 2.4) and examples of RDM conducted in research projects, as well as institutional approaches (section 2.5) will be outlined.

2.1   Management of research data

This chapter focuses on the state of the art of research data management (RDM). Initially, an introduction will be given to RDM (2.1.1). Section 2.1.2 provides an overview of RDM activities and the perspective of policy makers (e.g. funding agencies) with a global and national view. Following, the diversity of research data is described (2.1.3) and the research data lifecycle (2.1.4). Section 2.1.5 gives a brief introduction to RDM infrastructures and systems. Finally, the publishing and citation (2.1.6) as well as sharing and access (2.1.7) to research data are outlined.

2.1.1   Research data management – an introduction

Research data management (RDM), as described by PRYOR (2011a), is "[…] an active process by which digital resources remain discoverable, accessible, and intelligible over the longer term, a process that invests data and datasets with the potential to accrue value as assets enjoying far wider use than their creators may have anticipated. In the world of research, such a value-adding process is a significant contributor to the much desired achievement of impact." PRYOR (2011a: vii). CORTI et al. (2014) observe the term data management more from the practical perspective, for example, from the perspective of a researcher. They define that data management includes "[…] all data practices, manipulations, enhancements, and processes that ensure that research data are of a high quality, are well organized, documented, preserved, sustainable, accessible, and reusable." (CORTI et al., 2014: 2). In addition, LUDWIG & ENKE (2013) complement that RDM refers to all relevant measures, which are necessary to ensure that digital research data are available and

reusable. The required measures can vary strongly, depending on the purpose of use or re-use of the research data (e.g. re-use as working copy, re-use for future research).

RDM can be conducted for several reasons and purposes. KLUMP & LUDWIG (2013) name and characterize three types. (i) Research data can serve as a basis for a publication. Consequently, RDM can be used to document or verify publications. In addition, (ii) the actual research data can be published as a data publication. In this case RDM enables the publication and re-use of the data. Finally, (iii) research data can be managed in the context of a research project. In this case, RDM facilitates the scientific working with the data, like the improving of and re-use of the data for further research methods (e.g. data mining). In this context, the adequate documentation of the data with metadata is essential. Furthermore, WEICHSELGARTNER et al. (2011) state several reasons and benefits of the provision, management, and archiving of research data. Intrinsic reasons include the avoidance of redundant data collections, re-use of data (e.g. to conduct reanalysis), as well as replication of data. Furthermore, extrinsic reasons are outlined like the compliance of requirements (e.g. funder policies), the revision of published research results that are based on specific data and analysis, detection of scientific misconduct as well as enhancement of data sharing and citation within the scientific context. Furthermore, WEICHSELGARTNER et al. (2011) describe positive incentives for the researchers, like the increase of reputation, viability of research results, synergy with other researchers, as well as access to funding resources. Consequently, JENSEN (2011) agrees that RDM is essential in every project. A structured concept is the basis for the safeguarding of the research results during the project funding as well as the provision of the data to the research community.

In the context of publicly funded research projects, several funding agencies have a great interest in the management of the emerging research data (section 2.1.2). Consequently, some funders request the provision of a so-called Data Management Plan (DMP) with the grant application. DMPs are, for example, required by the DFG (2014b), the NATIONAL SCIENCE FOUNDATION (2011b), the NATURAL ENVIRONMENT RESEARCH COUNCIL (2014a), and the WELLCOME TRUST (2014). In a DMP, researchers have to outline the planned creation, management, documentation, sharing, and preservation of their data. The kind and coverage of a DMP is dependent on the requirements of the funding agency. Nowadays, several online tools like DMPONLINE (2014) or DMPTOOL (2014) are available to create DMPs. In addition, several checklists, for instance, from the DCC (2014a) or UK DATA SERVICE (2014b), are available to assist and guide the researchers. Further information about DMP is provided by DONNELLY (2011), JENSEN (2011), SALLANS & LAKE (2014), and CORTI et al. (2014).

In particular, in the context of collaborative research (e.g. instructions or research projects) the planning and management of the evolving research data is essential. CORTI et al. (2014) specify that research conducted in a collaborative environment (e.g. across various institutions and organizations) requires the sharing of data, documents, and other information in a well-managed, controlled, and structured manner (CORTI et al., 2014). In addition, KLAR & ENKE (2013) state that with regard to interdisciplinary, collaborative research environments and projects, there is a need to set up a research data infrastructure. These should serve cross-disciplinarily and support the research exchange of data between the researchers involved. Consequently, multidisciplinary large-scale research projects with focus on environmental sciences should establish a RDM infrastructure. This would improve the collaboration between the scientists, the linkage of their research results, and the development of synergies in a spatio-temporal context (MÜCKSCHEL et al., 2006). Further details about technical solutions of RDM infrastructures and systems are provided in section 2.1.5 and section 2.4.

Some challenges and problems are arising in the field of RDM. Many researchers agree to provide their data to a RDM system. In practice, as identified by NELSON (2009), only a minority of researchers successfully participates in the data provision. This varies in the different disciplines. The researchers argue with different reasons with a social and technical focus for not being able to provide and share their data. This includes ignorance of the infrastructure or archive that has to be used, loss of the data as well as lack of time. Further reasons have been identified by GRIFFITHS (2009), TENOPIR et al. (2011), and HARTTER et al. (2013). For

example, the lack of standards, the lack of funding resources or missing rights to make data public. Furthermore, a lack of experience or expertise in RDM is noticed as well as the fear of exploitation and competitive factors. The missing or limited reward for publishing and sharing data also has an effect. In addition, SALLANS & LAKE (2014) identify barriers of the RDM from the researchers' perspective. The key factors include the lack of training, the ignorance of best practices for RDM as well as insufficient guidance and support.

With regard to interdisciplinary, long-term research projects, MÜCKSCHEL et al. (2007) define challenges and problems in RDM. For example, a fundamental challenge while designing and implementing a RDM system is the insufficient communication between the scientists of different disciplines and the data manager responsible for the RDM system. As a consequence, a lack of acceptance by the researchers may result. Moreover, a low level of motivation by the researchers is identified with respect to providing their data and corresponding metadata to the RDM. To solve the identified problems and challenges in project related RDM systems, MÜCKSCHEL et al. (2007) provide several possible solutions. For example, the RDM should be integrated into the entire research project at an early stage and a continuous communication between the responsible parties involved is required. Moreover, the RDM system should be designed user-friendly (e.g. user-friendly user interfaces). In addition, to support the use by the researchers, the system has to be described well (e.g. user documentation, tutorials). Supplementary training and workshops should be provided. Finally, an internal data policy or guideline is recommended by MÜCKSCHEL et al. (2007). Similar system success factors and solutions are also identified by VAUGHAN (2001).

In order to support an infrastructure for RDM, for example in an institution or research project, several services have to be provided. These services are divided into three main categories, as shown in Figure 2-2.
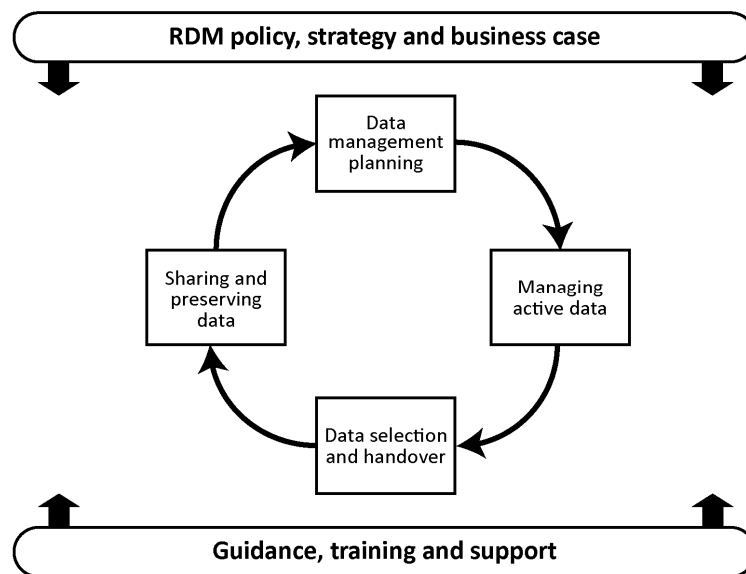
Figure 2-2: Components of RDM services, modified after JONES (2013)

The first category is the overarching governance framework, as described by JONES (2013). General regulations to set up RDM services are determined. The elements RDM policy, strategy, and business cases of a specific institution or research project have to be considered. For example, a data policy can define the core principals of RDM for an institution. Furthermore, extrinsic frame conditions of funding agencies have to be considered (section 2.1.2). The second category comprises all services and infrastructure, which is related to the research data lifecycle (section 2.1.4). This includes, (i) the provision of a Data Management Plan (DMP), (ii) the management of data during the active research phase, (iii) the selection of the data which should be stored on a long-term basis, and handover of the data, and (iv) finally, the preservation, sharing, and re-use of the selected data in a repository, archive, or data catalogue. The third category contains the provision of

guidance, training, and support for the RDM services and infrastructure usage. This can include the guidance and training of researchers to use the established RDM as well as the provision of helpdesk services and complementary advisory websites (JONES, 2013).

In the context of long-term research projects as well with regard to institutions, more policy makers are becoming aware of the importance and benefits of RDM. Consequently, many national, global as well as cross-national position papers, recommendations, and policies have evolved from a multiplicity of recent activities and initiatives. Some will be outlined in detail in the next section.

2.1.2 Research data management activities and policy makers

Many national, international, and cross-national initiatives, institutions and policy makers have emphasized the importance of scientific research data management (RDM) for publicly funded projects (e.g. DFG, 2014b; HOLDREN, 2013; WELLCOME TRUST, 2014). Most are aware of the importance of storing, sharing, and reusing research data. Within the last few years, several funding initiatives and instruments have been created to support the development of corresponding research data infrastructures. This chapter focuses on the main initiatives as well as on main activities of policy makers in the context of RDM. Initially, important global initiatives will be presented. Consecutively, national initiatives in Germany will be outlined with a main focus on the German Research Foundation (DFG) guidelines.

Global initiatives

At the international level, several research organizations, funding organization as well as political initiatives and institutions are aware of the need of RDM. They have issued recommendations, strategy documents, and guidelines. Subsequently, a short overview will be provided. Initially, regarding general global initiatives and then the focus will be on the federal state level.

In general, initiatives like the global Research Data Alliance (RDA, 2014), launched in March 2013, address the topic RDM. The aim of the RDA is to build technical and social bridges. The main vision of the RDA is to facilitate the open sharing and exchange of research data by researches and innovators. This should be irrespective of the technologies used, across disciplines and countries.

Furthermore, the Committee on Data for Science and Technology (CODATA) is promoting the handling of research data (CODATA, 2014). CODATA is an interdisciplinary organisation launched by the International Council for Science (ICSU) in 1966. The aim of CODTA is to support and improve the reliability, quality, accessibility, and management of research data.

In 2007, the issue of the *OECD Principles and Guidelines for Access to Research Data from Public Funding* (OECD, 2007) by the Organisation for Economic Co-operation and Development (OECD) had a major impact on RDM on the national level. For example, the OECD principles requested that their member states implement guidelines, which should be complementary to available national laws and policies, promote the global use, access, openness, and sharing of publicly funded research data in the member countries.

In addition, the importance of open data is also recognised by the G8 Science Ministers. They have released a statement in June 2013 with regard to the enhancement of the collaboration in the various fields of open scientific research data and their access on a global perspective (FOREIGN & COMMONWEALTH OFFICE UK, 2013). In addition, they signed the policy paper *Open Data Charter* to promote innovation, transparency, and accountability (CABINET OFFICE UK, 2013).

At the European level, RDM and research data infrastructures are of great interest. Some initiatives, recommendations and guidelines are available. A cross-national European initiative is, for example, KNOWLEDGE EXCHANGE (2014b), established in 2005. Currently five partner organisations are involved: the IT centre for Science in Finland, Denmark's Electronic Research Library, the German Research Foundation,

the Joint Information Systems Committee UK, and the Dutch SURF organisation. Knowledge Exchange is working on several activities. For instance, one working group is focusing on research data with the aim of coordinating and promoting efforts to make research data more sustainable, accessible, openly available, and reusable in the public domain (KNOWLEDGE EXCHANGE, 2014a).

In June 2008, the European Science Foundation (ESF) and the European Heads of Research Council (EUROHORCs) published *The EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build It* (ESF & EUROHORCs, 2008). In their vision, they demand the

> "[…] open access to the output of publicly funded research and permanent access to primary quality assured research data." (ESF & EUROHORCs, 2008: 2)

The visions and strategies let to essential developments. The *Science Europe Roadmap* issued in December 2013 by SCIENCE EUROPE (2013b) builds on this publication. Science Europe is an association of European Research Funding Organisations and Research Performing Organisations funded in 2011 (SCIENCE EUROPE, 2013a). One priority action area of the roadmap focuses on the access to research data. A corresponding specialized working group on research data will work on this topic.

With regard to e-infrastructure for scientific data in Europe, the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap, has supported the development of European research data infrastructures since 2006 (ESFRI, 2014). Furthermore, the High Level Expert Group on Scientific Data, in trusted by the European Commission, announced a vision for 2030 in the report *Riding the wave - How Europe can gain from the rising tide of scientific data* in October 2010:

> "Our vision is a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance." (HIGH LEVEL EXPERT GROUP ON SCIENTIFIC DATA, 2010: 4)

One aim in the European strategy *Digital Agenda for Europe* is the re-use of data from public resources (EUROPEAN COMMISION, 2014a). In this context, the European Commission issued the Commission Recommendation *On access to and preservation of scientific information* (EUROPEAN COMMISSION, 2012a) and the Communication *Towards better access to scientific information: Boosting the benefits of public investments in research* (EUROPEAN COMMISSION, 2012b) in 2012. Both focus on the management of data and recommend the open access of publicly funded data. The latter also gives recommendations about the RDM in the EU-funded framework program Horizon 2020. With regard to the Horizon 2020 initiative, the European Commission launched a pilot for open research data. Guidelines on the management of data, as well as on open access to research data and scientific publications were issued in December 2013 (EUROPEAN COMMISSION, 2013a, b).

At the federal state level, the United States of America show several initiatives and activities with regard to RDM. For instance, in March 2007 the *Cyberinfrastructure Vision for 21st Century Discovery* was released by the NATIONAL SCIENCE FOUNDATION (2007). Moreover, the BLUE RIBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS (BRTF-SDPA) (2010) outlines the importance of RDM in their report on *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information* in 2010. In addition, the Task Force on Data Policies of the National Science Board published a report on *Digital Research Data Sharing and Management* (NATIONAL SCIENCE BOARD, 2011) and discussed the importance of this topic. In addition, recommendations were provided to the National Science Foundation (NSF). Major federal funding agencies are also aware of the importance of research data management requirements. Already in 2003, the National Institute of Health released a policy on sharing research data (NIH, 2014). Furthermore, since January 2011, the NSF has included the requirement of a DMP into the grant application.

> "[…] proposals submitted to NSF must include a supplementary document of no more than two pages labelled *"Data Management Plan" (DMP)*. This supplementary document should describe how the proposal will conform

to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted." (NATIONAL SCIENCE FOUNDATION, 2011b)

The DMP should include details about the type of the data produced in the project, the data format used and the metadata standard, policies for sharing and access to the data as well as details about the re-use and archiving of the data (NATIONAL SCIENCE FOUNDATION, 2011a).

In terms of sharing and dissemination of the data, the NSF outlines that

"Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing." (NATIONAL SCIENCE FOUNDATION, 2013)

In addition, the White House Office of Science and Technology Policy has released a policy memorandum concerning the *Increasing Public Access to the Results of Scientific Research* (HOLDREN, 2013) in February 2013. In this context, results of scientific research include both, peer-reviewed publications and scientific data in digital form. This policy requires that:

"[…] each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government.." (HOLDREN, 2013: 2)

Finally, MULLINS (2014) provides an overview of the policy framework in the USA. Furthermore, DIETRICH et al. (2012) provides a detailed overview of data policies of funding agencies in the USA, as well as policy details in comparison.

In the UK, several initiatives and research funders are very active in the field of RDM, for instance, the Joint Information Systems Committee (JISC), the Research Council UK (RCUK) or the Digital Curation Centre (DCC). For example, the RCUK has issued the *Policy and Code of Conduct on the Governance of Good Research Conduct* (RCUK, 2013) in 2009, focusing on the preservation and management of research data. In addition, the Royal Society has published the report *Science as an open enterprise: open data for open science* (THE ROYAL SOCIETY, 2012). The report highlights the changes in sciences, gives recommendations on open data, and discusses the importance of RDM.

Many research funders in the UK have released data policies and require a DMP. According to JONES (2012), data policies are supported by several data centres in the UK. For example, the Natural Environmental Research Council (NERC) published a first data policy handbook in 1996 and revised this in 2010. Likewise, the Wellcome Trust has issued a *Policy on data management and sharing* in 2010 (WELLCOME TRUST, 2010). Further details about initiatives and policies are described by JONES (2011). An overview of data policies of UK funding agencies and their details in comparison is provided by DCC (2014d). In addition, JONES (2012) gives a detailed summary of the development in research funders data management and sharing policies for the UK.

In Australia, the *Australian Code for the Responsible Conduct of Research* (NHMRC, 2007) was issued in 2007 jointly by National Health and Medical Research Council (NHMRC), the Australian Research Council, and Universities of Australia. This code provides, for example, guidelines for the management of research data and primary materials and identifies appropriate responsibilities of institutions and researchers. For example, the code defines that

"The responsible conduct of research includes the proper management and retention of the research data. Retaining the research data is important because it may be all that remains of the research work at the end of the project. While it may not be practical to keep all the primary material (such as ore, biological material, questionnaires or recordings), durable records derived from them (such as assays, test results, transcripts, and laboratory and field notes) must be retained and accessible." (NHMRC, 2007: 2.1)

An initiative to support the RDM on the national level in Australia is the Australian National Data Service (ANDS). ANDS is a service with the aim to establish and provide a national research data infrastructure. This initiative is funded by the Australian Commonwealth Government. The aim of ANDS is to ensure and support that all Australian research data are well-managed, accessible, and reusable (BURTON et al., 2012; GROENEWEGEN & TRELOAR, 2013; TRELOAR, 2009; TRELOAR & WILKINSON, 2008).

At the global level, the importance of RDM is realized by universities as well. In particular, at British, American, as well as Australian universities the use of RDM policies is already very common. For example, the UNIVERSITY OF EDINBURGH (2011), UNIVERSITY OF OXFORD (2014), UNIVERSITY OF TENNESSEE (2009), JOHNS HOPKINS UNIVERSITY (2008), UNIVERSITY OF MELBOURNE (2005), and UNIVERSITY OF NEWCASTLE (2008) have a RDM policy. Further data policies from UK universities are provided by DCC (2014e).

In addition, a few journals mandate the archiving and sharing of data and provide data policies. These are for example, the open access journal PLOS One (PLOS, 2014), publications related to the American Geophysical Union (AGU, 2013) or Nature Publishing Group (NPG, 2014b).

Further information about global activities, initiatives, and data policies is provided by JONES (2011), PAMPEL & BERTELMANN (2011), VAN DER GRAAF & WAAIJERS (2012), and STRATHMANN (2013).

National initiatives

In Germany, several research organizations, funding organization, as well as political initiatives and institutions are aware of the need of RDM. They have issued strategy documents, recommendations, and guidelines. According to WINKLER-NEES (2013), the consequence of serious changes in the research environment and the provision of information for research have led to several discussions about strategic modifications and establishments of funding programs with a focus on scientific information infrastructures. This subsection will initially present important national initiatives and activities, as well as guidelines of funding agencies, like the German Research Foundation (DFG), in the context of RDM.

In general, the management of research data funded by public projects is affected by the independence and autonomy of researchers. The handling of their data is determined by the German Constitution article 5 paragraph 3 about the principles of academic freedom. This also facilitates the freedom and right not to publish research results (WINKLER-NEES, 2013).

With the announcement of the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* on October 22, 2003 (BERLIN DECLARATION, 2014), a new milestone was provided for the Open Access movement. The focus was mainly on open access to publications, but also the demand to open access for scientific research results and scientific knowledge was involved. The Berlin Declaration states that

> "Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material." (BERLIN DECLARATION, 2014)

In 2008, the discussion about the importance of RDM was taken on by the *Priority Initiative Digital Information* by the Alliance of German Science Organizations, which includes the biggest research organization in Germany (e.g. DFG, Helmholtz Association, Leibniz Association). In June 2010, the working group *Research Data* announced the *Principles for the Handling or Research Data* (ALLIANCE OF GERMAN SCIENCE ORGANISATIONS, 2010).

In 2010, the strategy paper *General Strategy for Information Infrastructure in Germany* (KOMMISSION ZUKUNFT DER INFORMATIONSINFRASTRUKTUR, 2011) was issued by the Commission on the Future of Information Infrastructure on behalf of the Joint Science Conference (ger.: *Gemeinsame Wissenschaftskonferenz des Bundes und der Länder, GWK*).

Moreover, the German Council of Sciences and Humanities (ger.: *Wissenschaftsrat, WR*) has issued some extensive position papers. For example, in 2011 recommendations were published with regard to scientific information infrastructures (WISSENSCHAFTSRAT, 2011a, b). In July 2012, the WR issued recommendations for future developments of the research information infrastructures in Germany through 2020 (ger.: *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020*) (WISSENSCHAFTSRAT, 2012).

As a consequence of several position papers of the WR, the establishment of a council responsible for information infrastructures (ger.: *Rat für Informationsinfrastrukturen*) was determined at the end of 2013 by the GWK and should be set up in 2014 (BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG, 2014; GWK, 2013). Furthermore, the establishment of framework requirements for an adequate handling of publications and data (open access) was declared in the recent coalition agreement between of the CDU, CSU and SPD for the current election period (BUNDESREGIERUNG, 2013).

In the context of German research organizations and funders, the announcement of the *Proposals for Safeguarding Good Scientific Practice* in 1998 (DFG 1998), the DFG provided a basis for the accurate handling of research data for a fixed period with regard to general, multidisciplinary demands. The recommendation seven on securing and storing primary data included the following statement:

> "Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin." (DFG, 1998: 8)

According to KLUMP & BERTELMANN (2013), the announcement of the proposals was a consequence of several scandals of scientific misconduct. The primary storage of the research data was originally not a purpose. Furthermore, WINKLER-NEES (2013) complements that another motivation was likewise the conformability and reduction of the research results to prevent scientific misconduct. In 2013, the DFG complemented the recommendation of 1998 and republished the proposals in an adjusted, updated version (DFG, 2013e). In the context of recommendation seven *Safeguarding and Storing of Primary Data*, only commentary notes were modified.

Arising from the DFG-position paper *Scientific Library Services & Information Systems - Funding priorities through 2015* (DFG, 2006), the service project section *Information Infrastructure* (so-called *INF*-project) was integrated in the DFG-funded coordinated program *Collaborative Research Centre (CRC)* and *CRC/Transregio* in 2007 (EFFERTZ, 2010). According to ENGELHARDT (2013), as of march 2013, 45 INF-projects were applied since 2007 and 27 of these were funded. Consequently, KLAR & ENKE (2013) complement that these are only approximately 10% of all funded CRC/Transregio research projects. CRCs are research facilities established at universities and can be funded for up to twelve years. CRCs are characterized as multidisciplinary, long-term research projects. Researchers from various disciplines at different institutions collaborate in sub-projects to achieve a result with regard to long-term, complex research. CRC can incorporate researchers from several universities and non-university research facilities. In contrast to a CRC, a CRC/Transregio holds strong, synergetic nationwide collaborations with researchers from other participating institutions to achieve a joint research result (DFG, 2014a). Within a CRC, the main purpose of a so-called INF-project is the management of all relevant data, which was collected by all researchers in the context of the CRC. The provision of an information infrastructure should support the long-term and systematic usage of the data beyond the funding duration. The *INF*-project has two aims. First of all, the research data have to be archived for at least ten years to fulfil the requirements of the *Proposals for Safeguarding Good Scientific Practice* (DFG, 1998). Secondly, scientific synergies should be generated between the researchers by means of collaborative research and communication platforms, which enable sharing and re-use of data. Local information facilities (e.g. university library or computing centre) should be involved in the data management. Already available technologies and tools should be given preference. INF-projects can apply funding for different objectives. For instance, they can focus on the design and implementation of a database, which is

able to handle all data generated by the CRC members. This database should assign metadata to the data and should be interoperable with other related repositories (DFG, 2013d).

In 2009, the DFG Subcommittee on Information Management, which is a part of the Committee of Scientific Library Services and Information Systems, issued the *Recommendations for Secure Storage and Availability of Digital Primary Research Data* (DFG, 2009b). These provide recommendations for researchers as well as information specialists to raise the awareness of the importance of this issue. Initially, the recommendations specify the term primary research as

> "[...] data that result during the course of scientific research, experiments, measurements, surveys or polls. They serve as the basis for scientific publications." (DFG, 2009b: 2).

The recommendations refer to the discipline dependency of research data management. Consequently, each scientific discipline has to establish the technical and organisational conditions, for example, with regard to sustainable storage and provision of the data. Depending on the discipline, scientists have to agree upon the data to be included (e.g. raw data, aggregated data). In addition, data should be assigned with a persistent identifier. National and international standards have to be considered for the storage of the data. Moreover, the storage structure has to enable personally labelling and assignment of the data to a researcher, associated with permission management. All data have to be described with accurate metadata, at least according to the Dublin Core Metadata Standard (section 2.2.2.1). As far as possible, scientist should make their quality controlled data freely available on a national basis.

Since 2010, the DFG has integrated the management of research data directly into the proposal preparation instructions (DFG, 2014b). The applicants are requested to provide a DMP, which should identify the appropriate measures that will be taken to handle the data. The instructions define:

> "If research data or information will be systematically produced using DFG project funds, describe if and how these will be made available for future re-use by other researchers. Please regard existing standards and data repositories or archives in your discipline where appropriate." (DFG, 2014b: 5)

In addition, the German Federal Ministry for Education and Research (BMBF) requests the applicants of funding to provide a *plan for realization*, which should give information about the handling of the collected data as well as information about re-use of the data and project results (WINKLER-NEES, 2012).

Further national initiatives and activities of other research organizations are outlined in detail by WINKLER-NEES (2013). For example, the Helmholtz Association of German Research Centres participates in the open access activities (HELMHOLTZ-GEMEINSCHAFT, 2014b) and provides a recommendation for good scientific practice (HELMHOLTZ-GEMEINSCHAFT, 2014a). Within the Leibniz Association, several individual approaches and initiatives exist. At the discipline-specific level, several research data centres have been set up (LEIBNIZ-GEMEINSCHAFT, 2014), for example, GetInfo (TIB, 2014a), the Portal for Science and Technology at the German National Library of Science and Technology (TIB) or the data centre West African Vegetation (SENCKENBERG, 2014), which is a data network for vegetation of West Africa at the Senckenberg Research Institute and Natural History Museum. In the framework of the Max Planck Society, the central Max Planck Digital Library (MPDL) was set up to handle publications as well as scientific raw data (MPDL, 2014). In addition, the Max Planck Society provides recommendations for safeguarding good scientific practice (MPG, 2009).

Furthermore, German universities are getting aware of the importance of RDM. For example, recently in May 2014 the German Rectors' Conference (ger.: *Hochschulrektorenkonferenz, HRK*) has released recommendations concerning the handling and management of research data (HOCHSCHULREKTORENKONFERENZ, 2014). In addition, the rectors' office of the University of Bielefeld has passed a resolution in November 2013 regarding to RDM (UOB, 2014b). Furthermore, German initiatives like the Network of Expertise in Long-Term Storage of Digital Resources (NESTOR) in Germany (NESTOR, 2014), a competence network focusing on specific subjects related to digital preservation, pushes the discussion on RDM.

More detailed information about national activities, initiatives, and data policies is, for example, provided by PAMPEL & BERTELMANN (2011) and WINKLER-NEES (2013). According to NEUROTH (2013), the abundance of position papers and recommendation reports in the various facilities shows, that with regard to information and research infrastructures, Germany is facing a realignment.

2.1.3 The diversity of research data

This chapter gives a short introduction to the subject of research data. The general terminology as well as an introduction to the documentation of research will be presented.

The general term *research data* is found, used, and described in many disciplines with different explanations. In the Anglophone literature, the terms *research data* and *scientific data* are common. In German literature, meanwhile the preferred term for research data is *Forschungsdaten*. They are characterized as data which have been produced in the context of a particular research question (BÜTTNER et al., 2011). KLUMP (2010) describes that scientists use different terms to describe their data depending on the research background. Common terms are, for example, measuring data, (scientific) raw data, empirical data, source data or primary data.

As a consequence of the use of data in the various disciplines, it is difficult to provide a uniform definition. According to the *OECD Principles and Guidelines for Access to Research Data from Public Funding* (OECD, 2007) research data are defined as

> "[…] factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated. This term does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g. laboratory samples, strains of bacteria and test animals such as mice)." (OECD, 2007: 13f)

GREENBERG et al. (2009) identify research data with regard to science as "[…] the essence of science - the holder of scientific truth when findings are reviewed." (GREENBERG et al., 2009: 196). BORGMAN (2012) uses a common definition for data, which is given in the National Academies of Science report where data are described as facts, letters, numbers, and symbols that define an idea, object, situation, condition, or other factors. Moreover, BORGMAN (2012) points out that the terms data and dataset sometimes conflate.

With regard to academic research, PRYOR (2011b) defines research data as "[..] the output from any systematic investigation involving a process of observation, experiment or the testing of a hypothesis, which when assembled in context and interpreted expertly will produce new knowledge." (PRYOR, 2011b: 3).

OSSWALD et al. (2012) complements that the term research data has to be redefined always in relation to the respective scientific discipline. Complementary, LUDWIG & ENKE (2013) and VAN DEN EYNDEN et al. (2011) describe digital research data as a valuable resource, created as a result of a scientific workflow, which usually requires much time, money and expertise.

HOX & BOEIJE (2005) state that research data can be either *primary* or *secondary*. Primary data are characterised as data which have been collected for a specific goal by using a certain procedure. Primary data originally collected for a specific purpose and then re-used for another research study are called secondary data. Primary data according to the DFG (2009b) are data, which "[…] represent a valuable pool of information acquired at great financial expense. Depending on research area and method, the data are reproducible or are based on non-reproducible observations or measurements." (DFG, 2009b:2). For instance, the DFG (2013e) gives examples for primary data, which are derived from observations, experiments, numerical calculations, measuring results, collections, study evaluations, cell cultures, material samples, archaeological finds or questionnaires.

In addition, research data appear in many different forms, as a consequence of the various disciplines. JOSHI & KRAG (2010) and BORGMAN (2012) point out that the form of data may vary significantly within the scientific communities, both digital and physical data can be found. As stated by GRIFFITHS (2009), researchers collect and create a variety of different data during their research study across the diverse disciplines and subjects. They use different methodologies and process for this purpose. For instance, DAVIS & VICKERY (2007) give the example that "[…] data and data sets include everything from geographic information systems or geospatial (GIS) data to genomic data to any data set supporting a scholarly publication, such as census data." (DAVIS & VICKERY, 2007: 1)

As already mentioned, research data comes in many different forms and there are various possibilities to classify research data. For example, the NATIONAL SCIENCE BOARD (2005) subdivides research data according to the nature of the data, their reproducibility, and processing level.

Nature of data The nature of data can be diverse and, for example, include images, audio streams, videos, animations, software, algorithms, models or simulation results.

Reproducibility level Regarding the reproducibility, data can be divided by their origin into three categories, which are observational, computational, or experimental data. The *observational* data include results from a specific place and date that cannot be recollected. For example, these are data of a direct observation of ocean temperature or weather observations. Consequently, these data are archived for an indefinite period. *Computational* data results from executing a computer simulation or model. These data can be reproduced, if the used hardware, software, and input data are extensively documented. The model output might only be preserved in some cases. *Experimental* data are results from laboratory studies or field experiments. In theory, these data can be accurately reproduced. However, in practice, the exact experimental results cannot be precisely reproduced because of unknown experiment and condition variables or prohibitive costs. Thus, it is necessary to ensure the long-term preservation of these data.

Processing level Finally, derivative data are produced by curatorial and processing activities. Raw data have been collected by sensors and instruments and then have been refined and analysed according to the research objective. The generated derivative data is generally more usable for further use. Consequently, in many circumstances it is warranted to preserve multiple forms of data.

Additionally, LUDWIG & ENKE (2013) and RIN (2008) distinguish the form of data, similar to the already described subdivision. The data form is divided according to content, generation, and acquisition of the data into the following five categories:

Experimental data These data can be reproduced in principal. In practice, the reconstruction is uneconomical or very difficult.

Models or simulations For simulated data it can be of importance to keep the model with the associated initial conditions of the simulation instead of the calculated output data.

Observations These data represent a specific phenomenon at a certain time or location. In general, these data contain a unique and irreproducible record of an event.

Derived data These are data derived by processing of (raw) data or linking of different data using a special method. The documentation of the origin and the derivation method is of relevance.

Canonical or reference data Data describing other data according to a specific standard or a translation of the data in a standard format is represented. The transition to metadata is smooth.

Depending on the form of research data, likewise the file format will change. Different file formats are available corresponding to the used software programme which created the file. According to CORTI et al. (2014), the method of data analysis and collection selected by a researcher is responsible for the chosen software and the appropriate data output format. Moreover, CORTI et al. (2014) recommend that researchers

should use (preferably non- proprietary) standard file formats, organise their data file in a well-defined structure, and rename the files in a sensible, logical manner. Examples for recommended file formats for long-term preservation, management, sharing, and re-use are provided by CORTI et al. (2014). In addition, best practice examples for choosing file formats and naming conventions in RDM are, for instance, provided by the UNIVERSITY OF OREGON LIBRARIES (2014) and the STANFORD UNIVERSITY LIBRARIES (2014). Both also highlight the importance of the documentation of research data and provide best practices. Subsequently, a short introduction will be given to research data documentation.

Documentation of research data

The documentation of research data is an essential duty. Well conducted documentation ensures that the data can be re-used and are interpretable (WEICHSELGARTNER et al., 2011). JENSEN (2011), HUSCHKA et al. (2011), and CORTI et al. (2014) emphasize that a detailed and clear description of data, complementation with contextual information, and additional annotation is necessary. This will ensure and is concurrently a requirement that the data can be understood and interpreted well and consequently can be used correctly by the secondary users. In this way, data becomes more user-friendly, sharable, and a long-lasting usability will be achieved. JENSEN (2011) and WEICHSELGARTNER et al. (2011) highlight that the documentation of the data should contain details about the project background and study design, context and condition, the method and procedure of data acquisition, the applied instruments as well as information about the data confidentiality, use, and accessibility.

According to VAN DEN EYNDEN et al. (2011), good data documentation should include information about:

- data collection context (e.g. project history, objectives, aim and hypotheses)
- data collection methods (e.g. data collection process, sampling methods, used instrument, hardware and software, resolution and scale, geographic and temporal coverage, used secondary data sources)
- dataset structure of data files, relationships between files, study cases
- procedures carried out (e.g. data validation, proofing, cleaning, checking, quality assurance)
- documentation of data file changes and data file versions
- restrictive covenant (e.g. access and use conditions) and confidentiality of data

Moreover, CORTI et al. (2014) describe the documentation of the data at the data-level. This may include, for instance:

- variables information (e.g. names, labels, descriptions, records, values)
- details about used codes and classification schemas
- definitions of used acronyms or specialised terminologies
- information about missing values (e.g. codes, reasons)
- details about derived data processes (e.g. collection, with algorithm or code)
- data listing of annotations for individuals, items or cases

Likewise, JENSEN et al. (2011) point out that, beside the genesis of the data and the used instruments, also the meaning of the data plays a central role within the documentation of the data. For example, the significance of a series of measurements and their values has to be clear before an analysis of the data can be conducted. For this purpose, a precise documentation of the data has to be provided. Moreover, the authors add that standardized metadata are a necessary requirement to document the research data and to ensure their permanent safeguard, re-use, and exchange.

However, documentation of research data also poses challenges. Data documentation is usually not included in the process of data publication and research. Consequently, this means an extra amount of work and time for the data creator. Usually, this will not be rewarded by the scientific community. Hence, RDM including the documentation of the research data is disregarded by the researchers (WEICHSELGARTNER et al., 2011).

Consequently, CALLAHAN et al. (1996) state that it is a key challenge to motivate people to document their data. Since data documentation is regarded as an onerous task by the scientist, this is often left undone. In chapter 2.2, the issue of data documentation by using metadata will be described in detail.

2.1.4 The research data lifecycle

Research data goes through a number of phases in its life time, as already illustrated in Figure 2-1. According to DCC (2014f), the ongoing activity of "[…] maintaining, preserving and adding value to digital research data throughout its lifecycle." (DCC, 2014f) is called digital (data) curation. This ensures the availability of the data for re-use and discovery in the future. Several data life cycle models have been developed in the different disciplines. For example, a popular data lifecycle model in the social sciences was developed by the Data Documentation Initiative (DDI, 2014) to manage and document data across the complete lifecycle, as well by the INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH (2012) and HUMPHREY (2008). In the field of environmental sciences, the United States Geological Survey (USGS) has created a science data lifecycle model (FAUNDEEN et al., 2014). Further data lifecycle models from the perspective of the researcher are presented by STRASSER et al. (2011), JOHNSTON (2014), DEELMAN et al. (2010), and the UK DATA ARCHIVE (2014). In contrast, the lifecycle of research data from the perspective of a curator is presented in the widely-used, comprehensive curation lifecycle model of the Digital Curation Centre (DCC, 2014b).

LUDWIG & ENKE (2013) use a simplified model to describe six single stages in a research data life cycle, which correspond to resulting duties. These cover:

Plan and create The process of data creation or collection should be planned reasonably. Data should be produced respectively to simplify its management. For example, current data formats and standards should be considered.

Select and appraise This stage includes the selection and evaluation of the data, which have to be preserved. In addition, the duration of storage has to be defined.

Ingest and acquire: This stage includes the preparation of the data to transfer them to an adequate storage environment, like a data archive. This includes, for example, the validation of the data, the generation of descriptive metadata, and the preparation of the release file.

Preserve The sustainable storage of the data is a main duty in the data management. Available challenges like size, amount, and access quantity of the data have to be considered. The storage should be conducted in an institution with expert knowledge, like a data centre or archive.

Maintain The stage should consider the maintenance of the data with regard to technological modifications as well as changes in the user target groups. Corresponding arrangements have to be taken.

Access and (re-)use This stage has to consider the access and use of the data. Data has to be searchable and interoperable. In addition, access restrictions have to be arranged.

Comprehensively, five duties and responsibilities are depicted by LUDWIG & ENKE (2013), which are important within the research data life cycle. They include (i) an organization responsible for the entire process of the data management and corresponding policies and (ii) the emerging costs which have to be considered in the funding. In addition, (iii) the applicable law and ethical principles (e.g. copyright, licensing), (iv) the accurate documentation with metadata, and (v) the usage of identifiers for research data are important duties.

A different approach to model the different stages of the research data life cycle is applied by TRELOAR & HARBOE-REE (2008), as presented in Figure 2-3.

Different Repositories for Different Domains
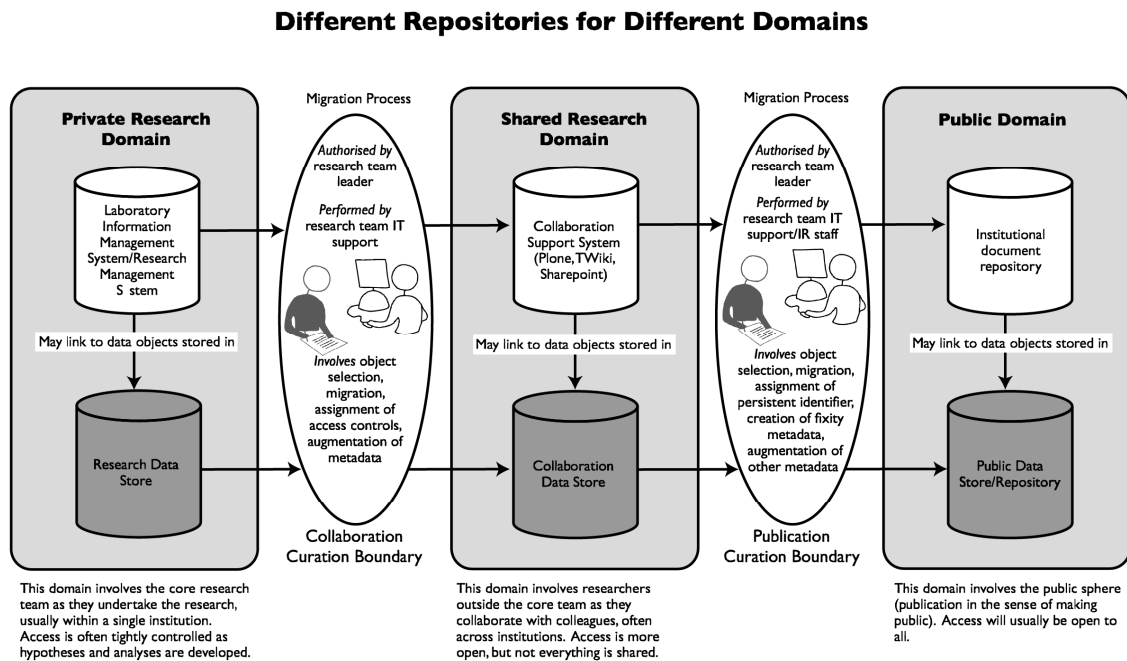
Figure 2-3: The data curation continuum (TRELOAR & HARBOE-REE, 2008)

The data curation continuum is divided into three domains. All three domains have various requirements, for example, with regard to data description and used technologies to store the data. The first domain in the data curation continuum is the *Private Research Domain.* This is the domain, where a research team is producing the data and the research results. All data is stored in the research data store, for example, a file system. Data in this domain is characterized as having less metadata, are continuously updated, less preserved, and with limited access. If the research team agrees to share the research results with other collaborative research teams, then the data has to migrate to the second domain, the *Shared Research Domain.* The data in this domain has to be stored in a collaborative data store. Depending on the file size and nature of collaboration, a corresponding data store has to be implemented. This has to facilitate greater structuring and access control. Only selected, static data is stored, which is characterized by more descriptive metadata, more preserved, and less restricted access. After finishing the research, a migration is carried out to the *Public Domain.* The resulting publications are available for public access, such as an institutional document repository. Underlying data, which are stored in a public data repository or store, can be linked to the publications, for example by using Digital Object Identifier (DOI). Data in this domain is characterized as being more described with metadata. The data are static, more preserved, and have open access (TRELOAR & HARBOE-REE, 2008). In addition, TRELOAR & HARBOE-REE (2008) point out that, theoretically, all three domains can be supported by a single repository. KLAR & ENKE (2013) discuss an expansion of the previous outlined domain model. Furthermore, they determine that projects, which cover more than two domains cannot be realized with two or three staff members.

The requirements of the various presented research data lifecycle models identify a need to develop and establish technical solutions, repositories or data centres to accurately store and preserve research data. The subsequent section addresses this issue.

2.1.5   Research data management infrastructures

In order to support the management of research data, several approaches are available for the provision and organization of an appropriate RDM infrastructures or systems. According to (RAZUM, 2011), RDM systems form the technical framework to collect, document, and provide research data, which have been created within a research group or initiative. This chapter gives a short overview.

The management of research data and the organisation of the data in a respective infrastructure are conducted by several institutions, organisations, and activities. Figure 2-4 presents the activities of managing research data in a four tier hierarchy.
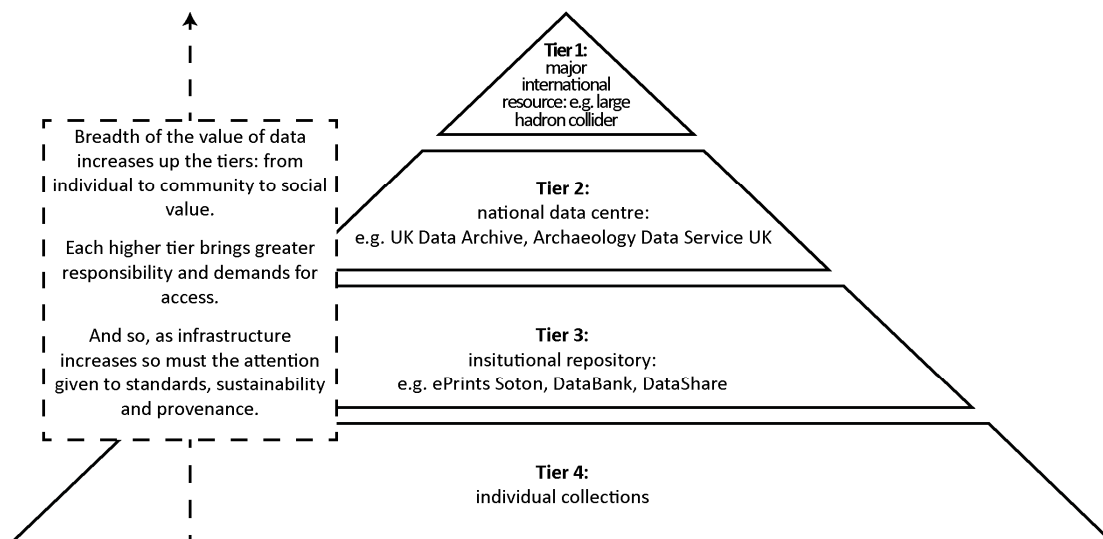
Figure 2-4: The data pyramid representing the hierarchy of data management infrastructures, modified after THE ROYAL SOCIETY (2012), previously explored by BERMAN (2008)

Tier 1 includes all institutions and activities that handle major international resources. For example, data generated by international programs, like the Large Hadron Collider, is considered. In addition, activities that curate large numbers of data on an international basis are involved (e.g. Worldwide Protein Data Bank, Dryad, PANGAEA). *Tier 2* comprises all initiatives on data management operating on a national level like data centres (e.g. UK Data Archive, National Geoscience Data Centre UK, Archaeology Data Service UK). *Tier 3* comprises all activities that are providing institutional repositories, like universities, individual research institutions as well as infrastructures containing research data generated within a research program. Finally, *Tier 4* includes all individual data collections provided by individual researchers or research groups. Each of the described tiers demands different support of infrastructure, financial support, usage of standards, considering of access restrictions, and sustainability of the data. All levels gain divers international reputation (THE ROYAL SOCIETY, 2012).

The different types of infrastructures supporting RDM as well as the influences of several communities involved have led to various terms in the different disciplines related to infrastructures for research data. According to MILLER (2011), some are used synonymously, like *digital library*, *digital collection*, *digital repository* or *digital archive*. Furthermore, repository types like *institutional repository*, *disciplinary repository*, *trusted repository*, and *open access repository* are common and cover different data and functionalities depending on the target group (ASCHENBRENNER & NEUROTH, 2011; SIMONS & RICHARDSON, 2013). In the field of geosciences the terms *geolibrary* and *spatial data infrastructure (SDI)* are used in this context (NOGUERAS-ISO et al., 2005). In addition, the term *community* or *project database* is common. Both are developed within a specific group or network of researchers with the purpose to disseminate or share their research results in the project or research community (LEONELLI, 2013). In addition, project databases developed within a collaborative

research project are characterized to have a limited service lifetime which corresponds to the funding period of the research project. Their purpose is to enable presentation, exchange, and search of research data within the research team (LOTZ et al., 2012). In the context of the described study in this thesis, the term *project database* was selected in 2007 to describe the data collection of the research project (section 4). Recently, the term *repository* is more common. Consequently, this term will be used in the context of this thesis as well.

A digital repository can be described with some characteristics, according to HEERY & ANDERSON (2005). In general, a repository contains digital content, which is managed by the repository architecture including metadata. In addition, a repository provides several services such as delivery, search, and access to content with appropriate access control features. Finally, a repository must be well-supported, well-managed as well as trusted and sustainable. Consequently, an institutional repository (IR) is defined as an online space of a specific institution or a group of associated institutions which enables the collection, preservation, and accessibility of research results in digital form. Usually, IR are established at universities as well as other research or academic institutions (MILLER, 2011). ASCHENBRENNER & NEUROTH (2011) list some functionalities which are covered by repository systems, like management of digital objects and corresponding metadata for identification purposes, crosslinking of digital objects, access to and re-use of research data, and the consideration of legal frameworks as well. In addition, ASCHENBRENNER & NEUROTH (2011) state that repositories can be designed with significant differences since the majority is adapted according to a specific research area, which requires a lot of individual design and adjustment.

For the creation of a digital collection or a data repository, several steps and functions have to be conducted according to MILLER (2011):

- Designing, managing, and administrating the project
- Development of detail collection (e.g. selection of resources which should be digitized)
- Digitizing, formatting, and saving of the created digital objects
- Design of the database and retrieval system
- Design and programming of the web-based user interface
- Design and creation of the metadata (description, cataloguing and indexing of resource)

In addition, in terms of establishing an institutional repository, SIMONS & RICHARDSON (2013) describe several aspects which have to be considered. These include, the establishment of a data policy as well as the identification of the repository scope, the needs of the users (e.g. with interviews, survey), the demands on the metadata schema, the requirements on the data storage, and the preferred way of data disposal in the repository. In addition, access restrictions, which are related to legal and ethical topics (e.g. access rights, copyright) and licensing of the data has to be considered. Workflows in the repository have to be defined, and the software for the repository has to be selected as well. The discovery of the data and the target group has to be discussed. Finally, it is important to find advocacies for the repository by means of promoting it.

Currently, there is a number of out-of-the box software available to set up repository systems. Some are presented in section 2.4. Moreover, KOUDELA et al. (2013) provide an overview of recent repository solutions. Derived from available software systems ASCHENBRENNER & NEUROTH (2011) outline that a repository system is arranged in a three-tier architecture. This includes:

Storage (physical) This layer provides the physical storage of the digital objects with corresponding metadata, usually organized in a file-based storage. The storage is either located on a local server or outsourced in a computing or data centre. The latter is recommended for huge data volumes

Object management (logical) This layer manages the objects (consisting of data and metadata, both linked) including versioning and linking to other objects as well as visualization and access.

Application (conceptual) This layer provides application features for the users (e.g. search, provision of data). The features are related to the repository context.

In the context of interdisciplinary, long-term research projects, it is the experience of many research projects with focus on earth and environmental sciences that existing software solutions are not fully complied with the special framework conditions (GÖTTLICHER et al., 2010; HEIMANN et al., 2010a; KUNKEL et al., 2013; MÜCKSCHEL et al., 2004; NADROWSKI et al., 2013; ZANDER et al., 2013). Consequently, those projects have decided to design their own software solutions instead of choosing available out-of-the box systems. LOTZ et al. (2012) identify several mandatory features which have to be supported by self-developed solutions, created in the context of collaborative research projects. This includes, for example, adequate management of user rights (e.g. access control) and implementation of web-based, user-friendly interfaces to support data disposal, download, and search. In addition, adequate metadata standards and requirements regarding the quality have to be considered, as well as the management of various data versions. Moreover, a cooperation with an institution, like a library or computing centre, might be fruitful for the development and implementation of the RDM system (DOWNS et al., 2007). In addition, BÜTTNER et al. (2011) outline that, besides the technical developments and concept, it is likewise important to raise and distribute the awareness of research data. For instance, introducing an adequate data policy at the domain level (private, shared or public) to increase the publishing of research data.

2.1.6 Publishing and citing research data

In recent years, the publishing of data became more important. For example, funding organisations, like the NSF (NATIONAL SCIENCE FOUNDATION, 2011b) mandated the publishing of research data. Likewise, a growing number of journals (e.g. PLOS, 2014) require the publishing of the underlying data with the submitted article. As stated by CORTI et al. (2014), the publishing of data has many advantages for the data creator. This includes the increase of the profile of the data and enables a proper citation. Consequently, published datasets can be acknowledged in publications as well as re-used in further investigations. This section will outline different approaches for data publication and citation of data by using the Digital Object Identifier (DOI).

Publishing of research data

At present, the scientific information workflow is often dominated by the publication of the research results in form of articles, published by a library, as shown in Figure 2-5 (left). According to KLUMP et al. (2006), in many cases the underlying research data of a publication are pimarily stored on private systems and are consequently lost for further re-use and reproduction of the study. These data are not going through the entire research data life cycle (section 2.1.4). Particularly in recent years, the handling of data has changed. The suggested approach of KLUMP et al. (2006), shown in Figure 2-5 (right), to publish data by data centres and web portals, and the inter-linking of these data with the accociated publication, is already partially realized in some disciplines nowadays (e.g in geosciences by PANGAEA (2014)).
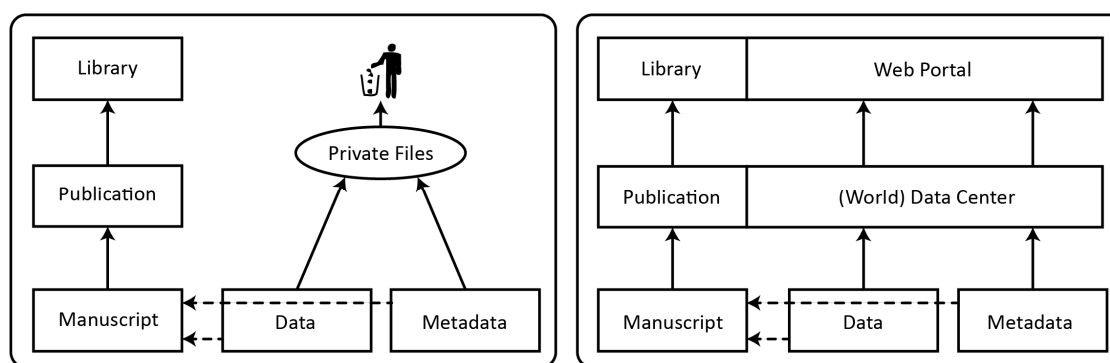
Figure 2-5: Current scientific information workflow (left) and modified handling of data (right), modified after KLUMP et al. (2006) and HELLY et al. (2003)

Research data can be variously published, as described by LAWRENCE et al. (2011), REILLY et al. (2011), and CORTI et al. (2014). In terms of research data with a focus on natural sciences, DALLMEIER-TIESSEN (2011) outlines three publication models, which also correspond to KRATZ & STRASSER (2014). These include (i) the publication of research data in a data repository, (ii) the publication of the data with a corresponding documentation in a data journal, and (iii) the publication of the data as supplementary material to a publication.

> **Data repository** The publication of the research data through a repository (e.g. PANGAEA, 2014; DRYAD, 2014a), has many advantages. As described by CORTI et al. (2014), a repository ensures the long-term preservation, backup, and secure access to the data. In addition, quality standards, the application of permanent identifiers, and licensing of the data are promoted.

> **Data journal** The publication of the data through a data journal, as a data article or data paper has recently increased (CORTI et al., 2014). In general, a data paper provides a documentation of the scientific valuable research dataset which describes, for instance, the methodology used to create the data. These papers go through a full peer-review process, like an ordinary research paper (DALLMEIER-TIESSEN & PFEIFFENBERGER, 2010). Exemplary data journals are the Earth Systems Science Data Journal (ESSD, 2014), the Geoscience Data Journal (GEOSCIENCE DATA JOURNAL, 2014), the Biodiversity Data Journal (BDJ, 2014), and the GigaScience Journal (GIGASCIENCE, 2014). In addition, Nature Publishing Group (NPG) has launched the new open-access, peer-reviewed, online-only publication portal Scientific Data (NPG, 2014a) in May 2014. With the exception of the GigaScience Journal, which has its own cloud repository for storing the associated data, all other data publishing journals demand the submission of the related datasets to a long-term, approved data repository such as Dryad Digital Repository (DRYAD, 2014a), PANGAEA (PANGAEA, 2014), NERC Data Centre (NERC, 2014b) or figshare (FIGSHARE, 2014). Some data journals are in the index of scientific publication databases. For example, ESSD is indexed in Scopus.

> **Data as supplementary material** When data is published as a supplement material to a publication, the data can be published through a repository related to the journal or an external repository (DALLMEIER-TIESSEN, 2011). Some repositories are related to publishers, for instance, PANGAEA is related to Elsevier (ELSEVIER, 2010), as well as Dryad is related to several other publishers and journals (DRYAD, 2014b). Usually, a dataset is linked and cited in a publication by a unique, persistent identifier.

Citation of research data using persistent identifiers

Unique, persistent identifiers are used to identify an object, person or class of objects (SIMONS & RICHARDSON, 2013). Examples for identifiers used with regard to research data are: the Uniform Resource Name (URN), Uniform Resource Locator (URL) or the most common Digital Object Identifier (DOI) (CORTI et al., 2014). The latter enables the accurate citation of data. The DOI system is adopted by several repositories including PANGAEA (PANGAEA, 2014) and Dryad Digital Repository (DRYAD, 2014a).

A DOI is an alphanumeric, globally unique name or string that identifies digital content with the general aim to link and cite electronic resources (BRASE & FARQUHAR, 2011). In 1998, the DOI was introduced by the International DOI Foundation (IDF, 2014). The IDF has official member registration agencies, for instance DataCite (BRASE, 2010). As described by BRASE et al. (2014), DataCite was founded in 2009 and a Global DOI Registration agency responsible for the registration of all scientific datasets was established. Currently, the German National Library of Science and Technology (TIB) is the managing agent of DataCite. Members of DataCite are, for example, the TIB, the German Leibniz Institute for the Social Sciences (GESIS), or the British Library. The registration of DOIs is provided through the DataCite members. For instance, the TIB has been a registration agency since 2005 and provides DOI registration services for over 40 data centres (e.g. German Research Centre for Geosciences (GFZ), PANGAEA, World Climate Data Centre). The focus of

the TIB is on data registration of research data, *grey* literature, no-textual material, and open-access journal articles (TIB, 2014b). According to BRASE (2013), the TIB is responsible for the DOI registration and storage of relevant metadata which are in compliance with the DataCite Metadata Schema (2.2.2.2). The data centre (or other trustworthy institutions) applying the DOI through the TIB is responsible for the storage and accessibility, quality assurance, and metadata creation of the relevant research dataset. Consequently, by resolving a DOI, the user is linked directly to the landing webpage of the DOI provided by the data centre. This webpage also includes the download links of the data and corresponding metadata. In detail, a DOI name comprises a predefined prefix (e.g. 10.5880), a slash, and an individual suffix (e.g. TR32DB.10), for example 10.5880/TR32DB.10. The persistent link to a dataset with an associated DOI is for instance http://dx.doi.org/10.5880/TR32DB.10. In combination with the creator, publication year, title, and publisher, the dataset can be cited according to the DATACITE METADATA WORKING GROUP (2011). For example:

> Curdt, Constanze (2014): TR32DB Metadata Schema for the Description of Research Data in the TR32DB. CRC/TR32 Database (TR32DB). Doi: 10.5880/TR32DB.10.

Further examples for the citation of research data in different disciplines is provided by BALL (2013b) and the CODATA-ICSTI TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES (2013). Furthermore, the Data Citation Synthesis Group of the Future of Research Communications and e-Scholarship (FORCE11) has published the *Joint Declaration of Data Citation Principles* (FORCE11, 2014) in February 2014.

Further initiatives with regard to data citation and the increasing of metrics of a datasets is the Thomson Reuters Data Citation Index (DCI), which makes data citable, reusable, discoverable, and interpretable (THOMSON REUTERS, 2014). In addition, the web-based tool ImpactStory (IMPACTSTORY, 2014) reports alternative metrics for many research objects, like articles, software, blog post, data sets, and posters. For example, download statistics are reported from blog posts, Wikipedia, Twitter, as well as data repositories, like Figshare and Dryad Digital Repository.

2.1.7 Sharing and access to research data

Data sharing in blogs or social media platforms, like Twitter, Facebook, Google+ or YouTube is very common (THALER et al., 2012). Most scientists agree to share their (raw) data or research results (MARX, 2012; NELSON, 2009). Though in practice, only a minority of scientists in various fields do share their data (VAN NOORDEN, 2013). Sharing of data varies in the different disciplines and communities. Some disciplines have a long history of sharing their data, like astronomy, oceanography or genetics (REICHMAN et al., 2011). For instance, sharing of DNA in genetics is very common, as is sharing of images of stars and galaxies from digital sky surveys in astronomy (VAN NOORDEN, 2013).

There are several reasons for avoiding data sharing. As described by HARTTER et al. (2013), the challenges of data sharing can be grouped in technical, socio-cultural, and ethical challenges. These include missing user participation, lack of time and funding, lack of resources, loss of data, ignorance of data archives or RDM systems as well as a lack of experience in RDM and usage of RDM systems. In addition, the fears of exploitation and competitive factors have an affect (GRIFFITHS, 2009; HARTTER et al., 2013; NELSON, 2009; TENOPIR et al., 2011). In addition, SALLANS & LAKE (2014) complements the lack of training and guidance as well as insufficient support, which is noticed from the perspective of the researchers. Furthermore, reasons for not sharing research data are missing or limited reward system for publishing and sharing data. As stated by BORGMAN (2010), in most disciplines scientist are only rewarded by publications. They are not getting reputation on data management. Consequently, REICHMAN et al. (2011) demands the establishment of a reward system, which should motivate scientists to share their data. Consequences of avoiding data sharing are empty archives or RDM systems (NELSON, 2009).

As described by BORGMAN (2012), there are several reasons and arguments for sharing research data. These include (i) the verification and reproduction of research results, (ii) the provision of public financed research results, (iii) enabling others to ask questions on the data, and (iv) advancing new innovation and research. In addition, an improved access to and sharing of public financed research data promotes new research, supports studies on measurements and data collection methods, supports the testing of alternatives or new methods of analysis and hypotheses, and facilitates the exploration of new topics, which were primary not envisioned (OECD, 2007). Furthermore, TENOPIR et al. (2011) complements that advantages of data sharing also involve the re-analysis of data and support different interpretations of the data, particularly in a multidisciplinary research environment. In addition, the re-collection of data is avoided or minimized, safeguard against scientific misconduct is supported, as well as data can be used to train new scientists.

There are many benefits for researchers to share their data and research results. As outlined by VAN NOORDEN (2013), some personal benefits arise by sharing data. They include the establishment of new links to colleagues, increase of citation, as well an improved viability. Consequently, successfully shared data is getting downloaded, re-used, and cited more often. Thus the researchers get more attention in the research society as well. Recent studies have showed that the citation of research articles increases, if the underlying research data associated with the article is publicly available and shared in detail (PIWOWAR et al., 2007). Consequently, the sharing of detailed research results rewards scientists who share their data. This should lead to a motivation to share research results more often.

As described by CORTI et al. (2014), sharing of files, data, documents, and information is especially essential in collaborative research environments. Often researchers have to share their data across various institutions or organisations. This has to be conducted in a well-organised, controlled, and managed way. In this context, several software solutions are used for the sharing of the data, like cloud-based solutions (e.g. Dropbox, iCloud, Google Drive, Google Docs), Virtual Research Environments (VRE) (e.g. Microsoft SharePoint), data repositories (e.g. DSpace, Eprints, Fedora, CKAN), as well as FTP (Secure File Transfer) servers (see also section 2.4).

In the framework of data sharing and data access data licensing also has to be considered. Several approaches are available to license research data, which is an instrument to permit re-use under certain constraints, such as a correct citation (BALL, 2013b). They are provided by various institutions and organisations. Examples of standard licences include the Creative Commons licence (CREATIVE COMMONS, 2014), the Open Data Commons licence (OPEN DATA COMMONS, 2014), and the Open Government Licence (NATIONAL ARCHIVES, 2014). In addition, some data repositories or RDM systems of research projects have developed their own licences, policies or provide terms and conditions for the access of their data collections. For instance, the UK Data Archive (UK DATA SERVICE, 2014a) provides terms and conditions to access their data, whereas the Archaeology Data Service requests a deposit (ARCHAEOLOGY DATA SERVICE, 2014). Furthermore, data policies or data agreements are used by the RDM systems FOR816dw data portal (RU816, 2007), TEODOOR portal (TERENO, 2014), PANGAEA data portal (GROBE, 2005), and the TR32DB data portal (CRC/TR32, 2011).

2.2 Introduction to metadata

Metadata are used for the documentation of (research) data and are, for example, very important for the re-use, search, and understanding of the described data. This chapter describes the documentation of research data by means of metadata. Initially, a short introduction to the basic principles of metadata will be presented (section 2.2.1). Subsequently, available metadata standards and schemas will be described (section 2.2.2). Finally, the importance of metadata for research data will be outlined (2.2.3).

2.2.1 Metadata

Metadata definition and function

Metadata is quite often simply described as *(structured) data about data*, *information about data* or *information about information* (BACA, 2008; FOULONNEAU & RILEY, 2008; HAYNES, 2004; HIDER, 2012; MILLER, 2011; VELLUCCI, 2000; ZENG & QIN, 2008). CAPLAN (2003) and GREENBERG (2005) state that the term *metadata* itself dates back to the 1960s, when Jack E. Myers coined it. In 1986, Jack E. Myers registered the term METADATA (spelled in all capitals or just an initial capital) as an U.S. trademark for his Metadata Company. Thereupon, the terms *metadata*, *meta data* and *meta-data* used in various spellings (all in lower case) were adopted by different computer and information science communities. With the growth of the new digital information technologies and resources by the early 1990s, the need and use of, and interest in metadata increased in a wide range of communities as well (SEN, 2004). For instance, HAYNES (2004) points out that the GIS community significantly participated in developing metadata standards to facilitate system interoperability in the late 1980s and early 1990s. Furthermore, the term metadata was adopted by the web community in the mid-1990s with the purpose to establish a metadata standard to describe web content. From this initial action, the Dublin Core Metadata Initiative (DCMI) evolved.

MILLER (2011) and HAYNES (2004) comment that the term metadata itself fragments into two words. The Greek prefix *meta* is specified with the meaning of *about*, *relating to*, *based on*, or *after*. The second word *data* has a Latin origin. This refers to information that can be processed by computers, as well as any kind of resource or information object composed of data also applying to non-digital data. In summary, metadata can be characterized as data describing data or information.

Metadata is used by several fields and stakeholders and has evolved from different communities with diverse discipline-specific objectives and backgrounds. Consequently, it is difficult to define it (CAMPBELL, 2005). YEUNG & HALL (2007a) describe metadata in general as

> "[…] a formally structured and documented collection of information about data that reveals minimally what is in the data, where the data originated from, who produced them, when they were produced and modified, why they were produced, and how the data can be obtained." (YEUNG & HALL, 2007a: 156).

GREENBERG (2003a) complements that metadata are "[…] structured data about an object that supports functions associated with the designed object." (GREENBERG, 2003a: 1876). The term structured data implies "[…] a systematic ordering of data according to a metadata schema specification […]" (GREENBERG, 2003a: 1876) and the object is specified as "[…] any entity, form, or mode for which contextual data can be recorded." (GREENBERG, 2003a: 1876). Any object or information resource can be described, located or explained with metadata. Consequently, it is easier to use, manage or retrieve the resource or object (NISO, 2004), to gather, find and maintain it over a long period of time as well. In addition, metadata help to connect related resources and to promote the resource and, consequently, reach the interested target audience (LUBAS et al., 2013). Thus, metadata is crucial to ensure that data or resources will be useful and accessible in the future (NATIONAL SCIENCE BOARD, 2005; NISO, 2004).

Another definition for metadata is provided by COYLE (2010). COYLE (2010) outlines that metadata are constructed, constructive, and actionable. Constructed metadata are a completely artificial invention and cannot be found in nature. Constructive metadata are created for a particular activity, a certain purpose or to solve a specific problem. Actionable metadata are intended to be useful in a particular way. Consequently, COYLE (2010) points out that good, functional metadata creation is significantly dependent on the understanding of the possible usage and the requirements that have to be fulfilled by the metadata.

HAYNES (2004) argues that metadata can be applied to describe different resources or information, like electronic/digital resources (e.g. text documents, digital images) or printed documents (e.g. journal, reports, books). The descriptive metadata can be stored separately in a database or it can be integrated in the

information resource. Furthermore, the author specifies that metadata can describe the format, content or attributes of information resources. Refering to scientific data, GRAY et al. (2005) add that "Metadata is the descriptive information about data that explains the measured attributes, their names, units, precision, accuracy, data layout and ideally a great deal more. More importantly, metadata include the data lineage that describes how the data were measured, acquired or computed." (GRAY et al., 2005: 3). WILSON (2014) emphasizes that a spatial component (the *where* element) has to be added to describe data with spatial information or metadata.

To understand the meaning of metadata, MILLER (2011) provides some examples of metadata. Moreover, he adds that everybody has encountered metadata in the daily life, without calling it metadata. For example, the descriptive information of a book can include the author, title, publisher, publishing year or page range. In addition, he cites the product information printed on food packaging. Here, information or details are specified about the ingredients of the food, which can also be described as metadata. Moreover, MILLER (2011) gives examples of metadata for digital files that are automatically generated during the production process of the file and are consequently embedded within the file details. For example, information about a document (e.g. title, author, keywords, subject, number of pages or words) is automatically generated by Microsoft Word as well as technical information about a digital image (e.g. resolution, recording date, colour mode, camera make and model) are automatically generated by a digital camera.

Metadata types and management

The wide use of metadata within various communities has led to a grouping of different types of metadata according to metadata functions. CAPLAN (2003) identifies three main categories of metadata. *Descriptive metadata* serve the purpose to discover, identify, and select resources. Furthermore, details to link, evaluate, and use the resources are included. Usually, descriptive metadata are publicly accessible, for example, in a search engine system. The aim of *administrative metadata* is to facilitate the resource management. This category can be subdivided into the subclasses *rights management*, *preservation*, and *technical metadata*. Details about the creation process and access restriction are included. Consequently, the access to administrative metadata is often restricted to a person in charge. *Structural metadata* indicate the internal structure or organisation of (compound) object. They are required to create a relationship between a set of related digital objects. However, GILLILAND (2008) depicts metadata in five categories to reflect the major issues of metadata functionality. *Administrative metadata* is used to administer and manage resources and objects. *Descriptive metadata* is used to describe and identify objects and related resources. *Preservation metadata* includes details related to an object or resource with focus on preservation management. *Technical metadata* provide information about the metadata's behaviour or system functions. *Use metadata* include details of the objects and resources about the type and level of use. In contrast, HAYNES (2004) introduces a *five-point model* based on the purposes of metadata to categorise the metadata functions. The five purposes of metadata cover: resource description, information retrieval, management of information resources, documenting ownership and authenticity of digital resources, and interoperability.

The metadata creation can be achieved through manual (human) and automated processes (SIMONS & RICHARDSON, 2013). The manual creation of metadata occurs when a person produces metadata to describe a digital resource. GREENBERG (2003b) identifies several classes of persons that are involved in the generation of metadata, e.g. technical metadata creators, professional metadata creators, content creators, or subject/community enthusiasts. Moreover, GREENBERG (2004) describes the automatic generation of metadata that is captured by machines and typically follows the extraction or harvesting approach. Many metadata creation tools have been developed by several metadata projects. In general, they support a specific metadata schema. These tools can be arranged into several categories that contain templates, mark-up tools, extraction tools and conversion tools (NISO, 2004).

Metadata can be managed in various ways. HIDER (2012) provides an overview of six steps that include (i) evaluation, (ii) input (to an information system), (iii) improving (adding or editing), (iv) efficient presentation, and (v) sustainable preservation of the metadata as well as conversion and standardization of metadata to support exchange and sharing of the metadata. Furthermore, MA (2006) introduces six steps that are necessary to implement metadata for a digital (library) project. This involves (i) analysing metadata requirements, (ii) adopting metadata schemas, (iii) creating metadata content, (iv) delivery and access, (v) evaluation of metadata, and (vi) sustaining metadata maintenance. With regard to metadata cataloguing in libraries, HUDGINS et al. (1999) illustrate another approach in planning, implementing, and evaluating metadata for a metadata project. In contrast, HAYNES (2004) outlines that the management of metadata has to be in accordance with the needs of the users and the served purpose. He describes the management of metadata according to the lifecycle of a metadata project in seven stages:

Analysing metadata requirements The main purpose(s) of the metadata has to be determined. This includes knowledge about the resource that has to be described and the employed information system as well as analysis of the user community, and existing standards.

Selecting and developing metadata schemas A metadata schema has to be chosen or developed appropriate to the framework requirements. Factors, like the user community, the nature of the described data, and available standards and conventions have to be considered.

Encoding and maintaining controlled vocabularies The used vocabulary of a field has to be defined and the encoding of the metadata schema. If controlled vocabularies are assigned to use, there are three options: select pre-existing standards to encode particular data types, adopt available external controlled vocabulary (e.g. thesaurus), or create an own controlled vocabulary according to the needs of the project.

Applying metadata (content rules) The metadata are applied to a resource (manually or partially automated). The use of cataloguing rules or conventions ensure the consistent form of the metadata.

Importing metadata (interoperability) The metadata of a resource can be imported from a third party (e.g. other source or repository). This requires quality control and interoperability of the metadata as well as compatibility of the used metadata standards. Mapping between the different schemas is provided by means of crosswalks and metadata registries.

Quality management The quality checking of the metadata (metadata content and administrative metadata) ensures that the metadata is accurate, consistent, reliable, and complete. Secure storage of metadata in a regular backed up system ensures the physical integrity.

Search aids and user education A successful, user-friendly presentation of the metadata improves their usability. For example, search and navigation facilities should fit to user needs. The metadata entry can be simplified by controlled vocabulary in conjunction with a drop-down list or a navigable classification schema. Users have to be led to the handling of metadata and the system operations.

Within the management of metadata, the use of metadata standards has an important role (DAY, 2005; GRANELL et al., 2009). In the following chapter, metadata schemas and standards will be described in detail.

2.2.2 Metadata schemas and standards

Metadata schemas (also called schemes) and standards form a vital requirement for the success of management systems, e.g. for digital preservation. They enable exchange and representation of metadata and ensure interoperability, compatibility, and integration (NEUROTH et al., 2010; VETTERLI et al., 2000). Metadata schemas and standards are developed for a specific purpose by a consultative process of experts from a special discipline. They can range from the purely local level (e.g. for a specific domain, institution, single collection, or particular project), widely shared from the national to the international level, according to

YEUNG & HALL (2007b). The terms schema, scheme and standard are often used interchangeably in general linguistic usage (HIRWADE, 2011; ZENG & QIN, 2008). In this thesis, the terms metadata schema and metadata standard will be used.

Metadata schemas or standards consist of a set of elements provided with the purpose of describing a particular type of information resource (NISO, 2004). According to CAPLAN (2003), metadata schemas can specify three aspects of metadata: semantics, content rules, and syntax. Semantics refer to the definition or meaning of the included metadata element and their refinements. Each element is specified with an explanatory name and definition and optional comment or notes. The metadata schema indicates whether a metadata element is repeatable or not and it determines whether an element is mandatory, optional or conditionally required. Content rules describe the constraints and guidelines for the selection and representation of metadata element values (attributes). For example, the types of values can be defined (e.g. date, text, number) and the standard format of the value. The syntax of the metadata schema describes the encoding of the metadata elements (e.g. in XML) with the purpose of providing an exchange format for interchanging metadata. Semantics, content rules, and syntax are not characterized by every metadata schema.

In contrast to a metadata schema, a metadata standard has been formally approved by an institution (SIMONS & RICHARDSON, 2013). There are a couple of national and international agencies, organisations, and initiatives, which are engaged in the standardisation of metadata, such as the International Organisation for Standardization (ISO), the American National Standards Institute (ANSI), the Open Geospatial Consortium (OGC), the World Wide Web Consortium (W3C), and the Dublin Core Metadata Initiative (DCMI) (YEUNG & HALL, 2007b).

Various communities, user environments, and disciplines have created a variety of different metadata-related standards and schemas to serve a specific purpose (HIRWADE, 2011). The multiplicity of available metadata standards and schemas is clearly visible in a glossary provided by RILEY (2009a). Metadata-related standards can be categorised in a fourfold typology (BOUGHIDA, 2005; ELINGS & WAIBEL, 2007; GILLILAND, 2008; MILLER, 2011; ZENG & QIN, 2008). The following categories are included: data structure standards (metadata element sets, metadata schemas; e.g. Dublin Core Metadata Element Set), data value standards (controlled vocabularies/lists, thesauri; e.g. Getty Thesaurus of Geographic Names), data content standards (cataloguing codes and rules, e.g. Anglo-American Cataloguing Rules), and data format/technical interchange/encoding standards (metadata schema encoding format; e.g. Extensible Markup Language). In contrast, RILEY (2009b) visualises the metadata universe according to the domain (e.g. datasets, geospatial data), community (e.g. archives, libraries), function (e.g. markup languages, controlled vocabulary), and purpose (e.g. preservation metadata, descriptive metadata). In addition, the *Disciplinary Metadata Catalogue*, provided by the UK Digital Curation Centre, lists a variety of available discipline-specific metadata standards as well as profiles, use cases, and tools. All resources are grouped according to discipline-specific metadata subject areas including Biology, Earth Science, Physical Science, Social Science& Humanities, and General Research Data (BALL, 2013a; DCC 2014c).

A single metadata standard or schema generally does not fulfil the specialized local needs and entire metadata requirements of a community, user group, repository, or application. To accommodate the needs and requirements, the establishment of an own adjusted metadata schema or *Application Profile (AP)* is a typical approach (HIRWADE, 2011). According to DUVAL et al. (2002), an AP combines or adapts metadata elements of one or more existing metadata schemas according to functional requirements and needs, while retaining interoperability with the used schemas. Consequently, an AP consists of an assemblage of metadata elements from the chosen metadata schemas (*mix and match*) that are combined to one new metadata schema. In addition, the characteristics of AP's are highlighted by HEERY & PATEL (2000). AP's (i) can use elements from one or more existing namespaces, (ii) cannot create new elements that are not specified in existing namespaces, (iii) may define permitted schemas or values (e.g. controlled vocabulary, formats), and (iv) can refine standard definitions.

The mapping of metadata describes the process of analysing and comparing two or more different metadata schemas (MILLER, 2011). In detail, the metadata elements, syntax, and semantics of the metadata schema are mapped (NISO, 2004). As described by WOODLEY (2008), the result of the mapping process is a textual and visual product, a so-called *crosswalk*. This crosswalk shows the equivalencies and relationships between the metadata elements in a chart or table. According to ZENG & QIN (2008), metadata crosswalks are a common method to enable interoperability among and between metadata schemas. In addition, metadata can be effectively converted by a system from one metadata schema to another. Several metadata crosswalks have been developed by different communities between various metadata schemas (CHANDLER et al., 2000; DAY, 2002; GODBY et al., 2004; NOGUERAS-ISO et al., 2004; WIRA-ALAM et al., 2012).

Metadata schema design and documentation

The documentation of the established metadata schema or application profile for a digital collection or repository is an essential key aspect within the design (MILLER, 2011). This documentation, also called data dictionary (LITWIN & ROSSA, 2011) or metadata usage guideline (FOULONNEAU & RILEY, 2008), is constructed and organised in many different ways. These may range from general to specific, informal to formal, or from a simple table to a very detailed description. In addition, the crucial purpose of schema documentation is depicted by MILLER (2011). Schema documentation serves as: (i) a guideline for the metadata creator, (ii) a record of local practices and decisions in the present and future, (iii) a resource for migration of metadata (e.g. to new software systems in the future, metadata harvesting), and (iv) further use in other institutions. Furthermore, MILLER (2011) states that the design of the schema/application profile varies immensely according to different aspects (e.g. scope and type of digital content that has to be covered, size of the digital collection, etc.). It is dependent on the understanding of the underlying, used metadata schema or standard. In addition, metadata schema design should ideally be performed simultaneously with the creation of the user interface (e.g. search, browse, display of metadata). Suggestions on building a metadata schema and where to start are provided by NISO (2008). According to MILLER (2011), the design of a metadata schema comprises the following steps:

Analysis of context, content, and users, as well as identification of functional requirements Analysis of the context within which the metadata schema has to be designed (e.g. institution, organisation). Determine the content (e.g. type, theme, and focus of digital objects) of the digital collection and the user groups and their needs. Identify a set of functional requirements according to the needs and requirements of the users that have to be performed by the user-interface.

Selection and development of the metadata element set Select one or more standardized metadata element set/schema according to the required needs and functions and/or develop an own element set. Two methods are available for the metadata design:

- A **general, cross-collection element set** applies to all collections of an institution. An existing, standardized element set is adapted in whole or in parts with/without additional elements from another element set or/and local established elements.
- A **collection-specific element set** applies to a set of local elements developed for a specific collection according to particular needs and functional requirements. After selecting an underlying base schema, selected elements are renamed with local collection-specific names and optionally complemented with other elements.

Establishment of specifications Specifications have to be developed for each metadata element. Commonly, the following specifications are included: obligation (metadata element is required / mandatory, required if applicable/available, optional/recommended), cardinality (also referred to as repeatability or occurrence of a metadata element), data value specification (encoding schema, controlled vocabularies), designation of field size, visibility, as well as searchability, indexing, and/or hyperlinking of the elements values.

Establishment of encoding schema and controlled vocabularies Select controlled vocabularies (e.g. thesaurus for a subject) and syntax encoding schemas (e.g. encoding for a date) for the values of individual elements. If necessary, own local vocabularies have to be developed to meet local needs.

Development of content guidelines Content or input guidelines have to be developed to specify the form of data content, e.g. handling of abbreviations, capitalization, acronyms, geographic names, missing or incomplete information.

Documentation of the schema The documentation of the metadata schema or application profile is the final and critical step. This includes for example the description of the element set, controlled vocabularies, encoding schemas, content guidelines, and specifications.

In the following, recent metadata standards, schemas, and directives are described that are used within this thesis.

2.2.2.1 Dublin Core

Dublin Core (DC) is one of the widest and most frequently used metadata schema in digital repositories (SIMONS & RICHARDSON, 2013), developed by the Dublin Core Metadata Initiative (DCMI, 2014). The initial *Dublin Core Metadata Element Set (DCMES)* was discussed and designed at a jointed Metadata Workshop of the Online Computer Library Centre/National Centre for Supercomputing Applications in Dublin (Ohio, USA) in March 1995, with the purpose of defining a simple, usable core set of metadata elements able to describe network accessible resources (CAPLAN, 1995). As the final workshop result, a set of thirteen metadata elements was proposed that was kept as small as possible, easy to understand by users, and flexible to describe digital resources from a variety of subject areas. In addition, the (further) development of the metadata elements follows principles like intrinsic, syntax independence, extensibility, repeatability, optionality, and modifiability (WEIBEL, 1995). Over the years, the initial developed DCMES has changed. The original elements were expanded to fifteen in DCMES version 1.1 (known as *simple* Dublin Core). There are no disclosures made by the DCMI concerning specifications, requirements or repeatability. Consequently, all elements may be used in any order, are repeatable as well as optional (MILLER, 2011). The DCMES Version 1.1 is approved in the standards documents ISO Standard 15836:2009 (ISO, 2014a), ANSI/NISO Standard Z39.85-2012 (NISO, 2012), and the Internet Engineering Task Force (IETF) Standard RFC 5013 (DCMI, 2012b; NETWORK WORKING GROUP, 2007) and is officially documented in detail at `http://dublincore.org/documents/dces/`. A listing of the fifteen DCMES elements (properties) is provided in Table 2-1 including the property name, definition, and comment according to DCMI (2012b).

The designation *elements*, initially used in the *simple* Dublin Core, was renamed to *properties*. In combination with guidelines on usage of controlled vocabularies and encoding schemas, the properties form one part of the *DCMI Metadata Terms* vocabulary set. The DCMI Metadata Terms enables the description of a resource and define partially values that may be used (LUBAS et al., 2013). An official, detailed documentation of the DCMI Metadata Terms is provided at `http://dublincore.org/documents/dces/`. According to DCMI (2011), the DCMI Terms distinguish four types of terms:

Properties resp. **Elements** are the *core* attributes of a resource that enable structured, uniform resource description. Formerly, they were known as *elements* in *simple* Dublin Core.

Classes group resources that have a certain property in common and consequently are put together in one concept. The DCMI Type Vocabulary is used in many cases for this purpose.

Datatypes resp. **Syntax Encoding Schemas** refers to the rules or standard that specifies the structure of a value.

Vocabulary Encoding Schema resp. **Concept Schema** defines the controlled vocabulary that should be used.

Table 2-1: Listing of the Dublin Core Metadata Element Set Version 1.1, modified after DCMI (2012b)

1.	contributor:	An entity responsible for making contributions to the resource.
		Comment: Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
2.	coverage:	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
		Comment: Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names (TGN). Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.
3.	creator:	An entity primarily responsible for making the resource.
		Comment: Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
4.	date	A point or period of time associated with an event in the lifecycle of the resource.
		Comment: Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding schema, such as the W3CDTF profile of ISO 8601 (W3CDTF).
5.	description:	An account of the resource.
		Comment: Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
6.	format:	The file format, physical medium, or dimensions of the resource.
		Comment: Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types (MIME).
7.	identifier:	An unambiguous reference to the resource within a given context.
		Comment: Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.
8.	language:	A language of the resource.
		Comment: Recommended best practice is to use a controlled vocabulary such as RFC 4646 (RFC4646).
9.	publisher:	An entity responsible for making the resource available.
		Comment: Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
10.	relation:	A related resource.
		Comment: Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
11.	rights:	Information about rights held in and over the resource.
		Comment: Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.
12.	source:	A related resource from which the described resource is derived.
		Comment: The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
13.	subject:	The topic of the resource.
		Comment: Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.
14.	title:	Title
		Comment: A name given to the resource.
15.	type:	The nature or genre of the resource.
		Comment: Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary (DCMITYPE). To describe the file format, physical medium, or dimensions of the resource, use the Format element.

All DCMI Metadata Elements and Terms are given a unique identity within the DCMI namespace. MILLER (2011) defines a namespace as "[…] an identifier of a virtual space in which the names of terms are officially defined." (MILLER, 2011: 49) The fifteen elements (properties) of the *simple Dublin Core* are defined in the DCMI namespace `http://purl.org/dc/elements/1.1/` and are abbreviated with the prefix *dc*. The properties of the Metadata Terms are defined in the DCMI namespace `http://purl.org/dc/terms/` and are shortened with the prefix *dcterms*. In addition, the properties of the DCMI Type Vocabulary are defined in the DCMI namespace `http://purl.org/dc/dcmitype/`.

Each term in the DCMI Terms vocabulary set is described with a minimal set of attributes including name, URI, label, definition, and type of term. Further attributes can be provided where applicable; for example: comment, references, refines, has range or version. An example for the description of the property `title` is shown in Table 2-2.

Table 2-2: DCMI Metadata Terms description of property `title` in the `/terms/namespace`, modified after DCMI (2012a)

Term Name	Title
URI:	http://purl.org/dc/terms/title
Label:	Title
Definition:	A name given to the resource.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/title
Version:	http://dublincore.org/usage/terms/history/#titleT-002
Has Range:	http://www.w3.org/2000/01/rdf-schema#Literal

An overview of the DCMI Metadata Terms (e.g. properties, vocabulary encoding schemas, syntax encoding schemas) is provided by DCMI (2012a).

2.2.2.2 DataCite Metadata Schema

The *DataCite Metadata Schema for the Publication and Citation of Research Data* is created by the members of the Metadata Working Group of the DataCite Consortium (DATACITE, 2014a). The schema is designed with the purpose of supporting the goals of DataCite to register DOIs (section 2.1.6). Moreover, the objectives of the metadata schema include: (i) a recommendation of a standard citation format for datasets, which is based on a small number of required properties for the identifier registration, (ii) a provision of interoperability with other metadata schemas for data management, (iii) a promotion of dataset discovery by using optional extra properties that enable a flexible resource description, and (iv) an application of controlled terms from DataCite and external vocabulary with regard to future services (e.g. discovery) (STARR & GASTL, 2011).

STARR & GASTL (2011) state that the first two DataCite Metadata Schema drafts were based on a basic metadata schema used by the German National Library of Science and Technology (TIB). The basic metadata schema of the TIB was developed in the framework of the DFG-funded project "Publication and Citation of Scientific Primary Data (STD-DOI)" (BRASE, 2004). The STD-DOI Metadata Schema is described in detail by KLUMP (2009). In July 2013, the most recent DataCite Metadata Schema Version 3.0 (DATACITE METADATA WORKING GROUP, 2013) was released. In this thesis, the DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011) released in July 2011 was applied.

The DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011) describes the core metadata properties which are selected to identify data in a consistent and accurate manner with the

purpose of citation and retrieval. In addition, recommended instructions of usage are provided. In the following, the Metadata Schema will be described in detail.

First of all, an overview of the entire DataCite Metadata Properties listed in a simple manner according to mandatory, optional, and administrative properties is provided in Table 2-3. The mandatory properties must be and the optional properties may be provided. The administrative properties are assigned to each metadata record by the managing agency responsible for the DOI registration.

Table 2-3: DataCite Metadata Properties listed according to mandatory, optional, and administrative properties, modified after DataCite Metadata Working Group (2011)

ID	Property
DataCite Mandatory Properties	
1	Identifier (with type attribute)
2	Creator (with name identifier attributes)
3	Title (with optional type attribute)
4	Publisher
5.	Publication Year
DataCite Optional Properties	
6	Subject (with schema attribute)
7	Contributor (with type and name identifier attributes)
8	Date (with type attribute)
9	Language
10	ResourceType (with description attribute)
11	AlternateIdentifier (with type attribute)
12	RelatedIdentifier (with type and relation type attributes)
13	Size
14	Format
15	Version
16	Rights
17	Description (with type attribute)
DataCite Administrative Metadata	
0.1	LastMetadataUpdate
0.2	MetadataVersionNumber

In addition, the DataCite Metadata Schema provides two recommended formats that should be used as DataCite citation. The first includes only the first five mandatory metadata properties, as follows:

Creator (PublicationYear): Title. Publisher. Identifier

In addition, a second citation is recommended, including two optional properties, as follows:

Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier

For example:

Waldhoff, Guido (2012): Enhanced Land Use Classification of 2009 for the Rur catchment. CRC/TR32 Database (TR32DB). doi: 10.5880/TR32DB.2

The DataCite mandatory and optional properties are then described in detail in tables in the metadata schema documentation. An example for the property `Title` is shown in Table 2-4. Each metadata property is described with several characteristics, for example the identifier (ID) and the DataCite property name. The property name (`Title`) is represented with a capital letter, whereas the corresponding attributes

(titleType) and child properties are represented with a lower case letter. In addition, a short definition of the property is provided. The relationship between the property and the associated attribute (A) or child (C) is specified in the column A/C. The frequency of occurrence (Occ) of each property or attribute/child is defined as follows: at least one is required and unlimited repeatable (1-n), at least one is required and not repeatable (1), optional and maximum one is required (0-1), and optional and unlimited repeatable (0-n). An attribute/child is required for a property, when the occurrence is provided with the notation Req. The notation Opt advises that the attribute/child is optional. Finally, allowed values, examples, or other constraints are specified (e.g. controlled lists, format, encoding).

Table 2-4: Description of the DataCite property title, modified after DATACITE METADATA WORKING GROUP (2011)

ID	DataCite-Property	Definition	A/C	Occ	Allowed values, examples, other constraints
3	Title	A name or title by which a resource is known.		1-n	The format is open.
3.1	titleType	The type of Title.	A	Opt	Controlled List Allowed values: Alternative, Title Subtitle, TranslatedTitle

At the end of the metadata schema, references to the corresponding DataCite Metadata Schema in XML are provided, including an example. Moreover, mappings from the DataCite Metadata Schema to other metadata elements are provided, for example, to simple Dublin Core and DCMI Metadata Terms.

2.2.2.3 ISO 19115:2003 Metadata Standard

The international standard, *ISO 19115:2003 Geographic information – Metadata*, was released as a standard in 2003 by the International Organization for Standardization (ISO). A corrigendum was published in 2006 as ISO 19115:2003/Cor 1:2006 (ISO, 2003, 2006). In 2005, ISO 19115:2003 was adopted by the European Committee for Standardization as EN ISO 19115:2005 (CEN, 2014a). According to NOGUERAS-ISO et al. (2005), the ISO 19115 standard defines a schema required for describing geographic services and information, for example, identification of spatial datasets, data quality, scope and coverage, spatial reference, and distribution of the data. The standard is applicable to describe geographic datasets and series, as well as single geographic features and their properties.

The entire metadata standard defines more than 400 metadata elements, including optional, mandatory, and conditional elements (YEUNG & HALL, 2007a). They are grouped in two levels according to ISO (2003) and KRESSE & FADAIE (2004): the 22 *core* metadata elements and the so-called *comprehensive* metadata elements. The core metadata element set (Table 2-5) contains mandatory (M), recommended optional (O), and conditional (C, mandatory under a certain condition) elements. They are at least required to describe and identify a dataset as well as to serve metadata applications, like data access, discovery, or transfers, and catalogue purposes. The comprehensive metadata elements cover the full list of metadata elements, defined in the standard.

ISO 19115:2003 is arranged in a hierarchical structure. The metadata schema is described in form of Unified Modelling Language (UML) diagrams in combination with a tabular data dictionary. Both together serve as a full description of the total metadata abstract model (LITWIN & ROSSA, 2011: 69ff). The data dictionary will be described in more detail later in the chapter.

The metadata elements of ISO 19115:2003 are divided in thematic packages that describe metadata sections. Each metadata section (equivalent to an UML package) contains one or more metadata entities (equivalent to UML classes). The metadata entities contain one or more metadata element (equivalent to UML class attributes). To meet the mandatory user or standard requirements, the metadata entities can be aggregated

Table 2-5: Core metadata elements of ISO 19115:2003 for geographic datasets with obligation and declaration of the specifying metadata entity and element, modified after ISO (2003)

Core Metadata (with Obligation)	Path (Metadata Entity and Element)
Dataset title (M)	MD_Metadata > MD_DataIdentification.citation > CI_Citation.title
Spatial representation type (O)	MD_Metadata >MD_DataIdentification.spatialRepresentationType
Dataset reference date (M)	MD_Metadata > MD_DataIdentification.citation > CI_Citation.date
Reference system (O)	MD_Metadata > MD_ReferenceSystem
Dataset responsible party (O)	MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty
Lineage (O)	MD_Metadata > DQ_DataQuality.lineage > LI_Lineage
Geographic location of the dataset (by four coordinates or by geographic identifier) (C)	MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or EX_GeographicDescription
On-line resource (O)	MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource
Dataset language (M)	MD_Metadata > MD_DataIdentification.language
Metadata file identifier (O)	MD_Metadata.fileIdentifier
Dataset character set (C)	MD_Metadata > MD_DataIdentification.characterSet
Metadata standard name (O)	MD_Metadata.metadataStandardName
Dataset topic category (M)	MD_Metadata > MD_DataIdentification.topicCategory
Metadata standard version (O)	MD_Metadata.metadataStandardVersion
Spatial resolution of the dataset (O)	MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance
Metadata language (C)	MD_Metadata.language
Abstract describing the dataset (M)	MD_Metadata > MD_DataIdentification.abstract
Metadata character set (C)	MD_Metadata.characterSet
Distribution format (O)	MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version
Metadata point of contact (M)	MD_Metadata.contact > CI_ResponsibleParty
Additional extent information for the dataset (vertical and temporal) (O)	MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent or EX_VerticalExtent
Metadata date stamp (M)	MD_Metadata.dateStamp

Table 2-6: Description of the metadata element `title`, part of the ISO19115:2003 metadata entity `CI_Citation`, modified after ISO (2003, 2006)

	Name / Role Name	Short Name	Definition	Obligation/ Condition	Maximum occurrence	Data type	Domain
360.	title	resTitle	name by which the cited resource is known	M	1	Character String	Free text

and repeated, as well as related to one or more entities. The metadata elements or entities can be mandatory, optional or conditional (ISO, 2003; LITWIN & ROSSA, 2011: 75ff). In addition, metadata entities are in relationship to or taken from other standards of the ISO 19100 family (KRESSE & FADAIE, 2004:90).

The characteristics of the metadata are described in detail in the data dictionary presented in a table. Each metadata entity and element is defined by seven attributes. This includes: (i) the name or role name assigning the label of the entity or element (entity names start with upper case letter), (ii) a short name of the entity or element, unique within the standard, (iii) a describing definition of the entity or element, (iv) the obligation or condition (M, C, O) to document the entity or element, (v) the maximum occurrence describing the maximum number of instances, (vi) the data type specifying the set of distinct values, (vii) and finally the domain indicating the covered line number of an entity or the allowed values of an element. This value can be limited to a data type or a code specified list. (ISO, 2003). An example for the description of the element `title` is shown in Table 2-6. The element `title` belongs to the metadata entity `CI_Citation`, described in the metadata section `Citation and responsible party information`.

2.2.2.4 INSPIRE metadata

Since 2007, the EU-Directive *INSPIRE (Infrastructure for Spatial Information in Europe, Directive 2007/2/EC of 14 March 2007)* (EUROPEAN COMMISSION, 2007a) of the European Parliament and of the Council is in force. This directive has the objective to determine a general regulation to establish an infrastructure for spatial information in the European Union with regard to community environmental politics and other activities with focus on the environment. To ensure that this target is achieved by the member states, the directive requires the establishment of common *Implementing Rules (IR)* of specific areas (e.g. Metadata, Data Specification, Network Services, Data and Service Sharing, Monitoring and Reporting). Moreover, 34 spatial data themes needed for environmental applications are defined in Annex I to III (EUROPEAN COMMISSION, 2007a). The INSPIRE-Directive is implemented into German law by the *Gesetz über den Zugang zu digitalen Geodaten (Geodatenzugangsgesetz, GeoZG)* (DEUTSCHER BUNDESTAG, 2009) and into federal German law in North Rhine-Westphalia by the *Gesetz über den Zugang zu digitalen Geodaten Nordrhein-Westfalen (Geodatenzugangsgesetz, GeoZG NRW)* (LANDESREGIERUNG NRW, 2009).

The crucial and priority role of metadata in the INSPIRE-Directive is emphasized by LITWIN & ROSSA (2011). Within this directive, the EUROPEAN COMMISSION (2007a) defines metadata according to Article 3, Section 6 as "[…] information describing spatial data sets and spatial data services and making it possible to discover, inventory and use them;" (EUROPEAN COMMISSION, 2007a: 5). Moreover, Chapter II of the directive is especially dedicated to the topic metadata. This determines that member states of the European Union are demanded to provide metadata for available spatial data sets and services corresponding to the defined spatial data themes. Details about the suggested content of the metadata are also specified. In addition, member states have to ensure that the metadata are complete and of sufficient quality. They are also requested to adopt Implementing Rules that should "[…] take account of relevant, existing international standards and user requirements, in particular with relation to validation metadata." (EUROPEAN COMMISSION, 2007a: 6). Finally, a schedule is described for the creation of the metadata according to the specified spatial data themes in the Annex.

As a consequence of the INSPIRE-Directive, the Implementing Rules for metadata are defined in two documents:

- Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata
- INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (ISO adopted by CEN)

The Commission Regulation No 1205/2008 (EUROPEAN COMMISSION, 2008), adopted and published in 2008, formally describes and enforces the implementation of the INSPIRE-Directive. Requirements to create and maintain metadata for spatial data sets, spatial data set series, and spatial data services referring to the mentioned spatial data themes are presented. The metadata implementation rules are described in the Annex of the Commission Regulation. Part A of the Annex defines the mainly used terms and interprets the validity of spatial data sets. Part B of the Annex lists and accurately describes 27 metadata elements in total, for example, including definition and value domain. These are grouped in ten categories (Table 2-7). Part C of the Annex discusses instructions for the metadata elements in terms of multiplicity and conditions. These are presented in two tables. One table summarises the metadata elements or groups related to spatial data services (including multiplicity and conditions). The other table is referring to metadata elements for spatial data sets or spatial data set series, presented in Table 2-7. In detail, the table includes the reference number, the name, the multiplicity, and a conditional statement of the metadata element. In addition, the table was extended with domain information specified for each metadata element from Part B of the Annex and, in addition from the Metadata Implementing Rules (see below in this chapter). Part D of the Annex determines six value domains referred to in the metadata element description in Part B that shall be used, for example: resource type, topic categories in accordance with EN ISO 19115, degree of conformity, or responsible party role (EUROPEAN COMMISSION, 2008). In summary, LITWIN & ROSSA (2011) state that the Commission Regulation meets the requirements of the INSPIRE-Directive. The metadata are specified at a minimum range. Consequently, resources can only be described at a very basic level.

The INSPIRE Metadata Implementing Rules define the technical guidelines that are necessary to implement the Commission Regulation No 1205/2008 using EN ISO 19115 and EN ISO 19119. The document was created and published in 2007 as a first version by the INSPIRE Metadata Drafting Team and the European commission Joint Research Centre. This first version was revised in 2008, 2009, 2010, and 2013. Further changes to the document are already announced. In this thesis, the INSPIRE Metadata Implementing Rules, Version 1.2 (DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE, 2010) published in 2010 were applied.

The INSPIRE Metadata Implementing Rules (IR) describe the practical, technical implementation of the INSPIRE-Directive. First of all, a list of indispensable, related documents is provided that have to be considered by using the IR, for example EN ISO 19115:2005 for Geographic Information – Metadata (CEN, 2014a), ISO 19119:2005 for Geographic Information – Services (ISO, 2005), EN ISO 19108:2005 for Geographic Information – Temporal Schema (CEN, 2014b), or ISO 639-2:1998 for Codes for the representation of names of languages – Part 2: Alpha-3 code (ISO, 2010), ISO 8601:2004 for Data elements and interchange formats – Information interchange – Representation of dates and times (ISO, 2013), ISO/TS 19139:2007 for Geographic Information – Metadata – XML Schema Implementation (ISO, 2014b) as well as associated corrigendum or extension of the documents. In addition, the IR involves three sections and a complementary annex which will be described in the following.

The first section describes the INSPIRE profile of ISO 19115 and ISO 19119. At first, the requirements of ISO 19115 Core are compared to the requirements of INSPIRE subdivided in spatial datasets and spatial dataset series as well as spatial data services. As a conclusion, it is stated in the IR that conformity to ISO 19115 metadata does not guarantee compliance with INSPIRE metadata requirements. However, applying INSPIRE metadata prevents conflicts with ISO 19115 metadata. Full compliance of INSPIRE metadata with ISO 19115 will be achieved by adding additional metadata elements.

The second section presents the basic mapping between INSPIRE metadata elements and the metadata elements defined in the ISO 19115/ISO 19119. For each INSPIRE metadata element, the mapping is summarized in a table, as shown in Table 2-8 for the metadata element `Resource Title`. The characteristic of the IR include the particular reference number to the paragraph in the Commission Regulation No 1205/2008 and the defined name of the metadata element. The obligation/condition and the multiplicity of the metadata element are specified.

Table 2-7: Metadata for spatial data sets and spatial data set series,modified after EUROPEAN COMMISSION (2008) and DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE (2010)

Multiplicity: 1 = mandatory, single-valued; 1..* = mandatory, multi-valued; 0..* = conditional, multi-valued

Ref.	Metadata elements	Multiplicity	Domain	Condition
1. Identification				
1.1	Resource title	1	free text	
1.2	Resource abstract	1	free text	
1.3	Resource type	1	defined in Part D.1	
1.4	Resource locator	0..*	character string	Mandatory if URL is available to obtain more details on resource.
1.5	Unique resource identifier	1..*	character string	
1.7	Resource language	0..*	conformity with ISO 639-2	Mandatory if resource includes textual information.
2. Classification of Spatial Data				
2.1	Topic category	1..*	defined in Part D.2	
3. Keyword				
	Keyword value	1..*	free text	
	Controlled vocabulary	0..*	controlled vocabulary title and reference date	Shall be provided, when controlled vocabulary is used
4. Geographic Location				
4.1	Geographic bounding box	1..*	west/eastbound longitudes, south/northbound latitudes (decimal degrees)	
5. Temporal reference				
5.1	Temporal extent	1..*	conformity with ISO 19108 and ISO 8601	At least one of the metadata elements shall be provided
5.2	Date of publication			
5.3	Date of last revision			
5.4	Date of creation			
6. Quality and Validity				
6.1	Lineage	1	free text	
6.2	Spatial resolution	0..*	integer (equivalent scale); numerical value associated with length unit (distance)	Mandatory for data sets if an equivalent scale or a resolution distance can be specified
7. Conformity				
	Specification	1..*	title and reference date	
	Degree	1..*	defined in Part D.5	
8. Constraints related to access and use				
8.1	Conditions for access & use	1..*	free text	
8.2	Limitations on public access	1..*	free text	
9. Organisation responsible for the establishment, management, maintenance & distribution of spatial data				
9.1	Responsible party	1..*	free text/character string	
9.2	Responsible party role	1..*	defined in Part D.6	
10. Metadata on Metadata				
10.1	Metadata point of contact	1..*	free text/character string	
10.2	Metadata date	1	conformity with ISO 8601	
10.3	Metadata language	1	conformity with ISO 639-2	

Table 2-8: Mapping of the metadata element `Resource title` of INSPIRE Metadata Implementing Rules (IR) and ISO 19115, modified after DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE (2010)

	Reference	Part B 1.1
IR	Element name	Resource title
	Obligation / condition	Mandatory
	Multiplicity	[1]
ISO 19115	Number	360
	Name	title
	Definition	Name by which the cited resource is known.
	XPath	identificationInfo[1]/*/citation/*/title
	Data type	CharacterString
	Domain	Free text
	Example	Image2000 Product 1 (nl2) Multispectral
Implementing instructions		None

Furthermore, the characteristics of ISO 19115 include the identifier number, the name, definition, the location, the data type, the domain value, and an example of the metadata element (Table 2-8). Some implementing instructions are provided, if necessary. In addition, for each metadata element an example is given in XML encoding.

The third section of the IR discusses a detailed mapping of the INSPIRE metadata elements and the metadata elements of ISO 19115 and ISO 19119. The result is displayed in a set of templates. Finally, the Annex of the IR specifies the XML encoding of the metadata element according to ISO/TS 19113 based on ISO 19115 and ISO 19119. Furthermore, instructions and examples for XML metadata sets are provided.

2.2.2.5 Ontology based metadata: Bibliographic Ontology and Event Ontology

The semantic web, as an extension of the world wide web, makes digital (research) data available for applications, instead of just linking documents with content that is not understandable by applications (NEHER & RITSCHEL, 2011). Thus, data and the corresponding relations within a specific domain are accurately and formally described by a shared vocabulary. This is achieved by the development of ontologies, as a model that structures and describes data. The ontologies can be built by different languages, such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL). The specific ontology is than referenced by a persistent URI (Uniform Resource Identifier) and the data is usually provided by XML (NEHER & RITSCHEL, 2011).

The Bibliographic Ontology Specification (BIBO) is the first Web Ontology Language (OWL) designed to describe resources for bibliographic references metadata. In the bibliographic community it is widely used, as stated by PERONI & SHOTTON (2012). According to D'ARCUS & GIASSON (2009), the Bibliographic Ontology provides main properties and concepts to describe bibliographic references and citations on the Semantic Web by using RDF (Resource Description Framework). For example, it can be used to document a classification ontology as a citation ontology or simply to describe all types of documents. The Bibliographic Ontology is designed with regard to existing metadata formats for document description. Moreover, it is pragmatic, straightforward, and expandable and, consequently, intended to be used on a wide scale. In addition, the Bibliographic Ontology can be also combined with other vocabularies like Dublin Core metadata elements or local extensions.

Furthermore, D'ARCUS & GIASSON (2009) outline that the Bibliographic Ontology URI (Uniform Resource Identifier) for the namespace is `http://purl.org/ontology/bibo/`. In detail, the specific content of the namespace and the description of all included classes, object properties, data properties, and individuals are described in the Bibliographic Ontology namespace document (BIBOTOOLS, 2014), such as `bibo:Book`, `bibo:Article`, or `bibo:Journal`.

The Bibliographic Ontology is related to the Event Ontology. The Event Ontology focuses on the notation `event` and the related terms to classify a space/time region. An event model includes location, time, factors, products, and active agents. The namespace URI of the Event Ontology is `http://purl.org/NET/c4dm/event.owl` and describes properties such as `event:place` or `event:time`. (RAIMOND & ABDALLAH, 2007a). In conjunction with the Event Ontology the Time Ontology (HOBBS & PAN, 2006) and the Timeline Ontology (RAIMOND & ABDALLAH, 2007b) are significant, both of which focus on temporal information.

2.2.3 Metadata for scientific research data

The documentation of research data, i.e. metadata, is a necessary key component in the context of research data management (e.g. DAY, 2005; GREENBERG et al., 2013; HIGGINS, 2007; MA, 2006), especially for long-term studies (KARASTI et al., 2010; MICHENER, 2006). LEVINE (2014: 136) states: "Metadata is as significant as the data itself. Producing meaningful metadata and making it available requires resources and expertise." As already mentioned in the previous section, research data should be described with information related to the context, content, structure, quality, source, and accessibility (MICHENER et al., 1997; NATIONAL SCIENCE BOARD, 2005). Furthermore, it is a necessary requirement to document the research data by using extensive, standardised metadata (ENKE & FRITZSCH, 2013; GRAY et al., 2005; QIN et al., 2012). In addition, JENSEN et al. (2011) suggest the use of national or international classification systems (e.g. ISO 639-1:2002 for language representation, Dewey Decimal Classification (DDC) for cataloguing) as well as standardised, controlled vocabulary.

Metadata enables scientists to describe, publish and share their data. Through the use of quality metadata it is possible to interact with the research data. For example, metadata enables browsing and searching for research data according to uncontrolled or restricted keywords, specific categories, a time period, geographic area, or specific metadata fields (e.g. author, title, subject, resource type). Metadata enables scientists to properly cite the data, as well as to use and understand the data in detail (BERMAN, 2008; MATTHEWS et al., 2010; PARK & TOSAKA, 2010). Furthermore, PHILLIPS et al. (2007) and GRAY et al. (2005) supplement that metadata can be leveraged for data analysis and through visualization. Thus, the content of the research data can be accessed more easily, better understood by users and programs, and linkages can be recognised as well. As a consequence, research data with (quality) metadata encourage persistent preservation and long-term re-use by the investigating scientist or further scientists, within or across scientific disciplines (ALLAN, 2011: 48; JENSEN et al., 2011; MICHENER, 2006; WILLIS et al., 2012).

The provision of good quality metadata is essential for the interaction with data or data collections (MILLER, 2011). BOSE & FREW (2005) state that metadata are often collected incomplete or in an unstructured form and then published. A study conducted by the RESEARCH INFORMATION NETWORK (2008) shows that the quality of metadata provided for research data is considerably variable. The range of the quality varies from enhanced, standardized metadata through semi- standardized to ad hoc labelling created by the researchers. Without good quality, insufficient or incomplete metadata or in case of inconsistencies with the file formats, it is not possible to interact with research data. Inadequate documented data will be lost for the science community and for future re-use. Scientist will not be able to discover the existence of these data (GRIFFITHS, 2009). Figure 2-6 illustrates the various processes of loss of information content of data and metadata through time.

To prevent the loss of information content of data and metadata, a management system or software to create extensive metadata documentations has an increased significance. JENSEN et al. (2011) emphasize that these applications should support input, management, and versioning of metadata. The applications should be adapted to the needs of the special scientific discipline. For the proper selection of the metadata schema to be used in the management system, KENNEDY (2008) provides nine guiding questions. Management systems or repositories that maintain data from various disciplines presumably support more than one metadata schema according to SIMONS & RICHARDSON (2013). Moreover, the metadata schemas have to be adjusted (modification or extension) according to the requirements of the research data that have to be described to suit the needs of the scientific disciplines(s) (QIN et al., 2012). The modification of the metadata with regard to a related research field also increases the user-friendliness and acceptance of the systems (MÜCKSCHEL et al., 2008).
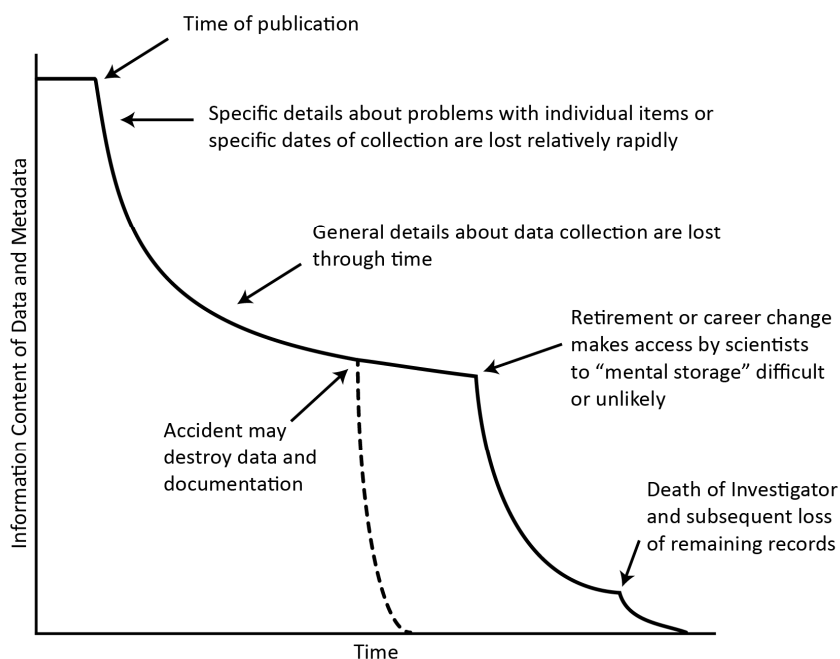
Figure 2-6: Degradation of information content over time, modified after MICHENER et al. (1997)

Based on the various subject-specific requirements on research data documentation and available metadata standards, several metadata frameworks have been formed. According to JENSEN et al. (2011), a metadata framework is a system that consists of metadata standards and norms, as well as a controlled vocabulary that complements one another. Moreover, a metadata framework is characterized to specify the metadata content on various levels. The documentation of research data has been already solved in various institutions or research projects by means of metadata frameworks or metadata management systems (e.g. DIEPENBROEK et al., 2002; GÖTTLICHER et al., 2010; LINNE, 2013; MICHENER et al., 2011; RUMSEY & JEFFERIES, 2013b; STOCKHAUSE & HÖCK, 2010; TAKEDA et al., 2010; WHITE et al., 2008).

2.3   Basic technical components

Research data management (RDM) systems form the technical framework to collect, document, and provide research data. These systems aim to collect all data which are created within a research group or institution (RAZUM, 2011). In general, a system is a unit of integrated hardware and software components, which serves for a specific purpose (VOGEL et al., 2009). As stated by RAZUM (2011), a RDM system provides features to store and search for data, as well as to manage user access rights. Furthermore, RAZUM (2011) outlines that the general features of and demands on the system are in particular dependent on the basic condition and used data types of the research project or institution. Nevertheless, general requirements on a RDM system

can be identified. These include for example: the linkage of research data with associated metadata for documentation, version control of the data, the support of multiple file formats, linkage of the research data, ownership management for datasets, registration of persistent identifiers, and support of access authorization. These requirements need to be regarded for the establishment of the architecture of the RDM system.

This chapter generally presents possible components of a RDM system. Initially, a short introduction will be provided on the client-server based system architecture (2.3.1). Subsequently, various data storage components (file- and database storage) will be described (2.3.2). Finally, web-technologies for the provision of web-interfaces (2.3.4) and an introduction to Geographic Information Systems (GIS) and web-based GIS (WebGIS) (2.3.5) are provided.

2.3.1   System architectures

Today, research data management (RDM) systems are usually based on multi-stage, distributed system architectures, like two-tier, three-tier or multi-tier (client-server) architectures (DIEPENBROEK & GROBE, 2007; KRALISCH & ZANDER, 2010; LOTZ et al., 2012). Consequently, this section will give a short introduction to client-server based system architectures.

In general, client-server models or architectures were developed to handle a large number of components in a computing environment, such as file servers, web servers, e-mail servers or database servers, as well as workstations, which are connected and communicate in a network. These servers can be accessed by a client or client machine which provides the local processing or user interface capability (ELMASRI & NAVATHE, 2014).

The server or server system is comprised of one or more shared computers and specific hardware and software. The server provides services (e.g. printing, file access, database access) to the client, which is an individual user's workstation. Server and client are connected by a network (ELMASRI & NAVATHE, 2014; OPPEL, 2009). According to SCHILL & SPRINGER (2012), the client-server model or architecture is an approach to structure distributed systems. This architecture can be organised in multi-level architectures, like two-tier, three-tier, or multi-tier (n-tier) client-server architectures (ALONSO et al., 2004). The two-tier architecture is based on two layers, the client and the server. For example, an application or user-interface program, running on a client-workstation is making requests on a centralized server (e.g. a database server). Two-tier architectures have the drawback of being inflexible and less scalable (ALONSO et al., 2004). The three-tier architecture solves this problem by an intermediate, extra layer between the client and the server – the server side application layer or middle tier (ELMASRI & NAVATHE, 2014; VOGEL et al., 2011). This architecture is used in many web applications. The three-tier architecture has some benefits, like centralized application logic in the middle tier, security, improved scalability (OPPEL, 2009), as well as a definition of interfaces between the single layers, and replication of the data in the middle tier (SCHILL & SPRINGER, 2012). As described by ELMASRI & NAVATHE (2014) with regard to the use of a three-tier architecture in a database environment, a web-interface or graphical user interface (client) sends a data request to the application server which accepts and forwards the query to the database server. The database server passes the result of the query via the application server to the client, where the data is presented or further processed. The three-tier architecture can be further extended to the n-tier or multi-tier (e.g. four-tier, five-tier) architecture, as the use of web-browsers became ubiquitous. For this purpose, the application or middle tier is extended with multiple layers (e.g. web servers) (OPPEL, 2009).

In the context of system architectures, the three-tier and n-tier architectures have prevailed. Likewise, information systems (ALONSO et al., 2004) and RDM system (DIEPENBROEK & GROBE, 2007; KRALISCH & ZANDER, 2010; LOTZ et al., 2012) are designed in this three-tier architecture. As described by ALONSO et al. (2004), SCHILL & SPRINGER (2012), and DE LANGE (2013), the following three tiers or layers are involved:

> **Presentation layer** involves components which are responsible for the presentation of the information or data (e.g. graphical user interface, web-interface).
>
> **Application layer (middle tier)** includes components which are responsible for the application logistic, like data processing (e.g. application server, web server, program).
>
> **Resource management layer (data tier)** comprises the components which are responsible for the management and storage of data (e.g. database, file system).

Likewise, RAZUM (2011) describes a similar three-tier architecture in the context of RDM systems. This structure is containing a persistence layer (storage of data, e.g. file or database based), a core layer (central features, e.g. metadata), and a service layer (expanding features, e.g. search, authentication).

In terms of software architecture for applications, a related three-tier architecture pattern is used, referred as the Model-View-Controller (MVC) principle, described in detail by STROBEL (2004). The aim of the MVC model is a flexible application design to enable subsequent extensions and modifications, as well as the reusability of single components (ILLIK, 2007). The MVC model is used for many software applications, such as web-applications. In this context, the three single components perform the following duties. The controller is responsible for the delegation and processing of the incoming requests. The controller passes the requests (e.g. for a parameter) to the model. The model has access to backend storage (e.g. based on file system or database) and handles the requests of the controller, like committing, writing or reading of data to the backend storage. The controller uses the responded data of the model component and passes these data or information to the view component. The view component represents the result in the client in a corresponding format (e.g. in HTML, JSON or XML) (HEIDERICH, 2009).

For the set-up of a system or software architecture, for example of a RDM or information system, usually several development steps have to be conducted (RAINER & CEGIELSKI, 2011; STAIR & REYNOLDS, 2010; WANG et al., 2007). According to ELMASRI & NAVATHE (2014), these steps include the subsequently phases: the analysis of the feasibility, the collection and analysis of the requirements, as well as design and implementation of the system. In addition, the acceptance of the system has to be tested and the performance has to be validated. Finally, the system will be released and has to be maintained.

In the following, several potential components of client-server-based system architectures are described, which are of importance for this thesis.

2.3.2   Data storage based on file systems

As outlined in the previous section, one component of a RDM system architecture can be a file system. Thus, the focus of this section will be on file-based data storage.

A file is a self-contained, fundamental organisational unit that is used to store a collection of data. It has a name for identification purposes and is located on a data medium (FISCHER & HOFER, 2011: 206). Files can be copied, deleted or renamed, and can only be read by certain methods or programs. A file system offers the persistent and secure storage of files. According to FISCHER & HOFER (2011), a file system is a model applied in an operating system for managing and storage of files on a secondary storage and further corresponding low-level processes. All files are summarized in directories and folders within the file system. They are arranged in a hierarchical structure to preserve an overview. Based on a root directory, the directories and folders are arranged in a branched out tree structure. A file can be accessed by specifying its name and the storage file path (BROY, 1998; HEUER & SIPPEL, 2004; LEVI & REMBOLD, 2003; MANDL, 2013; TANENBAUM, 2009).

HEUER & SIPPEL (2004) and MANDL (2013) point out that file systems can be distinguished between local file systems and network file systems or file servers. Local file systems are connected to a local peripheral device. In general, they are found on a magnetic data medium (e.g. hard disks, tapes) or they can be put down

on a CD-ROM. Examples for local file systems for Windows are: the File Allocation Table (FAT-12, FAT-16 and FAT-32) and the NT File System (NTFS). Examples for local file systems for Unix are: the Unix File System (UFS) or the Second and Third Extended File System (EXT2 and EXT3).

TROPPENS (2009) and MANDL (2013) describe that the natural extensions of local file systems are network file systems. Users or applications are able to access data files via a network protocol that are physically stored on a different computer (file server). Storage systems for typical Client-/Server applications are Network Attached Storage (NAS) and Storage Area Networks (SAN). Furthermore, TROPPENS (2009) and MANDL (2013) continue that a SAN is a network that can store a huge amount of data. The infrastructure of a SAN is characterized by a central dedicated storage network. This operates between servers and storage resources used by the servers (e.g. Disk Storage). The entire memory of the SAN is centrally managed and summarized to virtual units, even though the data is not physically located at the same location. SANs are primarily used to exchange data between computers (or servers) and storage devices. An advantage of SANs is the virtualisation of available storage devices. Thus, the distributed data can virtually be used as an individual hard disk. Nowadays, SANs are frequently used as storage system in large data centres and companies.

Furthermore, TROPPENS (2009) and MANDL (2013) point out that, in contrast to SAN, Network Attached Storage (NAS) is characterized by preconfigured file servers that are usually connected via Ethernet to a Local Area Network (LAN). The NAS can consist of one or more internal servers, includes preconfigured internal hard disk capacity, and usually uses a stripped-down or special operation system. NAS servers are very scalable and were especially developed for file sharing. They are employed to reduce the effort that is necessary for installation and administration of a dedicated file server. Therefore, the advantages of NAS are the provision of Plug & Play file systems as well as the significantly better optimization in opposite to generic operation systems. Disadvantages of a NAS are for example the upgrade of the path of the NAS server (e.g. a server cannot simply be replaced with a more powerful one). The most common NAS for Unix systems is the Network File System (NFS) developed by Sun Microsystems. The Common Internet File System (CIFS) is the corresponding NFS developed by Microsoft. Moreover, further NAS storage solutions are provided by several companies, like IBM (IBM, 2014), Synology (SYNOLOGY, 2014) or NetApp (NETAPP, 2014). According to HEUER & SIPPEL (2004), further distributed networked file systems are the Constant Data Availability File System (CODA, 2014) and the Andrew File System (AFS, OPENAFS (2013).

Furthermore, in large computing clusters, for example in the area of High Performance Computing (HPC), parallel file systems are used by theses HPC systems like Panasas's PanFS, Sun's Lustre, IBM's General Parallel File System (GPFS) or the Parallel virtual File System (PVFS) (ROSS, 2011). Nowadays, the virtualization of memory takes an important role, as well as cloud (LENK et al., 2009; MARIOS et al., 2009) and grid-based data storage (ASCHENBRENNER et al., 2011; LAWRENCE et al., 2009).

In contrast to file-based data storage, data storage based on database systems enables consistent storage that is available by an API (Application Programming Interface) and in particular allows multi-user access. Thus, a central data storage with an advanced user administration is possible and redundant storage of data is avoided (BRINKHOFF, 2013). The next section will focus on data storage based on database systems.

2.3.3 Data storage based on database systems

A Database System (DBS) is basically composed of a Database Management System (DBMS), application programs/queries, and (several) databases (DB), as presented in a simplified design in (Figure 2-7). According to BRINKHOFF (2013), a database is in a simplified way a collection of related, persistent, efficient, and consistent stored data that represent a part of the real world. Furthermore, BRINKHOFF (2013) explains that the DBMS is a software package (collection of programs) which enables, for example, the definition, creation, uniform description, secure administration, and access to a database. Input, reading, editing or deleting of data into the database is provided via the DBMS. ELMASRI & NAVATHE (2014) summarize this as provided

features of a DBMS, which are the processes of definition, construction, manipulation, and sharing of databases between several applications and users. By defining a database, the data type, constraints, and structure of the data (to be stored in the database) are specified. Additionally, metadata are stored in form of a database dictionary or catalogue, which include descriptive information or the database definition. The construction of the database describes the data storage process on a storage medium that is controlled by the DBMS. A DBMS provides the connecting user- or application program interface between the database and the programmers and users. CONNOLLY & BEGG (2005) complement that a Data Definition Language (DDL) is used to define the database (e.g. specification of data types and structure). In addition, a database can be manipulated through a Data Manipulation Language (DML) (e.g. insert, delete, update, and retrieve data from a database). In addition, KEMPER & EICKLER (2009) and KLEINSCHMIDT & RANK (2005) point out that there are different demands, characteristic properties or requirements on a DBMS. These are data independence, central data management, multi-user access, data security, data integrity, redundancy-free data storage, access control, processing of requests, and operating system independence. SUMATHI & ESAKKIRAJAN (2007) provide an overview of existing DBMS vendors and their products. This includes: IBM (DB2, Informix Dynamic Server), Microsoft (Access, SQL Server, DesktopEdition), Open Source (MySQL, PostgreSQL), Oracle (Oracle DBMS, RDB), and Sybase (Adaptive Server Enterprise, Adaptive Server Anywhere, Watcom). Database application programs enable users or programmers to interact with the database (e.g. create or maintain a database) by requesting the DBMS. Nowadays online application programs are common (CONNOLLY & BEGG, 2005).
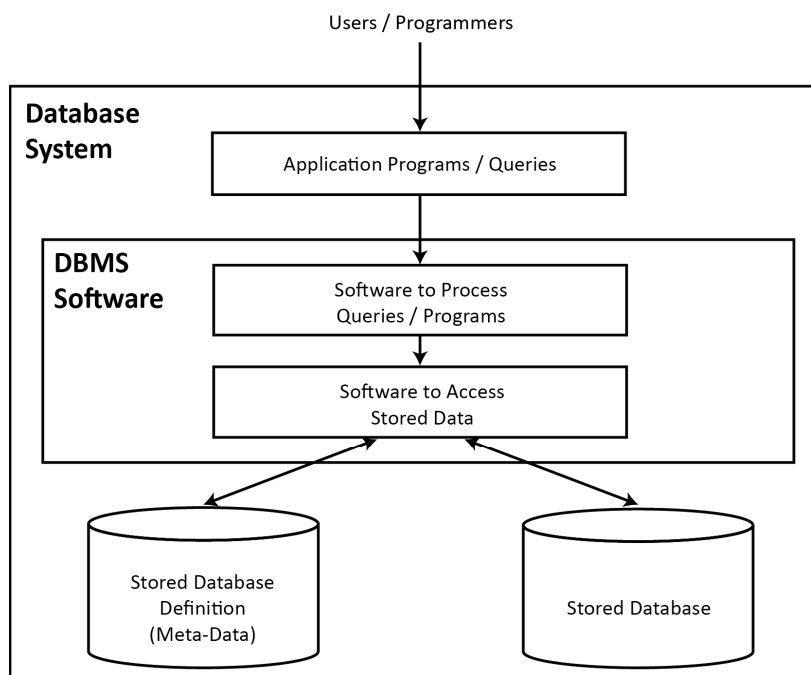
Figure 2-7: Simplified design of a database system environment, modified after ELMASRI & NAVATHE (2014)

DBS can be centralized and can comprise one logical database, which is managed by a single DBMS. In addition, Distributed DBMS (DDBMS) are available. They are characterized by managing distributed databases, which are physically distributed over a network of computers. Furthermore, parallel DBMS are available. These are designed to perform operations in parallel by running across multiple disks and processors with the aim to improve the performance (CONNOLLY & BEGG, 2005).

To ensure the logical and physical data independence, it is essential to use a three-schema architecture. Its goal is to separate the (physical) database and the user application. The three-schema architecture of a DBMS according to ANSI-SPARC (American National Standards Institute/Standards Planning and Requirements

Committee) is characterized in three levels (ELMASRI & NAVATHE, 2014; SCHUBERT, 2007; UNTERSTEIN & MATTHIESSEN, 2012):

Internal level The internal level describes the physical storage of the data in the database with the use of an internal schema. By means of a physical data model, the complete details of data storage as well as the access paths for the database are defined. In general, the data storage is realized as background data storage (on disk space). The internal level is managed by the DBMS.

Conceptual level The conceptual level covers the (logical) structure of the whole database by means of a conceptual (logical) schema. For example, tables, entities, data types or models, user operations, relationships, and constraints are described by means of a representational data model. The so-called database schema is defined. A database administrator or programmer can access the conceptual level via query languages like SQL.

External level The external level (or view level) defines the individual user-views on the database. This is realized by a number of external schemas. Therefore, each external schema determines the accessible information for the individual user (interesting parts of the database are visual and uninteresting parts are hided). The external schema is implemented and based on a representational data model. The external level can be accessed by a query language like SQL or via a graphical user interface (database application).

SUMATHI & ESAKKIRAJAN (2007) note that the database schema determines the design of the database. A schema is a collection of objects and provides a classification of these objects in the database. Database schemas can be specified in physical, logical, and subschemas. They can include, for example, tables, functions, and other objects. SUMATHI & ESAKKIRAJAN (2007) and YEUNG & HALL (2007a) state that a database model (also called data model) describes the structure of a database that represents the problem space of the real word. It makes use of various data abstraction techniques, as well as graphical notations and linguistic syntax, e.g. data types, constraints, and relationships. Consequently, the conceptual description of the database occurs though the database model. According to SUMATHI & ESAKKIRAJAN (2007), examples for conceptual data models are the Entity-Relationship (ER) model, the object-orientated model, the functional data model, the relational model, the network model, and the hierarchical data model.

Relational database

From the variety of models and schemes various DBS have evolved. For example, the Relational Database (RDB), first introduced and described by Codd (1970), is based on the relational model. It is used as a basis for many commercial systems. ELMASRI & NAVATHE (2014) name the following relational DBMSs (RDBMS): PostgreSQL and MySQL (Open-Source) (both open source systems), Informix Dynamic Server and DB2 (from IBM), Sybase DBMS (Sybase), Rdb and Oracle (from Oracle), as well as Access and SQLServer (from Microsoft). According to CONNOLLY & BEGG (2005) and UNTERSTEIN & MATTHIESSEN (2012), the relational model is based on the concept of relations. Consequently, the relational database is represented as a collection of relations. These relations are represented in tables that store data in a permanent, tabular form (Figure 2-8). The tables are composed of columns and rows. The column header of the table is named *attribute*. For each attribute a range of values, the so-called *domain*, is specified. This defines the set of allowed values (e.g. boolean, char, integer, numeric, string, date). The rows of a table represent a data *record* or so-called *tuple*. Each tuple is a collection of related data values that is distinct. Duplicate tuples are not allowed within a *relation*. The structure of a tuple has always the same structure, because all values of a column have the same domain.

Relational keys are used to uniquely identify tuples in a relation. According to CONNOLLY & BEGG (2005) following keys are available:

Superkey This key represents an attribute (or a set of attributes) that uniquely identifies a record in a relation. Not all included attributes in a superkey are necessary. A minimum superkey is called a candidate key.

Candidate key This key specifies a minimum superkey, chosen to identify a unique tuple in a relation is called a primary key.

Primary key This key is obligatory for each relation. A relationship between tuples of relations is represented by a foreign key.

Foreign key This key is defined as an attribute (or set of attributes) within a relation that matches a cardinal key (or primary key) of another relation.

Furthermore, CONNOLLY & BEGG (2005) and OPPEL (2009) describe that a linkage of the relations can be realized by various relationship types. In a *one-to-one* (1:1) relationship an instance of exactly one record can be associated with at most one record from another relation, and vice versa. The relationship is in both directions mandatory. In a *one-to-many* (1:* or 1:n) relationship an instance of one record from a first relation is associated with zero, one or many records from a second relation. In the reverse order any instance of the second relation can be associated with at most one record of the first relation. In a m*any-to-many* (*:* or n:m) relationship, any record of a first relation may be associated with zero, one or many records of a second relation, and vice versa. The cardinality defines the maximum number of instances possible for the relationship. Relations, cardinality, and relationship can be represented in an Entity-Relationship (ER) model.
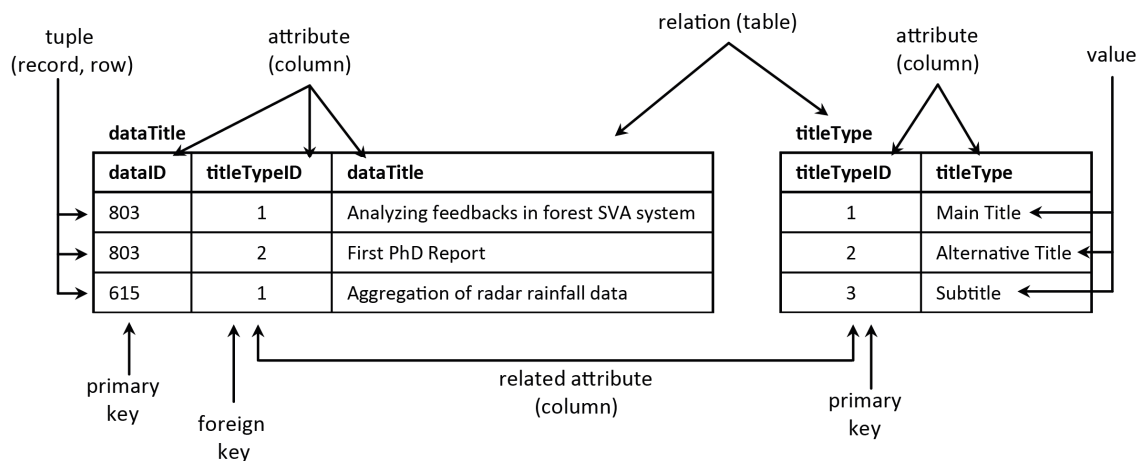
Figure 2-8: Elements of the relational database tables `Title` and `TitleType` (simplified representation)

The main language used to control, manipulate and define data in a relational data model is SQL (Structured Query Language). SQL has been standardized by the ANSI (American National Standards Institute) and ISO (International Organisation for Standardization) (OPPEL & SHELDON, 2009). According to SUMATHI & ESAKKIRAJAN (2007) three types of commands in SQL can be classified: Data Definition Language (DDL) commands (e.g. create, alter, drop), Data Manipulation Language (DML) commands (e.g. delete, insert, select, update), and Data Control Language (DCL) commands (e.g. grand, revoke).

Object-orientated database

Besides the relational databases also object-orientated databases have evolved. According to SUMATHI & ESAKKIRAJAN (2007) and OPPEL (2009), the concept of object-orientated databases dates back to the late 1970s. This was connected with the utilization of object-orientated programming languages, like C++, Java or C#. Object-orientated database technologies are characterized a as combination of database technologies and object-orientated programming. Object-orientated concepts make use of objects that represents a real-word thing and is described as a logical grouping of programming logic and related data. Each object is associated with variables and methods. Moreover, they are organized in a class hierarchy. The concept of the object-

orientated model was extended with the relational model. As a result, the object-relational database systems emerged. The first object-relational DBMS (ORDBMSs) was Informix, followed by Oracle and DB2. According to BRINKHOFF (2013), ORDBMS also serve as a basis for Spatial Database Systems. They are characterized as DBMS which support the storage and handling of spatial requests. Examples for ORDBMS with a spatial extension are: Oracle Spatial, Informix Spatial DataBlade and Geodetic DataBlade, as well as DB2 Spatial Extender.

Database design

The design of a database system is a part of the entire system development life cycle (SDLC), for example, of an information system, as described in OPPEL (2009) and CONNOLLY & BEGG (2005). The design of the database is a subarea of this. According to BLANSIT (2006), it is essential to accurately plan the structure of a database to create a useful and operating system. There are many variants of database design, for example characterized by TEOREY et al. (2011), ELMASRI & NAVATHE (2014), or PRÖLL et al. (2013). All approaches contain at least four main stages of database design. These are according to CONNOLLY & BEGG (2005), ELMASRI & NAVATHE (2014), and PRÖLL et al. (2013):

> **Requirement analysis** includes the process of collecting and analysing information about the issue of the real world that has to be represented with the database.
> **Conceptual design** comprises the process of constructing a conceptual model (e.g. using the entity-relationship model) that describes the analysed requirements of the database.
> **Logical design** contains the process of developing a logical data model. The developed conceptual model will be implemented into a database schema (e.g. relational database model).
> **Physical design** describes the process of specifying the implementation of the database (e.g. internal storage structure, file organisation, indexing).

One of the most popular and most widely used models for designing a database is the entity-relationship (ER) model (HALPIN & MORGAN, 2008). This is mainly predicted by a simplified graphical representation and the extensibility of the design, as well as on a safe construction (THALHEIM, 2000). According to FISCHER & HOFER (2011), an ER model is defined as a model to describe the principles of relationships of entities within and for a database. It specifies the association between the entities and denotes the number of conditions. Participating entities, relations, and cardinalities are represented.

Within the logical design of a relational database, the process of normalization has an important duty. CONNOLLY & BEGG (2005) point out that the technique of normalization ensures the correctness of the data model. Relationships between the attributes are examined. It is possible to use the normalization technique at any stage of the database design. OPPEL (2009) argue that the process of normalization is conducted in a series of steps. By applying various rules, the relations are reorganized into a normal form. Complementary, SUMATHI & ESAKKIRAJAN (2007), TEOREY et al. (2011), and UNTERSTEIN & MATTHIESSEN (2012) describe the normalization process in detail.

2.3.4   Web-based technologies and web design

The application of web-based client-server systems in different fields, such as tourism (e.g. hotel booking portals), interactive media (e.g. photo or video portals) or information services (e.g. weather or map portals) has increased in recent years. To provide these applications, special technologies are required (CHANTELAU & BROTHUHN, 2010). This chapter will provide basic information about technologies and design principles of web-based systems.

Web-based technologies

The International Network, abbreviated as Internet, is a worldwide network, where communication of the clients and servers (section 2.3.1) is provided by the Transmission Control Protocol/Internet Protocol (TCP/IP). This is the outcome of the computer network constructed in 1969 referred as the Advanced Research Projects Agency Net (ARPANET), which was first used by universities and research institutes after its release in 1972. For the allocation of a server, the Domain Name System (DNS) is used that connects a specific IP address (e.g. 129.128.1.2) to a domain name (e.g. www.google.de) (LEVI & REMBOLD, 2003). The Internet provides several services. The main services are: Electronic Mail (E-Mail; Simple Mail Transfer Protocol SMTP), File Transfer (File Transfer Protocol FTP), Terminal Emulation (Telnet; Telnet Protocol), Usenet News (Network News Transfer Protocol NNTP), and World Wide Web (WWW; Hyper Text Transfer Protocol HTTP; Hyper Text Transfer Protocol Secure HTTPS) (DE LANGE, 2013). The latter is a global information network, also well known by its short form *Web* established in the early 1990s (BERNERS-LEE, 1989). A few years later, the term *Web 2.0* was coined by Tim O'Reilly (O'REILLY, 2005). As stated by KAMAL (2002), the Web is a global collection of Websites, interconnected by the Internet. The Web uses HTTP or HTTPS for the communication between the web-servers and web-clients to transfer information (DE LANGE, 2013). In this context, a browser (e.g. Google Chrome, Microsoft Internet Explorer, Mozilla Firefox, Netscape Navigator, Opera, Safari), which is a client-based software application, communicates with the server (STROBEL, 2004). By means of a browser, a connection can be established to the server. Furthermore, the browser can request information from the server and display the results (e.g. webpage). The server is able to receive the request from the client and provide the corresponding information. For example, by means of the Hypertext Markup Language (HTML) a result (e.g. text, image) can be designed in form of a webpage and presented in a browser. Through the application of hyperlinks, a linkage can be created to other webpages or documents. This enables users to navigate to other webpages (BEKAVAC, 2013). The access to a resource (e.g. document, webpage) is carried out by a standardized, unique address or path on the Internet, called the Uniform Resource Locator (URL). A URL is an alphanumeric character string composed of the used protocol (e.g. http://), the domain name (e.g. www.google.de), and an optional file name (e.g. index.php) (GODBOLE & KAHATE, 2006).

As described by OPPEL (2009), webpages can be static and dynamic. Static webpages are indicated to remain the same and do not change frequently. In contrast, dynamic webpages are characterized to change on the fly according to a particular request to the server. In general, dynamic webpages have a static and a dynamic portion, the latter is created on a specific request on the server. Dynamic webpages are of special interest in the conjunction with databases. The growing importance of databases in combination with the Web was already noticed by BENN & GRINGER (1998), as well as corresponding technologies for web-based applications with databases are described by LOESER (1998).

On the client and the server side, several technologies and applications are necessary to create websites or webpages. Some applications and languages will be described subsequently. Initially, some client-based languages will be briefly described. In general, the markup language HTML is used as a basis to design web pages or user interfaces. The tag-based language enables the formatting and structuring of the documents, including text, images or hyperlinks. In addition, XHTML (Extensible HTML) is available, which is based on the Extensible Markup Language (XML). Furthermore, CSS (Cascading Style Sheets) can be applied to modify the layout of the website depending on the output device. The use of CSS allows the separation of layout and content. The programming language JavaScript (JS) facilitates the creation dynamic generated websites like a simple input of animated images or the evaluation of input forms (CHANTELAU & BROTHUHN, 2010). By means of AJAX (Asynchronous JavaScript and XML), which is a combination of several technologies, enables the asynchronous transmission of data between the client (browser) and server. Consequently, AJAX provides the opportunity to update certain areas of a HTML document without updating the entire webpage (DE LANGE, 2013). For the description of animations and two-dimensional vector graphics, the Scalable Vector Graphics (SVG), based on the XML, can be used (THIESING &

KORTEMEYER, 2008). Furthermore, by means of Java applets, which is a small application written in Java, a webpage can be designed dynamically from the client-side (POMASKA, 2012). Finally, ActionScript is another object-orientated programming language, used to create web pages (KOHL, 2011).

Likewise, server-side languages and interfaces for web applications are available. According to POMASKA (2012), these are programming languages, like Python, the web application framework Ruby on Rails, as well as the programming languages ASP (Active Server Pages), Java, Perl, and PHP (Hypertext Preprocessor). PHP is a server-side scripting language, which supports the use of variables, control structure, as well as various functions. Moreover, methods are available to connect to databases as well as to process XML. PHP is often used to create dynamic webpages. These are created on request of a browser which transfers the required parameters to the web-server. The server provides a response to the client which is composed of various resources. The PHP-script generates the HTML output according to the provided parameters. The result will be provided and presented in the browser (KANNENGIESSER & KANNENGIESSER, 2007; KOFLER & ÖGGL, 2010).

Besides the implementation of self-developed web applications by means of several client- or server-based languages, also computer programs are available, which enable a simple creation, publishing, and modifying of content in websites. These Content-Management-Systems (CMS) are used for the creation and maintenance of websites, in the past mainly in huge companies. CMS are designed to establish websites, which users can modify without special knowledge of programming languages. Examples for CMS are: TYPO3, MODX, Joomla!, Drupal, DotNetNuke, Plone, and eZ Publish (WENZ et al., 2011). As outlined by POMASKA (2012), the usage of weblogs or blogs has grown in importance in recent years. For the creation of blog websites, several blog services and systems can be used, such as Google-Blogger or WordPress. In addition, social networking services like Twitter, Myspace, YouTube, Flickr, Facebook, Google+, LinkedIn or Instagram enable the publishing and sharing of text, images or videos.

In addition, web technologies or applications are available to access data on a server. For example, software applications can be used to access PHP-files, HTML files, general data or documents located on a specific server. For instance, the software WinSCP (WINSCP, 2014) or FileZilla (FILEZILLA, 2014) can be used to upload or download files to/from a server. Both support the Secure File Transfer Protocol (SFTP).

Web design

The design of websites and the perception of their users have changed since the establishment of the Web in the 1990s POOKER (2009). For example, a study conducted by NIELSEN (1997) shows that many people (around 80 % of the study participants) only *scan* webpages. Therefore, it is necessary to implement simple webpages (e.g. with less text, highlighted keywords, images). In order to provide and design pleasant websites, it is important to follow recent standards and trends, as well as available guidelines. Subsequently, some web design principles will be presented.

At first, recent guidelines and national or federal law have to be considered. In the context of web design, the Barrier-free Information Technology Ordinance (BITV) (BUNDESMINISTERIUM DER JUSTIZ UND VERBRAUCHERSCHUTZ, 2011), which was issued in the framework of the Equal Opportunities for Disabled People Act (BGG) (BUNDESMINISTERIUM DER JUSTIZ UND VERBRAUCHERSCHUTZ, 2002), both entered into force in 2002, constitutes the legal basis for barrier-free websites in Germany. With regard to the federal state level, North Rhine-Westphalia has issued the Barrier-free Information Technology Ordinance North Rhine-Westphalia (BITV NRW) (LANDESREGIERUNG NRW, 2004). These outline principles which have to be considered for the sustainable creation of barrier-free website. The principles are based upon the Web Content Accessibility Guidelines (WCAG) (W3C, 2008) of the World-Wide-Web-Consortium (W3C).

They include principle groups in four categories:

Perceivability Components of user interfaces and their information must be presented to users in a perceivable way.

Operability Navigation and components of user interfaces must be designed operable.

Understandability Operations of user interfaces and their information must be presented in a common understandable way.

Robustness The application and the consent of the user interface must be designed robust to ensure the operation and reliability for prospective technologies and modified user agents.

In addition, the W3C provides recommendations with regard to the use of web technologies for web design and applications, semantic web, web architecture, XML Technology or authoring tools and usage of browsers (W3C, 2014).

According to POOKER (2009) and WENZ et al. (2011) usability and accessibility are important characteristics, which have to be considered during the development of the web design. The usability focuses on the orientation and the user guidance. For example, successful user guidance is characterized by well chosen, significant website titles, a meaningful headline hierarchy, clearly defined representation of user navigation and website content, as well as explicit identification of links and their link targets. The user perception of a website is an important criterion for the usability of the website according to KLUG (2006). The author makes some suggestions to enhance the usability and to increase the goodwill of website users. Thus, the first law of usability is "Don't make me think!" KLUG (2006: 10). Consequently, websites should be designed to answer all questions of users which they might have. Websites should contain all main things and content which the users would need to use the website. These features have to be easy and obvious to the users. In addition, the website should save the users' steps, wherever it is possible. Furthermore, websites should provide comfort features like printer-friendly webpages. Finally, websites must be tested before their launch to prevent errors.

The design or structure of a web page usually comprises several components as described by BALZERT (2004). A simplified view is presented in Figure 2-9.
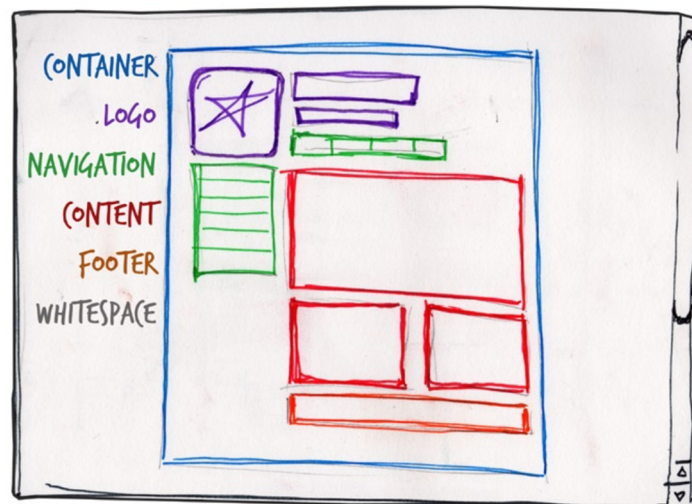
Figure 2-9: Structure of a website (BEAIRD, 2011)

According to BEAIRD (2011), the components of a web page are:

Container This component is the surrounding block which contains all elements of the web page. The width and height of the container can be static or dynamic. A dynamic container will be adjusted according to the browser window extent. A static container will remain at the same level regardless of the browser window.

Logo The logo component, also called identity block, should contain the logo and name of a company or project. This component fosters the recognition and emphasizes that the user is still on the same website.

Navigation The navigation has to be easy to find and easy to operate. It is usually located on the top or the left side of a website and represents the structure of the web page. The purpose of the navigation is to guide the user through the content of the web page (POOKER, 2009).

Content The content component is the most important part of a web page. Consequently, the main content has to be located in the centre. Thus, the user can quickly find the required information.

Footer The footer component usually contains copyright and legal references, contact information or links to the main content of the web page. In addtion, the footer represents the end of the web page, if content and footer are optically separated.

Whitespace The whitspace component refers to all areas of the webpage, which are not covered by text or images. White spaces are just as important as the content and are loosening the web page.

2.3.5   Introduction to GIS and WebGIS

In this section, a short overview of GIS and WebGIS as a component for RDM is presented. Geographic Information Systems (GIS) are a key component for the input, management, analysis, and presentation (IMAP-principle) of spatial data (DE LANGE, 2013). Likewise, GIS using web technology (WebGIS) are used in RDM for these purposes (DAMM & GÖTZE, 2009; MÄRKER & KANAEVA, 2010; MÜCKSCHEL et al., 2007; WILLMES et al., 2014; ZANDER et al., 2013).

The specific term *GIS* was first used by Roger Tomlinson in 1962, called the *father of GIS*, for computer-based information system working with spatial data in Canada (BILL, 2010; DE LANGE, 2013). According to DE LANGE (2013), a Geographic Information System (GIS) consists of four structural components. These are hardware, software, procedures, and data. Thus, GIS are capable to "[…] capture, store, edit, manipulate, manage, analyse, share, and display geo-referenced data." (PINDE & SUN, 2011: 4).

The digital data, which models real world appearances (DE LANGE, 2013), is one key component of a GIS. According to BILL (2010), all data have or are related to a geographic reference. Consequently, each specific feature or entity can be called spatial object or geo-object (DE LANGE, 2013). Data, which can be used in a GIS or WebGIS, are subdivided into different groups of data types:

Geometric data is mainly stored in two different data models, the raster- and vector data model (KAPPAS, 2012). A discretization of the mostly continuous appearance of a real world feature is used to store data in these data models. The vector data model stores every entity connected with a geographic coordinate, called node or vertice. These vertices can be connected to lines or polygons (closed lines). An efficient way in particular to store continuous data is to use a cell-based raster model. Starting with an origin, data are stored in cells with a matrix of specific rows and columns. This 2D structure is easy to use for programming and according analyses (KAPPAS, 2012).

Topologic data describes the adjacency and connectivity of geo-objects by a mathematical model. This topologic information is of interest for spatial queries and routing purposes (e.g. BAASER, 2010).

Attribute data are stored in tables, which are related to their spatial object by an identifier. This relational data model allows efficiently to add, store, and search specific spatial attributes and can be realized with database systems such as, Oracle or MS Access (KAPPAS, 2012). For extensive information object-relational and object-oriented data models are possible (section 2.3.3).

The displayed map results in a GIS are generated from the data, according to the chosen and linked style for point, lines, polygons and raster cells, as well as according annotations (BILL, 2010). Data are one component of a GIS. In combination with hardware, software, and users, data forms the four main components of a GIS (DE LANGE, 2013).

The spatial data are generally derived from analogue maps, remote sensing methods, surveying point data, and any further data which can be related to a spatial position or extent by integration through digitizing, scanning, and conversion (KAPPAS, 2012). According to BILL (2010) and KAPPAS (2012), each spatial dataset needs to be considered by its lineage, accuracy, consistency, and completeness. More information on different remote sensing methods, such as photogrammetry, laser scanning and radar remote sensing is given in LILLESAND et al. (2008), ALBERTZ (2013), and JENSEN (2014). As a further data provider, sensor networks can be integrated into GIS and WebGIS systems, which are already acknowledged by the OGC specifications Sensor Web Enablement (SWE) (BILL, 2010). These specifications contain, for instance, a schema and model for the sensors, the Sensor Model Language (SensorML) and corresponding services for the retrieval of specific observations, the Sensor Observation Service (SOS) and Web Notification Services (WNS), as well as Sensor Alert Services (SAS). The latter can be used, for instance, in warning systems for earthquakes, tsunamis or landslides (BILL, 2010).

Overall, the importance of GIS, spatial data, and geoinformatics as a discipline, for the society in terms of environmental studies, disasters, and diseases, has increased by time (DE LANGE, 2013; DICKMANN & ZEHNER, 2001; KAPPAS, 2012). Since 1995, the period of an open geomarket has evolved and WebGIS as well as geoservices are getting more important (BARTELME, 2005; DE LANGE, 2013). With the release of Google Earth in 2005, a "geohype" (DE LANGE, 2013: 3) has started and geoinformation is now ubiquitary (BILL, 2010). This development is embedded in the evolution of the Internet towards the web 2.0 (DE LANGE, 2013), with new technical developments, such as AJAX, user generated content, and application programming interfaces (APIs), as well as the relocation of data, software and processing to a *cloud* (section 2.3.4).

In terms of RDM, WebGIS are used for the visualization and analysis of research data with a geographic extent through the web (DAMM & GÖTZE, 2009; WILLMES et al., 2012). According to PINDE & SUN (2011), WebGIS are defined as GIS systems that use web technology. Several synonymously used terms exist, such as *online-GIS*, *Internet-GIS*, *web-mapping*, *Internet-mapping*, and *net-GIS* (DE LANGE, 2013: 3). However, minor differences occur (PINDE & SUN, 2011). For instance, Internet-GIS may use any Internet service (DE LANGE, 2013) not just the web technology of the Internet. However, these systems show a similar development as the Internet technology, starting with the first WebGIS in 1993 with a dynamic map viewer form the Xerox Corporation Palo Alto Research Center (PARC) (PINDE & SUN, 2011; PLEWE, 1997). The system was already capable of zooming, involved a layer selection, and converted map projections. Overall, GIS and WebGIS have evolved in accordance with the web technology and advances in computer and remote sensing technology since their early development (DE LANGE, 2013; PINDE & SUN, 2011). A web-based distribution of geographic information is nowadays a widely used approach (MITCHELL & EMDE, 2008). The most popular examples are Google Maps, Google Earth, NASA World Wind, and Microsoft Bing Maps.

The technical components of a WebGIS are a web application server and a client (two-tier system), which can be a web browser, a desktop application, or a mobile application (see section 2.3.1) (DE LANGE, 2013). The communication relies on the HTTP specifications, where the format of the request and response can be based on HTML, binary images, XML (eXtensible markup language), JSON (JavaScript Object Notification), Simple Object Access Protocol (SOAP), Representational State Transfer (REST), or others (PINDE & SUN, 2011). The two latter, SOAP and REST, are the most common implementations, where SOAP uses XML encapsulations and REST relies on the direct use of XML, JSON or other formats (PINDE & SUN, 2011). Additional tiers are added to this configuration, when the data is hosted at another location, or further services from other locations are used. The latter is possible, when the mashup approach is used, where

existing services from different servers are involved to generate a new application (PINDE & SUN, 2011). For example, the Google Maps API is often used (GOOGLE, 2014).

Possible server software for geospatial data content is the UMN MapServer (MAPSERVER, 2014), ESRIs ArcGIS for Server (older Versions: ArcIMS / ArcGIS Server) (ESRI, 2014), GeoServer (GEOSERVER, 2014), rasdaman (RASDAMAN, 2014) and its extended version, the EarthServer (EARTHSERVER, 2014). These servers host (geospatial) web services, which are programming interfaces for an interoperable use in a service-oriented architecture (SOA) (PINDE & SUN, 2011). Likewise, these services can be used in combination with and adjusted to the location of the user's device. The services with content altered to the specific location of the user are called Location Based Services (LBS) (BAASER, 2010; PINDE & SUN, 2011).

These geospatial web services can be distinguished by their function (PINDE & SUN, 2011):

> **Map services** respond to client request by images in different formats, such as PNG, JPG or TIF that show the selected extent and content. This is the most common type of service. The images can be dynamically rendered or preprocessed tiles of caches can be delivered, enhancing the performance. The latter is typically used for basemaps with little changes in time. Dynamic services are used for operational layers that change more often. Generally, the services support the querying of attributes and the identification of objects.
>
> **Data services** mainly allow editing of data by feature editing services, web editing and geographic sketching. In addition, search services exist which allow to find directly indexed data. Furthermore, image services are implemented to directly use raster data, such as remote sensing imagery or digital elevation data. These services allow, depending on their configuration, on-the-fly processing.
>
> **Analytical services** are subdivided in geocoding services or locator services that convert street addresses to coordinates, network analysis services that allow routing, the calculation of service areas (what can be reached in 5 minutes of driving?) and finding the closest point of interest (POI) in terms of location based services. In addition, geometric services enable to transform existing datasets by tasks such as merging and splitting, and calculate data by extents or buffering. Geoprocessing services calculate, depending on choices of the user, new data on the server, for instance complex models based on spatial data.
>
> **Metadata catalog services** exist which use published metadata of the aforementioned services. These metadata can be queried and the particular services can be selected.

All services can be hosted as standardized web services. Geospatial service standards are defined by the Open Geospatial Consortium (OGC) and the International Organisation for Standardization (ISO). This standardization allows establishing Spatial Data Infrastructures (SDI) on a regional, national and international level (PINDE & SUN, 2011). In Europe this is fostered by the Inspire Initiative (Section 2.2.2.4).

According to the above mentioned different types of geospatial web services, common standard web services are defined by the OGC. The web map service (WMS) is mostly used (BILL, 2010; DE LANGE, 2013; PINDE & SUN, 2011). This service allows three operations, the GetCapabilities-request for an XML-based view of the metadata of the service, the GetMap-request for the response of a particular image of the spatial data, and the GetFeatureInfo that allows retrieving specific information about a selected feature, usually in the GML format (DE LANGE, 2013). In addition, the Web Feature Service (WFS) allows editing of features, the Web Coverage Service (WCS) is returning raw raster data, Web Processing Services (WPS) allow executing of calculations on the server and retrieve the according results, and a Catalog Service for the Web (CSW) can be used for sharing metadata (DE LANGE, 2013).

All these services interact with clients. These clients can be web browsers, desktop applications, and mobile applications (PINDE & SUN, 2011). The web browser clients have been using static HTML or simple JavaScript, but nowadays rich and dynamic applications exist, based on Ajax, Flex, or Silverlight. Still, more

features are enabled on desktop clients, which range from very thick desktop software products, such as usual GIS Desktop software, like ESRI's ArcGIS for Desktop, to specific clients developed by Java, .NET or other languages. The most popular desktop client working with geospatial web services is Google Earth. Mobile clients can also use the browser-based applications, but also native applications are possible. These are all *apps* available for iOS or Android-based devices, such as Google Maps or Google Earth mobile. As aforesaid, in combination with the location of the client, location-based services are possible (PINDE & SUN, 2011).

The available possibilities allow numerous different applications. Mostly, applications of RDM are relying on web services with an OGC-standard served by the UMN MapServer (MÄRKER & KANAEVA, 2010; WILLMES et al., 2014; ZANDER et al., 2013) or GeoServer (KUNKEL et al., 2013). However, all WebGIS facilitate the analysis and visualization of spatial data, which is managed by a server.

2.4 Software solutions for research data management

This chapter gives an overview of available technical software solutions, which enable the setup of a repository out-of-the-box. In the literature, several studies are available that survey or compare available software solutions. For example, a detailed comparison between the digital library management software DSpace, EPrints, and Greenstone is provided by TRAMBOO et al. (2012). In addition, DOBRATZ (2007) provides a brief overview of the open source software DSpace, Eprints, MyCoRe, and Fedora. Further investigations on comparative studies of repository software tools are presented by OSI (2004), BORGHOFF (2005), PRUDLO (2005), and MARILL & LUCZAK (2009).

Table 2-9 provides an overview of the most common technologies used to set up a RDM system out-of-the box. For each software solution, example applications are presented. Most of the listed technologies originated from the digital library initiatives, like DSpace, EPrints or FEDORA (Flexible Extensible Digital Object Repository Architecture), in contrast to, for example, CKAN (Comprehensive Knowledge Archive Network). In the field of geodata metadata management, portal solutions like the open source software GeoNetwork are very common.

2.5 Research data management in research projects and institutions

Available approaches and solutions in research data management (RDM) are presented in this chapter. After an introduction, an overview will be given of available approaches. As already outlined in section 2.1.5, several activities grouped in several categories are available. Initially, examples of institutional approaches of RDM will be presented. These comprise all activities conducted on the institutional repository, national data centre, and major international resources level (Figure 2-4). Subsequently, examples of RDM are provided, which are conducted in research projects.

The need and importance to manage research data created by public funding is recognised by research instructions and authorities (section 2.1.2). This has led to a new request for proposals in this area by funding agencies. For example, in the UK, the Joint Information Systems Committee (JISC) has launched the digital infrastructure research data management programme (HODSON & MOLLOY, 2013; JISC, 2014a), covering the areas of research data management, research information management, and support of infrastructure for repositories and curation. In Germany, funding programs are, for example, available by the German Research Foundation (DFG) with regard to Information Infrastructures for Research Data (DFG, 2013a) and Virtual Research Environments (DFG, 2013b). Consequently, the number of projects and organisations providing data management repositories or infrastructures has increased since then. For example, in 2011 the DFG approved 27 projects to set up information infrastructures for research data (DFG, 2011). Within the funding of the large-scale, interdisciplinary, long-term research projects, so-called Collaborative Research Centres, the DFG supports the set-up of a service project section *Information Infrastructure (INF-*project*)* (section 2.1.2).

Table 2-9: Overview of available examplary RDM software solutions and repository systems

Software Name	Information	Website Data Portal	Example applications	Literature
CKAN	developed by Open Knowledge Foundation	http://ckan.org/	CKAN (2014a); DATA.GOV.UK (2014); WILLMES et al. (2014)	CKAN (2014b)
DSpace	open source, developed a Massachusetts Institute of Technology (MIT) / Hewlett-Packard (HP)	http://dspace.org	DSPACE (2014b); KUBEREK (2013); LINNE (2013); WHITE et al. (2008)	DSPACE (2014a); PHILLIPS et al. (2007); SMITH et al. (2003); TANSLEY et al. (2003)
EPrints	open source, developed by University of Southampton	http://www.eprints.org/	BROWN & WHITE (2013); EPRINTS (2014b)	BEAZLEY (2010); EPRINTS (2014a)
eSciDoc	Fedora based, developed by Max Planck Society and FIZ Karlsruhe	https://www.escidoc.org/JSPWiki/en/Software	DREYER & TSCHIDA (2009); JESCHKE et al. (2009)	RAZUM et al. (2007); RAZUM et al. (2009)
Fedora Commons Repository Software	open source, initially developed at Cornell University, University of Virginia	http://www.fedora-commons.org/	FEDORA (2014b); LAUDATIO (2014); WILSON et al. (2011)	FEDORA (2014a); LAGOZE et al. (2005)
GeoNetwork	open source, is part of the Open Source Geospatial Foundation	http://geonetwork-opensource.org/	FAO (2014); IMPETUS (2014); VBA-GEOPORTAL (2014)	GREENSTONE (2014)
INVENIO	open source, developed at CERN	http://invenio-software.org/	CERN DOC SERVER (2014); INSPIRE (2014); INVENIO (2014b); JUSER (2014)	INVENIO (2014a)
MyCoRe	open source, branch office at University of Hamburg	www.mycore.de	MYCORE (2014)	LÜTZENKIRCHEN (2002); OELTJEN (2007, 2012)

According to ENGELHARDT (2013), as of march 2013, 45 INF-projects have been requested since 2007 and 27 of these have been funded.

An overview of available research projects is provided for example by BALL (2009), who gives a brief introduction of project repositories and institutional repositories (e.g. national data centres, university repositories). In addition, LEONELLI et al. (2013) identify research data initiatives and data repositories with regard to plant sciences. In a survey conducted by HAAK MARCIAL & HEMMINGER (2010), the characteristics of 100 international scientific data centres or repositories are presented.

Furthermore, several web portals offer an overview of the heterogeneous landscape of available research data repositories. For example, the recently established project re3data.org (REGISTRY OF RESEARCH DATA REPOSITORIES, 2014) provides an overview of indexed research data repositories (PAMPEL et al., 2012; PAMPEL et al., 2013; VIERKANT et al., 2012). Furthermore, project and institutional research data repositories are presented in the OpenDOAR portal (DIRECTORY OF OPEN ACCESS REPOSITORY, 2014), the ROAR portal (REGISTRY OF OPEN ACCESS, 2014), as well as the Databib portal (DATABIB, 2014). An overview of earth science data and services is provided by the NASA (NASA, 2014).

RDM approaches in institutions

Approaches of institutional repositories and data management initiatives are available on a national and international basis. Besides discipline-specific data centres, RDM is also provided in higher education. For example, in the field of environmental sciences, discipline-specific data centres are the ISCI World Data

Centers for Climate (WDC-C), Remote Sensing of the Atmosphere (WDC-RSAT), and Marine Environmental Sciences (WDC-MARE). In the UK, there is a multiplicity of national data centres available for various disciplines (COLLINS, 2011). For instance, the NERC (Natural Environment Research Council) funds, for example, the British Atmospheric Data Centre (BADC), the Earth Observation Data Centre (NEODC), the Environmental Information Data Centre (EIDC), and the National Geoscience Data Centre (NGDC) (NERC, 2014b).

In the field of data management in higher education in Germany, several approaches and activities are available, for example at the universities of RWTH AACHEN (2014), UNIVERSITY OF BIELEFELD (2014a), HUMBOLDT UNIVERSITY OF BERLIN (2014), TECHNICAL UNIVERSITY OF BERLIN (2014), CHRISTIAN-ALBRECHTS-UNIVERSITY OF KIEL (2014), UNIVERSITY OF GÖTTINGEN (2014), LEIBNIZ UNIVERSITY OF HANOVER (2014), UNIVERSITY OF HEIDELBERG (2014), Karlsruhe Institute of Technology (KIT, 2014), and UNIVERSITY OF MANNHEIM (2014). Internationally, many more activities in RDM are available, especially in the UK and USA. For instance, RDM services are provided at PURDUE UNIVERSITY (2014), JOHN HOPKINS UNIVERSITY (2014), MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT, 2014), and MONASH UNIVERSITY (2014).

Table 2-10: Overview of exemplary institutional repositories for research data

Institution /Repository Name	Discipline	Implementation Characteristics	Website Data Portal	Literature
DANS (EASY)	Generic, Humanities	Fedora, triple stores technology	https://easy.dans.knaw.nl/ui/home	PALAIOLOGK et al. (2012), SCHARNHORST et al. (2012)
GESIS (Datorium)	Social Sciences	DSpace	https://datorium.gesis.org	WIRA-ALAM et al. (2012), LINNE (2013)
Dryad	Biosciences	DSpace	http://datadryad.org/	GREENBERG et al. (2009), VISION (2010), WHITE et al. (2008)
Figshare	Generic	Cloud based	http://figshare.com/	FIGSHARE (2014)
TU Berlin (DepositOnce)	Generic	DSpace	http://www.szf.tu-berlin.de/	KUBEREK (2012, 2013)
University of Edinburg (DataShare)	Generic	DSpace	http://datashare.is.ed.ac.uk	RICE (2007, 2014); RICE & HAYWOOD (2011)
University of Oxford (DataBank)	Generic	Fedora	http://researchdata.ox.ac.uk/	RUMSEY & JEFFERIES (2013a); WILSON & JEFFREYS (2013); WILSON et al. (2011)
University of Southampton (ePrints Soton)	Generic	EPrints, Three-layer metadata model e	http://library.soton.ac.uk/researchdata	BROWN & WHITE (2013); TAKEDA et al. (2010)
PANGAEA	Earth & Environmental Sciences	client-server architecture, SYBASE	http://www.pangaea.de	DIEPENBROEK & GROBE (2007); DIEPENBROEK et al. (2002); GROBE et al. (2006)
WDC for Climate (CERA Database)	Meteorology	Oracle 9i, ISO 19115/19139 Metadata	http://cera.wdc-climate.de/	STOCKHAUSE & HÖCK (2010); TOUSSAINT (1999); TOUSSAINT ET AL. (1999)
ZENODO	Generic	INVENIO	http://zenodo.org/	ZENODO (2014)

Table 2-10 provides an exemplary overview of institutional repositories including additional information. Besides initiatives in higher education, national repositories are also listed, such as the Dutch Data Archiving and Networked Services (DANS). In addition, discipline-specific repositories are listed, like the Dryad repository, PANGAEA (Data Publisher for Earth & Environmental Science), and Datorium (hosted at GESIS-Leibniz Institute for the Social Sciences). In addition, figshare and ZENODO provide a platform to easily share and publish all kind of multidisciplinary research results. Further information about mentioned out-of-the-box software tools are available in section 2.4.

RDM approaches in research projects

Several RDM implementation approaches of research projects are available, discipline-specific and multidisciplinary. Table 2-11 provides examples of German RDM repositories for research projects and includes further details for each project.

Table 2-11: Exemplary overview of research data repositories of research projects

Repository Name	Discipline	Implementation Characteristics	Website Data Portal	Literature
BEFdata	Biodiversity	Ruby on Rails, Ecological Metadata Language (EML)	http://china.befdata.biow.uni-leipzig.de/	NADROWSKI et al. (2013); NADROWSKI et al. (2012)
BExIS	Biodiversity	.NET framework, DB2 with pure XML	http://exploratories.bgc-jena.mpg.de/	HEIMANN et al. (2010a); HEIMANN et al. (2010b); LOTZ et al. (2012),
CRC806-Database	Archaeology, Paleoenviron-mental Sciences	Typo3, CKAN	http://crc806db.uni-koeln.de	WILLMES et al. (2012); WILLMES et al. (2014),
CRC 882 VRE	Social Sciences	Open Atrium for Drupal, DDI metadata standard	http://www.sfb882.uni-bielefeld.de/de/inf/vfu	FRIEDHOFF et al. (2013a, b)
FOR816dw	Biodiversity	Open source components, MySQL, Ecological Metadata Language (EML)	http://www.tropicalmountainforest.org/	GÖTTLICHER et al. (2010); LOTZ et al. (2012); NAUSS et al. (2007)
Past: GLOWA Volta Geoportal; Now: Volta Basin Authority Geoportal	Earth, Sciences, Social Sciences	Past: Web-Interface, Database Now: GeoNetwork, ISO 19115 Metadata	http://131.220.109.2/geonetwork/srv/en/main.home	SHUMILOV et al. (2008); VBA-GEOPORTAL (2014)
RBIS	Earth Sciences	PostgreSQL database, MapServer, ISO 19115 Metadata	leutra.geogr.uni-jena.de/brahmaRBIS/; leutra.geogr.uni-jena.de/vgtbRBIS/; http://leutra.geogr.uni-jena.de/dinarioRBIS	KRALISCH & ZANDER (2010); ZANDER et al. (2013); ZANDER et al. (2011)
ROAD	Archaeology, Paleoenviron-mental Sciences	MapServer, PHP, JavaScript, Typo3	https://www.roceeh.uni-tuebingen.de/roadweb	MÄRKER et al. (2011); MÄRKER & KANAEVA (2010); SCHMIDT & MÄRKER (2010)
TEODOOR	Earth Sciences	PLONE, Geoserver GeoNetwork	http://teodoor.icg.kfa-juelich.de/	KUNKEL et al. (2013); KUNKEL et al. (2012); SORG et al. (2013)
West African Vegetation Database	Biodiversity	Web-interface, MySQL	http://www.westafricanvegetation.org	JANßEN et al. (2011); SCHMIDT et al. (2012)

Examples include two *Information Infrastructure*-projects of DFG-funded Collaborative Research Centres. The CRC806-Database, developed in the framework of the CRC 806 "Our Way to Europe" and the CRC 882 VRE (Virtual Research Environment), established in the CRC882 "From Heterogeneities to Inequalities". Furthermore, two RDM projects are presented, which are funded in the DFG funding program of Research Units. The RU 891 "Biodiversity-Ecosystem Functioning" (BEF-China) handles their data in the BEFdata portal and the RU 816 "Biodiversity and Sustainable Management of a Megadiverse Mountain Ecosystem in South Ecuador" (ger.: "Tropicalmountainforest") in the data warehouse FOR816dw. In addition, the Biodiversity Exploratories Information System (BExIS) represents the RDM portal of the DFG-funded Priority Program "Biodiversity Exploratories".Furthermore, the Online Data Repository (TEODOOR) of the TERENO (TERrestrial ENvironmental Observatories) project is presented, funded by the Helmholtz Association of German Research Centres. The introduced River Basin Information System (RBIS) is an exception. This information and data management portal is applied in various research projects, for example the EU-funded BRAHMATWINN-project, the BMBF-funded DINARIO project, and the BMBF-funded LUCCi project. The West African Vegetation Database is an outcome of the BMBF-funded BIOTA (Biodiversity Transect Analysis in Africa) project and the EU-funded project SUN (Sustainable Use of Natural Vegetation in West Africa). ROAD (ROCEEH Out of Africa Database) is a project database for the ROCEEH (The Role of Culture in Expansions of Early Humans) project, funded by the Heidelberg Academy of Sciences. Finally, the BMBF-funded GLOWA Volta Geoportal is an exception. It was updated from the current implementation solution to the Volta Basin Authority Geoportal on the basis of GeoNetwork.

A further exception in RDM for research projects is demonstrated in the city of Kiel within the framework of the management of marine sciences. A joint team, the so-called "Kiel Data Management Team" (KDMT, 2014), was set up funded by the Helmholtz Association Centre for Ocean Research (GEOMAR), the DFG-funded cluster of excellence "Future Ocean", the CRC 574 and CRC 754. The aim is to establish one data repository which is responsible for the entire management of all data collected within the various research projects (SPRINGER et al., 2013).

In the following chapters, the focus will be on the RDM approach of the DFG-funded research project CRC/Transregio 32 (CRC/TR32) "Patterns in Soil-Vegetation-Atmosphere Systems". Initially, the used methods and components for designing and implementing the RDM system, referred to as TR32DB, will be described in the following chapter 3. Features supported by the TR32DB RDM system will be outlined (chapter 4). Finally, the used methods and results will be discussed including a comparison to previous presented RDM approaches (Table 2-10, Table 2-11).

3 Designing the TR32DB architecture: data, methods, and components

In order to collect, document, store, backup, share, and provide research data, which are created within the framework of the Collaborative Research Centre/Transregio 32 (CRC/TR32) project, a Research Data Management (RDM) system (section 2.1.5), the so-called CRC/TR32 project database (TR32DB), was designed. The TR32DB is a self-developed project database capable to support the CRC/TR32 scientists during the entire research life cycle (chapter 2, Figure 2-1) with appropriate features including data documentation (section 2.1.3), data publishing (section 2.1.6), and data sharing (section 2.1.7). For the design of the TR32DB architecture recent standards and principles were applied with respect to the usage of metadata (section 2.2), data storage and web-based technologies (section 2.3).

This chapter will introduce the design of the TR32DB architecture with the focus on the TR32DB Version 3.1 and the improved, updated Version 4.0. Initially, an overview will be provided about the background of the CRC/TR32 project and the data, created by or purchased for the project participants (3.1). Afterwards, the TR32DB architecture will be outlined (3.2), followed by a description of the single components. At first, the design and documentation of the underlying TR32DB Metadata Schema will be depicted (3.3). Subsequently, the methods used to implement the single technical components will be presented including the file-based data storage (3.4), the data storage based on a database (3.5), and the web-interface (3.6).

3.1 The CRC/TR32 project background

3.1.1 Project description

The Collaborative Research Centre/Transregio 32 (CRC/TR32) "Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation" (www.tr32.de) is funded by the German Research Foundation (DFG). The project funding of the CRC/TR32 started in 2007. Currently, the CRC/TR32 is situated in the second funding phase, which is operating from 2011-2014. There are three funding phases scheduled, each running for four years (Phase 1: 2007-2010; Phase 2: 2011-2014; Phase 3: 2015-2018).

The CRC/TR32 is an interdisciplinary, joint research project between the German Universities of Aachen, Bonn, Cologne, and the Research Centre Jülich. The involved research partners cover several scientific disciplines. These are for example soil and plant science, hydrology, geography, meteorology, geophysics, remote sensing, and mathematics. They work in different project sections on exchange processes between the soil, vegetation, and the adjacent atmospheric boundary layer (SVA). Their overall research goal is to yield improved numerical SVA models to predict CO_2-, water- and energy transfer by calculating the patterns at the various spatial and temporal scales. The hypothesis of the CRC/TR32 covers the explicit consideration of structures and patterns, which lead to a common methodological framework. The project participants focus their study on the catchment area of the river Rur, situated in western Germany, parts of the Netherlands, and Belgium. The field research of the project participants is focused on three sub watersheds that represent the three typical land use forms of the study area (forest, arable-, and grass land) (CRC/TR32, 2012).

The CRC/TR32 is subdivided into different project areas, the so-called clusters. The CRC/TR32 consists of four scientific clusters (A, B, C, D) and a central, coordination cluster (Z). The specific clusters focus their work on:

A: Characterization of flow, transport, and structure of unsaturated soils at the µm to m scale by non-invasive methods

B: Combining non-invasive measurements and modelling to analyse the spatio-temporal variability of water and matter fluxes at the field-scale

C: Analysing and modelling spatio-temporal variability of water and solute fluxes from micro- to macroscale catchments

D: Land surface - atmosphere exchange processes with a special focus on CO_2 and H_2O.

Z: Central Administration and Data Management

The scientific clusters differ by the sub-system (Soil-Vegetation-Atmosphere) and on the spatial scale range (local, sub-basin, and regional/basin), on which they concentrate (Figure 3-1).
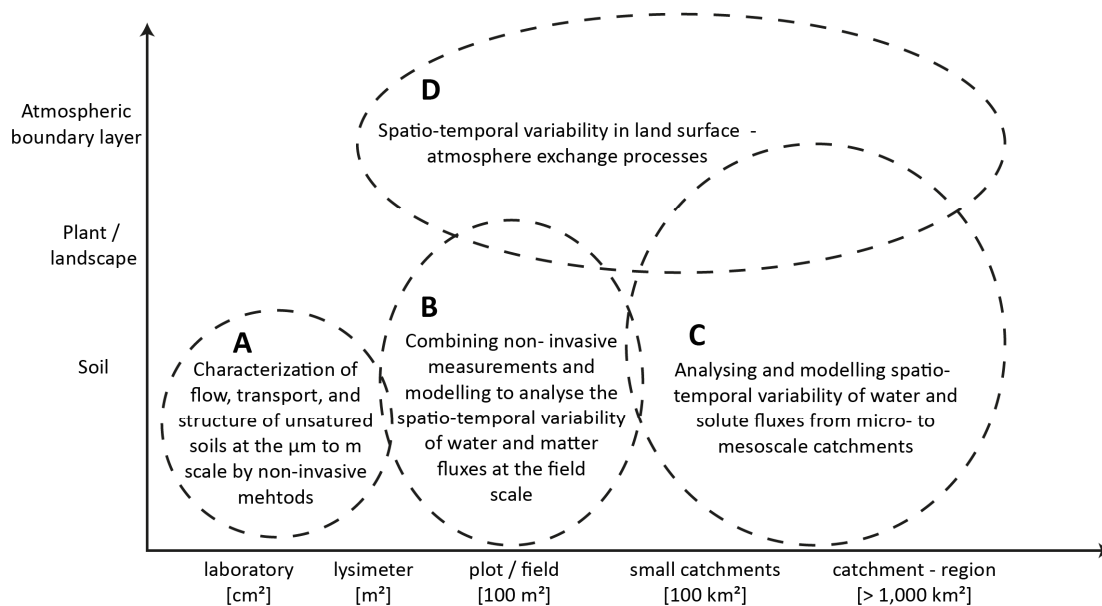
Figure 3-1: Focus of CRC/TR32 project clusters, modified after CRC/TR32 (2007)

Each cluster is split up into sub-projects (SP), for example A1, A2, B1 or C4. Currently, the DFG is funding 24 SPs in the second funding phase. Within each SP around 1-2 PostDocs and 1-2 PhDs plus Master/Bachelor students are working on the research goal of the CRC/TR32.

The CRC/TR32 participants produce a lot of different research data (section 2.1.3) as a result of the interdisciplinary background of the project. They work on the SVA topic with focus on various spatial scale ranges (laboratory, field plot, sub region, catchment region) and in various temporal scale ranges. To validate and also to be able to model research results, the project participants are dependent on adequate geodata of different scales, which were purchased within the framework of the CRC/TR32. The outcome of the project occurs in a huge amount of data in various file formats and file sizes. A detailed description of the CRC/TR32 project data will be given in the next chapter.

3.1.2 Project data

The TR32DB has to manage a great number of various data. These are data created within the research life cycle (section 2) of the CRC/TR32 by the project scientist and which are purchased from different institutions. These data will be described in the following chapter.

Scientific research data of the CRC/TR32

The CRC/TR32 scientists produce a multiplicity of heterogeneous data as a result of the interdisciplinary background during the project funding term. In accordance with the structure of the CRC/TR32, they focus their work on the SVA sub-system, various spatial and temporal scale ranges. Therefore, the scientists carry out research on the soil, vegetation, and the lower atmospheric boundary layer. They additionally work in various spatial scales ranging from the laboratory scale (µm - m), lysimeter (m²), plot/field scale (100 m²), sub catchment scale (100 km²), and catchment region scale (> 1000 km²). Furthermore, the CRC/TR 32 scientist do measurements or modelling in temporal scales ranging from the short time to the long-term scale, for example: minutely, hourly, daily, weekly, monthly, yearly , or for an entire decade (CRC/TR32, 2010).

The results from the scientists include laboratory methods, single or sensor networks, meteorological and hydrological monitoring, as well as airborne/spaceborne remote sensing. For example, the outcome of several field or flight measurement campaigns include results like wind temperature and air CO_2 concentration (SELBACH et al., 2010), soil water content and soil moisture (KORRES et al., 2010), soil CO_2 concentration (GRAF et al., 2011), leave characteristics and root growth (SCHICKLING et al., 2010), land use analyses (WALDHOFF & BARETH, 2009), and crop surface or volume models (HOFFMEISTER et al., 2008). In addition, some research partners produce results from laboratory measurement methods (BREEDE et al., 2011) or model simulations (SCHOMBURG et al., 2010; ZACHARIAS et al., 2012). Some results and measurements are shown in Figure 3-2.
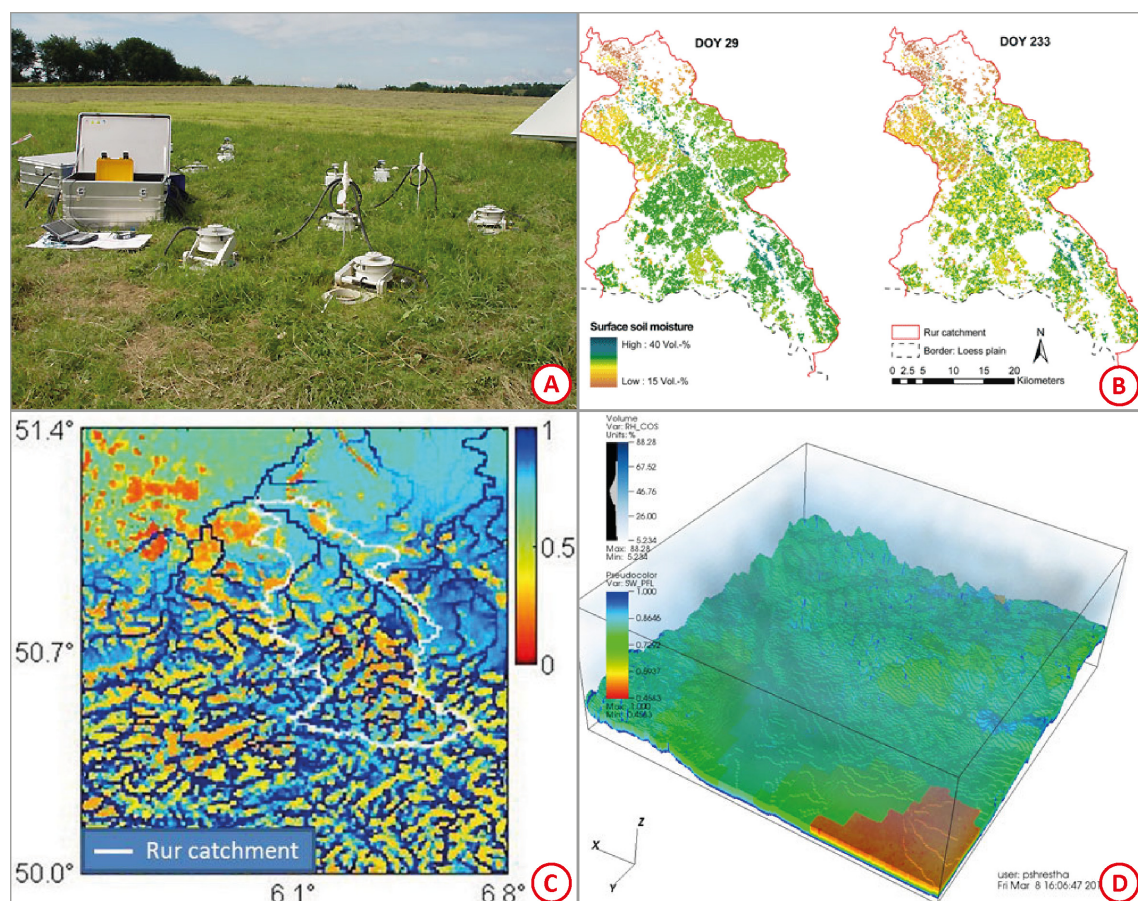
Figure 3-2: Exemplary research of the CRC/TR32: (A) measurements of soil carbon dioxide exchange processes; (B) modeled surface soil moisture; (C) relative surface saturation simulation; (D) simulation of lateral flow processes (CRC/TR32, 2014b)

As a consequence of the various results from field measurement and airborne campaigns, the corresponding analyses and model simulations, all project participants create a great number of publications, pictures,

presentations, and reports. The format and size of the CRC/TR32 data varies due to different research backgrounds of the project participants. Therefore, the project participants provide different data formats, such, as MS Excel, ASCII, NetCDF, binary format, PDF, JPEG, TIF, or geodata formats. In addition, the size of a single dataset varies from only a few kilobytes to several gigabytes.

Purchased data from organisations

To support and cover the needs of the project participants concerning their modelling and estimation purposes, several data were obtained from various institutions. For instance, adequate data that are related to a geographic reference (section 2.3.5), also called geodata, were purchased in different scales within the framework of the CRC/TR32. For example, these data were ordered from the district survey and mapping agency of North-Rhine Westphalia (ger.: *Bezirksregierung Köln*), the federal agency for cartography and geodesy (ger.: *Bundesamt für Kartographie und Geodäsie*), the district agency of geology (ger.: *Geologischer Dienst Nordrhein-Westfalen*), the district environmental agency of North-Rhine Westphalia (ger.: *Landesumweltamt Nordrhein-Westfalen*), the European environment agency, the United States Geological Service (USGS), and National Aeronautics and Space Administration (NASA). In addition, data from the German weather forecast agency (ger.: *Deutscher Wetterdienst, DWD*), the Meteomedia Group, the Eifel-Rur water resources association (ger.: *Wasserverband Eifel-Rur*), and the Erft water resources association (ger.: *Erftverband*) were obtained. These data include soil, hydrological, geological, land register, elevation, topographic, geomorphological, land use, remote sensing, and climate data. Subsequent, some of the data will be described, which are of importance for the study described in this thesis.

> **Cadastral map** (ger.: *Automatisierte Liegenschaftskarte, ALK*) was purchased from the district survey and mapping agency of North-Rhine Westphalia for three sub-regions of the CRC/TR32 study area. This map contains cadastral information of the parcels and buildings. The information can be used to identify plot number or area of a specific plot of land (BEZIRKSREGIERUNG KÖLN, 2014b).
>
> **Digital landscape model** (ger.: *Amtliches Topographisch-Kartographisches Informationssystem or ATKIS Basic-DLM*) was purchased from the district survey and mapping agency of North-Rhine Westphalia for the entire CRC/TR32 study region in 2007, provided in vector format, known for a high spatial accuracy and resolution that provides topographical information of prominent landscape features (BEZIRKSREGIERUNG KÖLN, 2014a).
>
> **Climate data** was obtained from the German weather forecast (DWD), the Meteomedia Group, Rur water resources association, and the Erft water resources association since 2007 yearly. These contain, for example climate parameters like cloud height, wind direction, temperature or the daily sum of sunshine duration (DWD, 2014).
>
> **Soil map information system** (ger.: *Informationssystem Bodenkarte*) obtained from the district agency of geology of North Rhine-Westphalia contains the digital, replenished, actual information of the according soil maps in a scale of 1:50,000. The data imply important soil parameters, such as soil type, field capacity, and cation exchange capacity (GEOLOGISCHER DIENST NRW, 2014).
>
> **Digital elevation model (DEM)** was purchased for the German parts of the Rur-catchment from the district survey and mapping agency of North-Rhine Westphalia. The official 5 m resolution model was filtered and combined by the Scilands Cooperation in a 10 m resolution model (BEZIRKSREGIERUNG KÖLN, 2014c).
>
> **Digital Orthophotos** were obtained from the district survey and mapping agency of North-Rhine Westphalia of the complete CRC/TR32 study area of the years 2007-2008 and 2010 (BEZIRKSREGIERUNG KÖLN, 2014d).

3.2   The TR32DB architecture overview

Within the framework of the CRC/TR32, the CRC/TR32 project database (TR32DB), was setup. The TR32DB is accessible online at www.tr32db.uni-koeln.de. The complete TR32DB is implemented in cooperation with the Regional Computing Centre of the University of Cologne (RRZK), where it is also hosted. The TR32DB is designed in order to ensure a centralized, sustainable storage and backup of all project data, which are created or purchased within the context of the CRC/TR32. This also includes the accurate presentation, access, and documentation of the data with detailed metadata, during the funding of the CRC/TR32 and beyond. For this purpose, a TR32DB RDM system was setup, which is based on a three-tier client-server architecture (section 2.3.1), presented in a simplified scheme in Figure 3-3.

The TR32DB three-tier architecture is organised in a data layer, application layer, and presentation layer. The data layer (resource management layer) included the persistent storage of the TR32DB data. A file-based data storage, operated by the Andrew File system (section 3.4), is provided. In addition, a further data storage, based on a MySQL database, is implemented (section 3.5). The designed TR32DB Metadata Schema (section 3.3) serves as the basis for the database structure. The application layer includes the application server components, like a map server or web server. These communicate with the data layer and presentation layer and perform the processing of the requests (e.g. data access). The presentation layer includes the web-browser based representation of the transferred information in the TR32DB web-interface including the internal WebGIS mapping component. The latter are described in section 3.6.

Figure 3-3: Simplified TR32DB three-tier architecture
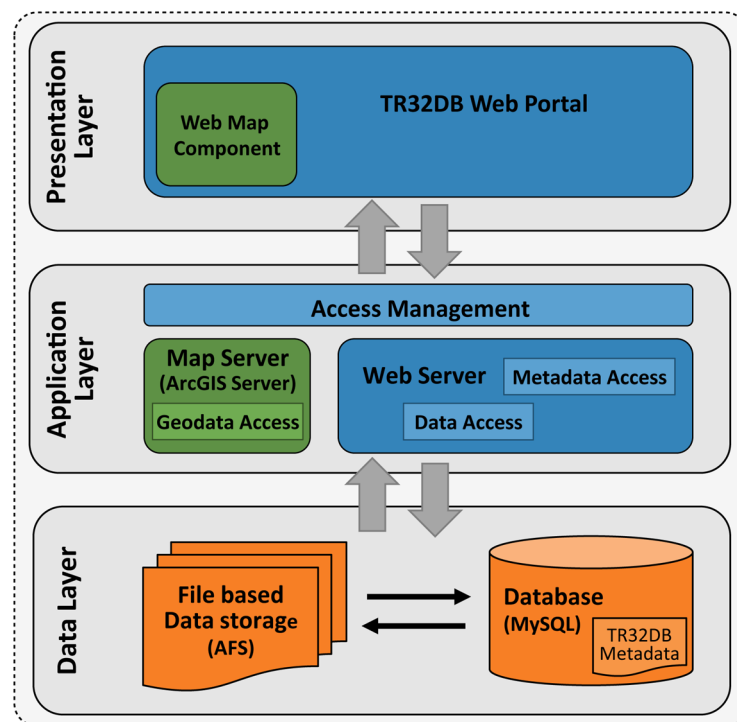
The documentation of the data and correspondingly the management of the metadata is an essential component of the TR32DB. Hence, at first, the focus will be on the design and documentation of the self-developed TR32DB Metadata Schema (section 3.3) and then the following paragraphs will describe the methods and design of the other main components of the TR32DB in detail.

3.3 The TR32DB Metadata Schema: design and documentation

This section focuses on the design and documentation of the TR32DB Metadata Schema. After a short introduction, the used metadata standards and schemas will be described, as well as the creation of the TR32DB Metadata Schema.

In order to describe and document all data (section 2.1.3) stored in the TR32DB with accurate, standardized metadata, a metadata schema or standard has to be chosen, extended or designed and finally documented. With regard to the interdisciplinary background of the CRC/TR32 and the amount of heterogeneous data created by the project participants (section 3.1), the application of one single metadata schema or standard to describe all types of data is not possible. Thus, a new metadata schema has to be provided to meet the requirements and to suit to the needs of the TR32DB data. In addition, it is important to apply a simple and user-friendly schema, which does not overburden the TR32DB users. Consequently, the new TR32DB Metadata Schema was designed and documented to support the metadata management in the TR32DB. The creation of the TR32DB Metadata Schema was carried out hand-in-hand with the implementation of the TR32DB user interface (e.g. enter, search, browse, display of metadata) and the design of the TR32DB database starting in 2007. This chapter outlines the process of creation.

The first step in the design of the metadata schema (section 2.2.2) was to analyse the framework conditions and current status concerning the content and context that should be considered in the metadata schema. Initially, the CRC/TR32 project proposal provided a basis. To archive a further overview, a survey was conducted among the participants of the CRC/TR32 in 2007. This survey (appendix A) focused on general questions about the participation with a centralized data management system, the data delivery (kind, format, file size), and the operating system and software used. The result of this survey showed the heterogeneity of research data expected to be stored and documented with metadata in the TR32DB. Besides measured or modelled research data from different disciplinary backgrounds with focus on Soil-Vegetation-Atmosphere (e.g. laboratory data, field measurement data, meteorological model output, point data, or time-series-data), also the handling of additional information (e.g. publications, thesis, pictures, and reports) produced by the project participants was requested. As a result of the variety of the different data, six main data type categories emerged: data, basic geodata, reports, presentations, publications, and pictures. Consequently, the TR32DB metadata schema had to be designed according to the interdisciplinary background of the CRC/TR32 data, and the demands of the users.

In a second step, it was necessary to get an overview of available metadata schemas and standards (e.g. RILEY, 2009a) with regard to all requested data, which could be used for the purpose to describe these data. In addition, existing self-developed metadata schemas and metadata profiles from other research projects were identified and inspected (e.g. BACKES et al., 2005; GÖTTLICHER & BENDIX, 2004; HOERNES & ROWECK, 2007; MÜCKSCHEL et al., 2005; SHUMILOV et al., 2008; TOUSSAINT, 1999). Finally, as a start, the Dublin Core Metadata Element Set (simple Dublin Core) (DCMI, 2012b)(section 2.2.2.1) was chosen as a *basic* metadata schema for the TR32DB describing all six data type categories. Dublin Core is a simple and widely accepted standard, covering all general requirements and information to describe all kind of data. In the course of the project duration, the simple Dublin Core was adapted with the DCMI Metadata Terms (DCMI, 2012a; section 2.2.2.1). Moreover, to accommodate the variety of requirements of the six chosen data type categories and the needs of the CRC/TR32, the base schema was further supplemented. By using a *mix and match* approach (section 2.2.2), a self-developed multi-level TR32DB Metadata Schema was developed, presented in Figure 3-4.

In order to meet the demands of the project background, CRC/TR32 specific properties were added to the metadata schema. Subsequently, general data type specific properties were complemented, as well as properties according to the special metadata standard (ISO 19115:2003 and INSPIRE Implementing Directives). Finally, to meet the demands of the project participants concerning the registration of DOIs, the DataCite Metadata Schema Version 2.2 (section 2.2.2.2) was also included in the TR32DB Metadata Schema.

In summary, the TR32DB Metadata Schema defines three levels of detail:

General properties are used to describe the data with the basic information. They are used to make the data citable and findable. For example: title, creator, or description of a dataset.

Project specific properties are used to describe the data with properties, which focus on specific needs and demands of the project background and topic. For example: keywords with focus on the CRC/TR32, status of the membership in CRC/TR32, TR32DB identifier for the dataset.

Data type specific properties are used to describe the data with properties, which are specific for each data type category. For example: publication type (e.g. article, event paper, and book section), lineage of dataset, and name of the instrument or modelling method used to create the dataset.
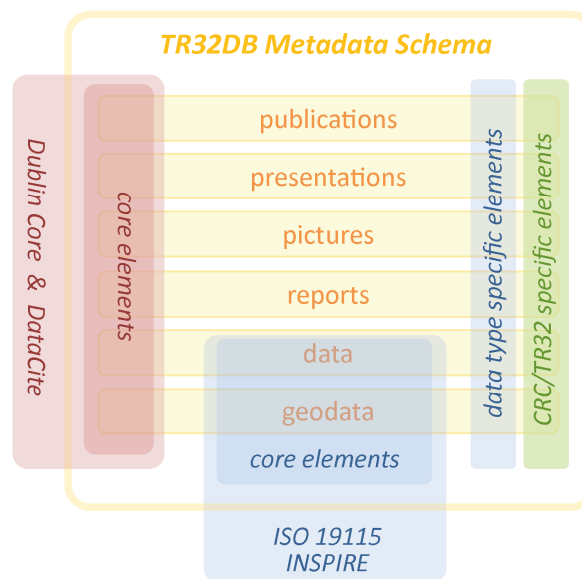
Figure 3-4: TR32DB Metadata Schema overview

In detail, for the creation of the single metadata properties or sub-properties of the TR32DB Metadata Schema, elements of the DCMI Metadata Terms (section 2.2.2.1) were combined with properties (elements) of the DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011; section 2.2.2.2), core elements of ISO19115:2003 Metadata Standard (ISO, 2003; section 2.2.2.3), as well as metadata elements according to the INSPIRE Implementing Directive (EUROPEAN COMMISSION, 2008) and the INSPIRE Metadata Implementing Rules (DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE, 2010) (section 2.2.2.4). Furthermore, the new metadata properties were added with focus on the CRC/TR32 background as well as with regard to the data type category requirements. They were added on demand by the project participants (e.g. instrument property) or with regard to an accurate citation of the datasets (e.g. journal name, volume and issue). For example, the TR32DB property `Creator`, presented in Table 3-1 is composed of some sub-properties adopted from Dublin Core, ISO 19115:2003, INSPIRE Implementing Directive, DataCite Metadata Schema Version 2.2 elements, as well as an additional own established TR32DB property (complete mapping see appendix D.4).

Besides an identifier, title and short definition for each TR32DB property and sub-property, obligations were defined. Three levels of obligations were allocated for single properties and sub-properties: mandatory (M), optional (O), and automatic (A). Mandatory properties must be provided, optional may be provided, and automatic properties are generated automatically by the TR32DB. To prevent the project participants being overtaxed with too much mandatory metadata properties, a minimum set of mandatory properties and sub-properties was determined. For this, only properties and corresponding sub-properties were selected that are essential for the description, citation (section 2.1.6) and findability of the data.

In addition, the repeatability or cardinality of each property and sub-property was indicated. In the first instance, only single use of the properties and sub-property was supported, with the exception of the property `keyword`. This was adapted. In the latest TR32DB Metadata Schema for each property and sub-property the occurrence is determined (e.g. optional and repeatable (0-n) or required and repeatable (1-n)).

Table 3-1: Mapping of TR32DB metadata property `Creator` and some associated sub-properties. Mandatory properties are highlighted.

ID	TR32DB property	DC Elements DC Terms	ISO 19115:2003	OB	INSPIRE	OB	DataCite	OB
7	Creator							
7.1	Creator-Person	dc.creator dcterms.creator					Creator	M
7.1.1	creatorRole		CL_Responsible Party.role	O	Responsible party role	M		
7.1.2	creatorFirst Name	dc.creator dcterms.creator	CI_Responsible Party.individual Name	M			CreatorName	M
7.1.3	creatorFamily Name	dc.creator dcterms.creator	CI_Responsible Party.individual Name	M			CreatorName	M
7.1.4	nameIdentifier						Name Identifier	C
7.1.4.1	nameIdentifier Scheme						nameIdentifierSch eme	C/ M
7.1.5	emailAddress		CI_Address. electronicMail Address	O	Responsible party	M		
7.1.6	TR32Member ship							

Moreover, for each property and sub-property a data value is defined (e.g. free text, date or controlled list). For example, date values should follow any format described in W3C Date and Time Format (W3DTF, 2014), e.g. YYYY-MM-DD. Controlled vocabulary lists refer either to a predefined list, provided from one of the metadata standards, metadata schemas etc., or to a self-established controlled value list. Predefined controlled vocabulary lists are applied for example for the sup-properties `contributorType` provided by DataCite Metadata Schema, `roleType` provided by ISO 191115 and INSPIRE Implementing Directive, or `language` provided by ISO 639-2. If predefined controlled vocabulary list are available in more than one of the used metadata standards or schemas, then the values are mapped to each other. For example, the sub-property `titleType` is mapped between elements of Dublin Core, DataCite Metadata Schema, and ISO 19115:2003. Self-established controlled lists were prepared with the expertise of the CRC/TR32 project members. They were interviewed to define relevant properties and content. For example, specific controlled lists were set up for significant keywords that focus on the CRC/TR32 topic, measuring instruments or modelling methods and corresponding output parameter used by the project participants, as well as significant CRC/TR32 locations and regions. In addition, the controlled vocabulary lists `creatorList` or `insitutionList` are created and continuously updated by the TR32DB users.

The final step was to describe the TR32DB Metadata Schema in a document. The documentation of the schema started at the End of 2013 and was published in April 2014. For this purpose, available examples of metadata schemas, application profiles, and metadata crosswalks documentation and guidelines were identified and inspected (e.g. DRYAD DEVELOPMENT TEAM, 2010; ENSOM & WOLTON, 2012; GIBBS, 2009; HAUSSTEIN et al., 2013; RICE et al., 2008; VIERKANT et al., 2012; ZENK-MÖLTGEN & HABBEL, 2012),

besides the used metadata schemas and standards. Since the examined examples are described very differently, a compromise was chosen to describe the developed TR32DB Metadata Schema. Initially, a general introduction to the metadata schema is provided. Afterwards, a clearly structured listing of all properties and sub-properties with a detailed description was established. Finally, a crosswalk to the used metadata standards, schemas, or implementing directives, as well as other related ontologies was described, as well as all available controlled vocabulary lists were presented in the appendix of the metadata schema. The described approaches are following the steps of metadata schema design according to MILLER (2011)(section 2.2.2).

3.4　File-based data storage and organisation

The TR32DB is developed in close cooperation with the Regional Computing Centre of the University of Cologne (RRZK, 2014a). The RRZK provides and administrates the infrastructure for the TR32DB. Therefore, initially the data storage environment of the RRZK will be described following by the TR32DB data storage organisation and implementation.

3.4.1　Data storage at the RRRZ

The data storage at the Regional Computing Centre of the University of Cologne (RRZK) is organized in different stages (KALLE & TIGGEMANN, 2007). These stages comprise the Storage-Area-Network (SAN) based Disk Storage, which provides the Andrew File System on top of the SAN. Another stage is the High Performance Computing (HPC) System, which incorporates the Lustre and PanFS file systems The Scale out File Services (SoFS) is based on the NetApp file system (KALLE, 2005; RRZK, 2014b, c; WONCZAK, 2014). A simplified overview of the single data storage stages in the RRZK is presented in Figure 3-5. Between the single storage systems, staging processes can be executed.

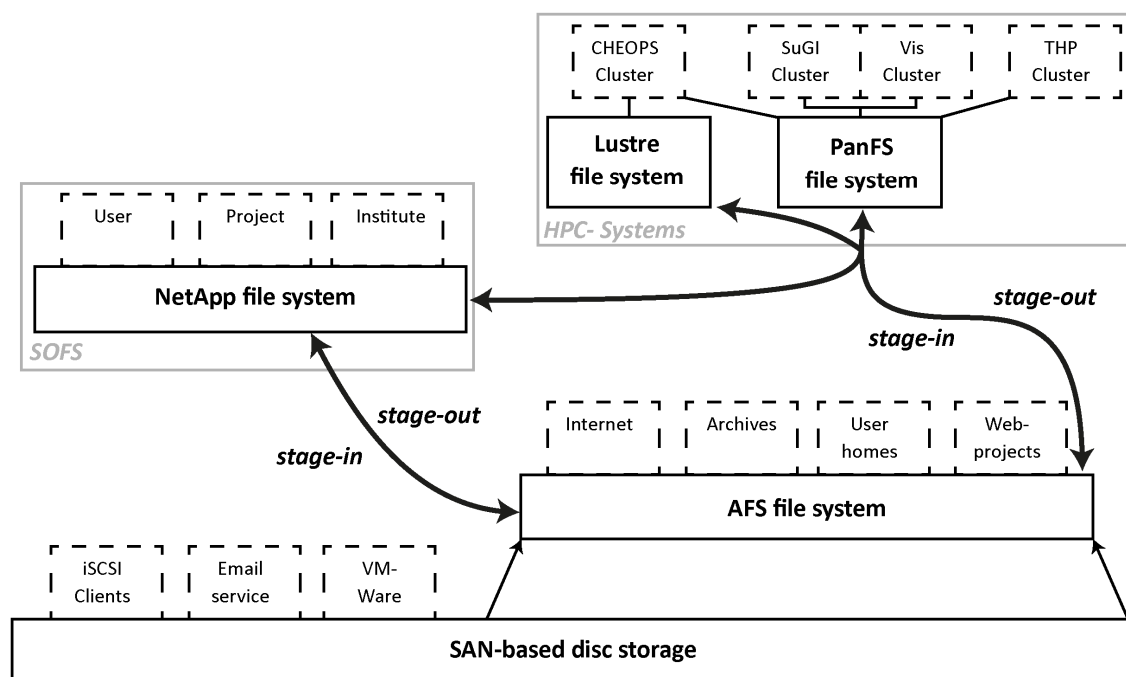

Figure 3-5: Storage implementation at RRZK, modified after KALLE (2005) and WONCZAK (2014)

This section will focus on the Andrew File System (AFS), because it is used within the infrastructure of the TR32DB. The implemented AFS System at the RRZK has a SAN-based Disk-Storage (section 2.3.2) as back-end storage (KALLE, 2005; WONCZAK, 2014). The AFS serves at the RRZK, for example, as back-end

storage for all web-projects including web-space hosted at the RRZK (e.g. www.uni-koeln.de, www.rrzk.uni-koeln.de, or www.tr32db.uni-koeln.de). In addition, the AFS manages the user home directories of all employees and students of the University of Cologne. Furthermore, the AFS serves or has served as an archive storage area for research projects (RRZK, 2013a) like Arachne (ARACHNE, 2013), CEEC (CEEC, 2014), Monasterium (MONASTERIUM, 2014), or Prometheus (KALLE & DIEKMANN, 2010; PROMETHEUS, 2014).

The Andrew File System (AFS)

The AFS is a distributed networked file management system based on the TCP/IP protocol and is designed on the basis of a client-server architecture developed in the early 1980s. BORENSTEIN (1996) and MORRIS et al. (1986) emphasize that at the end of 1982 the Carnegie Mellon University (CMU) started it in cooperation with IBM the *Andrew Project*. The project is named after the two benefactors of CMU Andrew Carnegie and Andrew Mellon. The aim was to develop a system to support the university's needs. For example, to provide a system to campus-wide easy share data (files) between the people and several departments of the University. LEONG (1994) provides further background information about the *Andrew Project*.

By far the most successful and significant component of the *Andrew Project* is the AFS. According to HOWARD (1988), the AFS was designed as follows: a system, which is compatible with UNIX at the system call level, uses whole files as the basic unit for moving and storing data and finally it is implemented with many small relation servers. In 1984 the first version of the system was online and in 1989 it was commercially available. Since 2000, the source code of the AFS is publicly available for community development and maintenance. This was the launching of the OpenAFS open-source project. In the following AFS will be used instead of OpenAFS as they are synonymous (MILICCHIO & GEHRKE, 2007).

MILICCHIO & GEHRKE (2007) and CAMPBELL (1998) point out that there are main benefits of AFS in comparison to other file systems like NFS. These are for example the location and system independence. The AFS can be accessed from various operation systems. It provides location independence and is for example accessible from a range of various systems like Microsoft Windows, UNIX, Linux, and MacOS X (OPENAFS, 2013). MILICCHIO & GEHRKE (2007) and CAMPBELL (1998) complement that another benefit is the high scalability. AFS can be used in small to very large environments. Furthermore, facilities for replication and backup are available. All data can be shared via a single worldwide namespace. Last but not least the centralized administration model, which is based on a distributed service, is another advantage. In addition, the AFS ensures an improved security model (access and group management) via an encrypted networked communication and Kerberos authentication service. For example the user administration is implemented via an Access Control List (ACL). The implementation of security aspects within AFS is described by SATYANARAYANAN (1989) in detail.

A high-level overview of the AFS architecture and the components is shown in Figure 3-6. CAMPBELL (1998) describes the communication of the several components as follows. The desktop client computer caches all information from the AFS onto its local disk. The caching reduces the network traffic and the server load enormously. The various servers respond to requests to store and fetch data. However, the central AFS file namespace is the most interesting part of the AFS architecture. All desktops and users can view the seamless file namespace provided by the servers. The directory names and files in the namespace are visible in a single, consistent directory structure and related to physical data on the servers. The files are physically located on disks and stored in logical containers, the volumes. According to SIDEBOTHAM (1986), a volume is a collection of files, which format a partial sub-tree of the file system hierarchy. All files of the volume are located on the same server. The amount of data that can be stored in a volume is limited to a given size, called quota. The default quota of a volume is 5 megabytes. This can be changed by the AFS administrator CAMPBELL (1998) and IBM (2000). MILICCHIO & GEHRKE (2007) and CAMPBELL (1998) complements, that it is possible within AFS to move volumes between file servers without interrupting the services of the user.

Due to the connection between volumes and path names, the enlargement of any area of the namespace is possible. In addition, the authors explain the necessity of databases servers, for example to store information about permissions and volumes. The collection of all file servers and database servers, which form a single administrative unit, is called an AFS cell. CAMPBELL (1998) names some global publicly accessible AFS cells, examples are:

- uni-hohenheim.de (University of Hohenheim)
- desy.de (Deutsches Elektronen-Synchrotron)
- rl.ac.uk (Rutherford Appleton Lab, England)
- cs.arizona.edu (University of Arizona, Computer Science Dept.)
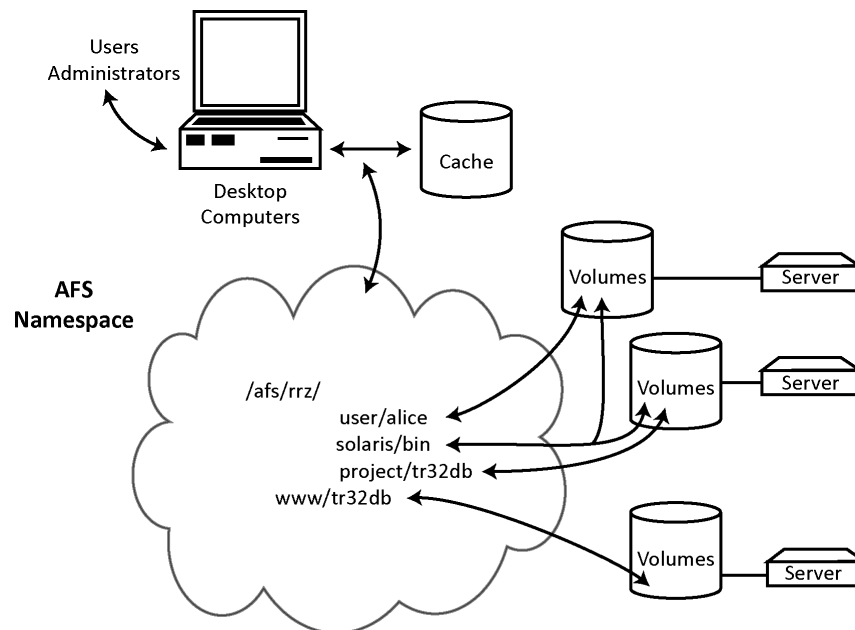
Figure 3-6: AFS Architecture, modified after CAMPBELL (1998)

The AFS cell of the RRZK is called `rrz.uni-koeln.de`. The corresponding shortcut is `rrz`. Therefore, the AFS file space or file path within the RRZK starts with `/afs/rrz.uni-koeln.de/...` or `/afs/rrz/...` . For example a user home directory is located at `/afs/rrz.uni-koeln.de/user/alice` or a web-project space at `/afs/rrz.uni-koeln.de/www/tr32db` (RRZK, 2013a).

To access a location in the AFS file space, a client requires a valid authentication credential provided by the database servers. In AFS, this is called a token, which is equivalent to a Kerberos ticket (MILICCHIO & GEHRKE, 2007). Additionally, CAMPBELL (1998) describes a token as a data object with an AFS identity and an expiration data. This time limitation is set by the AFS administrator. Within the AFS cell of the RRZK, the access to a location in the RRZK AFS file space requires a student-email account or a University account as authentication credentials provided by the RRZK. In consequence of a successful authentication, the user receives a valid token time limited to 25 hours by the RRZK AFS administrators. To receive information about the token like the expiring date, the command `tokens` can be used (RRZK, 2013a).

```
alice@dialog5$ tokens
Tokens held by the Cache Manager:
User's (AFS ID 68867) tokens for afs@rrz.uni-koeln.de [Expires Apr 1 13:07]
   --End of list—
```

The permission to access a directory or file in AFS is arranged by using Access Control Lists (ACL). For each directory, AFS stores an ACL that defines users and groups and the manner they are allowed to access the directory and all included files. For example, when a file moves to a new directory, the ACL of the new directory will be applied to the files. A `group` can be defined by a user and include both users and machines. Consequently, a user belonging to a group has the same ACL permissions awarded to the group. Reversely, users logged in to a machine that belongs to a group take over the same permissions granted to the group. By default, AFS defines three system groups: (i) `system:anyuser` includes everyone who can access the AFS cell, (ii) `system:authuser` includes all user from the local AFS cell with a valid token and (iii) `system:administrators` includes the system administrators of the AFS cell (CAMPBELL, 1998; IBM, 2000).

CAMPBELL (1998), IBM (2000), and MILICCHIO & GEHRKE (2007) describe that the AFS ACL permissions has by default seven standard rights. These can be divided into two groups: (i) applies permission to the directory and (ii) provide permissions to the files included in that directory.

The directory permissions including their abbreviated character code and description are:

> **`lookup(l)`** lists the content of the directory
> **`insert(i)`** creates new files or subdirectories' to a directory
> **`delete(d)`** removes files or subdirectories from the directory
> **`administer(a)`** changes the directory's ACL attributes

The file permissions including their abbreviated character code and description are:

> **`read(r)`** reads file content and query file status
> **`write(w)`** modifies file content
> **`lock(k)`** locks files for reading

In addition, the permissions characters can also be combined in an ACL entry. There is a shorthand notation for the following file permission:

> **`read`** is equivalent to rl (read, lookup) permissions
> **`write`** is equivalent to rlidwk (read, lookup, insert, delete, write, lock) permissions, all rights except a (administer)
> **`all`** is equivalent to rlidwka (all seven standard permissions)
> **`none`** removes all ACL permissions, leaving the user or group with no permission

The commands to list and modify the ACL will be described in the next section in detail.

3.4.2 TR32DB data storage: organisation and administration

The TR32DB data storage is the physical storage of all CRC/TR32 project data (section 3.1.2) that are created by the project participants (e.g. measured/modelled data, publications, etc.). It is carried out in the already established and described AFS. This file system was chosen in cooperation with the RRZK that provides and administrates the AFS infrastructure. In detail, the TR32DB data storage within the AFS is organized in a folder system or file directory. The organisation and design of the directory structure is developed in close cooperation with the RRZK. Experiences gained in further RRZK AFS projects, like Arachne (ARACHNE, 2013) or Prometheus (KALLE & DIEKMANN, 2010), as well as background information about the CRC/TR32 are used to develop the present AFS directory structure. The TR32DB AFS file space is accessible at `/afs/rrz.uni-koeln.de/project/tr32db`, subdivided in subdirectories.

These are:

> **Data** internal storage of purchased geodata, climate data, etc.
>
> **Phase(Phase_1,Phase_2,Phase_3)** sustainable storage of CRC/TR32 project data according to project funding phases in which they are created
>
> **Scripts** storage of AFS scripts, e.g. data moving scripts and corresponding log files
>
> **Share** storage of temporary exchange data
>
> **Upload(Upload_1,Upload_2,Upload_3)** temporary storage for uploaded CRC/TR32 project data (according to project funding phases in which they are created) before moving to corresponding sustainable `Phase` directory

The focus of the next sections will be on the subdirectories that handle the CRC/TR32 project data (`Phase`, `share`, `Upload`). These subdirectories are again subdivided in a specific hierarchy that corresponds to the already mentioned project structure of the CRC/TR32 in clusters (`A`, `B`, `C`, `D`, `Z`) and project sections (e.g. `A1`, `A2`, `B1`, `B2`, `C3`, `C4`, etc.). All `Phase` folders can be extended according to the next project funding phase. Furthermore, the `Phase` and `Upload` contain each subdirectory for the five data types, which can be stored. These are: `data`, `pictures`, `presentations`, `publications`, and `reports`. A special case appears in project sections Z1, where an additional subdirectory `basicGeodata` is added. These subdirectories are again subdivided in further directories starting with `000`, `001`, `002`, `003`, etc. The use of numbered folders for storing project files enables the scalability of the TR32DB data storage. Each of these directories finally contains one subdirectory for `metadata` storage (Figure 3-7).

The described file directory was set up by the AFS system administrator of the RRZK. The TR32DB default value for volume quota of the individual folders is set to 8 GB, as shown in Figure 3-7. As a consequence, the maximum size of files that can be stored in a specific folder is limited to 8 GB. This will be explained in detail giving three examples:

- The maximum file size in total that can be stored to a specific data type folder per project section in the `Upload` directory is 8 GB,
  e.g. `/afs/rrz/project/tr32db/Upload_1/A/A1/data`.

- The maximum storage of data files in each subdirectory of the `phase's` folders,
  e.g. `/afs/rrz/project/tr32db/Phase_1/A/A1/data/000` is 8 GB.

- In the `share` directory, each cluster can store 10 GB,
  e.g. `/afs/rrz/project/tr32db/share/A`.

Thus, in the second example, project A1 can theoretically store 8,000 GB (8 TB) of data type `data` in the `Phase_1` folder. As the storing folders for each data type can be expanded from 000 to 999, the same amount of data can be uploaded for all other `data` folders as well.

For the administration and navigation in AFS directories, CAMPBELL (1998) and MILICCHIO & GEHRKE (2007) name various AFS suites and commands. The AFS suite `fs` (file system) contains various subcommands to manipulate AFS files or the client view of AFS, e.g. the directory or volume properties of the AFS file space can be viewed or modified. For instance, the command `listquota` lists information about the volume quota of the current directory. The result represents the volume name, volume quota size, the amount of storage space used in the volume and the available room on the disk partition.

```
alice@dialog5$ cd /afs/rrz/project/tr32db/share/C
alice@dialog5$ fs listquota
Volume Name                 Quota      Used %Used    Partition
Hpro.tr32.sh.C            10000000   633034   6%          41%
```

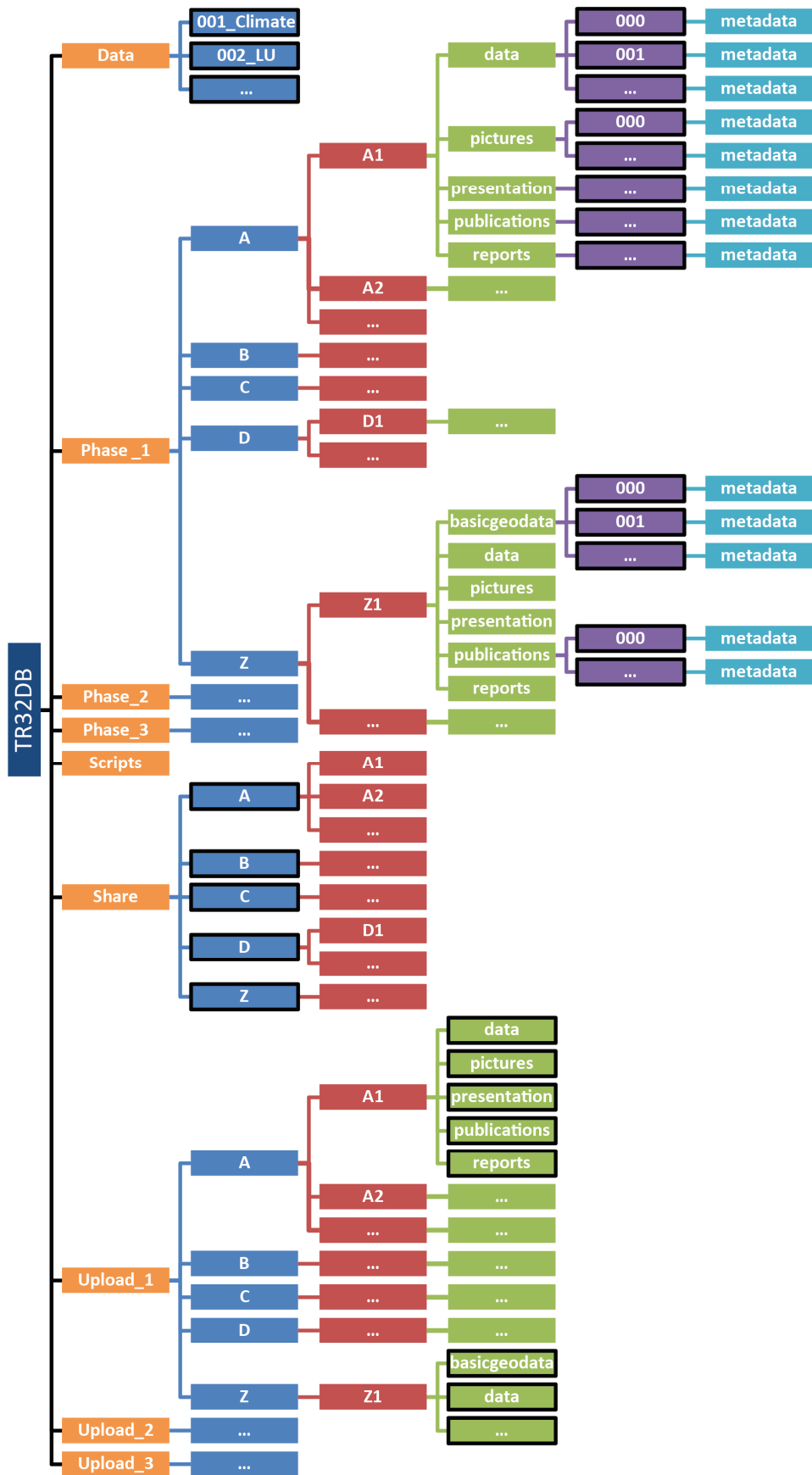Directory with volume. TR32DB default value quota 8 GB.

Figure 3-7: TR32DB data storage structure in the AFS

The access to the TR32DB AFS data storage directory is managed by various AFS groups. For this purpose three TR32DB user group types were set up by the AFS administrators. These are: the TR32DB admin group, the TR32DB project group, and TR32DB project section groups.

Admin group All TR32DB administrators are members in the admin group `webuser:tr32db_a`. This group has the `all (rlidwka)` permission for the entire TR32DB AFS directory. This enables all group members to access and modify adjustments at the TR32DB AFS directory.

Project group All TR32DB users are members in the project group `webuser:tr32db_project`. This group is used in the `upload` and `share` directory. In the `upload` directory all group members have permission to `lookup (l)` the TR32DB directory up to the cluster-level e.g. `/afs/rrz /project/tr32db/Upload_1/A/`. In the `share` directory all group members are able to `read (rl)` all data in the entire directory, e.g. data that are stored at `/afs/rrz/project/tr32db/share/C/C3`.

Project section group For every project section of the CRC/TR32 an AFS group was set up, e.g. `webuser:tr32db_a1`, `webuser:tr32db_c3`, etc. Every user of the TR32DB is allocated to at least one of the project section groups. The project sections groups are used in the `upload` and `share` directory. In the `share` directory, each project section group has the permission to `write (rlidwk)` data into the specific data type folder. In addition, in the `upload` directory, each project section group has the permission to `read (rl)` his project section folder and `write (rlidwk)` data into the specific data type folder. For example project participants from project section B1 are only allowed to upload their data to the directory `/afs/rrz /project/tr32db/Upload/B/B1/data` and `/afs/rrz /project/tr32db/share/B/B1`.

Furthermore, the TR32DB web server has the permission to access some parts of the TR32DB AFS directory. According to RRZK (2013c), the AFS web server user is called `webserver-tr32db`. Within the `phase` directory, the web server has the permission to `read (rl)` data in the entire directory and also `write (rlidwk)` data to each metadata subdirectory (e.g. `/afs/rrz /project/tr32db/Phase_1/B/B1/data/000/metadata`). In addition, the web server has permission to `read (rl)` all data in the entire `share` directory.

CAMPBELL (1998) and MILICCHIO & GEHRKE (2007) also describe `fs` commands to view or change the ACL. The subcommand `setacl` sets or modifies the AFS directory ACL. As shown below, the access permission of the TR32DB web server `webserver-tr32db` is set to `read` for the given directory. The command `listacl` lists the associated ACL of the directory. In this case, various groups and the TR32DB web server have different access permission to this directory.

```
alice@dialog5$ fs setacl
/afs/rrz/project/tr32db/share/C/C3 -acl webserver -tr32db read
alice@dialog5$ fs listacl
Access list for . is
Normal rights:
  webuser:tr32db_project rl
  webuser:tr32db_a rlidwka
  webadministrator:wm rlidwka
  system:administrator rlidwka
  webserver-tr32db rl
```

Furthermore, the `pts` suite is introduced by CAMPBELL (1998) and MILICCHIO & GEHRKE (2007). This suite manages group administration and membership. The subcommand `membership` lists all members of a group or all groups in which the user has a membership. With the subcommand `adduser`, a user or machine can be added to a given group. To delete or quit the membership of a user in a specific group, the

`removeuser` subcommand has to be used. Every TR32DB user has at least a membership in one of the TR32DB project section groups. As already mentioned, the user accounts are provided and managed by the RRZK. For example, the TR32DB user `alice` is added to the specific TR32DB project section group `webuser:tr32db_c3` by using the `adduser` subcommand. To show all current members of this group, the subcommand `membership` is used.

```
alice@dialog5$ pts adduser -user alice -group webuser:tr32db_c3
alice@dialog5$ pts membership webuser:tr32db_c3
Members of webuser:tr32db_c3 (id: -35077) are:
  alice
  bob
  peter
```

Within the environment of the TR32DB AFS storage directory, two project data moving shell scripts are operating. The first data moving script for data generated in funding phase 1, which are stored in the `Upload_1` directory, operates every night at 2 a.m. and moves data to the `Phase_1` directory. The second data moving script moves data, which are created in funding phase 2 and consequently stored in the `Upload_2` directory, to the `Phase_2` directory. This script operates every night at 3 a.m. Since both data moving scripts run at night, the data-upload to the upload directory is only limited per day. To run both shell scripts automatically, a cron job was set up. According to CAMPBELL (1998), a cron job is a command, configured to run at a certain time.

For example, the second data moving script moves the uploaded project data from the temporary `Upload_1` directory `/afs/rrz/project/tr32db/Upload_1/C/C3/data` to the final, corresponding, sustainable data storage in the `Phase_1` directory `/afs/rrz /project/tr32db/Phase_1/C/C3/data/001`, compare Figure 3-8.

Directory with volume. TR32DB default value quota 8 GB.

Moving of data from temporary to final data storage via script.
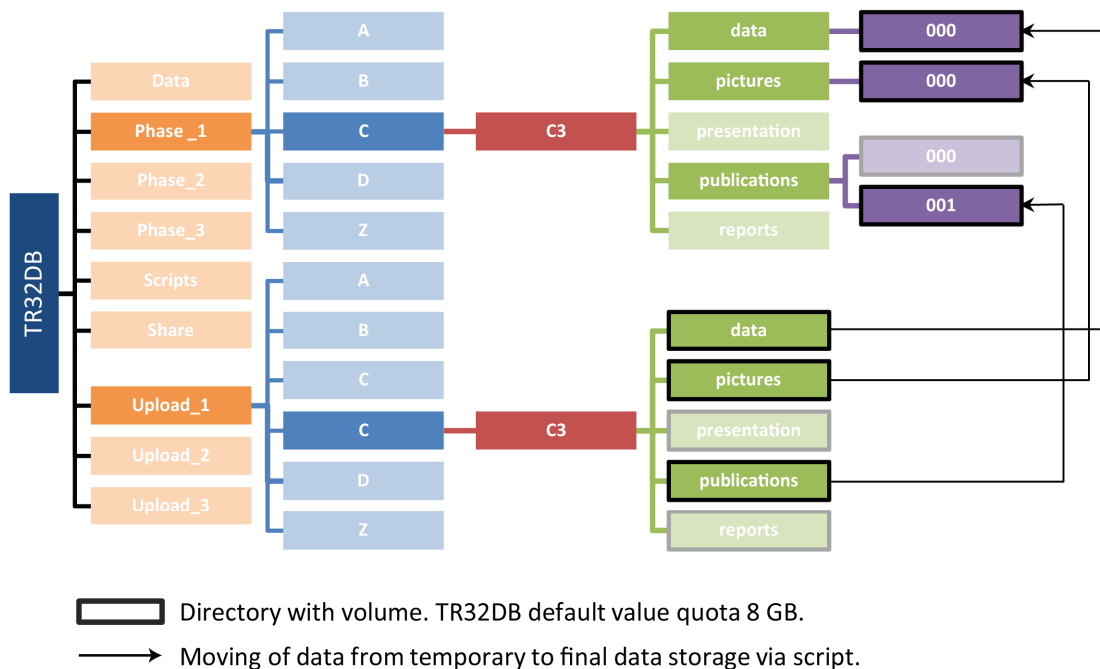
Figure 3-8: Moving of C3 data from temporary `Upload_1` directory to final `Phase_1` directory

When a moving script finds a file in the `Upload` subdirectory, it checks the available file space in the corresponding `Phase` subdirectories (e.g. `000, 001, 002,…`), where the file might fit in. If there is no file with the same file name stored in this directory, then the file will be moved to a folder with enough space left. In case there is already a file with the same filename stored in this directory, then the file will be moved to the

next subdirectory. In case there is no file space left in the corresponding `Phase` subdirectories, then the file will not be moved and stays in the `Upload` subdirectory. With a successful moving of each data file, a database entry is automatically generated. This contains information about the file and its provenance like: file name and extension, storage path, file size, data type (e.g. data, pictures, presentation, etc.), project section, funding phase and data upload date. Two log files document the details of the data moving process. These log files will be stored for two days and then transcribed.

A Backup of the entire TR32DB AFS directory (`/afs/rrz/project/tr32db`) is automatically running every night a 10 p.m. in the environment of the RRZK on tape storage. Again, a cron job is used to automate the process. At the RRZK, the already established IBM Tivoli Storage Manager (TSM) Backup-Archive Client is used (RRZK, 2013b).

3.5   Database-based data storage and organisation

The core of the TR32DB system is the database. The database is the physical storage of metadata of the CRC/TR32 project data (section 3.3 and 4.3), as well as of administrative details regarding the TR32DB system. The Regional Computing Centre of the University of Cologne (RRZK) provides and administrates the infrastructure of the TR32DB database. Besides the TR32DB database, further databases were set up and are used within the TR32DB system. A workshop database was set up to support the administration of workshops organised in relation to the CRC/TR32 project section Z1/INF and managed by the TR32DB web-interface. Furthermore, a climate database was set up storing climate data purchased from the DWD. This database is used to query and visualise DWD data (section 4.5.1.3 and 4.5.2.3). In this section, initially the generally used components are described. Subsequently, the design and organisation of the main TR32DB database (Version 3.1 and Version 4.0) will be presented in the following subsections.

In order to implement the database of the TR32DB system, the basic principles of database design were used, as already described in detail in section 2.3.3. The open-source relational database management system (RDBMS) MySQL (MYSQL, 2014b) was selected to set up the database of the TR32DB. The reasons for the selection of MySQL are the support and administration conducted by the RRZK and the great popularity of MySQL in combination with web applications. MySQL is a client-server system that is developed, distributed, and also supported by Oracle (MYSQL, 2014d).

Generally, the Structured Query Language (SQL) is used to design, query or update the TR32DB database. According to FEHILY (2008), SQL is a programming language that is simple, descriptive, embedded, interactive, standardized, and used to modify databases, their tables or data values. In the framework of this thesis, several SQL statements are used for database administration and web-interface programming with PHP (section 3.6). Some statements of SQL are consecutively listed:

> **CREATE TABLE** creates a new table in the database
> **ALTER TABLE** modifies the structure of an existing table
> **DROP TABLE** removes a table (including data) from the database
> **SELECT** queries data in the database
> **INSERT** inserts a new row into a table
> **UPDATE** updates a value in a table
> **DELETE** deletes a row from a table

These SQL statements are also used by the web-based interface phpMyAdmin and software MySQL Workbench that will be described in the following.

For an easier access and administration of the DBMS MySQL, the open source web-interface phpMyAdmin (PHPMYADMIN, 2014) is used. phpMyAdmin is able to perform various tasks that are available via a graphical user interface. This includes for example the creation, modification, or removal of databases. Database tables

can be created, copied, deleted, renamed, or edited. In addition, columns and rows can be added to the table, edited, or deleted. Moreover, phpMyAdmin offers the execution of SQL statements, the export and import of data files (e.g. csv, xml), and many further functions (DELISLE, 2012; KOFLER & ÖGGL, 2010; REIBOLD, 2013).

In addition, the software MySQL Workbench (MYSQL, 2014c) from Oracle provides a graphical user interface to work with MySQL. The main areas of functionality supported by MySQL Workbench are: modelling of data, development of SQL statements, administration of server instances, and data migration (ORACLE, 2014). In detail, MySQL Workbench facilitates, for example, the virtual design, modelling, creation, and management of databases using reverse (import) and forward (export) engineering. MySQL Workbench applies core modelling elements like tables, primary key, and relationships via foreign keys. In addition, it maintains the execution and creation of SQL query statements (MCLAUGHLIN, 2010; ORACLE, 2013b). Within the framework of this thesis, MySQL Workbench is used to visually design, model, and generate the database of the TR32DB system. In addition, it was applied to create complex entity-relationship (ER) models (section 2.3.2) and SQL statements for the TR32DB database.

3.5.1 Design and structure of the database Version 3.1

The structure of the TR32DB database Version 3.1 was initially designed and developed in 2008. Until 2013, this database was extended with modifications due to new requirements of the TR32DB. At the beginning of the database design, the general requirements that have to be represented in the database were analysed. Since the database should store the description of the stored data in terms of metadata, appropriate metadata standards and schemas (section 2.2) were considered. At that time the TR32DB Metadata Schema (section 3.3) was not yet designed. The database should also store administrative information, necessary to manage the entire TR32DB system (e.g. user administration). Consequently, for both requirements background information about the CRC/TR32 was observed. Besides the available proposal of the CRC/TR32 also results of the already mentioned survey (appendix A).were considered. Furthermore, general needs for the entire TR32DB system were contemplated.

In a next step, the relational database management system (RDBMS) MySQL was selected. The database structure was designed in a relational data model using the basic principles of database design described in section 2.3.3. Consequently, all analysed requirements were organised in tables including attributes, values, and relational keys. In addition, relationships between tables were considered. A preliminary, incomplete ER model of the TR32DB database structure from 2008 is presented in Figure 3-9. This model already represents a simplified user administration and metadata storage of the various data types. This structure was further modified to the final database structure Version 3.1 that will be described later in this chapter.

Finally, the database was physically designed and implemented in MySQL using the web-interface phpMyAdmin. This was in most instances executed by the graphical user interface. In addition, direct SQL commands were used to create database tables and their associated attributes. To create a database table, for example, the following simple SQL statement is used:

```
CREATE TABLE `subprojects` (
  `spID` tinyint(4) NOT NULL,
  `sp` tinytext NOT NULL,
  `sp_title` text NOT NULL,
  `sp_description` text NOT NULL,
  PRIMARY KEY (`spID`)
);
```
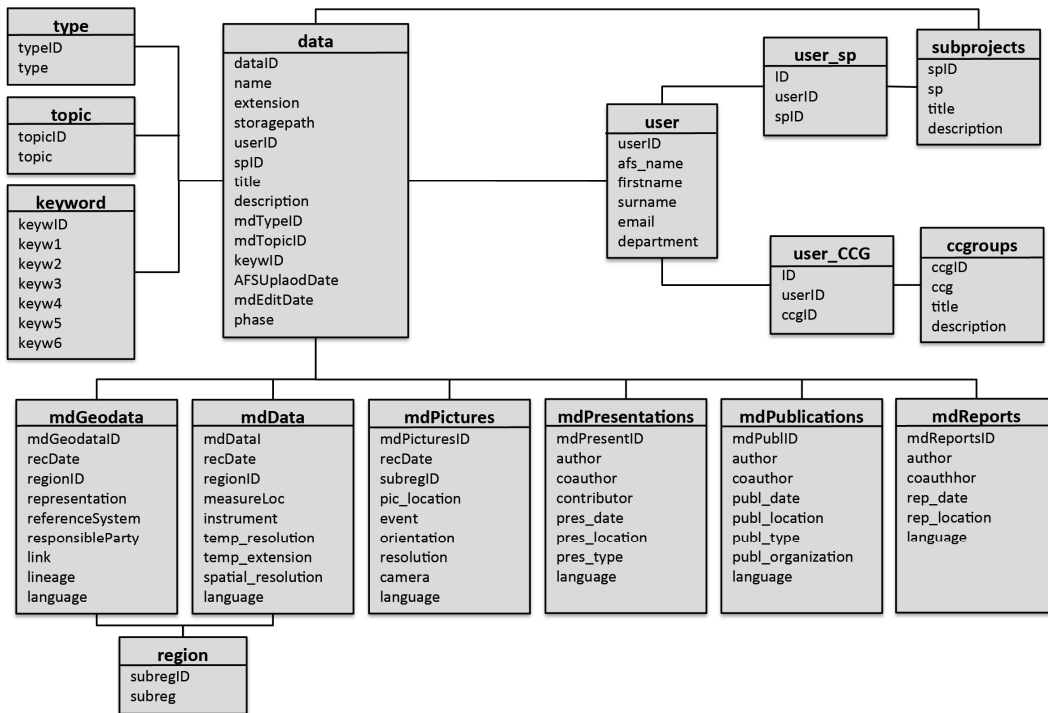
Figure 3-9: Preliminary ER model of the TR32DB database structure from 2008

By using this SQL statement, the table `sub-projects` was created in the TR32DB database. The table contains the attributes `spID`, `sp`, `sp_title`, and `sp_description`. For each attribute in the table, a single data type is determined, for example `tinyint`, `tinytext` or `text`. Some constraints are defined for the attribute that assigns rules for allowed values. The attribute `spID` is specified as primary key for the table. To prevent that `NULL`s are being inserted into the column values, all attributes hold a `NOT NULL` constraint. A detailed use of the SQL `CREATE TABLE` statement will be described in the next chapter.

The final structure of the TR32DB database Version 3.1, presented in detail in appendix C.1.1, contains 46 tables that are classified in ten groups (Figure 3-10). Only three tables of the database are connected with a relationship. Moreover, to sustain changes on some of the database tables, these are mirrored in the database to store the modification steps, such as `mdData`, `dataCreator` or `keyWords`. The mirrored tables are not represented in the database structure.
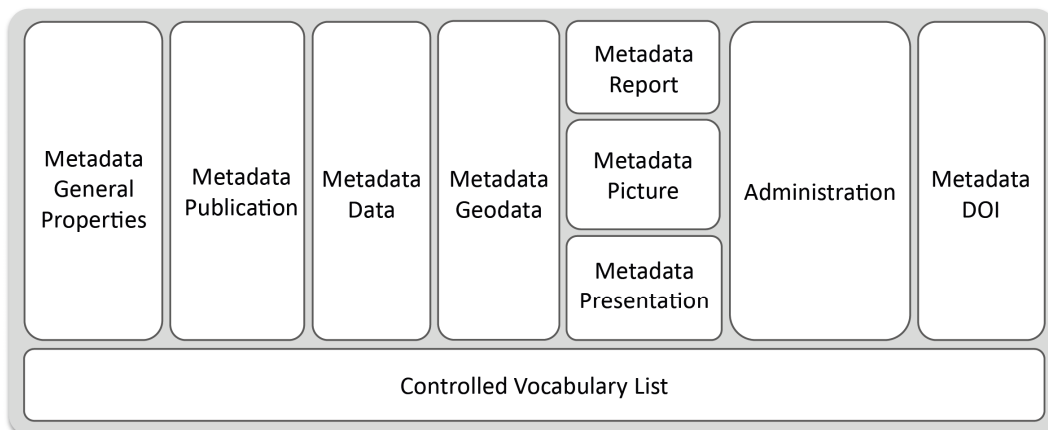

Figure 3-10: Simplified TR32DB database structure Version 3.1 (detailed structure in appendix C.1.1)

The TR32DB database Version 3.1 (appendix C.1.1) is organized in the following groups:

Metadata_GeneralProperties includes all metadata of a dataset that belong to the general properties. The table `data` is the central table in the database. This table comprises besides the administrative details of a file (e.g. TR32DB file identifier, storage path, file upload time, download permission) also descriptive details (e.g. title, description, citation advice). The table `dataCreator` enables the multiple storage of dataset creator. The table `keyWords` contains the descriptive keywords of the dataset that are limited to six per dataset.

Metadata_Publication covers the specific metadata of the data type Publication. The central publication table `mdPublications` determines all joined attributes of a publication. The special metadata details of a publication are stored in the three associated tables `mdPublEventPaper`, `mdPublArticle`, and `mdPublicationBookSection`.

Metadata_Data involves all tables of the database that store metadata of datasets from the data type Data. The table `mdData` includes the general metadata. The table `mdDataInstrument` and `mdDataInstrumentParamenter` are used to complement the associated instruments or models of a dataset.

Metadata_Geodata contains the table `mdGeodata` to store the specific metadata of all data from the data type Geodata. The table `dataOrganisation` enables the multiple storage of organisations that are corresponding to the dataset.

Metadata_Report is used to store the specific metadata of the data type Report.

Metadata_Picture comprehends the table `mdPictures`. This enables the storage of all specific metadata of a dataset of the data type Picture.

Metadata_Presentation covers the table `mdPresentations` to store metadata of a dataset from the data type Presentation.

Administration contains tables that are used to administrate the TR32DB (e.g. user administration). In addition, general tables are included to manage the news of the web-interface or to store the download count of datasets via the web-interface.

Metadata_DOI comprises all specific metadata of a dataset with a DOI.

ControlledVocabulayList includes a supplementary table. Some of the tables in this group have static values, while the number of values of other tables will be modified. They include defined controlled lists that are associated with a certain table. For example, the table `creator` is associated with table `dataCreator` via the attribute `creatorID`. Some tables are only used in connection with the TR32DB web-interface. For example, the table `kwList` is only used to represent values on the web-site, while the assumed corresponding table `keyWords` is not associated.

The final structure of the TR32DB database Version 3.1 is the base of the web-interface Version 3.1 (section 4.5.1).

3.5.2   Design and structure of the database Version 4.0

The structure of the TR32DB database Version 4.0 was designed and implemented in 2013. This database is a further development of the former database Version 3.1, described in the prior section. The design of the database was carried out hand-in-hand with the creation of the web-interface Version 4.0 and the TR32DB Metadata Schema (section 3.3). In addition, the database is the basis of the web-interface Version 4.0 (section 4.5.2).

The collection and analysis of the requirements for the design of a new database structure result from the experiences gained in designing the preliminary TR32DB database. In addition, the reconstruction was necessary, due to many implemented expansions of the previous version. Furthermore, new features and

functions in the TR32DB system were demanded by the users and also had to be considered. Moreover, modifications of the used metadata schemas (e.g. DataCite Metadata Schema Update) and an expansion with new metadata schemas or standards had to be included in the database structure.

The database structure was designed in a relational ER model using MySQL Workbench. The detailed ER model of the database Version 4.0 is available in appendix C.1.2. Basic principles of database design (section 2.3.3) were applied. The already existing database Version 3.1 was used as a foundation, rearranged, and extended with new details. Some old tables were used and modified, as well as new tables including attributes, values, and relational keys were created. For example, the central database table `data` from Version 3.1 was split up into the two new main tables `afsData` and `mdMetadata`. Remaining attributes of the `data` table were distributed to other tables. Furthermore, database tables of Version 3.1 describing controlled vocabulary lists were adapted and modified. For example, the database table `TR32Topic` from Version 3.1 was adapted to Version 4.0 by renaming the table name to `listTr32Topic`. The renaming ensures a consistent naming of the tables referring to controlled vocabulary lists. This offers an improved overview of the database structure and enables a simple distinction to other database tables. Available attributes were also renamed and supplemented with a new attribute `tr32TopicVal`. The `Val`-attribute always contains the value of the corresponding metadata standard or schema to ensure conformance.

In contrast to Version 3.1, all tables of the database structure Version 4.0 are connected with relationships in the ER model. Figure 3-11 shows a simple example of the related tables `Title` and `listTitleType` with their attributes and appropriate data types. The yellow key icon represents the primary key of the table (e.g. dataTitle.dataID, dataTitle.sequence, listTitleType.titleTypeID). Other attributes are indicated by colored diamond icons. A blue diamond icon indicates attributes that have a `NOT NULL` constraint (e.g. dataTitle.dataTitle). In contrast white diamond icons mark attributes with a default value `NULL` (e.g. `listTitleType.titleType`). A red diamond icon represents a foreign attribute (e.g. dataTitle.titleTypeID). Tables with a relationship are connected with lines. Identifying relationships are indicated with solid lines. Identifying relationships occur, when a foreign key is part of a primary key in a certain table. In contrast non-identifying relationships are presented with broken lines. They appear, when a foreign key is not part of a primary key of the table (Figure 3-11). The cardinality of the relationship is indicated by numbers and a diamond shape. *One-to-many* cardinality is represented with a black and white coloured diamond shape. A o*ne-to-one* cardinality is visualized with a white diamond shape. For example, a *one-to-many* relationship is presented in Figure 3-11.
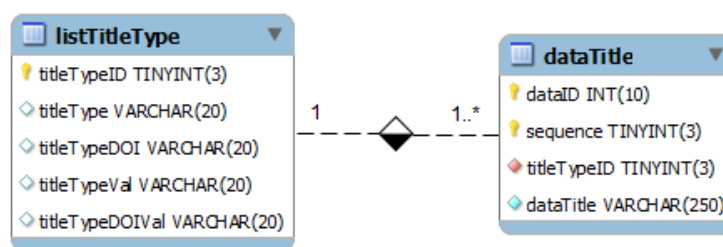
Figure 3-11: ER model of the related tables `listTitleType` and `Title`

Finally, the new database structure was physically implemented in the RDBMS MySQL with forward engineering using an SQL script. To create the SQL script, the final ER model of the TR32DB database Version 4.0 was exported by using the software MySQL Workbench. The resulting script was than imported to MySQL via the graphical web-interface phpMyAdmin. Consecutively, the SQL statement for the creation of the table structure for the two tables `dataTitle` and `listTitleType` is presented.

```
    CREATE TABLE IF NOT EXISTS `listTitleType` (
      `titleTypeID` tinyint(3) NOT NULL AUTO_INCREMENT,
      `titleType` varchar(20) DEFAULT NULL,
      `titleTypeDOI` varchar(20) DEFAULT NULL,
      `titleTypeVal` varchar(20) DEFAULT NULL,
      `titleTypeDOIVal` varchar(20) DEFAULT NULL,
      PRIMARY KEY (`titleTypeID`),
      UNIQUE INDEX `titleTypeVal_UNIQUE` (`titleTypeVal`),
      UNIQUE INDEX `titleTypeDOIVal_UNIQUE` (`titleTypeDOIVal`)
    ) ENGINE=InnoDB DEFAULT CHARACTER SET=utf8 AUTO_INCREMENT=5;
```

The statement `CREATE TABLE` creates both tables with their given name. The statement `IF NOT EXISTS` prevents the re-creation of the tables. In addition, both tables are created as `InnoDB` table, which enables the use of foreign key constraints. The default character set of the tables is set to the character encoding UTF-8.

```
    CREATE TABLE IF NOT EXISTS `dataTitle` (
      `dataID` INT(10) UNSIGNED NOT NULL,
      `sequence` TINYINT(3) UNSIGNED NOT NULL,
      `titleTypeID` TINYINT(3) UNSIGNED NOT NULL,
      `dataTitle` VARCHAR(250) NOT NULL,
      PRIMARY KEY (`dataID`, `sequence`),
      CONSTRAINT `fk_dataTitle_listTitleType`
        FOREIGN KEY (`titleTypeID`)
        REFERENCES `listTitleType` (`titleTypeID`)
        ON DELETE NO ACTION ON UPDATE NO ACTION,
      CONSTRAINT `fk_dataTitle_mdMetadata`
        FOREIGN KEY (`dataID`)
        REFERENCES `mdMetadata` (`dataID`)
        ON DELETE NO ACTION ON UPDATE NO ACTION
    ) ENGINE = InnoDB DEFAULT CHARACTER SET = utf8;
```

For each table several attributes with corresponding data types are defined. For example the data type `TINYINT(3) UNSIGNED` is chosen for the attribute `titleTypeID` in the table `listTitleType`. `TINYINT` is a numeric data type that requires values from 0 to 255 (unsigned). In contrast, the data type `INTEGER`, shorten to `INT`, enables values from 0 to 4294967295 (unsigned). `VARCHAR` is a string data type that can store a maximum of 255 values. The maximum length of the characters is defined in the bracket. For example, `varchar(20)` can store at most 20 characters. In the framework of this thesis, also the data types `DATE` and `DATETIME` are used to refer to temporal values. The `DATE` data type only specifies the date part in the value with the format `YYYY-MM-DD` (e.g. `2014-04-14`). `DATETIME` complements the time part to the date part with the format `YYYY-MM-DD HH:MM:SS` (e.g. `2014-04-14 20:15:00`). Further details about available data types in MySQL are described in detail at MySQL (2014a). To prevent that `NULL`s are being inserted into the column values, some attributes hold a `NOT NULL` constraint like `titleTypeID`. In contrast, other attributes hold `NULL` as a default value. Attributes, which own the additional characteristic `AUTO_INCREMENT` will increase to the next sequence by adding a new value to the table. The given number indicates the next value. For example, the attribute `titleTypeID` in the table `listTitleType` is defined as `AUTO_INCREMENT=5`. The next new value inserted in the table will automatically receive the value 5 as `titleTypeID`.

Furthermore, the relation keys are specified. The `PRIMARY KEYS` of the tables are listed (e.g. `listTitleType.titleTypeID`, `dataTitle.dataID`, `dataTitle.sequence`). In addition, two `FOREIGN KEYs` constraints are defined in the table `dataTitle` for the attributes `titleTypeID` and `dataID`. The attribute `titleTypeID` references the table `listTitleType` and the attribute `dataID` references the table `mdMetadata`. For both, the reference option `NO ACTION` is chosen for the update and

removing of values. In addition, indices were added to both tables. A `UNIQUE INDEX` is used for the values `titleTypeVal` and `titleTypeDOIVal`. This index ensures that the values for the attributes are distinct. Consequently, it is not possible to add a value that matches already existing values.

After creating the structure of the database in MySQL, the entire content from the previous database Version 3.1 as well as new content, for example resulting from the creation of the TR32DB Metadata Schema (section 4.3), had to be transferred to the new structure. This was primarily executed with SQL statements via the the graphical web-interface phpMyAdmin. In addition, new content available as comma-separated value (csv.) files, was imported via the graphical web-interface phpMyAdmin. Consecutively, examples of SQL `INSERT` statements for the input of new values to the tables `dataTitle` and `listTitleType` are presented.

```
INSERT INTO `dataTitle` (`dataID`, `sequence`, `titleTypeID`,
`dataTitle`) VALUES
(615, 1, 1, 'Aggregation of radar rainfall data'),
(803, 1, 2, 'Analyzing feedbacks in forest SVA system'),
(803, 2, 3, '1st PhD report, 2014');

INSERT INTO `listTitleType` (`titleTypeID`, `titleType`,
`titleTypeDOI`, `titleTypeVal`, `titleTypeDOIVal`) VALUES
(1,'Main Title','Main Title','',''),
(2,'Alternative','Alternative','alternative','AlternativeTitle'),
(3,'Subtitle','Subtitle','','Subtitle'),
(4,'Translated Title','Translated Title','','TranslatedTitle');
```

The statement `INSERT` inserts new tuples (records, rows) into the specified table, for example, the tables' `dataTitle` and `listTitleType`. The selected table and corresponding attributes are indicated. By using the `VALUES` syntax, multiple tuples can be inserted. The values for the tuples have to be declared in order by the indicted attributes. The statement fails with an error, if the declared attributes and indicated values do not match. Source of errors can be for example the inconsistency between the number of values and attributes or discrepancy with the data types.

The final structure of the TR32DB database Version 4.0, presented in detail in appendix C.1.2, contains 115 tables that are classified in seventeen groups (Figure 3-12). These groups are arranged similar to the TR32DB Metadata Schema Structure (see 4.3). With the exception of the table `news`, all other tables of the database structure are at least connected with one other table. Moreover, to sustain changes by the TR32DB users, some of the database tables, such as `mdMetadata`, `dataTitle`, `dataCreator` are mirrored in the database to store the modification steps. The mirrored database tables are always label with a `c_` at the beginning of the table such as `c_mdMetadata`, `c_dataTitle`, `c_dataCreator`. This ensures a consistent naming and separation from the other tables in the database. The mirrored tables are not represented in the database structure.

Throughout the entire database, a naming convention is used for the tables and attributes. The core tables storing the metadata details are characterized with `md`, for example `mdIdentification` or `mdResponsibleParty`. The attribute `dataID` serves as primary key in the `md` tables and refers to the central database table `mdMetadata`. A *one-to-one* relationship is applied between the `md` tables and the `mdMetadata` table. Supplementary tables, which are related to the core `md` tables, begin with `data`, for example `dataTitle` or `dataCreator`. These are implemented in a *one-to-many* relationship and enable multiple entries of tuples in the table. The combination of the attribute `dataID` and `sequence` serve as a primary key. Tables that correspond to DOI metadata, start with `DOI` (e.g. `DOItitle`, `DOIcreator`). All `DOI` tables are implemented in a *one-to-many* relationship. They include the attributes `DOIID` and `sequence`, which serve as a primary key. The overall tables referencing to controlled vocabulary lists start with `list`, for example `listTitleType` or `listCreator`. Different kinds of controlled vocabulary lists are maintained. Permanent lists do not change their content (e.g. `listIsoTopiCategory`). Values in dynamic lists are

expandable (e.g. `listCreator`). `List` tables are very often related to `md`, `data` or `DOI` tables, where the `list` attribute is defined as foreign key (e.g. `dataTitle.titleTypeID`, `dataCreator.creatorID`, `dataCreator.roleTypeID`). Some of the `list` tables are related, for example the table `listCreator` is related to the tables `listCreatorStatus`, `listCreatorScheme`, and `listInstitution`. Furthermore, the main attributes of a table in the database are always entitled to the table name, for example `listTr32Topic.tr32Topic` or `listTitleType.titleType`.
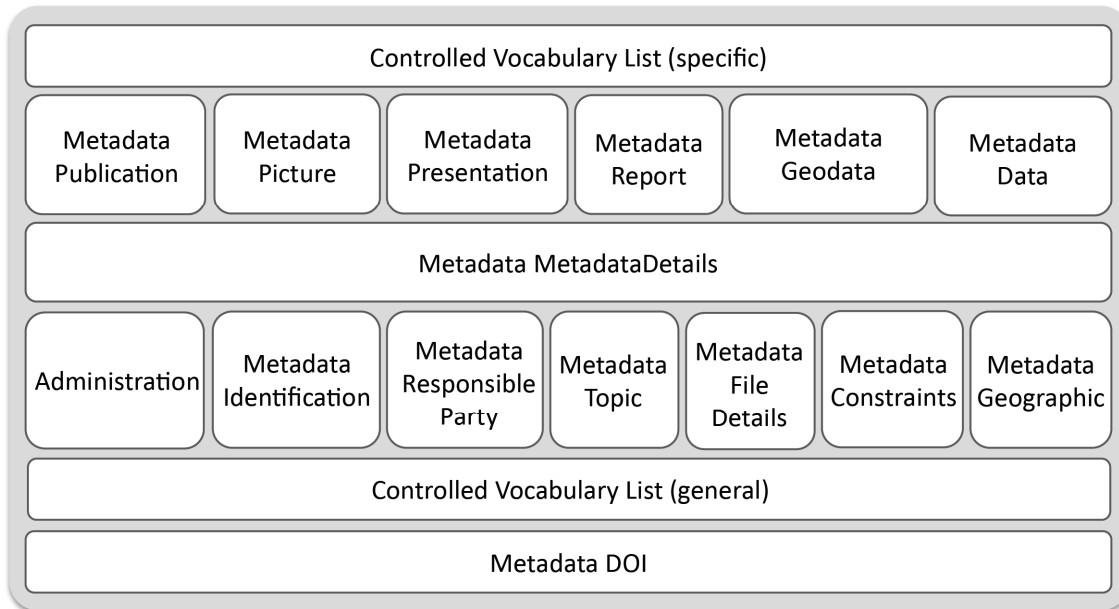

Figure 3-12: Simplified TR32DB database structure Version 4.0 (detailed structure in appendix C.1.2)

The TR32DB database Version 4.0 is organized in the following groups, which are presented in detail in the (appendix C.1.2):

Administration contains tables that are used to administrate components of the TR32DB. For example, the administration of the users is managed with the tables `user`, `userSubproject`, and `userAdmin`. Exemplary, the assignment of the users to a specific sub-project or the admin group is carried out here. General tables are included to manage the news of the web-interface or store the download count of datasets via the web-interface. In addition, the table `whoIsOnline` records the users of the web-interface.

Metadata_MetadataDetails comprises tables that store the so-called *metadata on metadata*. The tables `afsData`, `mdMetadata`, and `dataConformity` are included. The table `mdMetadata` is the central table of this group and the entire database structure. `mdMetadata` involves the attribute `dataID`, which is the identifier of a dataset in the TR32DB database. `dataID` is the relating attribute to the `md` and `data` tables, where it also serves as the primary key. In addition, the table `mdMetadata` comprises details about the metadata creation date and last update date, as well as the visibility of the metadata on the website. The table `afsData` includes information about the storage of the data file in the AFS, for example, the storage path, file size, file name, and upload date. In addition, the affiliation to a sub-project, funding phase, and metadata type are stored. The table `dataConformity` is not yet used. It is already designed for future application to store the conformity of the datasets to used metadata standards.

Metadata_Identification involves all metadata, which are used to identify a dataset in the TR32DB. The core table `mdIdentification` comprises attributes to store the citation advice for a dataset. In addition, the path to the additional describing PDF file is stored. For future use, the application of a graphic preview images is already considered in the table structure, but not yet

implemented. The tables `dataTitle` and `dataDescription` enable the storage of an identifying title and description. Details about an existing identifier or the relation to an identifier can also be specified.

Metadata_ResponsibleParty comprehends the metadata details concerning the responsibility of the data. The core table `mdResponsibleParty` contains only the attribute publisher. In addition, the dataset creator, a contributor, and an organisation can be specified in the tables `dataCreator`, `dataContributor`, and `dataOrganisation`.

Metadata_Topic contains all tables of the database that store metadata details, which describe a dataset with a topic or subject. The central table `mdTopic` includes the attributes `tr32TopicID` and `isoTopicCategoryID`, both related to a corresponding controlled list. In addition, the tables `dataKw`, `dataGemet`, `dataGemetInspireSpatialDataTheme`, and `dataDDC` are comprised. All facilitate multiple entries of tuples and are related to a particular controlled list, with the exception of table `dataGemet`.

Metadata_FileDetails includes tables and corresponding attributes to describe a dataset with metadata related to the file details. The central table `mdFileDetails` comprises attributes to specify the language, processing status, data type, file format and appropriate version. Furthermore, the tables `dataSize` and `dataDate` are included, which enable multiple tuple entries.

Metadata_Constraints contains only the core table `mdConstraints`. Here, attributes are specified, which enable the dataset description of access and use constraints, download permission and download information, as well as a licence. They are defined with the attributes `accessCon`, `accessUse`, `licenceID`, `dlPermID`, and `dlInfo`.

Metadata_Geographic involves the single core table `mdGeographic` to describe a dataset with geographic details. Geographic coordinates of a bounding box can be stored by using the attributes `north`, `east`, `south`, and `west`. Moreover, the `measurementLocationID` attribute is included, which is related to the table `listMeasureLocation`.

Metadata_Publication comprehends the specific metadata details of the data type publication. This group is an exception. Besides the main core table `mdPublication`, four further `md` tables are defined that describe the single requirements concerning the various publication types. The table `mdPublication` defines the attributes `publicationTypeID`, `publicationStatusID`, `year`, and `review` status. In addition, the table `PubEventPaper` determines attributes to describe a publication from the data type event paper. This includes details about the event (e.g. eventType, event name, event location), as well as information about the proceedings (e.g. proceedings editor, title, issue, volume). The table `mdPubBook` defines attributes to specify a publication from the data type book, for example series title, series editor, publishing city or book volume. The table `mdPubBookSection` includes attributes to describe a publication from the data type book section, for example book title, book editor, publishing city or page range. Finally, the table `mdPubArticle` contains attributes to define a publication from the data type article. The attributes article type, journal, issue, volume, and page range are, for example, determined.

Metadata_Picture covers the table `mdPicture` to define metadata details for a dataset from the data type picture. For example, general attributes with regard to the technical characteristics include resolution, image size, recording period or image orientation. In addition, details about the image recording or creation are defined, like recording place and method, as well as details about the event (e.g. event name, type, date period).

Metadata_Presentation is used to describe metadata of a dataset from the type of presentation. The table `mdPresentation` determines attributes like event type, presentation type, presenter, event name, location or event period.

Metadata_Report includes the table `mdReport`, which defines attributes to store metadata of a dataset from the data type report. The attributes like report type, report date, as well as the institution and city are specified.

Metadata_Geodata contains attributes to describe a dataset from the data type geodata. The core table `mdGeodata` includes, for example, the attributes lineage or temporal extent period. In addition, further tables with attributes are specified to meet the demands of the metadata standards. The tables `dataReferenceSystem`, `dataVerticalExtent`, `dataPresentationForm`, `dataMaintenanceInformation`, `dataSpaitial-Representation`, and `dataSpaitialResolution` are defined with corresponding attributes. All tables are optional and enable multiple entries.

Metadata_Data is used to describe metadata of a dataset from the data type data. The core table `mdData` is included with attributes like temporal extent period and lineage. In addition, the tables `dataInstrument` and `dataInstrumentParameter` are available, both related in a One-To-Many relationship, which enables multiple entries of tuples. Attributes about an instrument or model and corresponding parameters (e.g. `instrummentParameterID`, `temporalFrequency`, `resolutionDistance`) are specified.

Metadata_DOI comprises tables to store metadata of datasets with a DOI. The table `DOI` is the core table that is connected with the all other tables in this group. For example, the table `DOItitle` stores the title, the associated type of title and the sequence of a dataset. The creator(s) of the dataset and the corresponding sequence is retrained in the table `DOIcreator`.

ControlledVocabularyList (specific) covers all controlled vocabulary lists, which are necessary to support the defined attributes in the metadata groups with focus on the specific metadata details (`Metadata_Publication`, `Metadata_Picture` `Metadata_Presentation`, `Metadata_Report`, `Metadata_Geodata`, and `Metadata_Data`). The group contains 20 tables, for example listArticle, listPublicationType, listPresentationType, listReportType, listInstrument, and listInstrumentParameter. The controlled lists are related to at least one table. For example, the table `listReportType` is related to the table `mdReport`. The table `listEventType` is for example related to the tables `mdPresentation`, `mdPicture`, and `mdPubEventPaper`.

ControlledVocabularyList (general) includes supplementary tables for the groups `Administration`, `Metadata_DOI`, as well as such groups with focus on general metadata descriptions like `Metadata_Identification`, `Metadata_ResponsibleParty`, `Metadata_Topic`, `Metadata_FileDetails`, `Metadata_Constraints`, `Metadata_Geographic`, and `Metadata_MetadataDetails`. In total, 36 tables are determined. Some of the tables combine values from various metadata standards and schemas, like table `listDateType`. This table supports values from Dublin Core, DataCite Metadata Schema, and ISO 191115. The mapping of the metadata standards described in section 3.3 and 4.3 is implemented in a table structure. Besides the attributes that contain the metadata standard or scheme value (e.g. `dataTypeDOI`), additional attributes are defined, which include the values used for the visualisation on the web-site (e.g. `dataTypeDOIVal`).

The TR32DB database structure Version 4.0 is maintained and supported continuously. New requirements of the TR32DB system will be incorporated. Consequently, the database will be consecutively updated.

3.6   The TR32DB web-interface

The TR32DB web-interface is the graphical user interface (GUI) of the TR32DB system, hosted and administrated at the Regional Computing Centre of the University of Cologne (RRZK). The web-interface is accessible by a web-browser at `http://www.tr32db.uni-koeln.de` or `http://www.tr32db.de`.

The web-interface provides a web-based access to the project data stored in the file-based data storage (section 3.4) and the corresponding metadata stored in the database (section 3.5). The latter can be managed (e.g. input, edit, search) by the web-interface. In addition, further features are provided by the web-interface, like web mapping applications or registration of DOIs. In this section, a brief overview will be provided about the methods used to create the web-interface. Initially, general recommendations and guidelines will be outlined. Following, a brief outline will be presented of the design and used components to set up the previous web-interface Version 3.1 and the current web-interface Version 4.0. Finally, applied methods will be briefly described for the set-up of the integrated WebGIS and the registration of TR32DB DOIs.

In order to implement a web-interface, some general recommendations and guidelines (section 2.3.4) have to be followed. Initially, the Barrier-free Information Technology Ordinance North Rhine-Westphalia (BITV NRW) and the Web Content Accessibility Guidelines (WCAG) have to be considered with regard to the design and generation of the web-interface (e.g. use of standardized programming languages; design should be perceivable, understandable, operable, and robust). In addition, since the TR32DB web-interface is established in the environment of the RRZK, regulation and recommendations for the publishing of websites hosted at the RRZK have to be followed. This concerns rules about the barrier-free accessibility of the websites (e.g. layout, navigation, use of images and tables, browser technologies), general operating rules (e.g. liability and conformity to law), admission requirements (e.g. compliance of technical rules, obligations of webmaster), as well as legal advices (e.g. data protection) (RRZK, 2014d).

Web-interface Version 3.1 and Version 4.0

The TR32DB web-interface Version 3.1 and Version 4.0 were designed according to previous mentioned recommendations and rules, as well as according to usability principles and user guidance as described in section 2.3.4. In addition, demands and requirement of the TR32DB users were followed to create a user friendly GUI. The design of both web-interfaces is organized in several components, as shown in Figure 2-9 and their colour scheme is arranged in mainly yellow and green colours according to the logo of the CRC/TR32 (Figure 4-6 and Figure 4-41). Both web-interfaces are individually self-developed and are connected to the file-based data storage (section 3.4) and the database (section 3.5). The web-interfaces are generated on the basis of code especially written for this purpose by using languages like HTML or PHP. In both cases, a Model-View-Controller (MVC) principle (section 2.3.1) was applied. To provide secure access and communication to the internal part of the web-interface, for both versions the Secure Sockets Layer (SSL) encryption protocol, described by HEIDERICH (2009), is used. Subsequently, the set-up of the two web-interface version will be briefly described.

The web-interface Version 3.1 was developed on basis of Smarty (SMARTY, 2014), a web template system, which is based on PHP. In addition, the markup language HTML, the style sheet language CSS, as well as the programming language JavaScript were used. PHP and HTML were used to create the dynamic webpages, as described in section 2.3.4. Furthermore, CSS was applied to format and style the web pages (e.g. font, font color, font size, tables). In addition, JavaScript was used for some special features like a calendar or an integrated bounding box.

A simplified structure of the website scripts is presented in Figure 3-13. The structure is organized in a folder structure. The `docs` part contains the PHP files, organized according to special topics like `doi` or `search`. The corresponding template files, which are based on HTML, are located in the `smarty-folder`. The webpages are created from the PHP and template files.

The web-interface Version 4.0 is developed with the server-side scripting language PHP, the markup language HTML, the style sheet language CSS, as well as the programming language JavaScript. In this context, HTML and PHP are used to from the web pages. The web pages are created dynamically, as described in section section 2.3.4. In addition, CSS is applied to format and style the web pages (e.g. font, font color, font size,

tables lines). Furthermore, JavaScript is adopted for special functionalities such as a calendar, a map search, a spatial bounding box or an image gallery.
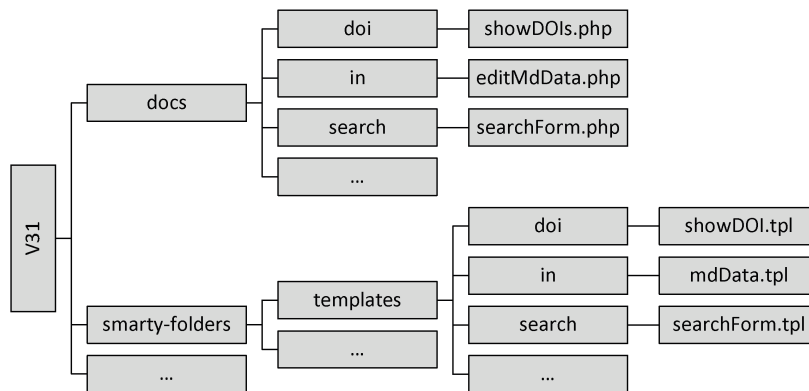

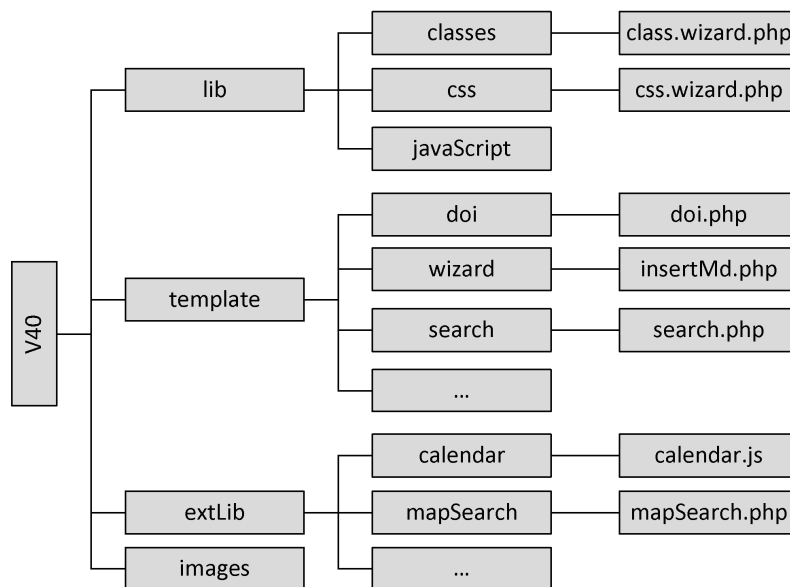Figure 3-13: Simplified part of website project structure Version 3.1

Figure 3-14: Simplified part of website project structure Version 4.0

A simplified structure of the website is presented in Figure 3-14. The website structure is organized in a well-arranged folder structure containing four main folders, which are subdivided in sub-folders according to a specific topic. These enable a complete separation of website content, structure, layout, external libraries, as well as images included in the web-interface. The library folder `lib` contains all PHP class files, which are used to manage the website. For example, the `css.wizard.php` styles the metadata input wizard (Figure 4-72 - Figure 4-59) or `class.wizard.php` is responsible for the checking of the single input fields. The `template` folder includes files, which provide the structure of the website. For example, the `doi.php` is responsible for the layout of the DOI data description webpage (Figure 4-54). In addition, external libraries are organized in the `exLib`-Folder, such as the `mapSearch.php`, which creates a map with overlaying markers using the Google Maps API (Figure 4-48). Finally, the images folder includes all images, which are displayed on the entire website like the header logo or footer logos (Figure 4-41).

TR32DB web mapping services and WebGIS

An important part of the TR32DB is the visualization, analysis, and supply of spatial data. Within the TR32DB several web mapping functionalities are used. These are the main TR32 Viewer (an integrated

WebGIS) (Figure 4-34 and Figure 4-86), the map search (Figure 4-48), and the bounding box map integrated in the metadata input form (Figure B-2) or wizard (Figure 4-64). The major TR32 Viewer and the bounding box map are only accessible by authorized users of the TR32DB.

All web mapping services (section 2.3.5) are hosted by using ESRI's ArcGIS Server, Version 10.0, which is installed on a virtual machine in the RRZK environment. The services are implemented for the different web mapping applications by using the Representational State Transfer (REST)–API (PINDE & SUN, 2011). The map services are generally capable of dynamic drawing, but also preprocessed tiled images are used for faster navigation. Furthermore, query and search functions for the attributes of the data are also available. The data is located on the mounted hardware drive of the virtual machine and are mainly administrated by a File Geodatabase (FGDB). A FGDB is a collection of files (spatial or non-spatial) in a folder on a mounted hardware, which enables the storage, management, and request of files (ESRI, 2012). In this FGDB, the vector data has been imported into different thematic feature datasets. Feature datasets consist of different feature classes, according to their theme and type of feature (point, line, area). For instance, all data describing the CRC/TR32, such as the Rur catchment boundary, the sub-catchment boundary (e.g. Kall), the measuring sites and areas (e.g. Selhausen), are combined in the `TR32 feature dataset`. Furthermore, the raster data is managed by a raster mosaic datasets in the FGDB. The original data is stored in folders, whereas feature attributes, locations, extents and overviews are directly stored within the FGDB. In previous versions of the TR32 Viewer, the ArcGIS Server Image Extension was used in combination with ArcGIS Server 9.3.

Table 3-2: Overview of the different types of web mapping services and data

Web mapping service name	Service type	Dataset Description	Tiled/tiles/amount
ALK	Map service	Cadastral map for three sub-regions.	Yes/194/232 MB
ATKISBasisDLM07	Map service	Digital landscape model for the entire Rur catchment of 2007	Yes/572/1.44 GB
ClimateStations	Map service	Climate stations of German weather forecast, Meteomedia Group, Rur and Erft water resources association since 2007 yearly.	No
ISBK50	Map service	Soil map information system of North Rhine-Westphalia in a scale of 1:50,000.	Yes/1808/5.88GB
Land Use Information	Map service	Land use information in 15 m resolution from each year of 2007 to 2012, derived by the multi-data approach involving remote sensing classifications (WALDHOFF et al., 2012)	
ScilandsDEM10m	Map service	Digital elevation model in 10 m resolution.	Yes/630/1.52 GB
TR32Specs	Map service	Specific project data: catchment and sub-regions boundaries, measuring sites, cosmic ray stations, climate stations of CRC/TR32, gauging stations, Eddy covariance stations, and radar stations.	
TR32webmap	Image Service	Basic project data, designed for the use in the map search and bounding box application	
DOP0708	Image Service	Digital orthophotos for the entire Rur catchment of the years 2007-2008	
DOP10	Image Service	Digital orthophotos for the entire Rur catchment of the years 2010	

All these data are stored and managed by the FGDB and are used to establish web mapping services, as well as image services (Table 3-2). The cartography, such as the colors of features and labeling, are adjusted in an ESRI ArcMap document (MXD file) and stored in a special document for the server, the map service definition (MSD file). These files can combine different data sources, for example from the FGDB, and each included dataset can be adjusted to specific needs. The ArcGIS Server uses the information in the MSD file to establish web mapping services, which draw images according to the request of the client and the settings of the map document and web mapping service. In contrast, image services can be established by the server, which directly rely on the chosen mosaic datasets. The latter shows additional functions that can be accessed by the clients (e.g. on-the-fly processing). The image services are used in this context for the delivery of the DOPs of 2007/2008 and 2010. These image services contain about 3,700 images with a size of ~ 287 GB. Overviews, which are images with a reduced resolution of a certain extent of several combined images, are automatically rendered and used to increase display speed. Several map services are pre-cached by the usual ArcGIS Online / Bing Maps / Google Maps-Schema that contains 20 different predefined scales. Tiles, each as PNG images with a 256 by 256 pixel size are rendered for the scales that are used by the aforementioned mapping services. An overview of the different types of web mapping services and the data (see also section 3.1.2) involved is presented in Table 3-2.

TR32DB DOI registration

For the persistently identification of data, which are stored in the TR32DB, a DOI (section 2.1.6) can be applied. By means of the DOI, the data can be accurately cited, for example, in a publication. The registration of a DOI has to follows several stages in the workflow, as presented in Figure 3-15. They will be described subsequently.
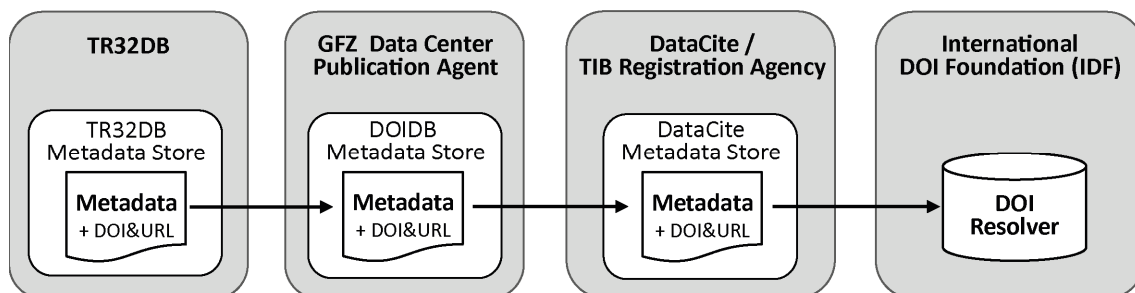

Figure 3-15: TR32DB DOI Registration Workflow

Initially, in stage one, the metadata provided by the TR32DB user (DOI applicant) (section 4.5.1.3 and 4.5.2.3) will be checked by TR32DB admin. These metadata have to be compliant to the DataCite Metadata Schema (section 2.2.2.2). In the context of the TR32DB, currently the DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011) is implemented. Supplementary to the provided metadata of the DOI applicant, further metadata will be added such as the DOI name. The name is comprised of the prefix 10.5880 and the individual suffix. The suffix is automatically generated, when the user applies for a TR32DB DOI (e.g. TR32DB.10). The suffix name can be modified by the TR32DB admin. For instance, for the application of DOI for proceedings articles of organized workshops (e.g. Kölner Geographische Arbeiten), the suffix is manually changed (e.g. TR32DB.KGA90.1). Besides the DOI name, furthermore the corresponding URL (e.g. http://www.tr32db.uni-koeln.de/search/view.php?dataID=518) is automatically added to the DOI metadata. In case, the metadata are complete and conform to the applied Metadata Schema of DataCite, they will be posted in XML format (Figure B-5) to the co-operating data centre of the Helmholtz Association Research Centre for Geosciences (GFZ). By using the REST (Representational State Transfer)-based API (application programming interface), the metadata will be stored in the DOIDB Metadata Store (GFZ, 2014). The DOIDB Metadata Store operates as a proxy service and is directly connected to the DataCite Metadata Store (DATACITE, 2014b). Consequently, by submitting metadata to the

DOIDB Metadata Store, the registration of the DOI is directly triggered at the DataCite Metadata Store. According to LUDWIG & ENKE (2013), the DOI name and the corresponding URL are finally forwarded to the International DOI Foundation (IDF). The registration of a TR32DB DOI is completed and the DOI is now resolvable. For example, the DOI `TR32DB.KGA90.1` can be resolved by using the address `http://doi.org/10.5880/TR32DB.KGA90.1` or `http://dx.doi.org/10.5880/TR32DB.KGA90.1`.

Overall, the TR32DB is responsible for the storage of the relevant dataset and the associated metadata including creation, storage, accessibility, and quality assurance. In addition, the allocation of a landing page is recommended by DataCite and is provided for TR32DB DOIs (Figure 4-15 and Figure 4-54). According to DATACITE (2012), the implementation of landing webpage is best practice for all registered users. This landing page should contain at least one of the following specifications: (i) a full citation of the registered data, (ii) a statement concerning the access to the data (e.g. details about usage restriction or download link to data), or (iii) associated metadata of the data.

4 The TR32DB: a project database for the CRC/TR32

In the previous chapters, the technical details of the CRC/TR32 Data Management System (TR32DB) (chapter 3) according to the state-of-art depicted in chapter 2 were presented. This chapter gives an overview of the functionalities of the TR32DB from the user´s perspective. Therefore, initially, the demands and requirements of the TR32DB are presented in section 4.1. The general structure of the TR32DB is introduced (section 4.2). Subsequent, section 4.3 focuses on the TR32DB Metadata Schema. Section 4.4 introduces the TR32DB storage and organisation of the project data. Finally, TR32DB web-interface will be presented in section 4.5. The web-interface Version 3.1 (section 4.5.1) and the updated, improved web-interface Version 4.0 (section 4.5.2) are described.

4.1 Demands and requirements of the TR32DB

Prior to designing and implementing a scientific data management for the CRC/TR32 (section 3.1), it is important to know the requirements, demands, and needs of the project and the project participants. To archive a first overview of the CRC/TR32 project participants' requirements and needs for a centralized data management system, a survey was carried out in 2007 (section 3.3). This survey focused on general questions about a centralized data management system, the data delivery (kind, format, file size), and the operating system and software used (appendix A).

In retrospect to the last years, three main requirements and demands on the scientific data management of the CRC/TR32 that have emerged and need to be considered. These are:

Amount of data The amount of data created by the project participants has to be considered. With regard to three scheduled funding phases, each four years, the project might operate for 12 years. In addition, the specific structure of the project has to be considered, which is divided in clusters and then subdivided in several project sections. In the course of the funding time of 12 years, approximately 60 project sections can be funded. Finally, the number of scientists working in the various project sections has to be considered. On average, around 1-3 PhD students or 1-2 Post-Docs are working in a project section per funding phase. On average, within 12 years of funding around 80 PhD students and 30 Post-Docs are working within the project and create data, as well as numerous Bachelor and Master Theses are produced.

Nature of the data Moreover, the nature of the data has to be considered. The CRC/TR32 has the different research fields (SVA) and thus, an interdisciplinary scientific background. The project participants create and collect very heterogeneous data in different spatial and temporal scales, as already introduced in section 3.1.2. They conduct, for example, measurements in a temporal resolution of minutes, hours, days, weeks, months etc. or in a spatial scale of points, plots, the sub-basin or the basin scale. Most of these data is created within a spatial domain. These data result from various field measurements, observations, or model simulations. Besides the collected or created research data, research results like publications, presentations, reports, or pictures need to be handled. Additionally, a multitude of purchased geodata has to be managed.

Format and size Finally, the format and size of the data varies due to the different research backgrounds. As a result, the project participants create and provide data in various file formats like MS Excel, ASCII, NetCDF, binary format, PDF, JPEG, or TIFF. The size of a single data file varies from a few kilobytes to several gigabytes.

Consequently, our CRC/TR32 Data Management System (TR32DB) was set up according to the needs and demands of the project data and background to handle a huge amount of very heterogeneous data. The design has considered recent IT developments and standards (section 3.4 and 3.5), as well as demands on documentation of the data with accurate metadata according to common, adequate metadata standards (section 3.3). The systems was developed and adjusted to requests of project participants (e.g. feedback on TR32DB trainings, project meetings). Identified problems and challenges in terms of empty archives, data publishing and sharing, as well as barriers of researchers (BORGMAN, 2012; GRIFFITHS, 2009; HARTTER et al., 2013; NELSON, 2009) (section 2.1) were considered. In addition, experiences gained by other interdisciplinary research projects in the field of environmental sciences (MÜCKSCHEL et al., 2007; NAUSS et al., 2007) were concerned.

Finally, the TR32DB was developed following specific DFG guidelines, already introduced in section 2.1.2, like the *Proposals for Safeguarding Good Scientific Practice* (DFG, 1998, 2013e). These recommend the archiving of the primary data (secure storage in a durable form) for at least ten years in the institution of their origin, when the data are a basis for a publication. In addition, the *Guidelines Collaborative Research Centres Programme* (DFG, 2013d) (previous: *Guidelines-Service-projects for information management and information infrastructure in CRC-INF* (DFG, 2009a)) has to be considered. These describe the tasks and duties of *Information Infrastructure* sub-projects (*INF*-projects). The main objectives are the fulfilment of the principles of *good scientific practice* and generation of synergies through the implementation of *data platforms* (e.g. sharing and re-use of data). For the creation of this data platform, existing local information facilities (e.g. university library or computing centre) and available tools and technologies should be used to archive and provide the created data beyond the funding duration. For instance, the focus of an INF-project could be on the development of a *project database*, which aims on the storage of the generated CRC data including associated metadata (DFG, 2013d). Subsequently, the implementation of the single components of the CRC/TR32 project database (TR32DB) will be described.

4.2   The TR32DB system architecture

As already introduced in chapter 3, the TR32DB system comprises three main components to manage the CRC/TR32 project data (Figure 4-1). These are a file system, a database, and a web-interface with an integrated web mapping application. The file system (section 3.4) is the physical storage of the CRC/TR32 project data, which is implemented and administrated within the AFS of the RRZK. The AFS is connected to the database by storing the AFS file path of the datasets in a MySQL database, as well as further details like type of data, corresponding sub-project and phase, file name and size, and upload time (section 3.5). The main task of the database is the storage and management of the entire metadata, which belong to the project data. In addition, the MySQL database has to handle administrative data, like user details or user rights. The self-developed TR32DB web-interface (section 3.6), accessible online at www.tr32db.de since early 2008, is the connecting component between AFS and MySQL database. Management, visualisation and publication of all created CRC/TR32 project data and purchased data (e.g. climate data or basic geodata) via metadata are the main duty of the web-interface. Some processed geodata, which are used for visualization purposes in the WebGIS (section 3.6), are stored in a virtual machine at the RRZK in a file-based storage and file geodatabase. The WebGIS is implemented as an alteration of the ESRI Flex Viewer and web mapping services are established by the ESRI's ArcGIS for Server technology (section 3.6).

TR32DB users, who want to store and publish their data in the TR32DB, have to upload their data to the data storage (AFS) in a first step. The data will be moved by the script, presented in section 3.4.2, to the final position in the AFS data storage overnight. In a second step, the user has to login to the TR32DB webpage the next day. He will be requested to add metadata for the file he has uploaded by the metadata input wizard of the webpage. After a successful submission of the metadata, the data file is directly available at the web-interface. Furthermore, webpage visitors and authorized users are able to search, request and download data

(depending on the download file permission). By means of the implemented internal Web-GIS, authorized users can visually search and explore existing geodata and their attributes. They can use a specific climate data tool, also implemented in the internal area of the web-interface or they can access data from the AFS share area by the webpage.
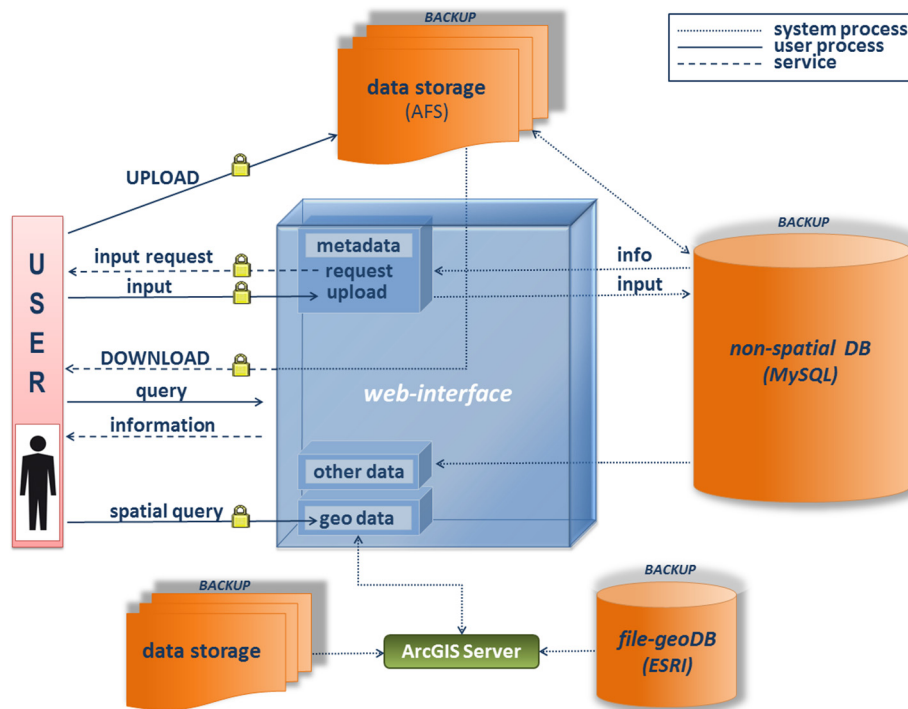

Figure 4-1: TR32DB components

There are some restrictions for the use of the TR32DB, which need to be considered. These are set by the administrator of the TR32DB for safety and technical aspects. Only authorized users (CRC/TR32 project participants) are allowed to store or share their project data within the TR32DB structure. The access to the TR32DB system is operated by user accounts of the RRZK (section 3.4.2). All employees of the University of Cologne receive this user account by default and accordingly all project members from other institutions have to apply for a user account of the RRZK (RRZK, 2013d) by the TR32DB administrators. This user account is by standard valid until the end of the recent project funding term of the CRC/TR32.

A detailed overview of the main workflows of registered users and all further users are presented in Figure 4-2 in addition to the previous described structure and workflows. To successfully submit data to the TR32DB registered users have to conduct two steps, as already introduced. In detail, at first, they have to login to the RRZK server and have to navigate to their specific storage area TR32DB storage. After copying the file to the server, they have to logout, and consequently finished the first step. This first step can be executed either for the upload of data to the data storage, as well as to store data in the share folder directory. To enter or edit metadata of a specific dataset the user has to access the TR32DB web-interface. After a login to the website, the user has to navigate to the user home, where he can select the dataset of interest, which should be complemented with new metadata or modified. Subsequently, the user can enter or modify the metadata of the chosen dataset. After checking the metadata, the user can submit and finally logout. Following the successful submission of the metadata, the registered user can execute an optional step and apply a DOI for his dataset by the web-interface. Initially, the user has to login to the web-interface, navigate to his user home and select the dataset of interest. Now, he has to complement the already available metadata, check and submit them, as well as finally log out again. All users can search for data, inspect all metadata and download the data, depending on the permission of the dataset. To perform this operation,

initially, the user has to search for the dataset by using different search options. After selecting the file of interest, the user can view the corresponding metadata of this dataset. Depending on the access authorization of the dataset, the user is able to download the dataset. The single stages will be described.
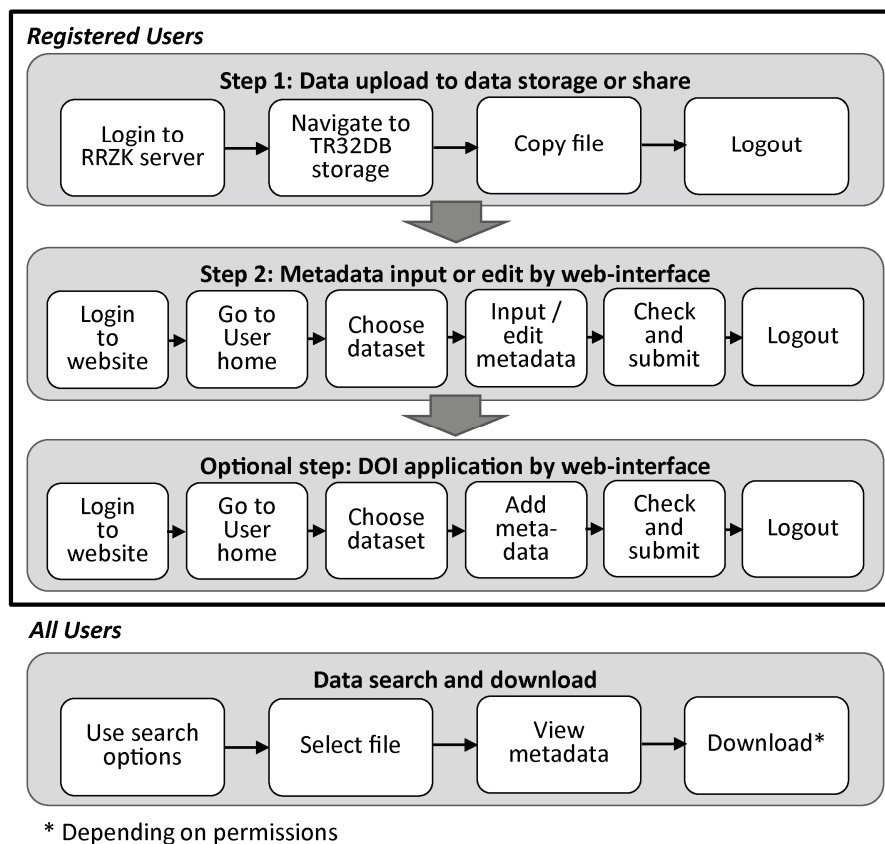

Figure 4-2: TR32DB user workflow for registerd and all users

4.3 The TR32DB Metadata Schema*

The self-developed TR32DB Metadata Schema was designed with the purpose to support the metadata management within the TR32DB. As explained in chapter 3.3, the schema is build up in compliance with the Dublin Core (DCMI, 2012a, b; section 2.2.2.1), the DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011; 2.2.2.2), the ISO19115:2003 Metadata Standard (ISO, 2003; section 2.2.2.3), as well as the INSPIRE Implementing Directive (EUROPEAN COMMISSION, 2008) and the INSPIRE Metadata Implementing Rules (DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE, 2010; section 2.2.2.4). The entire metadata schema is described in a detailed documentation and published online in April 2014 (CURDT, 2014b). The TR32DB Metadata Schema documentation specifies initially a general introduction of the metadata schema and, subsequently, a detailed description of the single TR32DB metadata properties and their attributes in structured lists. Finally, an appendix is presented including the mapping of the TR32DB metadata properties to other metadata schemas and standards, as well as the presentation of controlled vocabulary lists and their attribute values associated with the TR32DB metadata properties. The documentation of the metadata schema serves as a guideline for the users of the TR32DB, as a basic concept for the metadata management in the current Web-Interface

---

* The chapter is based on CURDT, C., 2014b. TR32DB Metadata Schema for the Description of Research Data in the TR32DB. Collaborative Research Centre/Transregio 32, Project Section Z1/INF, Institute of Geography, University of Cologne, Cologne, Germany. doi: 10.5880/TR32DB.10. (appendix D)

(Version 4.0, section 4.5.2), and as a technical documentation of the TR32DB system. In the following, the TR32DB Metadata Schema will be described, based on its documentation provided in appendix D.

The TR32DB Metadata Schema is set up on two levels, presented in Figure 4-3. The first level is the *general* level. This level includes metadata properties classified into seven categories: Identification, Responsible Party, Topic, File Details, Constraints, Geographic, and Metadata Details. The purpose of the general level is to describe all data with basic information. They are all equal for every data type. The second level is the *specific* level. This level contains the data type specific metadata properties. Currently, it is divided into six data type categories: Data, Geodata, Report, Picture, Presentation, and Publication. The category Publication takes a special position. This category is once again subdivided into the most common publication types: Article, Book, Book Section, and Event Paper. Event Paper can be applied to all kind of papers presented at an event like a conference, workshop, or project meeting. The purpose of the specific level is to describe the data with information, which are specific for one of the selected data types.
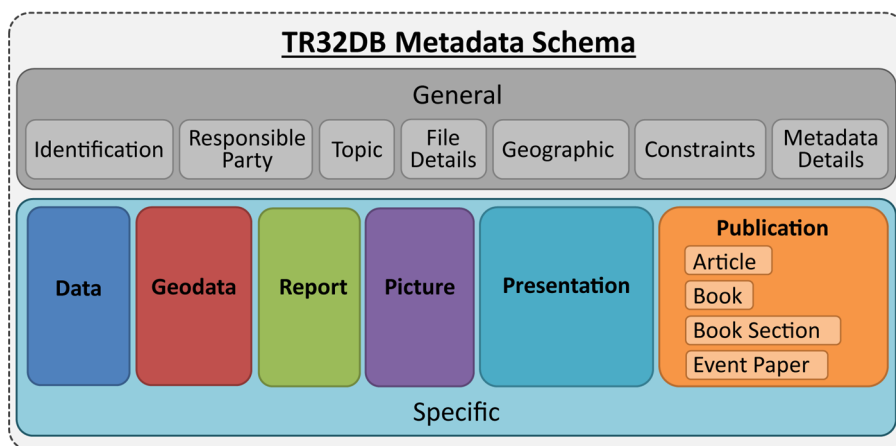
Figure 4-3: TR32DB Metadata Schema Structure set up in a general and specific level

Every data type category includes a certain number of properties and corresponding sub-properties. They are described in detail in a tabular form (e.g. Table 4-1) with the following attributes:

Identifier (ID) defines the number of the metadata property or sub-property. Properties are represented with a single number.

TR32DB property name is the label assigned to the metadata property or sub-property. All metadata property names start with a capital letter. All metadata sub-property names start with a lower case letter. Multiple words are concatenated to one single word, with each sub-word beginning with a capital letter.

Definition describes the metadata property or sub-property.

Occurrence (OCC) indicates the quantity or cardinality constrains of the metadata property or sub-property. The following selection is available:

0-n means metadata property or sub-property is optional and repeatable

0-1 means metadata property or sub-property is optional, but not repeatable

1-n means metadata property or sub-property is required and repeatable

1 means metadata property or sub-property is required, but not repeatable

Obligation (OB) describes the level of creation for each metadata property or sub-property. This is distinguished in:

Mandatory (M) properties or sub-property must be provided

Optional (O) properties or sub-property may be provided

Automatic (A) properties or sub-property are automatically generated

> **Notes** of a metadata property or sub-property provide information about the allowed values, examples, or other constraints, for example data value type (e.g. free text, date, controlled vocabulary), syntax encoding scheme (e.g. encoding for a date), or details on the obligation level.

Table 4-1 provides an example of the description of the property `Title` and the corresponding sub-property `titleType`. Both are associated with the category `Identification`. The `Title` property is mandatory and is required at least one time, but can be repeated as often as required. A free text can be used for this property. The sub-property `titleType` is also mandatory, but may occur only once. This sub-property demands the use of a value from the associated controlled vocabulary list `TitleType`.

Table 4-1: Description of the metadata property `Title` and the sub-property `titleType`

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Identification					
1	Title	A name given to the dataset. Typically a title will be a name by which the dataset is formally known.	1-n	M	Free Text Example: Enhanced Land Use Classification of 2012 for the Rur catchment
1.1	titleType	The type of Title.	1	M	Controlled List: **TitleType** See underline appendix for values and definitions.

The controlled vocabulary list `TitleType` (Table 4-2) is presented in the appendix of the metadata schema documentation (appendix D.5, Table D-54). The attribute values of the controlled list are composed of various metadata schemas and standards. The header line of the table indicates the origin of the value name. For the list `TitleType`, values of the Dublin Core (DCMI, 2012a, b), the DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011), and the ISO19115:2003 Metadata Standard (ISO, 2003) are specified. A mapping has been conducted of the available values. The used TR32DB value contains values of the just mentioned metadata schemas and standards and is extended with the own value `mainTitle`. In addition, a definition of each single TR32DB values is provided.

Table 4-2: The controlled list `TitleType` (appendix D.5, Table D-54)

ID	Value name TR32DB	Value name (dcterms)	Value name (DataCite)	Value name (ISO)	Definition
1.	mainTitle				Main title of the dataset.
2.	alternativeTitle	alternative	alternativeTitle	alternateTitle	Alternative title of the dataset.
3.	subtitle		subtitle		Subtitle of the dataset.
4.	translatedTitle		translatedTitle		Translated title of the dataset.

In contrast, self-established controlled value list are also provided in the TR32DB Metadata Schema. They are created with focus on the background of the CRC/TR32 or with regard to the demands of the data type categories. Some of the controlled lists are marked with an asterisk *. They are continuously updated by the CRC/TR32 project participants. For example an excerpt from the controlled list `MeasureLocationTR32` is presented in Table 4-3 (appendix D, Table D-33).

In addition, self-established controlled lists are provided with focus on the background of the CRC/TR32 that are not described in the documentation of the TR32DB Metadata Schema. In general, these lists are also marked with an asterisk * and are also continuously updated. In contrast, these controlled lists are more frequently and contain a huge amount of values. Consequently, they are not described in detail in the

documentation of the TR32DB Metadata Schema. Instead they are available online and accessible for the TR32DB users. For example the controlled list `InstitutionList` is available at `http://www.tr32db.de/institutionList`, the list `InstrumentList` is available at `http://www.tr32db.de/instrumentList`, and the list `keywordList` is accessible at `http://www.tr32db.de/keywordList`.

Table 4-3: Excerpt from the controlled list `MeasureLocationTR32` (appendix D.5, Table D-33)

ID	Value name TR32DB	Definition
1.	Wüstebach	TR32 measure/model location Wüstebach located in Erkensruhr.
2.	Rollesbroich	TR32 measure/model location Rollesbroich located in Kall.
3.	Merken	TR32 measure/model location Merken located in Ellebach.
4.	Merzenhausen	TR32 measure/model location Merzenhausen located in Ellebach.

Moreover, controlled lists are provided that are predefined by a standard or organisation. For example the controlled list `Language` is available at `http://www.tr32db.de/LanguageList` and the controlled list `MimeType` is accessible at `http://www.tr32db.de/mimeTypeList`.

In the following, the content of the single metadata property categories of the TR32DB Metadata Schema will be described in detail. The numbers (identifiers) indicate the corresponding entry in the TR32DB Metadata Schema description (appendix D.3).

General level

The category `Identification` includes metadata properties with focus on the identification of the described data (appendix D.3.1). This category specifies the following properties:

1. Title (with titleType) (M)
2. Description (with descriptionType) (M)
3. Identifier (with identifierType) (O)
4. Relation (with identifier, identifierType, relationType) (O)
5. CitationAdvice (O)
6. Additional Description (with descriptionFileName, graphicFileName) (O)

The properties `Title` and `Description` are both mandatory. They are required at least one time, but can be repeated as often as demanded. They are used to describe the data with basic textual information. Further details can be added by using other optional properties of the `Identification` category. For example, by using the property `Relation`, a relationship from the described data to another resource or identifier can be established. This requires the declaration of the sub-properties: `identifer`, `identiferType`, and `relationType`. The `identifier` represents a unique value identifying the data (e.g. 10.5880/TR32DB.1, 0454-1294). The `identifierType` specifies the type of identifier. Here, a value is demanded from the controlled list `IdentifierType` (e.g. DOI, ISBN). In addition, the type of relation is required. A value form a controlled list `RelationType` is requested (e.g. isVersionOf, isPartOf). By using the property `CitationAdvice`, additional information can be provided to cite the described data. Moreover, the described data can be optionally extended with additional information using the `AdditionalDescription` sub-property. It is possible to add an optional extra description text and graphic file name. Information about the described data can be added, which is not covered by other metadata properties. For example, detailed background information about the used measurement or modelling approach, the used method, or considered study site. This enables the better understanding of the described data.

The category `Responsible Party` covers metadata properties with focus on the responsibility of the data (appendix D.3.1). This category specifies the following properties:

7. Creator (including CreatorPerson and CreatorOrganisation) (M)
8. Contributor (with contributorName, contributorType, nameIdentifier, nameIdentifier-Schema) (O)
9. Publisher (M)

The property `Creator` describes the person(s) or organisation(s) that are associated with the creation of the data. It is mandatory and has to be used at least once and can be repeated as often as required. This property is subdivided into the sub-properties `CreatorPerson` and `CreatorOrganisation`. `CreatorPerson` is used by default to describe data, while the use of `CreatorOrganisation` is required and mandatory, when claiming conformance to ISO 19115 and INSPIRE. Consequently, the use of `Creator-Organisation` is currently only usable to describe data from the data type category `Geodata`. The property `Creator` requires the mandatory sub-properties `creatorFirstName` (e.g. John), `creatorFamilyName` (e.g. Public), and `emailAddress` (e.g. john.public@gmail.com) that enable all free text. The sub-properties `TR32Membership` and `creatorRole` demand a value according to the controlled vocabulary lists `RoleType` (e.g. Author, Owner, PointOfContact) and `CreatorStatus` (e.g. FullMember, NoMember). While the `CreatorStatus`-list is a self-established controlled value list, the `RoleType`-list is a predefined controlled vocabulary list provided by ISO 19115 and INSPIRE Implementing Directive. In addition, the optional identifier for the creator can be added by using the free text sub-property `nameIdentifier` and corresponding controlled-list property `nameIdentifierScheme`. For example, the creator can be an identifier with the name '0000 0001 1701 487X' and the corresponding value 'ISNI' (International Standard Name Identifier). In addition, the `CreatorPerson` property requires the mandatory sub-property `CreatorInstitution`, describing the institution details of the data creator with the following sup-properties `universityName`, `instituteName`, `streetName`, `postcode`, `city`, `country`, and `website`.

The property `Contributor` describes a person or institution that is responsible for making contributions to the data and is an optional property within the category `Responsible Party`. It may be repeated as often as required. Details like `contributorName` (e.g. John Smith), `contributorType` (e.g. Distributor, PrincipalInvestigator), `nameIdentifier` (e.g. B-5321-2013), and corresponding `nameIdentifierScheme` (e.g. ResearcherID) can be added in a similar way to the just described `Creator` properties.

Finally, the publisher of the data is requested via the property `Publisher`. The publisher is the organisation responsible for making the data available. The default value is 'CRC/TR32 Database (TR32DB)'. In case of the data type category `Publication`, the publisher should be named according to the publishing house of the publication, for example 'Elsevier Science, Amsterdam, The Netherlands'.

The category `Topic` contains metadata properties with the aim to describe the data with a subject or theme (appendix D.3.1). This category specifies the following property:

10. Subject (with TR32topic, TR32keyword, GEMETThesaurus, topicCategory, INSPIRETheme, and DDC) (M/O)

The property `Subject` contains various sub-properties to describe the data with terms according to a specific theme topic. The sub-properties `TR32topic` and `TR32keyword`, both self-established controlled value lists are mandatory. The `TR32topic` defines a subject describing the CRC/TR32 specific topic. Available values are represented in the controlled list `TR32Topic`, for example 'atmosphere', 'land use' or 'remote sensing'. The controlled value list for the sub-property `TR32keyword` is continuously updated according to the requirements of the CRC/TR32 participants. Therefore, this list is an exception and available online at http://www.tr32db.de/keywordList. Examples include: 'Canopy', 'Carbon' or 'Cloud Coverage'. In addition, optional sub-properties can be added. For example, these are subjects describing the

data taken from the General Multi-Lingual Environmental Thesaurus (GEMET) (EIONET, 2012a) (e.g. climate, crop rotation, classification) or the Dewey Decimal Classification System (DDC) (OCLC, 2014) (e.g. 550 Earth Sciences, 630 Agriculture). Moreover, subjects taken from ISO 19115 and/or INSPIRE Implementing Directive can be added. The sub-property `topicCategory` enables to describe data according to a category provided by ISO 19115 and INSPIRE. This value is mandatory, when claiming conformance to ISO 19115 or INSPIRE. The corresponding controlled value list is called `IsoTopicCategory`. Example values are 'GeoScientificInformation', 'Environment' or 'ClimatologyMeterologyAtmosphere'. Likewise, the sub-property `INSPIRETheme` (EIONET, 2012b) is defined that is also mandatory if claiming conformance to INSPIRE. The corresponding controlled value list is called `INSPIRETheme` and includes values like 'Atmospheric conditions', 'Land cover' or 'Land use'. With the exception of the sub-property `TR32topic` and `topicCategory`, all other subjects can be repeated as often as needed.

The category `File Details` includes various properties with focus on the format and file details of the described data (appendix D.3.1). This category specifies the following properties:

11. DataType (M)
12. DataFormat (with dataFormatVersion) (O)
13. DataSize (with dataSizeType) (O)
14. Date (with dateType) (M)
15. Lanuage (M)
16. DataStatus (M)

The properties `DataType` (e.g. Dataset, Text), `Language` (e.g. English), and `DataStatus` (e.g. Raw Data, Completed) are all mandatory and are required only one time. All three apply controlled lists. The property `Date`, also mandatory, but required at least one time and repeatable as often as required, is composed of the date-format property `Date` (e.g. 2014-04-04) and the controlled-list sub-property `dateType` (e.g. Available, Created, Issued). The properties `DataFormat` (e.g. PDF; JPEG; MS EXCEL 2007) and `DataSize` (15 Pages; 10 Datasets) are both optional and may be provided as often as needed. Both include a free text and a corresponding controlled-list value.

The category `Constraints` comprises various properties with regard to access and use constraints of the described data (appendix D.3.1). This category specifies the following properties:

17. Download (with downloadPermission, downloadInformation) (M/O)
18. Constraints (with accessAndUseConstraints, accessLimitations, licence) (O)

The property `Download` is divided into the sub-properties `downloadPermission` and `downloadInformation`, whereby the former is mandatory and the latter is optional. Both are not repeatable. The sub-property `downloadPermission` demands a value from the associated controlled list `DownloadPermission` (e.g. OnlyTR32, Free). The chosen value determines who is able to download the described data. Additional details concerning the download can be provided by using the free text sub-property `downloadInformation`. Further information concerning the access and use, as well as access limitations can be provided via the optional free text sub-properties `accessAndUseConstraints` (e.g. According to the TR32DB data policy agreement) and `accessLimitations` (e.g. no limitations). Both are not repeatable. In addition, a licence can be added by using the optional, non-repeatable sub-property `licence`. The associated controlled list `License` contains the available licences values such as 'TR32DB Data Policy Agreement' or 'Creative Commons - Attribution 3.0 Unported'.

The category `Geographic` comprises various properties with regard to the geographic background of the described data (appendix D.3.1).

This category specifies the following properties:

19. MeasuringSite (with measuringRegion, measuringLocation) (M)
20. GeographicBoundingBox (with westBoundLongitude, eastBoundLongitude, northBoundLatitude, southBoundLatitude) (O)

The property `MeasuringSite` is used to describe the place, where the described data is measured or modelled or focuses on. This is subdivided into the sub-properties `measuringRegion` (e.g. Ellebach, Erkenruhr) and `measuringLocation` (e.g. Merken, Selhausen). Both are mandatory, required only once, and are associated with a corresponding controlled list. Furthermore, an optional geographic coverage can be declared via the property `GeographicBoundingBox` expressed in terms of geographic coordinates given as a bounding box.

The category `Metadata Details` comprises various properties with focus on the creation of the metadata of the described data (appendix D.3.1). This category specifies the following properties:

21. TR32DBIdentifier (with TR32DBUrl) (A)
22. TR32MetaDataType (A)
23. MetadataCreator (with creatorFirstName, creatorFamiliyName, academicTitle, eMailAddress, phone, fax, TR32subproject) (A)
24. MetadataCreatorInstitution (with creatorFirstName, creatorFamiliyName, academic-Title, eMailAddress, phone, fax, TR32subproject) (A)
25. MdCreationDate (A)
26. MdLastUpdateDate (A)
27. MdWebVersion (A)
28. MdLanguage (A)
29. FileInformation (with filename, fileExtention, fileFormat, fileSize, fileUploadTime, fileSubproject, fileFundingPhase) (A)
30. Conformity (with specificationTitle, date, dateType, degree, explanation) (A)

The properties of the category `Metadata Details` are all automatically created by the TR32DB System. Most of them appear once and are not repeatable. A unique identifier for the described data is presented with the property `TR32DBIdentifier` (e.g. 787, 632, 724). The property `TR32MetaDataType` demands the use of the corresponding controlled list `TR32MetaDataType` that includes the data type categories (e.g. data, geodata, picture, publication). Depending on this value, the data type specific metadata properties will be adjusted. The properties `MetadataCreator` and `MetadataCreatorInstitution` describe the contact details of the person responsible for creation or updating the metadata information. The creation and last update date of the metadata are presented in the date-format properties `MdCreationDate` (e.g. 2014-04-01) and `MdLastUpdateDate` (e.g. 2014-04-04). The property `FileInformation` contains details about the described file like `filename` (e.g. Land_Use_Classification_2013), `fileExtention` (e.g. .zip, .pdf) or `fileSize` (e.g. text/css, image/jpeg, application/pdf). Finally, the property `Conformity` describes details about the conformity of the provided metadata to a selected specification. Available specifications are listed in the corresponding controlled list `ConformitySpecification` such as 'DCMES', 'ISO 19115' or 'DataCite'.

Specific level

Categories within the specific level contain properties of metadata according to the requirements of a specific file type. They will be subsequently described.

The category `Data` comprises various additional properties with regard to the creation and background of the described data from the data type category `Data`.

This category was created to fulfil the data type requirements and specifies the following properties (appendix D.3.2):

31. TemporalExtent (with startDate and endDate) (M)
32. Lineage (O)
33. MeasuringInstrumentModelMethod (Instrument with: instrumentName, instrument-Model, instrumentManufacturer, manufacturerRegisterdOffice, manufacturerWebsite; InstrumentParameter with: ParameterName, parameterUnit, ResolutionDistance, resolutionDistanceUnit, TemporalFrequency, temporalFrequencyUnit) (O)

The property `TemporalExtent` is the only mandatory property within the category `Data` and is not repeatable. This property describes the time period covered by the described data and consists of the sub-properties `startDate` (e.g. 2014-04-01) and `endDate` (e.g. 2014-04-04). Both are mandatory and have the date format. The property `lineage` may be added optionally. This free text property is not repeatable and should specify the process history and/or overall quality of the described data, for example details about validity or quality assurance. The property `MeasuringInstrumentModelMethod` (e.g. Temperature Sensor) is also optional, but repeatable as often as demanded. It should define the measuring instrument or modelling method used to create the described data. Besides details about the instrument or method, e.g. instrument name, model specification or manufacturer can be named. Furthermore, optional information about the output parameter (measured or modelled) may be added by using the sub-property `InstrumentParameter`. This sub-property is repeatable as often as needed and defines various sub-properties, for example, describing the name or unit of the parameter (e.g. Temperature, °C; Total Plant Weight, cm).

The category `Geodata` covers various additional properties with regard to the creation and background of the described data from the data type category `Geodata` (appendix D.3.3). This category was created to fulfil the requirements of the spatial related data and with regard to the core metadata of ISO 19115 and INSPIRE Directive. The following properties are specified:

34. TemporalExtent (with startDate, endDate) (O)
35. VerticalExtent (with minimumValue, maximumValue, referenceSystem, referenceSystemCode, referenceSystemName) (O)
36. Lineage (O)
37. Spatial Resolution (with scale, distanceNumber, distanceUnit) (O)
38. SpatialRepresentationType (O)
39. PresentationForm (O)
40. ReferenceSystem (with referenceSystemCode, referenceSystemName) (O)
41. MaintenanceInformation (with maintenanceFrequency, maintenanceNote) (O)
42. Scope (O)
43. InitiativeType (O)

The entire complementing properties of the category `Geodata` are optional. As already described for the category `Data`, the optional properties `TemporalExtent` and `Lineage` may be defined once. Furthermore, information about the vertical domain of the described data can be extended. The spatial resolution of the described data may be provided. This may be expressed either as a scale factor or using a ground distance value with the corresponding unit of measurement. The used spatial reference system is specified in the property `ReferenceSystems`. This requires the sub-properties `referenceSystemCode` (e.g. 32632) and `referenceSystemName` (e.g. WGS 84/UTM zone 32N), both represented with a controlled list. The frequency of maintenance and updates of the described data and complementary notes may be specified with the property `MaintenanceInformation`. The scope of the data and the type of initiative under which the data was produced may be added with the support of the predefined controlled lists `ScopeCode` (e.g. dataset, series) and `InitiativeType` (e.g. campaign, experiment).

The category `Report` comprises various additional properties with regard to the creation, background, and citation of the described data from the data type category `Report` (appendix D.3.4). This category was created to fulfil the data type demands and specifies the following properties:

44. ReportDate (M)
45. ReportType (O)
46. ReportCity (M)
47. ReportInstitution (M)
48. ReportVolume (O)
49. NumberOfPages (O)
50. PageRange (with startPage, endPage) (O)
51. FurtherInformation (O)

The category `Report` specifies mandatory and optional properties. Mandatory properties, required to accurately cite a report, are: `ReportDate` (e.g. 2014-04-01), `ReportCity` (e.g. Cologne, Gemany), and `ReportInstitution` (e.g. University of Cologne, Institute of Geography). All are not repeatable. Furthermore, optional properties may be added. The optional, non-repeatable property `ReportType` is related to the self-defined controlled vocabulary list `ReportType`, which includes for instance 'PhDReport', 'MasterThesis' or 'Minutes'. Moreover, the volume of the report, as well as the number of pages, and the page range of the report including start and end page can be added. Details about the report, not covered by other metadata properties, can be provided with the free text property `FurtherInformation`.

The category `Picture` contains various additional properties with regard to the creation, background, and citation of the described data from the data type category `Picture` (appendix D.3.5). This category specifies the following properties:

52. RecordDate (with startDate, endDate) (M)
53. RecordPlace (M)
54. RecordMethod (O)
55. Orientation (O)
56. Resolution (O)
57. Size (with width, height) (O)
58. CopyrightInformation (O)
59. Event (with eventType, eventName, eventLocation, eventWebsite, eventPeriod (with startDate, endDate)) (O)

The mandatory properties for the category `Picture` are `RecordDate` (e.g. 2014-04-01) and `RecordPlace` (e.g. Selhausen, Germany). Both are not repeatable and are required to accurately cite the described data. Moreover, optional properties can be provided. For example, the orientation of the picture can be added by using a value from the controlled list `Orientation` (e.g. Horizontal/Landscape), the resolution can be provided in dots per inch (dpi) such as '200' dpi, the size can be specified for the width and heights in pixels (e.g. 150 x 300) and copyright information can be complemented. In addition, details about the event, where the picture was taken, can be indicated.

The category `Presentation` includes various additional properties with regard to the creation, background, and citation of the described data from the data type category `Presentation` (appendix D.3.6). This category specifies the following properties:

60. Presenter (O)
61. PresentationDate (M)
62. PresentationType (M)
63. Event (with eventType, eventName, eventLocation, eventWebsite, eventPeriod (with startDate, endDate)) (M/O)

To accurately describe and cite data from the category `Presentation`, three mandatory properties have to be used. These are the not-repeatable properties `PresentationDate` (e.g. 2014-04-01) and `PresentationType` that require a value from the controlled list `PresentationType` (e.g. Keynote, Talk, Poster), and `Event`. Optional, the presenter of the presentation may be complemented, also not repeatable.

The category `Publication` is an exception and subdivided in the sub-categories `Article`, `Book`, `BookSection`, and `EventPaper`. The properties of each sub-category distinguish various properties. These will be described below.

The sub-category `Publication-Article` comprises various additional properties with regard to the citation of the described data from the data type category `Publication` (appendix D.3.7). This category specifies the following properties:

64. Status (M)
65. Review (M)
66. Year (M)
67. PublicationType (M)
68. ArticleType (O)
69. PublicationSource (with PublicationSourceWebsite) (M/O)
70. Volume (O)
71. Issue (O)
72. NumberOfPages (M)
73. PageRange (with startPage, endPage) (M)

The sub-category `Publication-Article` defines seven mandatory properties that are necessary to accurately cite the described data, all not-repeatable. The properties `Status` and `Review` status of the publication require a value from the associated predefined controlled list `PublicationStatus` (e.g. inPrint, Accepted, Published) and `PublicationReview` (e.g. PeerReview, NoPeerReview). Furthermore, the year of publication has to be added, as well as the type of publication linked. The latter is linked to the corresponding self-predefined controlled list `PublicationType` (e.g. Article, Book, BookSection, EventPaper). The value `Article` must be chosen. In addition, the name of the publication source (e.g. journal or magazine name) has to be provided with an optional website. The number of pages and the page range of the article are also mandatory. The volume and issue of the publication source are optional.

The sub-category `Publication-Book` includes various additional properties with regard to the citation of the described data from the data type category `Publication` (appendix D.3.7). This category specifies the following properties:

74. Status (M)
75. Review (M)
76. Year (M)
77. PublicationType (M)
78. SeriesTitle (O)
79. SeriesEditor (O)
80. City (M)
81. Volume (O)
82. Edition (O)
83. NumberOfPages (O)
84. PageRange (with startPage, endPage) (O)

Again, the mandatory properties `Status`, `Review`, `Year`, and `PublicationType` have to be provided. The mandatory value for the `PublicationType` is `Book`. Furthermore, the `City` property, describing the

location where the publisher is located, is mandatory (e.g. Cologne, Germany). In addition, the optional properties may be provided, for example series title and editor, as well as volume, edition, number of pages, and page range of the book.

The sub-category `Publication-BookSection` specifies various additional properties with regard to the citation of the described data from the data type category `Publication` (appendix D.3.7). This category specifies the following properties:

85. Status (M)
86. Review (M)
87. Year (M)
88. PublicationType (M)
89. BookTitle (M)
90. BookEditor (M)
91. City (M)
92. SeriesTitle (O)
93. SeriesEditor (O)
94. Volume (O)
95. Chapter (O)
96. NumberOfPages (O)
97. PageRange (with startPage, endPage) (M)

The category `Publication-BookSection` specifies eight mandatory properties to accurately describe and cite the described data. Again, these are the properties `Status`, `Review`, `Year`, and `PublicationType` with the mandatory value `BookSection`. In addition, the provision of the title, editor, and publishing city of the book are mandatory, as well as the page range of the book section. Details about the series title and editor, volume, chapter, and number of pages may be provided optionally. The page range is mandatory.

The sub-category `Publication-EventPaper` contains various additional properties with regard to the creation and citation of the described data from the data type category `Publication` (appendix D.3.7). This category specifies the following properties:

98. Status (M)
99. Review (M)
100. Year (M)
101. PublicationType (M)
102. Event (with eventType, eventName, eventLocation, eventWebsite, eventPeriod (with startDate, endDate)) (M/O)
103. Proceedings (ProceedingsEditor, ProceedingsTitle, ProceedingsURL, Volume, Issue) (O)
104. NumberOfPages (O)
105. PageRange (with startPage, endPage) (M)

The sub-category `Publication-EventPaper` defines six mandatory properties. Again, the properties `Status`, `Review`, `Year`, and `PublicationType` (with the mandatory value `EventPaper`) have to be provided and are not-repeatable. Details about the page range of the described publication have to be provided. Furthermore, information about the event that is associated with the described publication has to be specified, for example event name, location, type or duration. If available, further optional details can be given like the number of pages of the publication or indications about the proceedings (e.g. editor, title, website, volume, issue).

Schema Mapping

A mapping between the TR32DB metadata properties and elements of four metadata schemas and standards was conducted. This mapping was used to create the TR32DB Metadata Schema and it represents the interoperability to the applied metadata standards and schemas (section 3.3). The following metadata schemas and standards are included in the mapping: Dublin Core (DCMI, 2012a, b), DataCite Metadata Schema Version 2.2 (DATACITE METADATA WORKING GROUP, 2011), ISO19115:2003 Metadata Standard (ISO, 2003), and INSPIRE Implementing Directive (EUROPEAN COMMISSION, 2008) along with the INSPIRE Metadata Implementing Rules (DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE, 2010). In addition, further metadata schemas are considered with regard to future use of the TR32DB Metadata Schema. For example, various metadata properties of different ontologies are included like the Bibliographic Ontology (D'ARCUS & GIASSON, 2009).

For each property of the TR32DB Metadata Schema the mapping is conducted. The entire result is presented in a tabular form available in the documentation of the TR32DB Metadata Schema (appendix D.3.7). The summary table provides an overview of the values of the applied metadata standards and schemas that match the single TR32DB metadata properties and sub-properties. Besides the name of the value, the identifier number (ID) of the respective metadata standard or schema is also specified, as well as the particular obligation (O.). In case of the mapping to INSPIRE, the ID is referenced to Part B of the INSPIRE Implementing Directive. All mandatory properties or sub-properties are highlighted in a specific colour. Some abbreviations are applied to in the summary table of the mapping to mark the used metadata schema or ontology. For example, the Dublin Core Metadata Element Set Version 2.2 is shortened to *dc* and the DCMI Metadata Terms are abbreviated with *dcterms*. The used ontologies are likewise abridged, for example the Bibliographic Ontology (bibo), the Event Ontology (event), and the Time Ontology (time). The abbreviation and the appropriate value are indicated.

As an example, Table 4-4 presents the mapping of the property `Title` and the corresponding sub-property `titleType`, both assigned to the category `Identification`. The available values of the various metadata schemas and standards are mapped to `Title` and `titleType`. In addition, corresponding identifiers are specified, as well as the obligation. Finally, the mandatory properties are highlighted.

Table 4-4: TR32DB metadata property mappingfor the property Title and the sub-property titleType. Mandatory properties are highlighted.

ID	TR32DB property	Dublin Core	ID	ISO	O.	ID	INSPIRE	O.	ID	Data Cite	O.	Other Elements
Identification												
1	Title	dc.title dcterms: title	360	CI_Citation. title	M	1.1	Resource title	M	3	Title	M	
1.1	titleType	dcterms. alternative	361	CI_Citation. alternateTitle	O				3.1	Title Type	C	

The supplementary Table 4-5 shows the mapping of the TR32DB property `NumberOfPages` and `PageRange` with the corresponding sub-properties `startPage` and `endPage`. Both are assigned to the category `Publication` and the sub-category `Article`. In this example the compliance is only achieved with `Other Elements`. For example, the TR32DB property `NumberOfPages` can be mapped to the Bibliographic Ontology element `bibo:numPages`.

Table 4-5: TR32DB metadata property mapping for the property `NumberOfPages` and `PageRange`

ID	TR32DB property	Dublin Core	ID	ISO	O.	ID	INSPIRE	O.	ID	Data Cite	O.	Other Elements
Publication												
Article												
72	NumberOfPages											bibo:numPages
73	PageRange											bibo:pages
73.1	startPage											bibo:pageStart
73.2	endPage											bibo:pageEnd

4.4   The TR32DB storage and organization of data

TR32DB users who want to store their data in the TR32DB data storage (section 3.4.2) should follow some pieces of advice from the TR32DB administrators. Before uploading data to the TR32DB, it is recommended to process and organise the files (section 2.1.3). For example, measured data can be sorted according to the measurement location, the used measurement instrument or the measuring time extent. The data should be compressed in a common standard format like .zip, .rar, or .tar.gz. Particularly, a large number of files with a small file size from continuous measurements or modelling should be grouped and compressed into one single folder, for example a daily, a weekly, or a monthly series of measurements. Furthermore, it is necessary to arrange the data according to the sort of data types that can be stored. These are data (e.g. created or collected results from measurements or modelling), pictures (e.g. single pictures or sets of pictures from field measurement campaigns, project meetings etc.), presentations (e.g. presentations presented at conferences, workshops, etc.), publications (e.g. publications created by the project participants), and reports (e.g. internal reports, student reports, hints and tips, transcripts, etc.). Finally, all data files should be appropriately entitled.

Furthermore, all TR32DB users have to consider some administrative and technical restrictions of the TR32DB data storage. They are just allowed to upload their data in their specific sub-project subdirectory (e.g. A1, A2, B1, C3, C4, etc.) of the particular funding phase (Phase 1, Phase 2, Phase 3) in which the dataset was created. In addition, the upload of data is limited to 8 GB per data type of a project section per day, since the data moving script operates over night (section 3.4.2). Therefore, the maximum file size of a single dataset is currently approximately 7 GB. However, a maximum file size of 4 GB is recommended for single files. In exceptional cases, the upload of increased datasets is possible in prior consultation with the TR32DB administrator.

In the following section, the upload of the project data to the TR32DB data storage will be described. In general, the user has to follow four steps to upload his data to the TR32DB data storage, as illustrated in Figure 4-4.
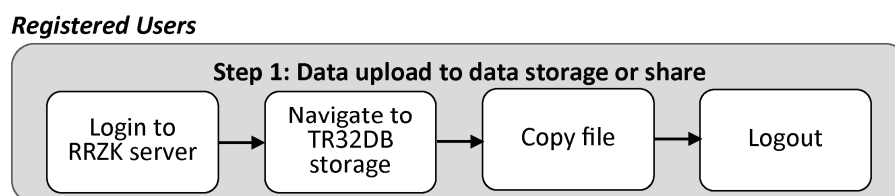
Figure 4-4: TR32DB Data storage workflow (Step 1)

Because of the various disciplines in the CRC/TR32, different operating systems are used by the project participants. Consequently, the upload of data to the TR32DB system will be exemplarily described by the widely used operating systems Microsoft Windows and Linux.

4.4.1 Data storage for Windows users

Windows users have various possibilities to connect to the TR32DB AFS data storage (section 3.4.2) in the RRZK and upload their data. They can use different SSH clients to access the AFS storage. For examples they can use the OpenAFS Client (OPENAFS, 2013) or common SFTP or SCP software like FileZilla (FILEZILLA, 2014) or WinSCP (WINSCP, 2014). Below, the access to the TR32DB AFS data storage will be exemplary described using the WinSCP software.

Following the TR32DB tutorial *Instructions for storing data in the TR32DB using MS Windows* (CURDT, 2013b) available via the internal TR32DB webpage, the first demand is to install the WinSCP software. The second step is to log on to the RRZK Server (`dialog.rrz.uni-koeln.de`) with the RRZK user account by using the SFTP protocol. After a successful login, the user can set-up his WinSCP software user interface by adding for example bookmarks to his frequently used directories like the directory of his specific sub-project subdirectory in the TR32DB AFS `Upload` or `Share` directory. The data path to this subdirectory is always set up in the same order that is `afs/rrz/project/tr32db/Upload_<funding phase>/<cluster> /<project section>`. For example data path directories can be added to the bookmarks like `/afs/rrz.uni-koeln.de/project/tr32db/Upload_1/C/C3` to access the upload folder of the sub-project C3 in phase 1 and `/afs/rrz.uni-koeln.de/project/tr32db/Upload_2/C/C3` for the according upload folder of C3 of phase 2. Furthermore, `/afs/rrz.uni-koeln.de/project/tr32db/ share/C/C3` has to be used to access the appropriate share folder of C3. It is important to keep in mind that by default users just have access permission to their sub-project subdirectory of the funding phase, in which they are working.

As already introduced, the storage of data files in the TR32DB depends on the project section and data type. By using WinSCP, the user can simply copy and paste or drag and drop the desired file from his home directory (left) to the chosen TR32DB subdirectory (right). As shown in Figure 4-5, a publication is moved from the home directory to the data path `/afs/rrz.uni-koeln.de/project/tr32db/Upload/ Z/Z1/`.

The data storage in the TR32DB `share` directory operates like the already described data upload to `Upload` directory by using copy and paste or drag and drop to a modified data path. For example a user-defined file can be moved from the home directory to the data path `/afs/rrz.uni-koeln.de/project/tr32db/ share/Z/Z1`.
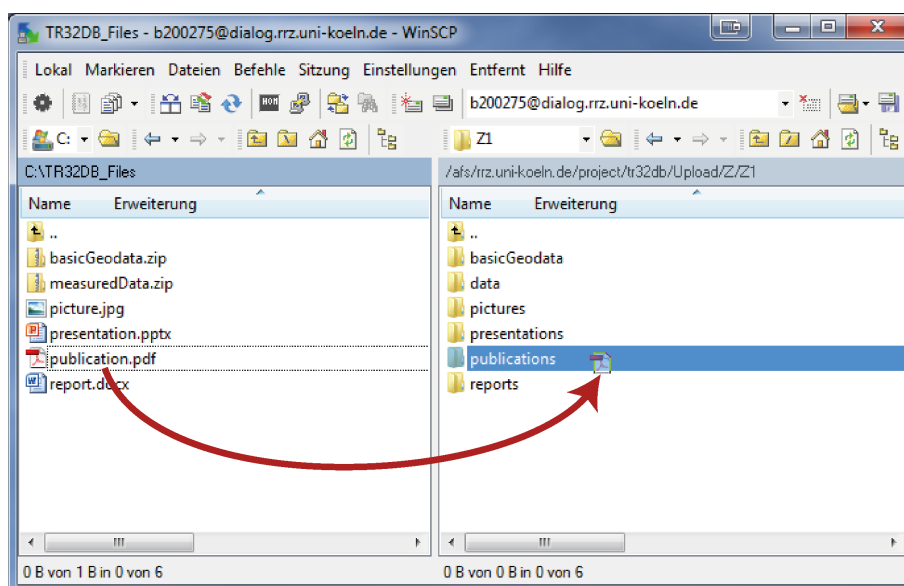
Figure 4-5: Data Upload to TR32DB AFS data storage, exemplary for a publication uploaded to project section Z1

4.4.2 Data storage for Linux users

Correspondingly, Linux or Unix users have different options to store their data in the TR32DB AFS data storage. A descriptive TR32DB tutorial *Instructions for storing data in the TR32DB using Linux* (CURDT, 2013a) is available via the internal TR32DB webpage for user support.

(i) To securely copy a file from the local home directory to the appropriate remote RRZK directory, the user has to use the command scp (secure copy). In general, the command has to include the file name, user name, server name and the remote directory, where the file should be copied. For example, if the user smithj wants to copy the file test.txt from his local directory to the remote directory /afs/rrz.uni-koeln.de/project/tr32db/Upload/B/B1/data, he would type the following command line into his shell:

```
smithj@agro601:~$ scp temp_profiles_2009.zip smithj@dialog.rrz.uni-
koeln.de: /afs/rrz.uni-koeln.de/project/tr32db/share/B/B1/data
smithj@dialog.rrz.uni-koeln.de's password:
temp_profiles_2009.zip          100%   3675KB       262.5KB/s     00:15
```

After submitting the command, the user password will be requested.

(ii) In the second option, the user will connect with the RRZK server (dialog.rrz.uni-koeln.de) first by using the ssh command and then copy the file using the scp command. In general, the ssh command has to include the user name and server name, e.g. user@dialog.rrz.uni-koeln.de. In the given example, the user smithj has to indicate the ssh command first and complete it by typing smithj@dialog.rrz.uni-koeln.de. After submitting, the user password is required.

```
smithj@agro601:~$ ssh smithj@dialog.rrz.uni-koeln.de
smithj@dialog.rrz.uni-koeln.de's password:
```

Finally, the chosen file can be copied by using the already described scp command.

4.5 The TR32DB web-interface

The webpage of the TR32DB is the connecting component of the entire TR32DB data management structure and the user interface to access the TR32DB features. The web-interface is accessible online since summer 2007 at http://www.tr32db.uni-koeln.de or http://www.tr32db.de. The main task of the webpage is the management of data created by the CRC/TR32 participants by means of their descriptive metadata. This includes search, visualisation, and metadata management functionalities. Besides general available features accessible for every visitor of the webpage (e.g. data search via metadata), the TR32DB also provides internal functions (e.g. editing of metadata, a climate data section, an integrated Web-GIS or the access to the TR32DB Share). In addition, web-interfaces are available for administration purposes (e.g. user management) for the TR32DB administrators. It is essential to develop a safe and stable web-interface with regard to licenses of purchased data and unpublished data files of CRC/TR32 participants, as well as management, download or visualization of large data sets.

Based on changing user requirements and needs, as well as developments of the Internet technology, the TR32DB is maintained on-going and consequential modified since the launch of the website in summer 2007. Therefore, this section will first focus in detail on the functionality and features of the TR32DB web-interface Version 3.1 and the associated administration web-interface (section 4.5.1). Subsequent, the TR32DB web-interface Version 4.0 will be presented, which is an improved, updated Version of the prior web-interface enhanced with additional features and a new layout (section 4.5.2).

4.5.1 Web-interface Version 3.1

This section focuses on the design and functionality of the TR32DB web-interface Version 3.1. Initially, the general design will be described, following by an overview of available user and admin user applications.

4.5.1.1 Web design

The web-interface Version 3.1 has been available online since the spring of 2011. This version is a further development of the originally launched TR32DB website version of 2007. The underlying database Version 3.1 of the website is described in section 3.5.1. The web-interface is designed and implemented according to previous mentioned recommendation rules, principles, and methods (section 2.3.4 and 3.6). The website has a simple layout colour-coordinated with the logo of the CRC/TR32, basically in green and yellow colours (Figure 4-6).

Two menu bars bestride the layout of the webpage. The user is able to access general information about the CRC/TR32 and TR32DB on the `general menu bar` on the top. The `data menu bar` (Side menu) on the left side deals with the project data of the CRC/TR32 and is divided in two sections. Section B facilitates organisation and search for project data via various predefined search queries or a combined search form. Section A offers complementary features with regard to project data for registered users. The `header` of the webpage is dominated by a title of the webpage "SFB TR32 – Database" and the logos of the participating institutions of the CRC/TR32. The `footer` of the webpage is formed by a tiny map showing the locations of the visitors to the TR32DB webpage provided by ClustrMaps (CLUSTRMAPS, 2014). In addition, copyright information and the version of the webpage are provided. The main `content` part of the webpage is changing according to the chosen menu point. In the following, the single parts of the web-page will be described in detail.
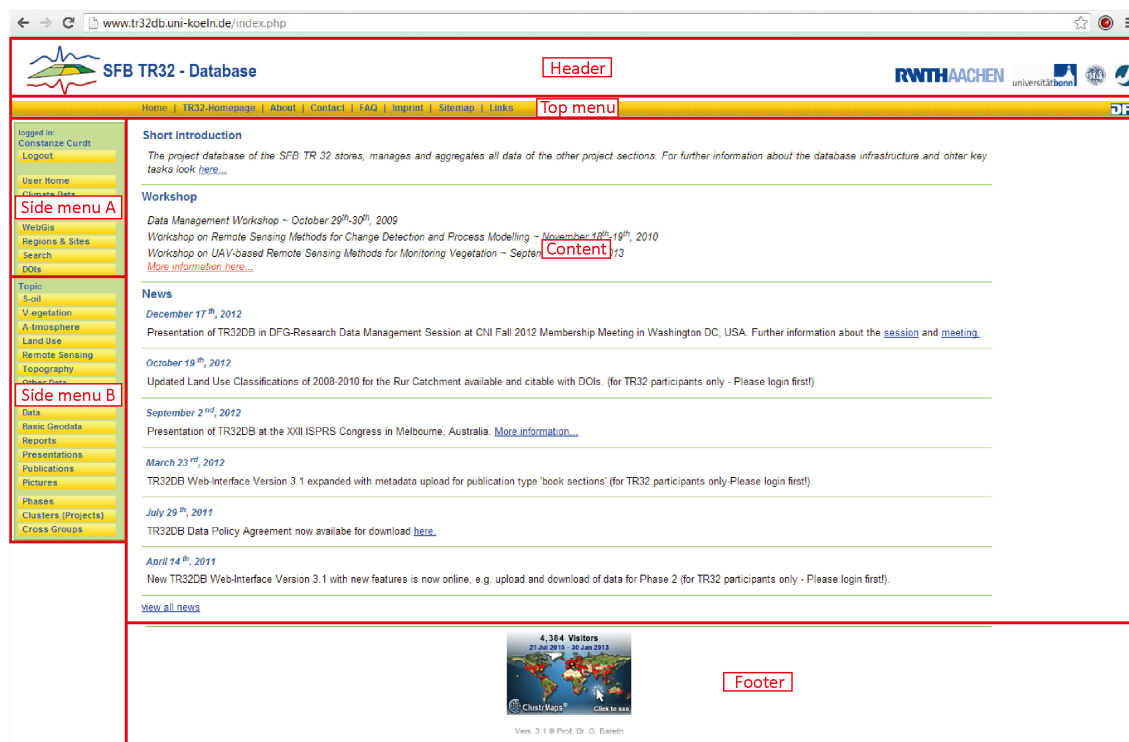
Figure 4-6: TR32DB web-interface start page Version 3.1

Startpage

The main content part of the start page (Figure 4-6) is basically divided into 3 parts: a short introduction to the TR32DB, information and linkage to Workshops organized and related to the TR32DB, and a news section.

Short introduction The short introduction links to the content of the `About`-webpage.

Workshops This section is linked to an overview webpage of all workshops, already organized or currently arranged by the CRC/TR32 INF/Z1 project sections. Additional details about each workshop are provided. For example, materials and further information of past workshop like posters, poster abstracts, pictures or the workshop program is accessible for download. In addition, for workshops in preparation, visitors can access information about the workshop location, travel or accommodation information, as well as important dates or author guidelines for papers and poster presentation guidelines. Moreover, they are able to register for the workshop via the webpage. Content for the workshop webpages (e.g. poster title, abstracts) or the workshop registration is provided by a TR32DB-workshops-database.

News The news section displays the latest news related to the TR32DB webpage, like update details for geodata, participation of the TR32DB at conferences or modification of the TR32DB webpage. The link `view all news` enable the display of the entire available news. Content for the news section is provided by the TR32DB database (section 3.5.1).

General menu bar

All website visitors are able to access general information at the `general menu bar` (top menu). These are:

Home The website visitor can navigate back to the TR32DB start page

TR32-Hompage This menu item links to the official CRC/TR32 web site located at http://www.tr32.uni-koeln.de.

About This section offers general information about research data management in interdisciplinary research projects as well as details about the implementation of the TR32DB.

Contact Information about the contact persons of TR32DB is provided here.

FAQ A couple of frequently asked questions are answered in this section to support the use of the TR32DB. In addition, the CRC/TR data policy agreement (CRC/TR32, 2011) is accessible for download, a list of CRC/TR32 keywords, as well as all TR32DB tutorials are available for download for registered users. For example, the tutorial *TR32 database (TR32DB) Web-Interface user guide (Version 3.1)* (CURDT, 2011) is provided, which describes the range of functions of the web-interface.

Imprint This section provides imprint details of the TR32DB and refer to the imprint of the University of Cologne at http://www.uni-koeln.de/uni/impressum/.

Sitemap Here, a hierarchically overview and direct access is given of all webpages provided by the TR32DB web-interface.

Links A list of links to related institutions and research projects is available here.

Data menu bar

All website visitors are able to search for CRC/TR32 project data at the `data menu bar` (section B). Predefined queries enable the search for project data in the TR32DB according to:

Topic Search for data according to the selected CRC/TR32 topics (soil, vegetation, atmosphere, land use, remote sensing, topography, and other data).

Data Type Search for data according to selected data types (data, basic geodata, reports, presentations, publications, and pictures)

Regions & Sites Search for data according to selected measurement/modelling regions and sites of the CRC/TR32, for example: Erkensruhr (Wüstebach), Ellebach (Merzenhausen, Selhausen, Merken), or Kall (Rollesbroich).

Phases Search for data according to the funding phases, e.g. Phase 1 (2007-2010) or Phase 2 (2010-2014).

Clusters (Projects) Search for data according to the cluster groups and project sections of the CRC/TR32, for example Cluster C and associated project sections C1, C3, C4, and C6.

DOIs Search for data that own a TR32DB DOI.

An advanced data **search** form enables all visitors to join various queries with each other. These are for example a full-text search, the funding phase, CRC/TR32 keywords, and measurement/modelling region and site are available to combine. The predefined and advanced data search will be described in detail in section 4.5.1.2.

All registered users are able to access further features provided by the `data menu bar` (section A). These are:

Login/Logout To access the internal features of the TR32DB, a registered user has to login to the web-interface. Likewise he has to logout to quit the session.

User Home This section provides various features to administrate the project data of a certain user via their metadata. Besides the upload of metadata for new datasets, viewing or editing of metadata of already available datasets, it is also possible to apply for a DOI for a dataset (section 4.5.1.3).

Climate Data The climate data section provides an overview about available climate data purchased within the CRC/TR32 (section 3.1.2). Information about data agreements and publication notices, detailed descriptions, important links, as well as download of data is offered. In addition, the feature `online query of climate data` according to a climate station and temporal extent is available, as well as the feature `draw climate diagrams` according to climate station, climate parameter, and temporal extent. This will be described in detail in section 4.5.1.3.

Share The share section lists all data that is currently stored in the TR32DB `share` data storage (section 3.34). The aim of this storage is the secure and temporary exchange of project data between the project sections. This will be described in detail in section 4.5.1.3.

WebGIS This section is automatically forwarding to the integrated TR32DB WebGIS, called the TR32 Viewer. Purchased geodata and geodata of project participants (section 3.1.2) is visualised here. Further information and features of the WebGIS will be described in section 4.5.1.3.

4.5.1.2   General web-interface features

All users or visitors of the web-interface are able to use general web-interface features in order to explore and download data of the TR32DB. For this purpose certain steps have to be followed (Figure 4-7).
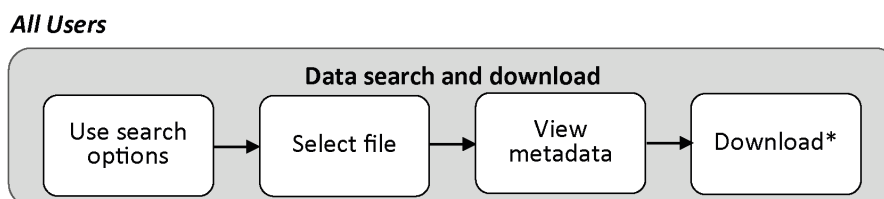
Figure 4-7: TR32DB user workflow for data search and download using the web-interface

These steps include the use of specific search functions (e.g. predefined search for data, combined search for data), selection of data files, exploration of data via metadata, as well as download of data depending on permissions. In the following, these features will be described in detail.

Predefined search for data

The predefined data search enables all visitors of the web-interface a fast and easy search for project data in the TR32DB according to determinate categories. The search for data operates via the metadata of the dataset. As already described in the previous section, the following predetermined categories are accessible via the left data menu bar: topic, data type, regions and sites, phases, clusters (projects), and DOIs.

In the following the process to use the various predefined data search opportunities will be described in detail.

> **Topic** The predefined search for data according to the category `topic` is subdivided in seven sub-categories related with the CRC/TR32 research focus. These are soil, vegetation, atmosphere, land use, remote sensing, topography, and other data. To search for data from a specific sub-category, the visitor has to select a sub-category form the data menu bar. As a result he will receive a list of all datasets in the TR32DB, which are assigned with the chosen topic.

> **DataType** The predefined search for data according to the category `data type` is subdivided in the sub-categories: data, basic geodata, reports, presentation, publication, and pictures. To search for data from a specific data type sub-category, the website visitor has to select a sub-category from the data menu bar. He will receive a list of all dataset in the TR32DB, which are assigned with the chosen data type, as a result.

> **Regions and Sites** The predefined search for data according to the category `regions and sites` offers a search according to the CRC/TR32 measurement or modelling locations (sites) and regions. To search for data in this category, first, the website visitor has to select this section from the data menu bar, he will receive an overview list of all available CRC/TR32 measurement or modelling locations and regions (Figure 4-8). Now, he is able to choose his preferred region or site. As a result, he will receive a list of all datasets in the TR32DB, which are assigned with the chosen region or site.
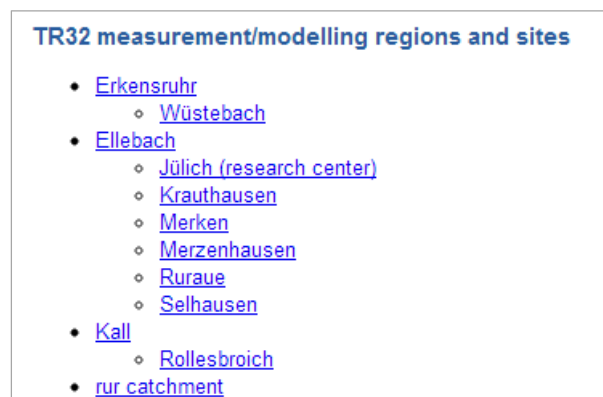
Figure 4-8: Predefined search for data according to regions and sites sub-categories

> **Phases** The predefined search for data according to the category `phases` offers a search according to funding phases of the CRC/TR32, when the dataset was created. To search for data in this category, the visitor of the website has to select this section from the data menu bar. The visitor will be forwarded to an overview list of all funding phases of the CRC/TR32 (Figure 4-9). In the next step, the preferred funding phase has to be chosen. A list of datasets assigned with the chosen phase will be displayed as a result.
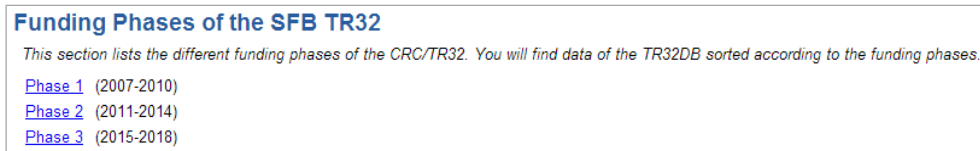
Figure 4-9: Predefined search for data according to funding phases category

Clusters (Projects) The predefined search for data according to the category `clusters (projects)` provides a search according to the cluster groups or project sections, who created the dataset. The website visitor will receive an overview list of all clusters and projects sections of the CRC/TR32 (Figure 4-10), by selecting the clusters (projects) section category from the data menu bar. Now, the preferred project section has to be selected. As a result, a list of datasets assigned with the chosen project section will be available.

Figure 4-10: Predefined search for data according to clusters (projects) category

DOIs The predefined search for data according to the category `DOIs` enable search for datasets, which own a DOI and can be cited with it. To search for data with a DOI, the category DOIs has to be selected from the data menu bar. As a result, a list of all datasets in the TR32DB with a DOI is shown (Figure 4-11). By clicking on the DOI number of a dataset, the website visitor will be forwarded to the overview webpage of the chosen DOI (Figure 4-15).
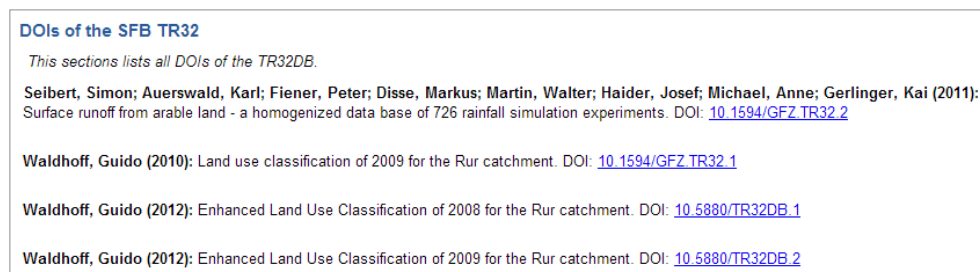
Figure 4-11: Predefined search for data according to DOIs category

Advanced Data Search

The advanced data search enables visitors of the website to search for data in the TR32DB by combining various queries with each other. The data search operates via the metadata of a dataset. By selecting the search section from the data menu bar, the website visitor will be forwarded to a search form (Figure 4-12). Here, it is possible to choose specific values from selective lists or enter terms in input fields. For example, the visitor is able to use a full-text search by typing in a specific term (e.g. cloud coverage), select the requested data type, the corresponding funding phase, the time period (e.g. 10.05.2009-10.11.2009), when the dataset was created, a CRC/TR32 keyword (e.g. Sunshine Duration), the dataset creator (e.g. Jan Schween) or the region and site, where the dataset is geographically located (e.g. Ellebach). A list of all matching datasets will be displayed as a result below.

Figure 4-12: Advanced data search with exemplary values and search result

Presentation of Data

All data of the TR32DB are displayed on the web-interface by means of their metadata. The representation of the metadata depends on the required data category, as well as on the requested data type. In general an overview list of all datasets is provided as a result of a predefined or combined query. By default, 20 datasets are shown in the overview table per webpage. To continue to the next page, the arrow next to the page number has to be selected. All data assigned with the required category are displayed in a table, for example, according to the data topic atmosphere (Figure 4-13). The header of the table will be slightly adapted to the data query, e.g. type will be replaced by topic.

The columns of the table may be sorted by selecting the underlined column name in alphabetic or numeric order, e.g. ID, title, or type. In addition, a selection of the funding phase, when the dataset was created is possible, by choosing `Phase 1, Phase 2` or `All Phases`. By clicking on the `info` button, the metadata details of the dataset will be shown in a pop-up window. The metadata are arranged in a list according to the mulit-level metadata approach depending on the data type (section 3.3). First, the basic metadata details are displayed, followed by the specific metadata of each data type. For example, Figure 4-13 represents the metadata of a data type data. In this case, in addition to the basic and specific metadata, also additional metadata are available. By clicking on the `plus` symbol, additional details will be displayed. For example, details of the data creator, metadata creator, parameter and instrument. When there is an additional description file of the dataset available, a PDF download-button will be shown. All visitors are able to download and view the PDF by clicking on the button. Only authorized users have the permission to download the data file. This depends on the dataset download permission, set by the metadata creator of the dataset. If the download is not possible, the advice 'download not possible' will be displayed. Otherwise, authorized users have to agree to the 'CRC/TR32 Data Policy Agreement', for example concerning the use and adequate citation of the dataset, before they are able to download the dataset.

Figure 4-13: Two-step presentation of metadata according to the topic `atmosphere`. The general list in the background provides basic information of the dataset according to the search or predefined selection. A click on the info-button reveals detailed information of the chosen dataset, ordered by the multi-level metadata approach from general to specific attributes. The plus symbol shows further information. Additional information in a PDF-File can be viewed and the data can be downloaded according to rights and permissions.

The presentation of datasets according to the category `clusters (projects)` forms an exception. In addition to the overview list of the assigned project data, it is possible to `show` or `hide` further details of the corresponding project section, e.g. project title, principal investigator and instruction. This content is provided directly by the official CRC/TR32 Joomla CMS (www.tr32.uni-koeln.de). Thus, double data maintenance of the project section information is avoided (Figure 4-14).

Figure 4-14: Representation of data according to project section `A2`

In addition, the presentation of datasets according to the category DOIs forms an exception. Each dataset with a DOI has a permanent, separate, unique webpage (e.g. http://www.tr32db.uni-koeln.de/DOI/doi.php?doiID=3) that displays some associated metadata describing the identified dataset (Figure 4-15).

All website visitors are able to view the metadata details. The overview webpage acts as a landing page of the DOI. Therefore, selected metadata details of the identified dataset like citation, DOI name, related DOI(s), title, abstract, resource type, contact, and publication date are displayed. Clicking on related DOI names refers to the landing pages of the indicated DOIs. Further details of the contact person will be presented by selecting the contact name. Moreover, metadata details from the TR32DB are accessible via the `info`-button. The download permission of the dataset is defined by the contact person and applicant of the corresponding DOI. If a download is possible, then the authorized user has to agree to the *CRC/TR32 Data Policy Agreement*

(CRC/TR32, 2011), as well as to accomplish adequate reference and citation of the dataset, when it is used in a publication or presentation. Otherwise, if the download is restricted to participants of the CRC/TR32, there will be an appropriate notification to log in to the TR32DB or to contact the TR32DB administrators.
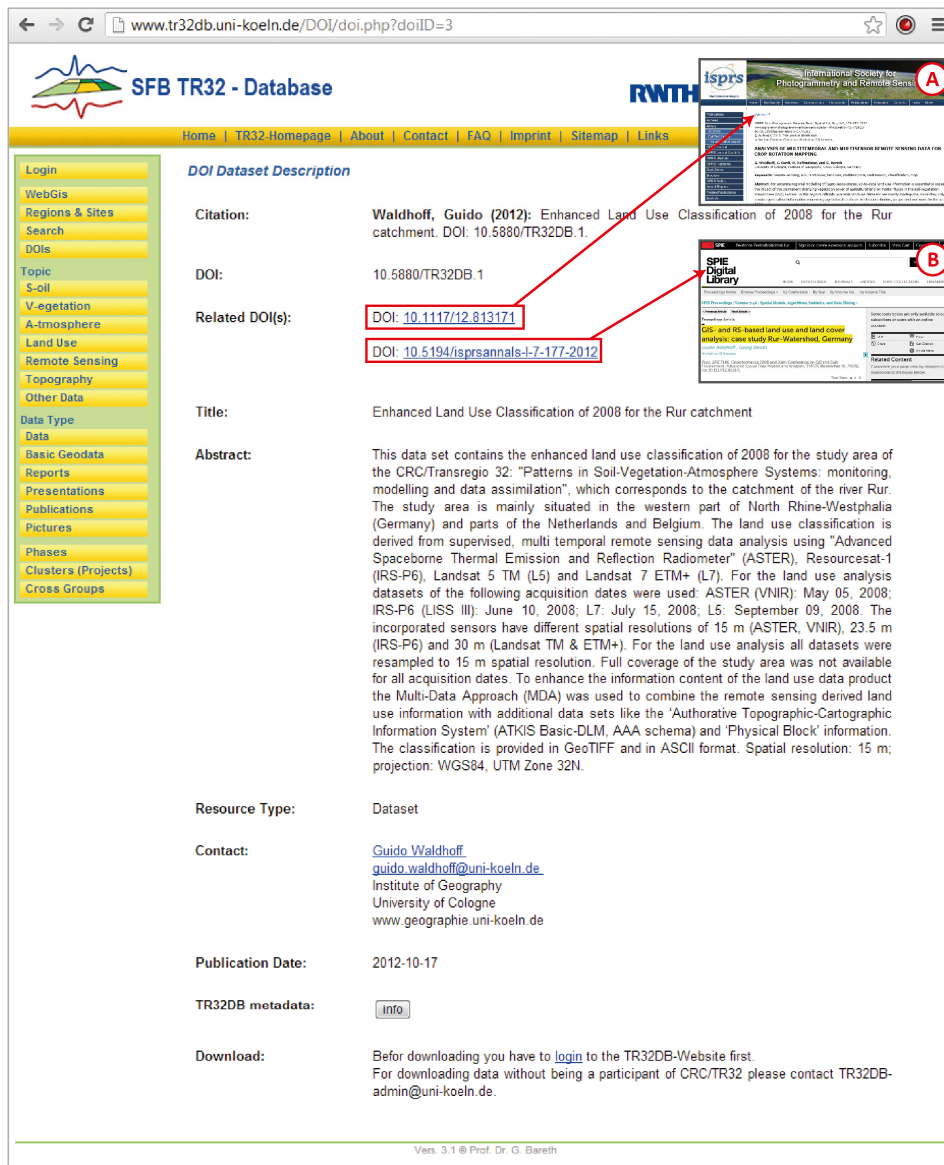
Figure 4-15: Data Description of DOI, (A) and (B) link to related identifiers webpages

4.5.1.3 Internal web-interface features

The internal features of the web-interface are only accessible for registered users. Part of these features is, for example, the metadata administration, the share area, climate data tools or the WebGIS. They will be described in detail now.

Web-Interface Login and User Home

The web-interface login enables all registered users of the TR32DB to access the internal web-interface features. By selecting login from the data menu bar, a log-in window will appear (Figure 4-16).
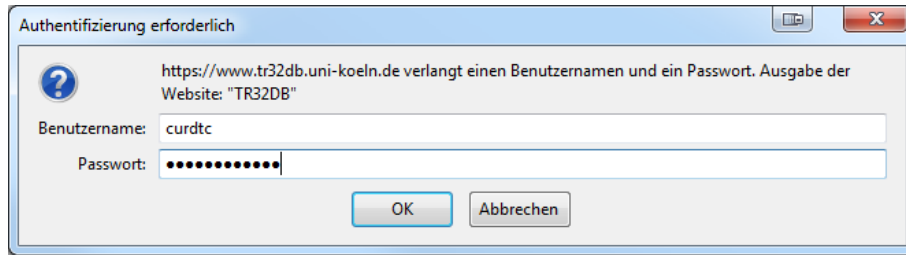
Figure 4-16: Log-in window of TR32DB web-interface

The user has to enter and submit his user name and password. As already stated, the login to the TR32D operates by user accounts of the RRZK (section 3.4). A successful log-in will expand the data menu bar with the internal features (e.g. User Home) and enables the user to access them. In addition, the `login` menu item on the data menu bar will change into a `logout` item, as well as the name of the currently successful logged in user will be added.

By selecting the menu item `User Home` from the data menu bar, several features to administrate the metadata of a dataset are provided (Figure 4-17).
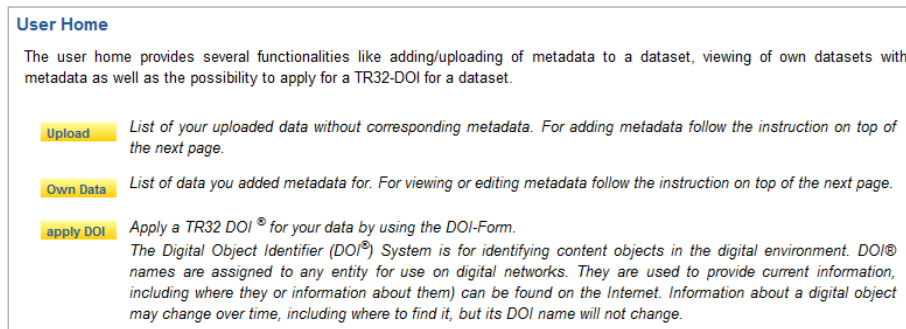
Figure 4-17: TR32DB web-interface `User Home`

These are `Upload, Own Data` and `apply DOI`. The favoured feature is accessible by selecting the appropriate yellow button. The `Upload` feature enables a user to add metadata to an already uploaded dataset of the data storage (see 4.3.). By selecting the `Own Data` feature, a registered user will receives a list of all datasets he has already added metadata to. He is able to view metadata details of a dataset, as well as edit its metadata. By choosing the `apply DOI` feature, a registered user will also receive a list of all datasets he has already added metadata to. Again, he is able to view the metadata details. In addition, he is able to apply for a DOI for a specific dataset.

Metadata administration for datasets

The administration of metadata for each dataset is an important part of the TR32DB. Therefore, one of the main relevant, internal web-interface features is the submission and editing of corresponding metadata to a dataset. To input or edit metadata to a dataset, the registered user has to follow six steps (Figure 4-18).

Figure 4-18: Web-Interface Metadata input or edit workflow

After a successful `login` to the web-interface, the registered user has to select the item `User Home` from the data menu bar and choose the `Upload` feature. As a result, the user will receive an overview table of all datasets without corresponding metadata that are stored in the TR32DB data storage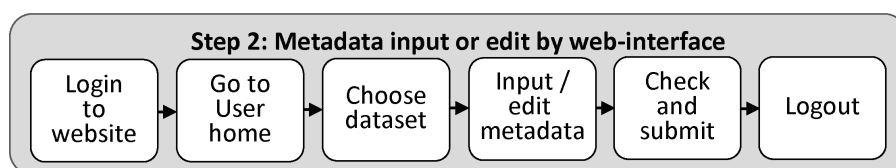, assigned to his project section and the appropriate funding phase (section 4.4). The overview list is sorted according to the funding phase and project section. Besides the internal identifier number, also file name, data type, project section and upload time to the data storage is shown. Before the dataset is accessible via the web-interface, the user is requested to deliver metadata to a specific dataset.

Overview of data without corresponding metadata

This table contains all datasets of **your sub-project** without corresponding metadata that are stored in the TR32DB. You have the possibility to add metadata (edit button) to a specific dataset. Please select a dataset to add corresponding metadata.

Phase 1

Subproject: Z1

ID	name	type	sp	time	
236	Radar2012.zip	data	Z1	2012-08-01 13:47:19	Edit
238	RadarData2011.zip	data	Z1	2012-08-01 13:47:19	Edit
208	MeasurementSelhausenzip	picture	Z1	2011-06-21 22:57:57	Edit
209	Presentation_TR32DBzip	presentation	Z1	2011-06-21 22:57:57	Edit
224	GeographieBook.pdf	publication	Z1	2009-03-24 13:47:19	Edit
207	SVA_Reportzip	report	Z1	2011-06-21 22:57:57	Edit

Figure 4-19: Overview of datasets without corresponding metadata, which can be assigned with metadata by clicking on the corresponding edit button

By selecting `Edit` for a specific dataset, the user will be forwarded to a metadata input form. This is arranged in a simply structured form that will be adjusted according to the chosen data type. The form is arranged according to the TR32DB multi-level Metadata Schema (section 4.3). Therefore, the form offers input fields for `general properties`, and `specific properties` depending on the data type. Apart from mandatory input fields, marked with an asterisk, also optional input fields are provided.

To enter all data into the requested fields, the user has two options to complete the metadata input form. These are to enter all data manually to the input form or to use the feature `Inserted Metadata` shown in Figure 4-20.

Metadata for Data

chosen data: sodar_sel_201006.zip

required fields

Inserted Metadata

| (Choose...) | ▼ |

(Choose...)
[73] ALS data - PPP-method... (ALS_PPP)
[74] ALS data - GNSS-method... (ALS_GNSS)

Metadata inserted by your subproject. You can choose this if the metadata you want to insert, is similar or equal to older metadata used in your subproject.

Figure 4-20: Metadata input by using the feature `Inserted Metadata`

This feature enables the use of already inserted metadata of another data file as a template. The selection menu offers a list of datasets of the current logged in user, project section, and the current chosen data type that already owns metadata. For example, if a user is a member of the project section B1 and wants to submit metadata to a dataset with data type `data`, then the drop down list will just contain datasets of project section B1 from the data type `data`. By choosing a specific dataset form the drop down list, represented by dataset number and title, the input form will be filled automatically with the requested metadata. It is now possible to add or change corresponding metadata for the current dataset.

After the successful input of all metadata to the input forms, the user has again two options to continue. By choosing the `reset` button, all data in the input form will be deleted. By selecting the `submit` button, the metadata in the input forms will be checked. In case, a false value is submitted or a required field is missing, an error will occur. The error message will specify the value that has to be changed. After successful

examination of all input fields, the metadata will be posted to the database and the dataset with the according metadata will be accessible online as described in the previous section.

The following sections will describe various, available metadata input forms in detail. All metadata input forms, for example Figure 4-21, are subdivided into three columns. The first column names the input field title, the second column provides the input field, and the third column offers a complementary descriptive explanation and help. To avoid faulty entries and to enable a user-friendly, simplified metadata input, four types of input fields are provided. These are:

Text box The text box offers a single-line input field that a user can enter text.

Text area The text area or multiline text box enables a user to add multiple lines of text with automatic word wrapping. The multi-line text box offers a scroll function that enables to handle a large amount of text more easily.

Select menu The select menu creates a drop-down list that enables a user to select a value from a list.

Calendar The calendar input field offers the input of a date by using a pop-up calendar that will appear by clicking in the corresponding input field. By choosing the appropriate date, this will be inserted into the input field in a pre-defined order.

The metadata input form for *general properties* enables the input of general metadata attributes (Figure 4-21 and Figure B-1) according to the TR32DB Metadata Schema (section 4.3). This input form is the same for every data type with the exception of the data type geodata (Figure 4-21). The dataset can be described with various details by entering information in a text box or text area, as well as by using the calendar field or choosing a value form a select menu. In the following, the general metadata properties that can be added to a dataset are described. The statement in the square brackets [ ] describes the input field type. The asterisk in brackets (*) marks mandatory input fields. Corresponding example values are represented with single quotation marks ' '. In case of combined input fields the example values are merged with a vertical bar |.

Figure 4-21: Exemplary metadata input form `General Properties` (partial view). The complete general properties for all types of data are shown in Figure B-1

Title [text box](*) Main title of the dataset has to be inserted.

Description [text area](*) A corresponding abstract, describing the dataset, has to be added.

Creator-person name [select menu](*) Details of the creator-person (first name, family name, and e-mail-address) can be chosen from the drop down list. If the creator-person is not available in the drop-down list, the value `add new creator` has to be selected. Through this, a new input field will appear to enter the details of the new creator-person (name, family name, e-mail). By using the `insert` button and agreeing to the corresponding security request, the values will be submitted to the TR32DB database, pre-selected in the drop down list, and also available in the next drop down list. By selecting the `add creator` button, another creator-person will be added. Respectively, a creator can be removed by using the appropriate `delete creator` button.

Creator-person role [select menu](*) An appropriate creator-person role has to be selected from the drop-down menu for each creator-person. For example, values like owner, distributor, or point of contact can be chosen to clarify the creator role.

Creator-institution [select menu](*) Details of the appropriate institution have to be added to each chosen creator from the drop-down list. If the creator-institution is not available in the drop-down list, the value `add new institution` has to be selected. Through this, a new input field will appear to enter the details of the new creator-institution (institute name, university name, website). By using the `insert` button and agreeing to the corresponding security request, the values will be submitted to the TR32DB database, pre-selected in the drop down list, and will also be available in the next drop down list.

Contributor [text box] Details of the dataset contributor can be added. The dataset contributor is a person and/or institution responsible for making contributions to a dataset.

Publisher [text box] Details about the organisation responsible for making the dataset available can be inserted. By default, the denoted term for publisher is 'University of Cologne, Institute of Geography, TR32DB'. This can be changed by the metadata editor. In particular, this has to be changed for the data type publication to the publishing house in charge.

Citation advice [text area] An advice to cite the dataset can be inserted into the citation advice text area field.

Identifier [text box][select menu] An corresponding identifier (if already available) for the dataset can be added by using the `add identifier` button. Here, the identifier name has to be entered to the text box and additionally the identifier type has to be selected from the drop-down list, e.g. '10.5880/TR32DB.3'|'DOI'. By selecting the `add identifier` button, it is possible to insert another identifier to the dataset. Respectively, the identifier can be removed by using the appropriate `delete identifier` button.

Relation [select menu][select menu] A reference to a relation dataset in the TR32DB can be added by using the `add relation` button. Two drop-down lists will appear. The first list contains a list of datasets, represented by the ID and title of the dataset, appropriate to the project section of the registered user. The second select menu includes a list of the relation types. To enter a relation, the related dataset has to be chosen from the first drop-down list and the appropriate relation type has to be chosen, e.g. '[237] radar data'|'is version of'. By selecting the `add relation` button again, it is possible to insert another related dataset. Respectively, the relation can be removed by using the appropriate `delete last relation` button.

Creation date [calendar](*) The creation date of the dataset has to be chosen from the calendar, appearing by clicking into the calendar field.

Publication date [calendar](*) The publication date of the dataset has to be chosen from the calendar, which appears by clicking into the calendar field. By default the value of the calendar is set to the current date.

File size [text box][select menu] The amount of data included in the dataset has to be specified. For this purpose, a data unit has to be chosen from the drop-down list and supplemented with the corresponding value, for example '5'|'pages' or '10'|'datasets'.

Dataset type [select menu](*) The type of data has to be selected from the drop-down list. The recommended and default setting for the data type data is the value 'dataset'. The value 'dataset collection' is recommended for a dataset series. Moreover, the value 'text' can be used for the data type publications and reports, as well as the value 'image' for the data type pictures.

Format [select menu][text box](*) To add the format and version of the dataset, a format has to be selected from the drop-down list and can be complemented with the format version, for example 'MS Excel'|'2007', or 'ASCII'|' '. If the format of the dataset is not listed in select menu, value `add new format` has to be chosen. Though this, a new input field will appear to enter the new format to the list and an optional format version.

MIME type [select menu][select menu] The MIME (Multipurpose Internet Mail Extensions) Media Type is composed of two parts: the content type and the subtype. To add the format of the dataset corresponding MIME type, first, a content type has to be selected from the first drop-down list. The second select menu will automatically adjust according to the appropriate subtypes. Again, a value has to be selected from this list. As a result, the following values can be selected: 'application'|'zip','image'|'jpg', or 'text'|'xml'.

Dataset language [select menu](*) The primary language of the dataset has to be added. A language can be chosen from the drop-down list. The preselected value is English.

TR32-topic [select menu](*) A CRC/TR32 specific topic has to be chosen from the drop-down list, e.g. 'atmosphere', 'soil', 'land use', or 'remote sensing'.

TR32 keywords [select menu](*) At least one keyword has to be selected from the CRC/TR32 specific keyword drop-down list to describe the dataset. A maximum of six keywords can be chosen. By selecting the keywords from the drop-down list, they will be added to the closest text area.

GEMET thesaurus [select menu][text box](*) To describe the dataset with keywords according to the GEMET thesaurus, a value has to be selected from the drop-down list. To receive a selective list, a term has to be entered into the input field. As a result, a drop-down list will be displayed containing all keywords of the GEMET Thesaurus that contain the entered term. After choosing a keyword from the selective list, this will be added to the text box field by using the `add keyword` button. Again, at least one keyword has to be selected and a maximum of six keywords is possible.

Status [select menu](*) The processing status of the data has to be specified by choosing a value from the drop-down list, for example 'completed'.

Permission [select menu](*) The download permission has to be added to the dataset. For this purpose, a value has to be selected from the drop-down list, for example 'onlyTR32'.

Subsequent to the general properties, data type `specific properties` have to be added. The metadata input form for specific properties enables the input of specific metadata attributes depending on the data type according to the TR32DB Metadata Schema (section 4.3). Information can be entered by using again input fields like text boxes, text areas, calendar fields, or select menus. In the following, the specific metadata properties that can be added to a dataset are described depending on the various data types.

The `specific properties` for the data type `data` are (Figure 4-22):

Measurement/model region [select menu](*) The measurement region or modelling domain of the dataset has to be added by selecting a value from the select menu, e.g. Germany, NRW, rur catchment, Ellebach or Kall.

Measurement/model location [select menu](*) The location, where the dataset was measured or the extent of the modelling has to be specified. For example, values like measuring location like Merken, Merzenhausen or Kall can be chosen from the drop-down list.

Temporal extent (begin/end) [calendar](*) The time period covered by the content of the dataset has to be added. By clicking on the calendar-button, a single calendar will open, where the appropriate start or end date can be chosen. To enter a date including time to the dataset, the clock-button has to be clicked on. The appropriate date and time can be chosen and will be added.

Instrument details [select menu][text box] To insert an associated measuring instrument or modelling method to a dataset, the `add instrument` button has to be chosen. A drop-down list will appear, offering a list of measurement instruments and modelling methods that can be chosen, e.g. 'Air Temperature/Humidity Sensor (HygroClip)'. If the corresponding instrument is not available in the list, the value `add new instrument` has to be selected. Through this, a new input field will appear to enter the details of the new instrument and describing information about it, e.g. equipment group or method name, model name, manufacturer, registered office and its website. By clicking on the `insert`-button, the new instrument will be added to the list and will be automatically preselected after the reload of the webpage. By selecting the `add instrument` button again, it is possible to insert another instrument. Respectively, the instrument can be removed by using the appropriate `delete instrument` button.

Instrument resolution distance [text box][select menu] A resolution distance can be added to the instrument by entering a distance to the text box and selecting a unit of length from the drop-down list, e.g. '2'|'cm'.

Instrument temporal frequency [text box][select menu] The temporal frequency of the measurement or modelling can be declared by entering the value for the frequency to the text box and closing the appropriate unit from the select menu, e.g. '5'|'Hz'.

Instrument parameter [select menu][select menu] To add the measured or modelled output parameter, the `add parameter` button has to be chosen. Through this, two select menus will be available. First, the parameter has to be selected from the first drop-down list, e.g. 'Cloud Cover', and then the corresponding parameter unit can be chosen from the second drop-down list, e.g. '%'. If the requested parameter or unit is not available, the value `add new parameter` or `add new unit` has to be chosen from the select menu. An input field will appear to enter the missing parameter and/or unit. By clicking on the `insert`-button, the new parameter and/or unit will be added to the list and will be automatically preselected after the reload of the webpage. By selecting the `add parameter` button again, it is possible to add another parameter. Respectively, the parameter can be removed by using the appropriate `delete parameter` button.

Additional description By choosing the `search` button, an additional description can be added to describe the dataset in more detail. This file has to be in PDF-format with a maximum file size of 5MB.

Additional Metadata according to INSPIRE directive can be added for the data type `data,` as also partially shown in Figure 4-22. These are:

INSPIRE topic category [select menu] A topic category value can be chosen from the drop-down list according to the INSPIRE directive or ISO 19115, e.g. 'climatologyMeterologyAthmosphere'.

Geographic bounding box [text box][map] The extent of the dataset in geographic space can be added to the dataset. The coordinate values in decimal degree for the north bound latitude, west bound longitude, east bound longitude, and south bound latitude can be entered into the text

box. Moreover, a map window is provided that enable the automatic input of the coordinates of the chosen geographic extent. The map extent can be changed by moving and zooming.

Lineage [text area] A description of the process history and/or overall quality and validity of the dataset can be entered to the lineage text area.

Figure 4-22: Exemplary metadata input form for `Specific Properties` for data type `data` (partial view), the complete input form is presented in Figure B-2

The `specific properties` for the data type `report` are:

Report date [calendar](*) The date of the report has to be added. By clicking on the calendar-button, a single calendar will open, where the appropriate start or end date can be chosen. To enter a date including time to the dataset, the clock-button has to be clicked on. The appropriate date and time can be chosen and will be added. By default the value of the calendar is set to the current date.

Report type [select menu] The type of report can be added by selecting a value from the drop-down list. For example, values like 'book review', 'conference review', 'minutes' or 'PHD report' can be chosen.

The `specific properties` for the data type `presentation` are:

Presenter [text box] A presenter of presentation can be entered to the presenter text box.

Presentation date [calendar](*) The date of the presentation has to be added. By clicking on the calendar-button, a single calendar will open, where the appropriate start or end date can be chosen. To enter a date including time to the dataset, the clock-button has to be clicked on. The appropriate date and time can be chosen and will be added. By default the value of the calendar is set to the current date.

Presentation type [select menu] The type of presentation can be chosen from the select menu. For example, values like 'talk', 'poster', 'keynote', or 'other' can be chosen.

Presentation event [select menu] The event type can be selected from the drop-down list. For example, values like 'conference', 'workshop', 'project meeting', 'meeting', 'cross group meeting', or 'other' can be chosen.

Event title [text box] The title of the event can be entered to the event title text box, e.g. 'European Geosciences Union General Assembly 2010'.

Location [text box] The location, where the event took place, has to be entered into the texbox, e.g. 'Vienna, Austria'.

Event period (begin/end) [calendar][calendar] The event period, when the event took place has to be entered. By clicking into the input field, a single calendar will open, where the appropriate start or end date can be chosen.

The `specific properties` for the data type `picture` are:

Record date (begin/end) [calendar](*) The record date of the picture, including beginning and ending, has to be added. By clicking on the calendar-button, a single calendar will open, where the appropriate beginning and/or end date can be chosen. To enter a date including time to the dataset, the clock-button has to be clicked on. The appropriate date and time can be chosen and will be added. By default the value of the calendar is set to the current date

Location [text box](*) The location, where the event took place, has to be entered into the texbox, e.g. 'Selhausen, Germany'.

Region [select menu] The region, where the event took place, can be entered to the texbox, e.g. 'Ellebach, Germany'.

Event [text box] The event, where the picture was taken, can be entered into the text box, e.g. 'Laser Scanning Measurements May 2012'

Orientation [select menu] The orientation of the picture can be added by selecting a value from the drop-down list. For example, the values 'vertical/portrait' or 'horizontal/landscape' can be chosen.

Resolution [text box] The resolution of the picture can be entered in dots per inch (dpi) to the text box, e.g. '200'.

Size [text box][text box] The size of the picture (horizontal x vertical) can be entered, e.g. '1680' x '1050'.

Camera [text box] The model or the camera can be entered into the text box, e.g. 'Olympus E-620'.

Copyright information [text area] Information about the copyright of the picture can be entered into the text box, e.g. 'Copyright 2012 CRC/TR32. All rights reserved.'.

Additional description By choosing the `search` button, an additional description can be added to describe the dataset in more detail. This file has to be in PDF-format with a maximum file size of 5MB.

The metadata input form for the `specific properties` for the data type `publication` is an exception. This input form is divided into two parts. The first part offers input fields to describe general publication properties. The second part of the input form facilitates specific input fields with focus on the publication type. The input form will be adapted depending on the chosen publication type. Figure 4-23 provides an overview of the general specific publication properties and the specific properties of the publication type `book section`.

The `specific properties` for the data type `publication` will be subsequently described in detail. The general specific properties for the data type `publication` are:

Year [text box](*) The year, when the publication was published, has to be entered into the textbox, e.g. '2008'.

Status [select menu](*) The status of the publication has to be added by selecting a value from the drop-down list. For example, values like 'in print', 'in review', 'published', 'submitted' or 'unpublished' can be chosen.

Review [select menu](*) The information about the review status of the publication has to be selected from the select menu, e.g. 'yes' or 'no'.

Type [select menu](*) The type of publication has to be chosen from the drop-down list. For example, values like 'article', 'book section' or 'event paper' are available. The subsequent input form will be adjusted according to the selection made.

Figure 4-23: Exemplary metadata input form for `Specific Properties` of publication type `book section`

The `specific properties` of the publication type `article` are:

Article type [select menu](*) The type of article has to be added by selecting a value from the select menu. For example, values like 'journal', 'magazine', 'newspaper' or 'electronic' can be chosen.

Publication source [text box](*) The name of the publication source has to be entered to the text box, e.g. 'International Journal of Digital Curation'.

Publication source URL [text box] The corresponding website of the publication source can be entered into the text box, e.g. 'http://www.ijdc.net'.

Publication place [text box] The appropriate place of publication can be added to the text box, e.g. 'Bath, UK'.

Volume [text box] The volume of the publication source can be entered into the text box, e.g. '7'.

Issue [text box] The issue of the publication source can be added to the text box, e.g. '2'.

Pages [text box] The number of pages of the publication can be added to the text box, e.g. '13'.

Page range [text box] The range of page of the publication can be inserted into the text box, e.g. '68-80'.

The `specific properties` of the publication type `book section` are:

Book title [text box] The title of the book can be entered into the text box, e.g. 'Proceedings of the Data Management Workshop'.

Place published [text box] The place of publication can be added, e.g. 'Cologne, Germany'.

Book editor [text box] The editor of the book can be inserted into the text box, e.g. 'Curdt, C., Bareth, G.'.

Chapter [text box] The book chapter of the publication can be added, e.g. '5'.

Pages [text box] The number of pages of the publication can be entered into the text box, e.g. '6'.

Page Range [text box] The range of pages of the publication can be inserted, e.g. '131-136'.

Series Title [text box] The title of the book series can be added, e.g. 'Kölner Geographische Arbeiten'.

Series volume [text box] The volume of the book series can be entered, e.g. '90'.

The `specific properties` of the publication type `event paper` are:

Event Type [select menu](*) The type of event has to be added by selecting a value from the select menu. For example, values like 'conference', 'workshop' or 'other' have to be chosen.

Event Name [text box](*) The name of the event, where the paper was presented, hast be entered to the text box, e.g. 'XXI. ISPRS Congress'.

Event Place [text box](*) The place, where the event took place, has to be inserted, e.g. 'Beijing, China'.

URL of Event [text box] The corresponding website of the event can be added, e.g. 'http://www.isprs.org/congresses/beijing2008/'.

Date of Event (begin/end)[calendar][calendar] The time period covered by the event can be added. By clicking on the calendar-button, a single calendar will open, where the appropriate start or end date can be chosen. To enter a date including time to the dataset, the clock-button has to be clicked on. The appropriate date and time can be chosen and will be added.

Editor [text box] The editor of the event proceedings can be entered, e.g. 'Chen, J.'.

Proceedings Title [text box] The title of the event proceedings can be added, e.g. 'International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science'.

Volume [text box] The volume of the event proceedings can be added to the text box, e.g. 'XXXVII'.

Issue [text box] The issue of the proceedings can be inserted into the text box, e.g. 'Commission V, Part B4'.

Pages [text box] The number of pages of the publication in the proceedings can be added to the text box, e.g. '6'.

Page Range [text box] The range of page of the publication can be inserted into the text box, e.g. '131-136'.

After a successful input of metadata for a dataset, the metadata creator has the option to edit these. For this purpose, the user has to navigate to the `User Home` and select the feature `Own Data`. As a result, the user will receive an overview table. This table lists all datasets with metadata that have been inserted by the user (Figure 4-24).

Figure 4-24: Overview of datasets with corresponding metadata of current user (partial view)

Figure 4-25: Modification of metadata from the data type `publication` (partial view)

The listing can be sorted according to the dataset identifier number, the data title, the chosen topic, the data type, and date of metadata entry. By choosing the `info` button of a specific dataset, the metadata details of this dataset are presented in a pop-up-window, as already described in a previous section 4.5.1.2. To modify the metadata of a specific dataset, the `edit` button has to be chosen. The user will be forwarded to the metadata input form. In contrast to the metadata entry as described previously, all input fields and choice boxes are already completed with the available metadata of the chosen dataset (Figure 4-25). The user is now able to change the metadata and resubmit them again. The user is able to repeat the procedure as much as he wishes. The modification of the metadata is immediately visible on the website. All changes of the metadata are sustained as a precaution in the database.

DOI Application

TR32DB users are able to apply a TR32DB Digital Object Identifier (DOI) (section 2.1.6) for their dataset, when they have successfully completed the upload and adding of metadata for a dataset (step one and two of the workflow as shown in Figure 4-4). The application of a DOI is only permitted for datasets, which do not have a DOI yet. To apply for a DOI for a dataset, the registered user has to follow six general steps (Figure 4-26).
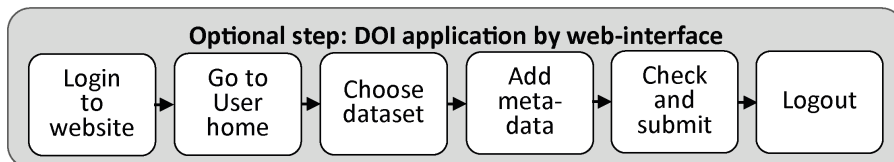
Figure 4-26: DOI application workflow (optional step)

At first, the user has to log in to the web-interface. He has to navigate to the `User Home` and select the feature `Apply DOI`. The appropriate button is not available, in case a DOI has already been applied. By choosing the `Apply DOI` button, the user will be forwarded to an overview table. This table lists all datasets of the user, which are associated with metadata inserted by the user. Again, this list can be sorted, for example according to the identifier number or dataset title. Details of this dataset are likewise presented in a pop-up-window by selecting the `info` button, as already described in detail. To apply a DOI for a specific dataset, the `apply DOI` button has to be chosen (Figure 4-27).

Figure 4-27: Overview of datasets with metadata of current user to apply a DOI (partial view)

The user will be forwarded to a DOI application overview website. Some guidelines and conditions have to be considered in the context of the DOI application process. The user has to accept particular conditions to continue the application, presented in Figure 4-28. For example, the DOI applicant is pledged to control the quality of the dataset. In addition, the TR32DB system is in authority to provide the infrastructure for the sustainable storage of the data and guarantee the long-term accessibility of the DOI. Both, DOI applicant and TR32DB system are responsible for the quality control of the descriptive metadata.

The metadata input form for the DOI application is shown exemplary as an excerpt in Figure 4-28. The entire DOI application input form is presented in Figure B-4. Existing metadata of the specific dataset are

already completed in the input form. The DOI metadata input form is a reduced entry form, in comparison with the previous described general metadata input. The input form is arranged according to the requirements of the STD-DOI Metadata Schema (KLUMP, 2009). For example, general metadata details are included like title, description, creator or contributor. All mandatory input fields are marked with an asterisk. Incorrect values in an input fields cause error notifications and have to be adjusted. After the user has successfully completed the metadata input form, in a next step, the data have to be posted to the TR32DB database by using the `submit` button. All following steps to receive a DOI are executed by the TR32DB administrator and will be described in the section 4.5.1.4.

Figure 4-28: Exemplary DOI application form (partial view), the complete form is shown in Figure B-4

Temporary Share

By selecting the menu item `Share` from the data menu bar the user will receive the `Temporary Share` website (Figure 4-29). On this website, the user is able to access all data that is currently stored in the corresponding temporary TR32DB data storage `share` directory (section 3.34). Initially, general information about the temporary share is described. Subsequently, the procedure to store and exchange data is listed. In the bottom part of the website a listing of all current research projects of the CRC/TR32 is listed. By clicking on one of the project sections names, the content of the particular `share` folders will be listed. The listing includes the title of folders created in the respective sub-project directory and the appropriate files included in this folder. They files are represented by their name and file size. To download a file, the `download` button has to be selected. For example, Figure 4-29 shows all data that are currently stored in the temporary share folder of project section B1. In this folder, a subfolder `Graf` was created, in which some files were stored. By selecting the `download` button below the file `ruebe_bellartz_5mStation_CR3000_Slow.txt`, the data will be downloaded.
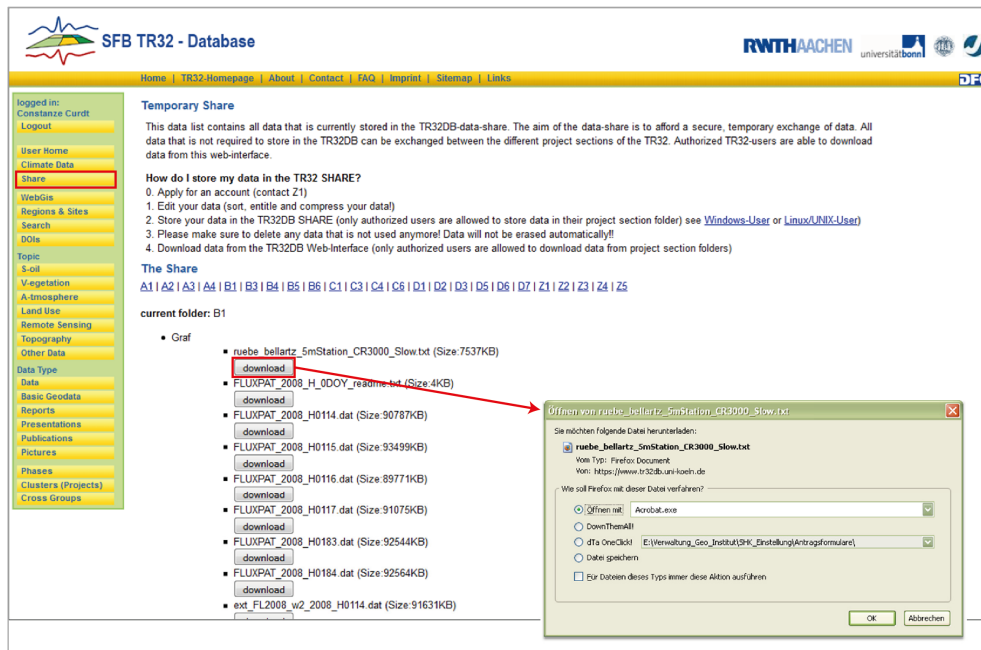
Figure 4-29: Web-interface presenting temporary share

Climate Data Section

The climate data section will be displayed by selecting the menu item `Climate Data` from the data menu bar. This website gives an overview of all available climate data, which was purchased in the framework of the CRC/TR32 (section 3.1.2). Climate data was purchased from the German Weather Service (DWD), the network of the Meteomedia Group, the Eifel-Rur Water Resources Association, and the Erft Water Resources Association. Initially, the overview website (Figure B-6) provides general information, like the data agreement with the different organisations and publication notices for download. Furthermore, detailed descriptions of the climate data is given for download, for example data format description or code tables. The section `data sets` lists all data organised by various institutions and provides download links for the single data sets. All data from the DWD will be provided for download on a sub-webpage via links. In addition, an online query of climate data purchased from the DWD is available via the Climate Data overview website. Two features are available: `query climate station data` and `draw climate diagrams`. Both will be described consecutively.

The feature `query climate station data` enables an easy and quick search for climate data purchased from the DWD according to a climate station and a chosen temporal extent (Figure 4-30). To execute a search for data, at least one or multiple climate data stations have to be selected from the selective list (e.g. Aachen, Aachen-Merzbrück). In addition, a period of time has to be chosen (e.g. 01.01.2009-01.01.2009). Both values have to be chosen carefully to receive a reasonable result. Finally, the `submit` button has to be selected to execute the search.

Figure 4-30: Search for DWD data by using the feature `query climate station data`

The query will be posted to the climate data database, described in detail in section 3.5.1. The result of the query will be presented in a table, as shown in Figure 4-31. In addition, the result can be exported as text file by using the `Export Data` link. By selecting the link `Abbreviations`, a PDF-file with abbreviations to the climate data will open.

Figure 4-31: Search result of feature `query climate station data`

The feature `draw climate diagram` enables the query of DWD data and the representation of the result in a chart (Figure 4-32). Initially, at least one or multiple climate data stations have to be selected from the selective list (e.g. Aachen, Aachen-Merzbrück). In addition, the climate parameter has to be chosen from the drop down list (e.g. temperature) and the period of time (e.g. 01.01.2009-01.02.2009). The query will be executed by selecting the `submit` button.

Figure 4-32: Search for DWD data by using the feature `query climate station data`

The query will be posted to the climate data database, which is described in detail in section 3.5. The result of the query is represented in a curve chart, as shown in Figure 4-32. The chosen parameter will be presented on the vertical axis including the appropriate unit (e.g. temperature in °C). The horizontal axis represents the chosen time period (e.g. 01.01.2009-01.02.2009). The selected climate data stations (e.g. Aachen, Alzey, Arnsberg-Neheim, Ahaus) are represented in the legend and allocated with various colours. By selecting single values on the curve of a climate station, the corresponding accurate value will be displayed.
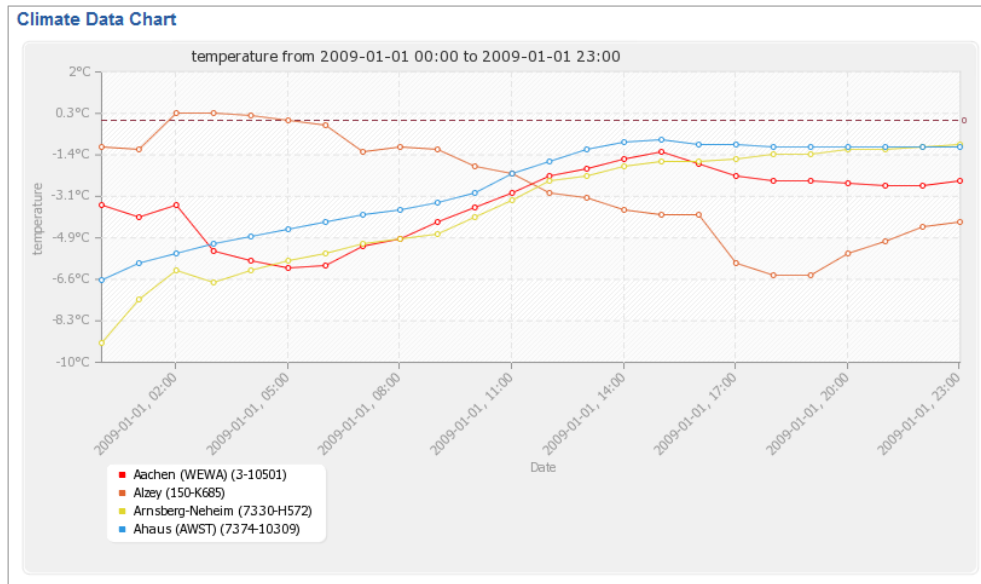
Figure 4-33: Chart result of using the feature `query climate station data`

WebGIS Version 2

By selecting the menu item `WebGIS` from the data menu bar, the user will automatically be forwarded to the TR32 Viewer, a WebGIS (Figure 4-34). The TR32 Viewer was implemented by using the ArcGIS Viewer for Flex, described in section 3.6. Flex provides a pleasant and fast visualization of data. The main task of the TR32 Viewer is the visualisation of some purchased basic geodata and a few provided geodata of project participants (e.g. land use data). Moreover, several measurement stations of the CRC/TR32 are visualised, as well as the DWD stations. The WebGIS is only accessible for authorized TR32DB users, due to access and license restrictions concerning the purchased geodata issued by the responsible organisations.

For example, Figure 4-34 represents the land use classification of the Rur catchment (Waldhoff, 2012; Waldhoff & Bareth, 2009) and the digital landscape model. In addition, the climate station Broichtal is visible with corresponding properties. Furthermore, the overview map, the DWD Stations listing, and the Table of Contents is visible.

In general, the TR32 Viewer has a main menu bar containing the four items `map` (globe symbol), `navigation` (compass symbol), `tools` (cube symbol), and `help` (question mark symbol). By moving over one particular symbol, the corresponding features (widgets) are displayed and can be selected. The chosen feature will be displayed (e.g. Overview Map, Table of Contents). In addition, the north arrow and general navigation tools are available (e.g. map zoom slider, scale bar, scroll and close sidebar). The main features of the TR32 Viewer are subsequently described, according to their affiliation to a feature group:

> **Map Zoom Slider** This tool is displayed on the left site of the map. It enables the zoom in and out by either clicking on the chart or by holding the triangle and drag it up and down. The current scale will be displayed by clicking on the triangle. Furthermore, zoom in and out is provided by using the mouse wheel.

Scale Bar This feature is displayed on the bottom left side of the map. The scale of the map is presented in kilometres/meters or respectively in miles/feet.

Scroll and Close Bar This menu bar is located on the right hand site of the map. It enables scrolling up and down the sidebar located at the right hand site using the up and down arrows. By using the two right-facing arrows, the sidebar will be closed.

Table of Contents (Map) The `Table of Contents` displays all available layers. The user is able to switch on the various layers (e.g. land use data, regional soil type map, digital landscape model, digital cadastral map, digital orthophotos, CRC/TR32 boundaries and sites). In addition, the user is able to change the transparency of each single layer on the fly.

Overview Map (Map) This widget adds an overview map to the map. The current spatial extent of the map is displayed as a red rectangle. By moving the red extent rectangle in the overview map, the user can navigate on the map.
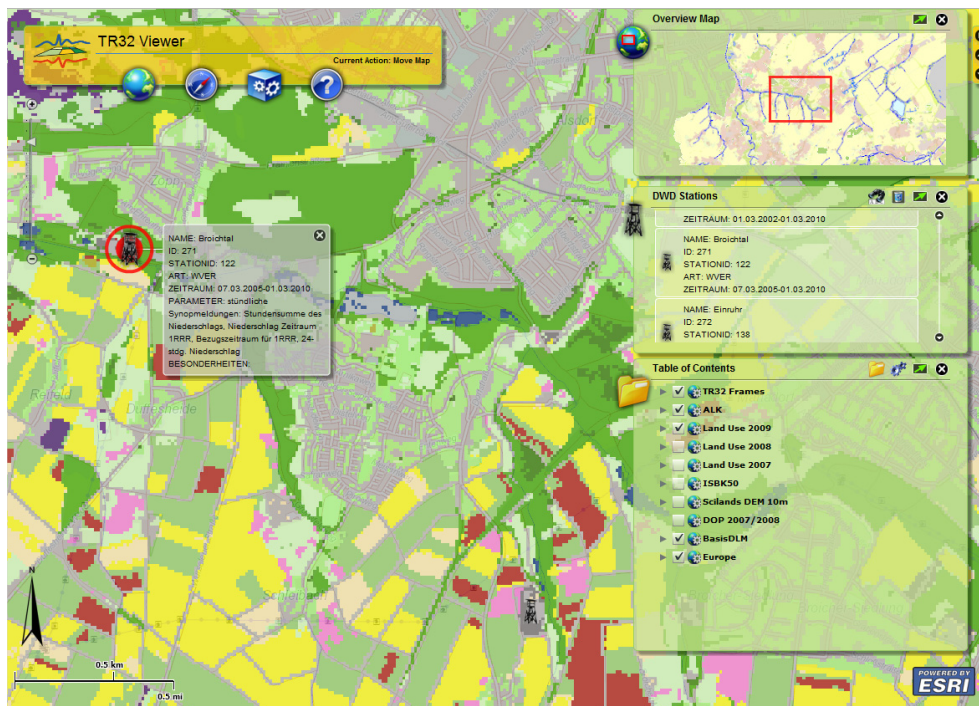
Figure 4-34: The TR32 Viewer Version 2

Bookmarks (Map) The `Bookmarks` widget displays a collection of predefined map view extents (e.g. from measurement site of the CRC/TR32). The user is able to create and add new bookmarks by choosing the function `Add Bookmarks`.

Print (Map) To print the visible map display the `Print` widget has to be used. A title and subtitle for the document can be added.

Zoom In (Navigation) To zoom in at a defined map extent, this feature can be used. First, the `Zoom In` button has to be chosen and then the part of interest on the map has to be selected.

Zoom Out (Navigation) By selecting the Zoom Out feature, the user can zoom out to a defined map extent. First, the zoom our button has to be chosen and then a part on the map has to be selected.

Full Extent (Navigation) The map will be reset to the initial zoom state and default map area by selecting the `full extent` button.

Re-Center Map (Navigation) By choosing the `Re-Center Map` feature, the view window of the map can be adjusted by panning. Click and hold on the map and then move the mouse to the preferred map extent.

Mouse Coordinates (Navigation) By choosing this widget, the latitude and longitude of the current place of the pointer mouse on the map will be displayed. The coordinates will dynamically change by moving the pointer mouse on the map.

Locate (Tools) The `Locate` widget enables the user to search for a location according to an address or by longitude and latitude.

DWD Stations (Tools) To display the available DWD stations on the map, the `DWD Stations` tool can be used. A list of all DWD stations will be displayed in the widget window. By clicking on a DWD station on the list, the map display zooms to this location. By moving over the DWD station with the pointer mouse, details of the station will be displayed in a pop-up window (e.g. station name, measuring period, parameters, measurement gaps). To search for a specific station in the list, the user can use the `filter` function and enter the particular station name. As a result, the user will receive a list of matching DWD stations.

Draw (Tools) This widget enables the user to highlight or comment on important locations on the map. The user is able to draw shapes (e.g. point, line, freehand line, polygon, freehand polygon, text) in various colours and sizes.

Identify (Tools) To identify a feature on the map, the `Identify` feature can be used. By clicking on the map, all attributes of a layer will be displayed.

Help (Help) The `Help` feature will forward the user to a detailed online help, describing all features of the TR32 Viewer .

Home (Help) The user will be forwarded to the website of the TR32DB by selecting the `Home` function (CURDT, 2009).

About (Help) The `About` widget provides information about the TR32 Viewer (e.g. development version and team).

4.5.1.4 Administrative web-interface features

The TR32DB Admin-Tool Website provides several features to administrate the TR32DB system Version 3.1. This website is only accessible for the administrators of the TR32DB. The operations of the Admin-Tool are available via a side-menu. The main operations of the Admin-Tool will be consecutively described in detail, as well as other features.

User Management

The administration of the users of the TR32DB system is managed by the TR32DB Admin-Tool. By selecting `User Admin` from the side menu, several operations can be chosen: `insert new user`, `view user`, `insert new subproject`, `view subprojects`, and `select user and subprojects`. To set up a new user account for the TR32DB, the operation `insert new user` has to be selected. An input form will be displayed (Figure 4-35). Contact details about the user have to be entered like the user name, provided by the RRZK (section 3.4.2), first name, last name, email address or the respective institution. To add a new institution to the selective list, the function `add new institution` has to be selected from the institution selective list. An input from will appear to enter the details of the institution. In addition, optional details can be added. The affiliation to a sub-project has to be completed by choosing the project section and the funding phase from the selective list. Multiple entries are possible. Depending on the selected affiliation to a sub-project and funding phase, the user will be able to edit metadata for a dataset or download datasets from the TR32DB (section 4.5.1.3). By using the `submit` button, the data provided to the input fields will be entered to the database tables `user` and `userSP` (section 3.5.1).

In addition, all users of the TR32DB are listed in an overview table, by using the function `view user`. Contact details of a single user can be modified by selecting the `edit` operation from the overview list. The feature `insert new subproject` enables the input of a new sub-project to the database. The entire list of sub-projects is visible by selecting `view subprojects` and can be modified by choosing the `edit`

button. Finally, the `select user and subprojects` operation enables a combined search for user details. For example, a user can be selected according to the funding phase and/or the sub-project. In addition, the affiliation of a user to a funding phase or sub-project can be queried.

Figure 4-35: User Administration using the TR32DB Admin-Tool

News Management

The news section, displayed on the start page or the TR32DB web-interface is managemed by the TR32DB Admin-Tool. The content of the news is accessed from the associated database table `news` (section 3.5.1). Differtent operations are available from the news administration: `view all, new entry,` and `manage news.` By selecting the `view all` feature, an overview of the TR32DB news is provided, which can display `current news, older news,` and `all news`. The operation `new entry` enables the input of a new news statements to the database. Attributes like the news date, content text or content-links will be posted to the database by selecting the `insert` button. Finally, the operation `manage news` can be chosen (Figure 4-36). This feature is responsible for the representation of the news on the startpage of the TR32DB web-interface. All data that own the status `active` are visible on the start page of the website. In contrast, data that holds the status `disable` is initially not displayed. Disabled news will only be represented on the startpage of the TR32DB website, when the the operation `view all news` is used.

Figure 4-36: News Administration using the TR32DB Admin-Tool

DOI Administration

The administration of applied TR32DB DOIs (section 4.5.1.3) is managed by the TR32DB Admin-Tool. By selecting `DOI Admin` from the side menu, an overview of all DOIs created by the TR32DB is displayed (Figure 4-37). This table lists the following details for each dataset with applied DOI: internal TR32DB DOI identifier, user applying the DOI, the DOI identifier name, the TR32DB data identifier number, title of the dataset, and status of submission to DataCite. The table can be sorted according to the user, DOI identifier, TR32DB data identifier, title or submit status. The content of the table is accessed from the associated database table `DOI` (section 3.5.1).

DOIs created by TR32DB

DOI ID	user	DOI	dataID	title	submitted to datacite
49	Waldhoff	10.5880/TR32DB.9	801	Land use classification 2012 of the Rur Catchment	Submit Now
48	Waldhoff	10.5880/TR32DB.8	547	Enhanced land use classification 2013 of the CRC/TR32 measurement region Selhausen/Merken	Submit Now
35	Curdt	10.5880/TR32DB.KGA92.2	539	Potentials of Low-cost Mini-UAVs	yes
34	Curdt	10.5880/TR32DB.KGA92.3	540	Current flood mapping capabilities using synergies from radar altimetry and MODIS multi-spectral satellite imagery and the need for a SWOT mission	yes
33	Curdt	10.5880/TR32DB.KGA92.8	528	Agricultural Monitoring with Spaceborne X-band SAR Data	yes
32	Curdt	10.5880/TR32DB.KGA92.4	541	3D terrestrial laser scanning for field crop modelling	yes

Figure 4-37: DOI Administration: Overview List

To apply a DOI for a dataset at DataCite, the `submit Now` link has to be used in the DOI overview table. The admin user will be forwarded to a website displaying the metadata of the chosen dataset in XML format. The XML structure is automatically generated by accessing the appropriate database tables in the `Metadata_DOI` group (section 3.5.1). A section of the XML structure is displayed in Figure 4-37. The complete XML structure of this example is presented in Figure B-5. The XML structure is conform to the XML schema of the DataCite Metadata Schema Version 2.2 (section 2.2.2.2). The content of the XML structure was provided to the database by a user applying the DOI (section 4.5.1.3), as well as automatically by the TR32DB system (e.g. DOI identifier or publisher).

```
Download XML
Generiertes XML:

<?xml version="1.0" encoding="utf-8"?>
<resource xmlns="http://datacite.org/schema/kernel-2.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://datacite.org/schema/kernel-2.2
  <identifier identifierType="DOI">10.5880/TR32DB.9</identifier>
  <creators>
    <creator>
      <creatorName>Lussem, Ulrike</creatorName>
    </creator>
    <creator>
      <creatorName>Waldhoff, Guido</creatorName>
    </creator>
  </creators>
  <titles>
    <title>Land use classification 2012 of the Rur Catchment</title>
  </titles>
  <publisher>CRC/TR32 Database (TR32DB)</publisher>
  <publicationYear>2013</publicationYear>
  <subjects>
    <subject>Agriculture</subject>
    <subject>ATKIS</subject>
    <subject>Land Cover</subject>
    <subject>Land Use</subject>
    <subject>Remote Sensing</subject>
    <subject subjectScheme="ISO 19115:2003 MD_TopicCategoryCode">GeoScientificInformation</subject>
  </subjects>
```

Figure 4-38: Overview of Metadata in XML Format according to DataCite Metadata Schema V 2.2, the complete XML file is presented in Figure B-5

By using the `download XML` link on the overview page, the displayed xml file can be downloaded. It is now the task of the admin user to check the accurateness of the provided metadata details. If the XML content is correct, the DOI can be applied. For this purpose, the link `DOI registration and metadata upload` at the bottom of the XML overview page has to be used. The XML file, as well as the DOI identifier name and the corresponding URL of the DOI are submitted to the DOIDB Metadata Store at the GFZ

Potsdam (GFZ, 2014) using the DOIDB API (section 3.6). A server error will be presented, in case the submitted details are incorrect. The DOIDB operates as a proxy service, which is connected to the MetaData Store of DataCite (DATACITE, 2014b). By submitting metadata to the DOIDB Metadata Store, a live registration of the DOI is triggered at DataCite. The registration of a DOI is consequently completed.

Metadata Administration for datasets from Data Type Geodata

The administration of data from the data type geodata is provided by the TR32DB Admin-Tool and divided into the menu options `Geodata Input` and `Geodata View`. Both will be described consecutively.

By selecting the menu option `Geodata Input`, an input form for the data type `geodata` is provided. Basically, this input form is set up like the one for the data type `data`, described in detail in section 4.5.1.3. This is arranged according to the multi-level TR32DB Metadata Schema (section 4.3) into two parts: `general properties` and `specific properties`. The input form for geodata enables the input of `general`, `data specific`, and `geodata specific properties`, whereas the latter both are allocated to `specific properties`. As presented in Figure 4-36, all required input fields are marked by an asterisk. Again, two options are available to enter the metadata into the input form. The manual input and the feature `Inserted Metadata`. Likewise, the input form is subdivided into three columns: the title of the input field, the metadata input field itself, and a complementary, descriptive help statement. Again, the introduced four types of input fields are provided: text box, text area, select menu, and calendar, all described in detail in section 4.5.1.3.

Figure 4-39: Part of metadata input form for geodata and `Inserted Metadata` function

In contrast, to the usual, already described `General Properties` metadata input form (section 4.5.1.3), the one for `geodata` is extended with subsequently described properties:

user [select menu](*) A corresponding, responsible user for the metadata input has to be selected from the drop-down list.

dataset name [text box](*) The exact file name of the dataset has to be inserted.

extension[text box](*) The file extension of the dataset has to be added.

creator-organisation [select menu](*) Details of the creator-organisation have to be chosen from the selective list. If the creator-organisation is not available in the drop-down list, the value `add new organisation` has to be selected. Through this, new input fields will appear, into which the new creator-organisation can be entered. For example, attributes like organisation name, department, street, post code, city, URL, contact email, phone, fax, and contact person can be added. By using an insert button, the values will be submitted to the TR32DB database.

creator-organisation role [select menu](*) An appropriate creator-organisation role has to be chosen from the selective list for each added creator-organisation.

Subsequent to the general properties, the `data specific properties` have to be added. In addition to the already described properties measurement/model region and location and temporal extent (section 4.5.1.3), the following properties are included:

spatial resolution-scale [text box] The equivalent scale of the data can be entered, for example 1:'1000'.

spatial resolution-distance [text box][select menu] A spatial resolution distance can be added by inserting a resolution distance to the text box and selecting a unit of length from the drop-down list, e.g. '50'|'m'.

measurement/model frequency [text box][select menu] A frequency of measurement or modelling can be included. A frequency value has to be entered into the text box and an appropriate unit has to be chosen from the selective list, e.g. '10'|'minutes'.

Finally, the `geodata specific properties` may be declared. Besides the already explained properties INSPIRE topic category, geographic bounding box, and linage (section 4.5.1.3), the subsequent properties are also available:

reference system identifier [text box] Details about the vertical coordinate reference system may be added. This text box request the identifying EPSG code of the reference system, e.g. '32632'.

reference system name [text box] The corresponding name of the reference system can be declared, e.g. 'WGS 84/UTM Zone 32N'.

representation [select menu] The representation form of the dataset can be selected from the drop-down list, e.g. 'vector data', 'raster/imagey'.

All data provided in the input forms will be checked after using the `submit` button. In case, an invalid value was entered to an input form, an error message will be displayed. This message will specify the invalid value. After successful correction of the inserted data, the metadata of the entire fields will be posted to the database. The metadata details of the dataset are on-the-fly online available.

By selecting the menu option `Geodata View`, an overview table for all data from data type `geodata` is provided. This table lists the following details for each geodata dataset: the TR32DB data identifier, the file name, the associated TR32 topic, the metadata creator, the upload date of metadata, and the title of the dataset. The `info` button of a specific dataset shows all available metadata of the selected dataset in a pop-up-window. By `selecting` the edit button, the admin user will be able to modify the metadata of a geodata dataset. The admin user will be directly forwarded to a metadata edit input form. This is equal to the geodata metadata input form. The incorrect metadata can be modified. Moreover, missing details can be completed. By using the `submit` button, the data will be uploaded to the database, after successful validation.

Other features

More features are available on the administration website and accessible via the side menu, besides links to the Admin-Tool Home, the TR32DB Website and the TR32 official website. Additional features will be described subsequently.

Data View This feature enables the listing of all data stored in the TR32DB. The displayed data can be filtered to view all data, data without metadata, and data without metadata, stored in the TR32DB. The listing can be sorted according to the Identifier, the dataset title, the data type, the metadata creator, the sub-project, and the data upload to the data storage. In addition, by using the `info` button, the entire metadata of a dataset can be viewed.

Instrument Admin This section enables the administration of the controlled list `instruments`, `instrument parameter`, and `instrument parameter unit`. For each controlled list, an overview list is provided, as well as an insert and edit input form. For example, a detailed listing of the controlled list instruments is provided including the attributes instrument identifier, instrument equipment, model name, manufacturer, registered office, and a corresponding URL. By choosing the `edit` button for a single instrument, the attributes of this instrument can be modified (Figure 4-40). By using the function `insert new instrument`, an input form will be displayed, into which the attributes of the instrument can be entered. By using the `submit` button, the data will be directly posted to the database.

Figure 4-40: Modifying attributes of an instrument

Keyword Admin This section enables the administration of CRC/TR32 keyword list. New keywords can be inserted into the list by using the `insert new keyword` option. An input form will be provided, into which the keyword name and a description of the keyword can be entered. The data will be posted to the database by using the `submit` button. An edit function is available to modify existing keywords. In addition, an overview list of all keywords is provided and can be sorted according to the keyword identifier, keyword name or description.

Dl- Counter This section lists all details, which are collected by downloading a dataset from the TR32DB. These are the user Identifier, the data identifier, the file path of the dataset including file name, and the download time. The information is accessed from the associated database table `dlcounter` (section 3.5.1)

4.5.2 Web-interface Version 4.0

This section gives an overview over the design and functionality of the TR32DB web-interface Version 4.0. Initially, the general organisation will be presented, followed by an overview of available user and admin user features.

4.5.2.1 Web design

The TR32DB web-interface Version 4.0 is available online since the autumn of 2013. This version is an extended, updated version of the prior described Version 3.1 (section 4.5.1). In general, the main functionality of the prior version was adopted, modernized, and extended with new features, as well as with a new layout. The underlying database Version 4.0 of the web-interface is described in detail in section 3.5.1. The web-interface Version 4.0 is designed and implemented, just as the prior version according to previous mentioned recommendation rules, principles, and methods (section 2.3.4 and 3.6). The website has a simple layout colour-coordinated with the CRC/TR32 logo presented in green and yellow colours (Figure 4-41).

Figure 4-41: TR32DB web-interface start page Version 4.0

The structure of the web-interface is based upon the prior Version 3.1. Again, two main menu bars bestride the layout. The `general menu bar` on the top enables access to general information about the TR32DB and CRC/TR32. The `data menu bar` (Side Menu) on the left side deals with the administration of and access to the CRC/TR32 project data. Again, this menu bar is divided into two sections. Section `B` offers search and access to project data. Section `A` facilitates further features for internal users and admin users. The `header` is dominated by the web-interface title and the CRC/TR32 logo. The `footer` was rearranged. Now, it contains the logos of the main participating institutions of the CRC/TR32. In addition, contact details are provided. The `content` was also redesigned in comparison with the Version 3.1. This part will be change

according to the selected menu feature. Consecutively, the individual parts will be described with focus on the improved features.

Startpage

The content of the start page was rearranged and extended by new sections. The start page is now divided into six parts. They will be subsequently described.

Welcome introduction A short introduction to the TR32DB is given in this section.

Workshops This section lists all workshops, organized by the CRC/TR32 project section Z1/INF. The content of this section was adopted from the prior web-interface Version 3.1 and is describe in section 4.5.1.1.

News This section lists the latest news related to the TR32DB, as already described in detail in section 4.5.1.1.

Advanced search The full-text search section was added with the recent web-interface Version 4.0. The search enables a text search for data within metadata properties title and description of a dataset.

Recently uploaded data The most recent five datasets uploaded to the TR32DB are listed in this section. They are represented by a symbol, indicating the type of data, and the title of the dataset. By clicking on the dataset title, the user will be forwarded to the entire metadata of the dataset.

TR32DB visitors This sections displays a tiny map, provided by ClustrMaps (CLUSTRMAPS, 2014), which includes the location of the visitors of the TR32DB, as already described in detail in section 4.5.1.1.

General menu bar

The `general menu bar` (top menu) provides general information about the TR32DB and CRC/TR32 for website visitors. The menu structure of the previous web-interface was streamlined. The existing eight menu tabs were reduced and combined to five new menu tabs. These are:

Home To navigate to the TR32DB start page, the `home` tab can be used.

CRC/TR32 Home By using this tab, the user will be forwarded to the official CRC/TR32 website (http://www.tr32.uni-koeln.de).

About Details are provided about the background of the TR32DB as well as a shot overview of design and functionalities. The `about` section of the previous Version 3.1 was adapted and updated with new information.

FAQ Information is provided to solve basic questions and problems. This section was adapted from the prior Version 3.1, modified, and updated with new details. For example, the access to available lists is provides used with the metadata input form, for example a list of all available creators, CRC/TR32 keywords, institutions, instruments and instrument parameter. In addition, the *CRC/TR32 Data Policy Agreement* (CRC/TR32, 2011) is available for download. Furthermore, tutorials are provided for download to support the use of the TR32DB. For example, a user guide for the web-interface Version 4.0 is available (CURDT, 2013c).

Contact/Imprint Contact and imprint information is provided in this section. The sections `contact` and `imprint` from the previous Version 3.1 were combined, adapted, and updated.

Data menu bar

The `data menu bar` (side menu) is mostly adopted form the prior Version 3.1, updated, and extended. Consequently, part B of the side menu bar facilitates search for project data. Three types of project data search are available via the `data menu bar`:

Predefined search (browse by) Search for data according to topic, data type, measurement/ modelling region, funding phase, cluster (project section), and DOI.

Advanced Search Search for data by means of a combined, advanced data search form, described in detail in section 4.5.2.2.

Map Search Search for data by means of a map, described in detail in section section 4.5.2.2.

The first two search types were transferred from the `data menu bar` of Version 3.1, described in detail in section 4.5.1.2, reorganised and partially renamed. The map search menu item was newly implemented in Version 4.0.

Part A of the side menu bar facilitates further features to access internal project data, as well as to manage data and administrate TR32DB details depending on the user status. This part of the menu is only visible after login to the TR32DB. The menu bar will be expanded automatically after login. The following menu elements are available.

Login/Logout User access to internal features of the TR32DB by login. Likewise, the logout quits the access.

Admin Home Administration and management of TR32DB features and content (e.g. user administration, DOI administration). Only administrator users have access.

User Home Management of project data of a logged in user (e.g. metadata input and edit, DOI application).

Climate Data Information and access to climate data purchased within the CRC/TR32.

Share Provision and access to temporary stored data in the TR232DB share storage.

WebGIS Access to TR32 Viewer, which visualises most of the purchased basic geodata.

All menu elements were adopted from the previous Version 3.1, described in detail in section 4.5.1.2, except for the element `Admin Home`. This element was newly implemented in Version 4.0. All features provided by the menu items will be described in detail in section 4.5.2.3 and section 4.5.2.4.

4.5.2.2 General web-interface features

All users or visitors of the web-interface Version 4.0 are able to access and use general web-interface features in order to explore and download data of the TR32DB. For this purpose certain steps have to be followed (Figure 4-42). These steps involve the use of specific search functions for data (e.g. predefined search, advanced search, map search), selection of data files, exploration of requested data via metadata, as well as download of data depending on permissions. The basic functionality was adopted from web-interface Version 3.1 (section 4.5.1.2), modified and updated. Subsequently, the general features will be described.
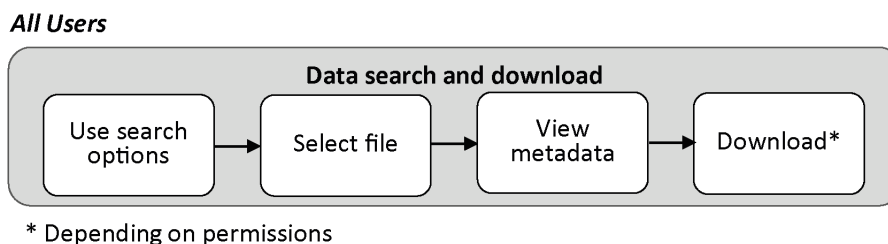
Figure 4-42: TR32DB user workflow for data search and download using the web-interface

Predefined search for data

The predefined search for data enables all visitors of the web-interface an easy and immediate search for project data according to the chosen category. The search is executed via the metadata of a dataset. As already mentioned in a previous section, the access to the various search queries is accessible through the `data menu bar`. A predefined search is available for the categories: topic, data type, regions (sites), phases, clusters (projects), and DOIs. The general search query is described in detail in section 4.5.1.2. Therefore, the description of the various predefined search options will be short and focused on the modification.

> **Browse by Topic** By selecting the predefined search for data according to a CRC/TR topic (e.g. soil, vegetation, atmosphere), a list of all datasets, assigned with the selected topic is presented (Figure 4-49).
>
> **Browse by Type** By choosing the predefined search for data according to a data type (e.g. data, reports, publication), a list of all datasets, assigned with the chosen data type will be displayed.
>
> **Browse by Region** By applying the predefined search for data with focus on a CRC/TR32 predefined region or site, initially a listing of all available regions and sites will be presented (Figure 4-43). After selecting a preferred region or site, the result of the query will list all assigned datasets.

Figure 4-43: Predefined data search for regions and sites of the CRC/TR32 (partial view)

> **Browse by Phase** By selecting the predefined search for data according to the funding phase of the CRC/TR32, initially an overview of the available funding sites is presented (Figure 4-44). After choosing a funding phase, a list of assigned datasets will be displayed.
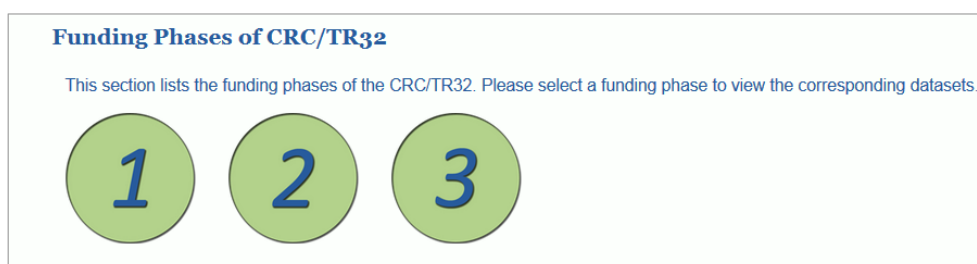
Figure 4-44: Predefined data search according to funding phases

> **Browse by Cluster** By choosing the predefined search for data with focus on the cluster or project section, at first an overview of all CRC/TR32 clusters is presented with corresponding project sections (Figure 4-45). By selecting a project section, the user will receive a list of all data referring to the query.

Cluster and Project Sections of the CRC/TR32

This section lists the Cluster Groups of the CRC/TR32 with the corresponding project sections. By selecting a project section, you will receive a list of all availabe datasets in the TR32DB related to the choosen project section.

A

A1 - Pore structure, moisture, and water transport in soil by NMR

A2 - Characterization of flow and transport properties of unsaturated soils using non-invasive methods and numerical modelling at the µm to dm scale

A3 - Inverse modelling of soil hydraulic property patterns from non-invasive electrical measurements

A4 - Multiphase flow simulations coupled to NMR and SIP signatures in reconstructed natural porous media using Lattice Boltzmann methods

Figure 4-45: Predefined data search according to a cluster or project section (partial view)

Browse by DOI By selecting the predefined search for data with an DOI, an overview of all data with a TR32DB DOI will be displayed (Figure 4-46). The user will be forwarded to the overview webpage of the DOI, by choosing the DOI name, for example 10.5880/TR32DB.KGA90.10. The metadata overview will be described later in this chapter.

List of all TR32DB DOIs

This sections lists all dataset in the TR32DB with a DOI sorted by surname of first creator:

Bareth, Georg; Bolten, Andreas; Bendig, Juliane (2011): Potentials of Low-cost Mini-UAVs. Geographisches Institut der Universität zu Köln - Kölner Geographische Arbeiten. DOI: 10.5880/TR32DB.KGA92.2

Baru, Chaitanya K. (2010): Integration or Mashup? Challenges in Bringing Together Heterogeneous Scientific Data. Geographisches Institut der Universität zu Köln - Kölner Geographische Arbeiten. DOI: 10.5880/TR32DB.KGA90.2

Figure 4-46: Predefined data search according to DOI (partial view)

Advanced search for data

The advanced data search enables a visitor of the web-interface to search for data by combining different request with each other. This search is also executed via the metadata of datasets. The combined, advanced data search of the prior Version 3.1, described in detail in section 4.5.1.2, was mostly adopted, updated, and extended with new query options and a new layout. By selecting the menu item `data search`, the user will receive an input form (Figure 4-47).

Data Search

Search for data in the TR32DB via a search form. Complete the form with your request and press **submit**. As a result of the query, you will receive a list with all dataset that match your query.

Please consider: The values of the query will be joined during the search with AND!

Data Search	
Full-text search:	land use classification
Data Type:	Data
TR32 Topic:	Land Use
Keyword:	Land Use
Project Section:	choose...
Funding Phase:	choose...
Creator:	Waldhoff, Guido (guido.waldhoff@uni-koeln.de)
Region / Site:	--RurCatchment--
Instrument:	choose...
Parameter:	choose...
	Submit

Figure 4-47: Advanced data search with exemplary values

This enables mainly selective lists, except one text box. Further information about a selective list or text box is shown by moving or clicking on the " i "-symbol. The user is able to enter terms into the text box or select

values from the drop-down lists. By using the submit button, the query will be executed and the user will receive a result. All queries in the search form will be combined with AND. Consequently, the visitor will only receive a result that matches all chosen values. For example, the web-interface visitor can enter a specific term to the full text search (e.g. land user classification). In addition, a data type can be chosen (e.g. data), a TR32 Topic (e.g. Land Use), the creator of the data (e.g. Guido Waldhoff), as well as the measurement or modelling region or site (e.g. Rur Catchment).

Map search for data

The map search enables the search for project data by means of a map. This feature was newly implemented in the web-interface Version 4.0. The feature is available by selecting the Map Search item on the data menu bar. As a result, the web-interface visitor will receive a website including a map (Figure 4-48). In addition, some general information is given above the map describing the functionality of the map and its symbols. The representation of the map can be selected (e.g. base map, terrain map, satellite map with and without labelling) by choosing Map or Satellite. Furthermore, TR32DB Web Map Services are complemented to the map and presented depending on the zoom level (e.g. Rur catchment and sub-catchment boundaries, measurement stations, and boundaries of study sites). The visitor can pan and zoom to the area of interest by using the zoom slider, pan wheel, zoom tools or by using the mouse wheel. By selecting Zoom to Area, the map will zoom to the start extend.

A red marker pin represents a dataset with geographic coordinates. Consequently, datasets of the TR32DB without geographic coordinates will not be displayed. The location of the red pins represent the calculated centre point of given bounding box coordinates. Depending on the zoom-level, these maker pins will be clustered to yellow, blue or red circles. The corresponding number in the circles represents the included number of marker pins. The visitor has to zoom-in or selecting a cluster to devolve the circle. Red marker pins that are located nearby or have the same coordinates will fan out into a circle representing all included pins by clicking on a certain pin. To receive further information about a red marker pin, the visitor has to click on the pin. The TR32DB identifier, dataset title, data type, and TR32 topic of the selected dataset will be represented. To access the entire metadata detail of a dataset, the visitor has to click on the corresponding dataset title.

Figure 4-48: Map search for datasets in TR32DB at Selhausen area, Germany

Presentation of data

Data stored in the TR32DB is represented on the web-interface by means of their metadata. The representation of the metadata on the website was totally rearranged in the web-interface Version 4.0. Subsequently, the new presentation of datasets will be described.

The result of a predefined search or combined query is always displayed in an overview list of datasets that match the request. The design of all outputs websites of search results is identical, except from the query `Browse by DOI`. The default search result has always the same structure in the content, as presented in Figure 4-49. The properties will subsequently be described in details.

- **Title** The title defines the content of the search result. The headline and the corresponding description text will be modified according to the query (e.g. Search Result: All Data from Topic - Remote Sensing, Search Result: All data from Data Type - Publication).
- **Sorting** All datasets available in the listing can be sorted in order by: the TR32DB internal dataset identifier, the data type, the creator surname, the TR32 Topic, and the last upload date of metadata. By clicking on the symbols ∨ or ∧, the single dataset will be sorted. By default, the listing is sorted downwards in order by the TR32DB dataset identifier.
- **Filter** All datasets available in the listing can be filtered according to the funding phase, in which they were created. For example, datasets can be listed for all funding phases, phase 1, and phase 2. By selecting the desired phase, the output list will be changed.

Figure 4-49: Presentation of search results according to the topic `remote sensing`

- **View** The number of displayed datasets in the listing can be changed (e.g. 5, 10, 15, 20, 50, 100). The default number of datasets listed is five. By selecting a desire value number, the list of the datasets will be changed.
- **Navigation** To navigate through the search result list, two options are available. To navigate to the next page of the dataset listing, the page navigation has to be used. By clicking on the desired page number, the dataset listing will change. To navigate from the bottom of the website to its

top, the horizontal navigation has to be used. By clicking on the `top` ∧ symbol, the top of the website will be displayed.

Dataset Output All datasets related to a certain query are listed in the search result list. Each dataset is represented in a green box. This box provides an overview of the dataset and is divided into a header and content part. The header includes the TR32DB identifier and the title of the dataset. The content contains the corresponding keyword(s), the dataset creator(s), the last upload date of the metadata details, the data type of the dataset, as well as the TR32 topic. In addition, a logo is displayed representing the associated data type.

Dataset Features There are several features available for each dataset in the overview listing depending on the access rights and the availability of the underlying file (Figure 4-50). The features are: `View Metadata`, `Additional File`, and `Download Data`. The full available metadata of a dataset will be displayed by selecting the dataset title or relevant logo, as well as by choosing the `view metadata` button (A). If an additional descriptive PDF file is available, this can be downloaded by selecting the PDF symbol (feature `Additional File`) (B). To download the underlying file of the dataset, the floppy disk symbol has to be chosen (feature `Download Data`). The download of the file is available, depending on the download permission of the dataset. An appropriate note about the download will be displayed by moving the mouse pointer over the floppy disk symbol (D).
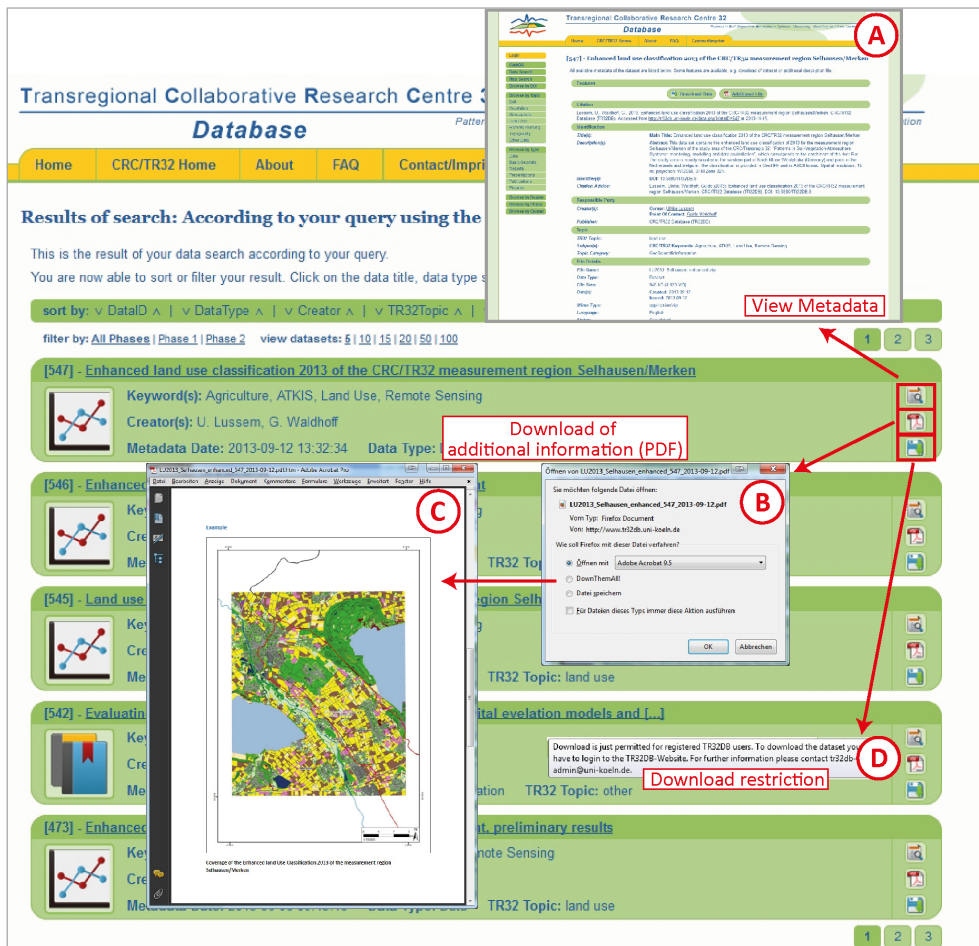
Figure 4-50: Presentation of Dataset Features; (A) view metadata of dataset, (B) download additional PDF file, (C) view additional PDF file, and (D) additionally information with download information for dataset

By downloading a dataset from the TR32DB, the user will receive a download advice (Figure 4-51) that has to accept before the actual download window will pop up. The user is able to save or open the dataset. The download of a PDF is available without download advice information.
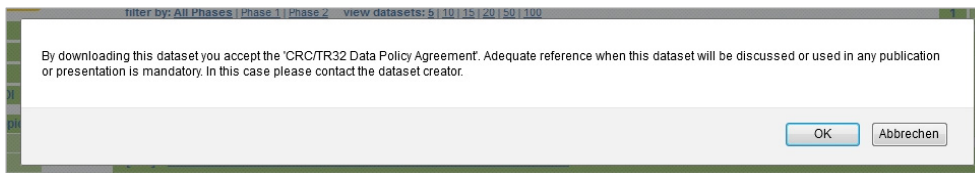
Figure 4-51: Download advice has to be accepted before downloading of dataset

The download of the data file and the descriptive PDF will be logged automatically in the TR32DB database. Available download details like the TR32DB data identifier (e.g. 547), TR32DB user identifier, the IP address (Internet Protocol address), the storage path of the data in the TR32DB data storage including file name (e.g. /afs/rrz.uni-koeln.de/project/tr32db/Phase_2/C/C3/pictures/000/Fotos_2011.zip), as well as the download time will be saved (e.g. 2014-04-01 15:15:22).

The metadata view of each dataset (Figure 4-53) can be viewed by all visitors of the website. Each metadata view website has its own persistent, identifying URL (uniform resource locator), for example `http://www.tr32db.uni-koeln.de/data.php?dataID=547`. All URLs are set up in the same structure. Merely, the TR32DB dataset identifier (`dataID`) at the end of the URL will be modified. The metadata view is structured thematically in a listing, arranged according to the TR32DB Metadata Schema (section 4.3) displaying all available metadata details of a dataset.

Initially, the header of the metadata view displays the TR32DB data identifier and the name of the title of the dataset (e.g. [547] Enhanced land use classification 2013 of the CRC/TR32 measurement region Selhausen/Merken).

Afterwards, the available features of the dataset will be presented. These depend on the user rights and the availability of the dataset file. Currently, a maximum of four features is available (Figure 4-52). These are:

Download Data This feature enables the download of a dataset file, depending on availability and access rights as described previous.
Additional File This feature enables the download of an additional, descriptive file in PDF format (if available) as described previous.
Edit Data This feature enables the modification of metadata and is only available for the creator of the metadata (section 4.5.2.3).
Apply DOI This feature enables the application of a TR32DB and is only available for the creator of the metadata (section 4.5.2.3).
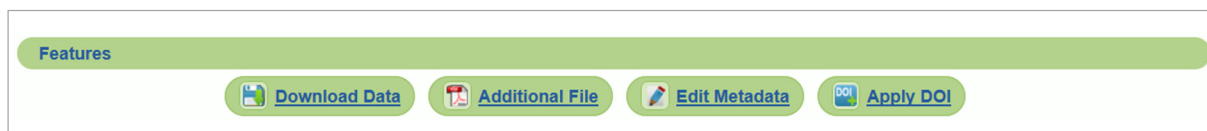
Figure 4-52: Maximum metadata features of a dataset available at the metadata view

Subsequently, a citation advice is provided for all datasets. This is automatically created by means of the available metadata of a dataset. The citation advice is modified according to the data type. In case of `publication`, the citation advice is adjusted according to the publication type. Basically, the structure of the citation advice is currently following the reference style of Elsevier (e.g. ELSEVIER, 2014). The reference style for datasets from the data type `data` follows the recommended format of DataCite (DATACITE METADATA

WORKING GROUP, 2011) and was adjusted for the requirements of the TR32DB. The citation advice can be adapted to any other reference style.

Consecutively, the metadata of the dataset are shown. They are thematically structured into several sections arranged according to the metadata categories introduced in the TR32DB Metadata Schema (section 4.3). At first, the general metadata categories are shown. They are extended to the specific metadata categories depending on the data type. In general, all available metadata properties and sub-properties are displayed. The sub-property (e.g. `titleType`) value is always shown in front of the property (e.g. `title`) value. Multiple uses of properties are sequently presented. Some of the metadata details are linked to further information. In the following, a short overview will be given about each single section.

Identification This section represents available metadata related to the metadata category `Identification`. This are, for example, details about the dataset title and corresponding title type, description and associated description type or identifier and appropriate identifier type. An identifier with the identifier type value DOI or URL will be linked to a corresponding website (A), as shown in Figure 4-53.

Responsible Party This section displays all available metadata of the metadata category `Responsible Party`. These are, for example, metadata about the creator of the dataset as well as the publisher. In case of data type `geodata`, the creator organisation and creator person are shown. The creator value is always linked to an associated TR32DB website representing the details of the creator (B), as shown in Figure 4-53.

Topic This section shows all available metadata details of the metadata category `Topic`. These are for example TR32 topic, TR32 keywords or topic category. All subjects that can be chosen a number of times, are grouped and presented among subject(s) including their sub-property name, for example CRC/TR32 Keywords, Gemet Thesaurus or DCC. All values of these subjects are displayed successively.

File Details This section lists all available metadata details of the metadata category `File Details`. For example the file name, file size, and corresponding dates are displayed. The mime type will be automatically presented, depending on the extension of the file.

Constraints This section displays all available details form the metadata category `Constraints`. For example, the download permission, access and use constraints, and the chosen licence is presented. The licence value is linked to the corresponding content (C), as shown in Figure 4-53.

Geographic This section represents all metadata information of the dataset, which is related to the metadata category `Geographic`. For example, the measurement or modelling region and site can be listed, as well as the bounding box coordinates. In case bounding box coordinates are allocated, a map will be displayed showing the centre point of the bounding box on the map with a red marker pin. The image `no map data` is presented, when geographic coordinates are unavailable.

Specific Information This section lists the data type specific metadata and will be adjusted according to the data type of the dataset. Metadata of the category `Data` describe, for example, the temporal extent of the data and the lineage. Metadata details of the category `Geodata` include, for example, a reference system, spatial representation or spatial resolution. Metadata of the category `Report` contain for example the report date, type, city or institution. Metadata details of the category `Picture` display, for example, the record place, image orientation, file size or resolution, and recording extent. Metadata of the category `Presentation` describe the presenter, presentation date and type, as well as details about the corresponding event. The metadata of the category `Publication` are dependent on the publication type. For example, details about the publication and review status, publication year, publication type, publication source, series editor, series title, proceedings title and editor, event details, volume, issue or page range can be displayed.

Figure 4-53: Exemplary metadata view for Data Type Data with linking to (A) the identifier of the dataset, (B) the contact details of the creator, (C) the choosen licence, and (D) the contact details of the metadata creator

>**Metadata Details** This section displays all metadata details available for the dataset of the metadata category `Metadata Details`. For example, the creator of the metadata, the metadata creation date, as well as last upload of metadata, metadata language and version is presented. The metadata creator is always linked to an associated TR32DB web-interface (D), as presented in Figure 4-53.

Finally, at the bottom of the metadata view, the available features of the dataset are presented again, as already previously described in this section.

The data description of a single DOI (Figure 4-54) can be seen by all visitors of the TR32DB web-interface. It serves as a landing page for the DOI and will be displayed after resolving the DOI. The website provides an view of a single dataset. The general structure of the website was adapted from the web-interface Version 3.1, extended and updated with new metadata details and layout. Each DOI description website has its own, persistent, unique URL (e.g. http://www.tr32db.uni-koeln.de/DOI/doi.php?doiID=48). The URLs of the DOIs are all set up in the same structure. Only the DOI identifier (`doiID`) value changes at the end of the URL.

The data description view is always set up in the same structure and lists all available metadata associated with the selected TR32DB DOI. Initially, the title gives an advice about the described dataset including the DOI name (e.g. Data Description of DOI 10.5880/TR32DB.8). At first, the citation of the dataset is presented following the recommended format of DataCite (DATACITE METADATA WORKING GROUP, 2011). Subsequently, metadata details of the dataset are presented, for example, title, description, contributor or subject. All available metadata properties and corresponding sub-properties are listed, for example, the values of the property `title` and the corresponding value of the sub-property `titleType`. The sub-property is always presented in from of the property (e.g. Main Title|Enhanced land use classification 2013 of the CRC/TR32 measurement region Selhausen/Merken). Multiple uses of properties are presented in the following. Some metadata details are linked to complementary information. For example, the related identifier with the identifier value URL or DOI is always linked to the associated website (A and B), as shown in Figure 4-54. The value of the licence is also always linked to the corresponding describing content of the licence (C), as presented in Figure 4-54.

Finally, further information is provided past the listing of the metadata details. Contact details of the DOI applicant are provided. The value of the contact is always linked to an associated TR32 website that represents the details of the creator (D), as shown in Figure 4-54. These typically include the creator's name and membership in a CRC/TR32 project section, the associated institution and address information, as well as the mailing address and telephone number. Moreover, a link is provided to the entire metadata detail of the TR32DB (E). By selecting the view metadata link, the user will be forwarded to the appropriate TR32DB website (Figure 4-54). Moreover, the dataset can be downloaded, depending on the chosen download permission during the DOI application and the access authority of the user or visitor of the TR32DB. In case, the download is only permitted for registered TR32DB users, a note will be displayed with appropriate, explanatory references and contact information (Figure 4-54). If a download of the dataset is possible, a download link is provided. By selecting this link, the user will receive a download notice (Figure 4-51). By accepting the conditions, the actual download window will pop up and enable the user to save or open the dataset. By accessing the download window, the download will be logged automatically in the TR32DB database. Available download details like the TR32DB data identifier, TR32DB user identifier, the IP address (Internet Protocol address), the storage path of the data in the TR32DB data storage including file name, and the download time will be saved.

Figure 4-54: Data Description of DOIwith linking to (A) the first related resource of the dataset, (B) the second related resource of the dataset, (C) the chosen licence, (D) the contact details of the metadata creator, and (E) the TR32DB metadata view

4.5.2.3   Internal web-interface features

The internal features of the web-interface Version 4.0 are only accessible for registered users. The administration of metadata, the application of DOIs, the access to the temporary share and climate data, as well as the WebGIS belongs to the internal features. The basic functionality was adopted from web-interface Version 3.1 (section 4.5.1.3), modified and updated with new layout and features. They will be described subsequently.

Web-interface login and user home

The access to the internal web-interface features is only provided for registered users of the TR32DB. To login to the web-interface, the user has to select the `login` item on the `data menu bar.` A log-in window will appear (Figure 4-16). The user has to enrol, as already described in detail in section 4.5.1.3. After a successful login, the `data menu bar` will be expanded with the internal features, as shown in Figure 4-41 and described in section 4.5.2.1. The name of the successfully logged in, current user will be displayed.

By accessing the menu item `User Home` from the `data menu bar`, the user will be forwarded to his `User Home` (Figure 4-55), where he is able to use different internal features. A descriptive text provides support. Two main functions are provided. These are `Data Administration` and `DOI Administration` both containing two features each. The `Data Administration` contains features to manage the metadata of the

datasets by using the features `Insert Metadata` and `View Own Metadata`. The `DOI Administration` includes features to apply and manage the datasets with TR32DB DOIs by using the features `Apply DOI` and `View own DOI`. All features will be described in detail in the next section.
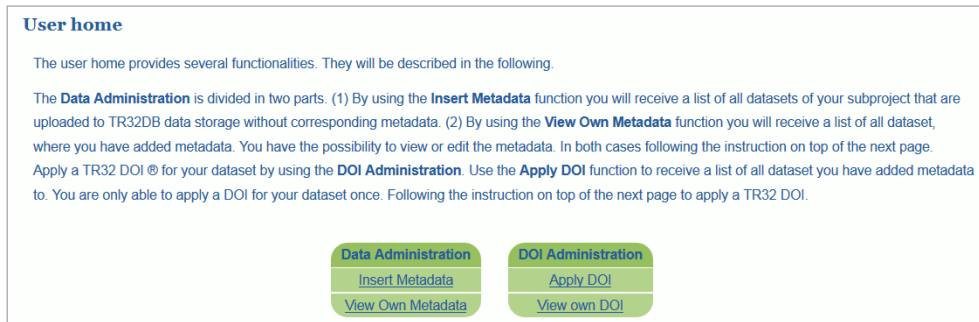
Figure 4-55: TR32DB User Home of Web-Interface Version 4.0

Metadata administration for datasets

One of the main important tasks of the TR32DB is the administration of metadata. The general administration of the metadata was adapted from Version 3.1, modified, updated, and extended. As already described in detail section 4.5.1.3, the administration of metadata has to follow six general steps (Figure 4-56).
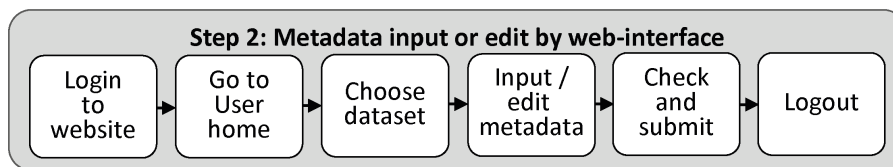
Figure 4-56: Workflow of metadata input and edit (step 2)

After a successful `login` to the web-interface and navigation to the `User Home`, the user has to select the feature `Insert Metadata`, to enter metadata for a file that has already been stored in the TR32DB data storage (section4.4). The user is forwarded to a website, which gives an overview of all datasets in the TR32DB without corresponding metadata of sub-projects related to the user currently logged in. The datasets are organised according to the funding phase and project section into several listings. For each dataset, there are the following details presented in the listing: the TR32DB internal identifier, the name of the file, the time of upload to final data storage, and the data type. To publish and make a dataset visible at the web-interface, metadata have to be added.

To add metadata for a dataset, the user has to select the `insert` button of the required dataset. The user will be forwarded to the next website that displays an input wizard for the metadata (Figure 4-58). In addition, already available metadata details of the chosen dataset are displayed above the wizard. These include the TR32DB dataset identifier (e.g. 847), the file name (e.g. LU2010_Selhausen_enhanced.zip), the data type (e.g. data), the upload time of the file to the TR32DB final data storage (e.g. 2013-09-12 03:00:03), and the file size in kilobyte (e.g. 945 kB).

The input wizard is adapted to the different types of data stored in the TR32DB and is arranged according to the TR32DB Metadata Schema (section 4.3). As presented in Figure 4-58 to Figure 4-72, the input wizard is divided into eight tabs. These are: a `Template` tab, six tabs to describe general metadata properties (`Identification, Responsible Party, Topic, File Details, Constraints, Geographic`), as well as a `Specific` tab to enter specific properties depending on the selected data type. The user is able to navigate through the single tabs by using the `<previous and next>` buttons, as well as by clicking on the single tab names. The metadata input in the single tabs is implemented by using four input

field types: text box, text area, selective menu, and calendar. All are described in details in section 4.5.1.3. Mandatory input fields are marked with an asterisk (*). Selective lists are always mandatory, if they are used in combination with another selective list, text box, text area or calendar. Further information and comments about a single input fields or selective lists is available by moving the mouse pointer over the " i "-symbol. Moreover, examples values are presented. By clicking on the " i "-symbol, the info window will remain. By selecting the " x " symbol in the window, it will be closed. Subsequently, the single tabs of the wizard will be described in more detail. Example values for input fields are represented with single quotation marks ' '. In case of combined input fields the example values are merged with a vertical bar | .

Figure 4-57: Overview of metadata without metadata of sub-projects related to the current user

The tab `Template` (Figure 4-58) enables the user to re-use available metadata of another dataset. This feature was adapted from the web-interface Version 3.1 (section 4.5.1.3), modified and updated. The template feature facilitates and accelerates the metadata input of similar datasets, like time series of measurements with the same instrument and parameter. To use a template of an existing dataset, a dataset has to be chosen from the selective list. This listing contains all datasets, which own corresponding metadata that was provided by the current user for the chosen data type. The datasets are represented in a listing by their dataset name and their corresponding TR32DB identifier in order by the name of the dataset. For example, all datasets from the type `data` of the user currently logged in are presented in case the user is providing metadata for a new `data` dataset (e.g. [547] Enhanced land use classification 2013 of the CRC/TR32 measurement region Selhausen/Merken). By selecting a dataset from the drop-down list, the dataset overview metadata will be displayed. The user is able to access the entire metadata of the dataset by using the `view metadata` feature. After accepting the choice of the template, the available metadata of the dataset will automatically fill the input fields of the wizard. Metadata that was inserted into the input fields in the several tabs before executing the template feature will be overwritten. The user is now able to continue the metadata input. Inserted metadata can be completed, modified or added by the user. In case no template is used, all metadata have to be inserted manually into the input forms.

Figure 4-58: Metadata input wizard tab `Template`

The tab `Identification` (Figure 4-59) enables the user to enter metadata details in order to identify the dataset. For example, the user has to specify a name or title for the dataset and choose a corresponding title

type (e.g. Enhanced land use classification 2013 of the CRC/TR32 measurement region Selhausen/Merken|Main Title). Multiple entries of a title are offered by using the add button.

Figure 4-59: Metadata input wizard tab Identification with exemplary values

A short description of the dataset has to be inserted with a corresponding description type (e.g. This data set contains the enhanced land use classification of 2013 for the measurement region Selhausen/Merken of the study area of the CRC/Transregio 32|Abstract). If an identifier like a DOI or ISBN is already available for the dataset then this can be added (e.g. 10.5880/TR32DB.8|DOI). Moreover, a reference from the described dataset to a related resource can be specified by inserting the name of an identifier, and selecting the identifier type and type of relation to the related resource (e.g. 10.5880/TR32DB.6|DOI|isVersionOf). Finally, an optional advice to cite the dataset can be complemented (e.g. Lussem, Ulrike; Waldhoff, Guido (2013): Enhanced land use classification 2013 of the CRC/TR32 measurement region Selhausen/Merken. CRC/TR32 Database (TR32DB). DOI: 10.5880/TR32DB.8). Multiple inputs of values are enabled for the properties title, description, identifier, and relation.

The tab Responsible Party (Figure 4-60) facilitates the user to add metadata details describing the responsibility of the dataset. The creator(s) of the dataset and his creator role have to be specified. The creator can be defined as the main researcher involved in producing the dataset, a person responsible for the content of the resource, respectively the author of a publication. The creator(s) have to be entered in the correct order to ensure the correct citation of the dataset. At least one creator has to be specified. A creator and the corresponding creator role can be selected from a drop-down list (e.g. Guido Waldhoff|Point Of Contact). In case, the desired creator is not available in the listing, a new creator can be added by using the new button (Figure 4-60). An input form will be displayed with the required information (e.g. first name, surname, e-Mail, TR32 membership, institution) to add a new creator to the list. By using the insert button, the entered details will be checked. In case, all required details are provided and the creator is not yet available in the listing, then the creator details will be submitted to the database and respectively to the creator listing. An institution can be added to the institution listing in a similar manner by selecting add new institution from the institution listing. Again, an input form will be expanded and allows the entry of institution details, like university name, institution name, corresponding address and website. Furthermore, a

dataset contributor with a corresponding contributor role may be added in the tab `Responsible Party`. A contributor is a person or institution, which is responsible for making a contribution to the dataset in supplement to the creator(s). A contributor's name can be entered to the text box and a corresponding contributor type has to be selected from the selective list (e.g. CRC/Transregio 32, Sub-Project Z1/INF|Producer, German Research Foundation (DFG)|Funder). Finally, a publisher has to be specified, which is described as the responsible organisation or person for making the data available. By default, the mandatory `publisher` property will be displayed at the tab `Responsible Party` with the default value 'CRC/TR32 Database (TR32DB)'. Only in case of data type `publication` and `report` the publisher input field is not listed in this tab. It is relocated to the tab `Specific.` Multiple inputs of values are enabled for the properties `creator` and `contributor`.

Figure 4-60: Metadata input wizard tab `Responsible Party` with exemplary values

The tab `Topic` (Figure 4-61) enables the user to add metadata details to describe the dataset with various subjects. The user has to select a CRC/TR32 topic from the selective list (e.g. Remote Sensing, Land Use, Vegetation). In addition, the user has to choose at least one TR32 keyword, which focuses on the CRC/TR32 research topic, from the selective list (e.g. Land Cover Mapping, Land Use). Moreover, keywords describing the dataset according to the GEMET Thesaurus can be chosen from a drop down list (e.g. land use, land use classification). A selective list will be displayed, after entering a correct term into the text box. Finally, a value from the topic category (according to ISO 19115 and INSPIRE directive) listing can be selected (e.g. geoscientificInformation) as well as value from the DDC (Dewey Decimal Classification System) drop-down list (e.g. 550 Earth Sciences, 633 Field & plantation crops). Multiple inputs of values are enabled for the properties `TR32 keyword, Gemet-Thesaurus,` and `DCC`.

The tab `File Details` (Figure 4-62) enables the user to enter information about the file or format details of the dataset. A data type has to be selected from the selective list. For example, the value `dataset` can be chosen for data from the data type `data`. The value `text` can be selected for a publication. In addition, the storage format and corresponding version of the dataset can be inserted. The format can be selected from the drop-down list (e.g. MS Excel) and the associated version (e.g. 2010) has to be entered into the text box. In case, the file format is not available in the listing, then a new value has to be added by using the `new` button. An input form will be displayed to enter the format name (e.g. PNG), the corresponding category (e.g. image)

and an optional description (e.g. Portable network graphics). Furthermore, the size of the data can be added by inserting the amount of data and selecting the corresponding data size type (e.g. 5|Datasets, 8|Pages). Consecutively, at least one date, referring to the dataset, has to be added with an appropriate date type. A date has to be inserted by using the calendar, appearing by clicking into the calendar field (e.g. 2014-04-01). By default, the value of the calendar is set to the current date. The data type has to be selected from the drop-down list (e.g. Created, Date Accepted). Finally, the primary language of the dataset (e.g. English, German) and the processing status of the dataset (e.g. Completed, In Process, Raw Data) has to be selected from the appropriate drop-down list. Multiple inputs of values are enabled for the properties `size` and `date`.

Figure 4-61: Metadata input wizard tab `Topic` with exemplary values

Figure 4-62: Metadata input wizard tab `File Details` with exemplary values

The tab `Constraints` (Figure 4-63) enables the user to enter details concerning access and user of the dataset. Initially, a download permission has to be selected from the drop-down list (e.g. Free, Only own sub-project, Only TR32). Depending on the chosen download permission, the access to the dataset will be

restricted for visitors of the TR32DB web-interface. Additional details about the download of the dataset can be provided in a text box. Furthermore, conditions that apply to access and use of the dataset can be specified. In case, none of the conditions apply, the value 'no conditions apply' should be used. The value 'conditions unknown' should be used, when the conditions are unknown. In addition, limitations on public access of the dataset can be added and corresponding reasons. If there are no limitations on the dataset, the value 'no limitations' should be used. Finally, a licence can be applied for the dataset by selecting a licence name from the drop-down list. Currently, licenses are available from Creative Commons (CREATIVE COMMONS, 2014), Open Data Commons (OPEN DATA COMMONS, 2014), as well as the *CRC/TR32 Data Policy Agreement* (CRC/TR32, 2011) (e.g. TR32DB Data Policy Agreement, Creative Commons-Attribution 3.0 Unported).

Figure 4-63: Metadata input wizard tab `Constraints` with exemplary values

The tab `Geographic` (Figure 4-64) enables the user to complement the metadata of the dataset with geographic details. Initially, a measuring or modelling region or location has to be selected from the drop-down list to describe the place or spatial region where the dataset was measured, modelled, or collected or focused on (e.g. Rur Catchment, Merken, Ellebach). In addition, a geographic coverage of the dataset can be added. This is expressed in terms of geographic coordinates given as a bounding box. By enabling the choice box, a map will be displayed. By default, the area of the Rur catchment is shown. Depending on the zoom level, further details like the shape of the study sites (e.g. in Selhausen or Wüstebach) will be visible. In addition, four text boxes are available above the map that represent the coordinates of the north bound latitude, south bound latitude, east bound longitude, and west bound longitude of the map. The coordinates in the bounding boxes will be modified automatically by panning or zooming of the map. The dataset will be displayed and searchable via the `Map Search` feature from the `data menu bar`, if the user has enabled the choice box.

After successful completion of the general metadata properties, the specific metadata properties can be added into the `Specific` tab. This tab will be adapted depending on the data type. The current type of data is displayed in the metadata details displayed above the wizard, as well as in brackets in the `Specific` tab. Subsequently, the single `Specific` tabs will be described in more detail.

The tab `Specific (Data)` (Figure 4-65) enables the user to add specific metadata details with focus on the data type `data`. Initially, the user has to insert the period of time covered by the dataset (e.g. measurement or modelling period). This includes a start and end date. One day events have the same start and end date. By

Figure 4-64: Metadata input wizard tab `Geographic` with exemplary values

Figure 4-65: Metadata input wizard tab `Specific (Data)` with exemplary values

clicking on the calendar field, a calendar appears on which the time and date can be selected (e.g. 2009-01-01 19:57 - 2009-12-27 19:57). A statement on the processing history or the overall quality of the dataset can be inserted to the lineage text area. This could also include details about the quality assurance or validity of the dataset, legal validity or details about versioning of the dataset. Finally, the measuring instrument or modelling method can be described, which was used to create or is included in the current dataset. To add an instrument to the dataset, the `add` button next to the instrument has to be used. A drop-down list appears with available measurement instruments and modelling methods. The user can now select an instrument from the list (e.g. Ecohydrological model (DANUBIA)). If the desired instrument is not yet available in the listing, the `new` button next to the instrument has to be used. An input form will expand to enter metadata details of the desired instrument or model (e.g. instrument name, instrument model, manufacturer, registered office, and manufacturer website). By using the `insert` button, the instrument will be available in the listing.

Furthermore, a measured or modelled output parameter can be added to the instrument by using the `add` button next to parameter. A drop-down list with currently available parameters will be displayed as well as a selective list with corresponding parameter units. The user is able to choose a parameter and the appropriate unit (e.g. Leaf Area Index|m²/m²). In addition, the measured or modelled resolution distance including the corresponding unit (e.g. 150|m) and the temporal frequency with corresponding unit (e.g. 1|day/s) can be added. In case, the desired parameter and unit is not available in the listing, then both can be supplemented using the same method as already described for the instrument. Multiple inputs of values are enabled for the properties `instrument` and `parameter`.

The tab `Specific (Report)` (Figure 4-66) enables the user to enter specific metadata details with focus on the data type `report`. Some mandatory properties have to be inserted, which are required for the accurate citation of the dataset. A date of report (e.g. 2013-08-28) has to be entered using the calendar field. A type of report (e.g. PhD Report, TR32 Fellow Report, Bachelor Thesis) has to be selected from the drop-down list. In addition, the corresponding city and country where the report was created (e.g. Bonn, Germany) and the appropriate institution (e.g. Meteorological Institute, University of Bonn) has to be entered, as well as the publisher. The default value for publisher is 'CRC/TR32 Database (TR32DB)'. Furthermore, optional properties can be added. Besides the volume of the report, also the number of pages (e.g. 8) and the page range (e.g. 1 – 8) can be inserted. Finally, further additional information about the report can be provided in a text area (e.g. TR32 Student Report Phase II).

Figure 4-66: Metadata input wizard tab `Specific (Report)` with exemplary values

The tab `Specific (Presentation)` (Figure 4-67) enables the user to define specific details with focus on the data type `presentation`. Some mandatory properties have to be specified. They are required to appropriately cite the dataset. The presenter of the presentation may be defined (e.g. Constanze Curdt). Furthermore, the specification of presentation date (e.g. 2009-06-17) by using the calendar field function is mandatory, as well as the selection of the presentation type (e.g. Poster, Talk, Keynote) from the drop-down list. Supplementary information can be provided with regard to the event and where the presentation was presented. The event name (e.g. GSDI 11 World Conference), the event location (e.g. Rotterdam, The Netherlands), as well as the duration of the event (e.g. 2009-06-15 – 2009-06-19) are mandatory. In addition, optional detail like the event type can be selected from the drop-down list (e.g. Conference, Workshop, Project Meeting) and the event website can be provided by a text box (e.g. http://www.gsdi.org/gsdiconf/gsdi11/).

Figure 4-67: Metadata input wizard tab `Specific (Presentation)` with exemplary values

The tab `Specific (Picture)` (Figure 4-68) enables the user to describe the dataset from the data type `picture` with specific metadata details. Two mandatory properties are specified. The user has to enter details about the recording or creation period of the image. This includes a beginning and end date (e.g. 2011-04-18 – 2011-08-01). In addition, the recording place, where the picture was taken or that is displayed in the image (e.g. Merzenhausen, Germany) has to be specified. Furthermore, optional metadata details can be provided. Some details can be inserted using text boxes. For example, the recording method can be described by specifying the make and model name of a camera or a used software (e.g. Pentax Optio W30), the resolution of the picture in dots per inch (e.g. 300), the picture or image size (e.g. 3072 x 2304), as well as information about the copyright (e.g. Copyright 2011 TR32. All rights reserved.). The orientation of the picture can be chosen from the drop-down list (e.g. Horizontal/Landscape, Vertical/Portrait). Finally, information about the event, where the image or picture was taken, can be provided. This includes the type of event (e.g. Field Campaign, Project Meeting), which can be selected from a drop-down list. Furthermore, other details can be inserted via text boxes, like the name of the event (e.g. CRC/TR32 Field Measurement Campaign), the location (e.g. Merzenhausen, Germany), the event website (e.g. http://tr32.uni-koeln.de/index.php/monitoring), and the duration of the event (e.g. 2011-04-18 – 2011-04-18).

The tab `Specific (Publication)` enables the user to add specific metadata details with focus on the data type `publication`. This tab is an exception and is subdivided in two parts. The first section provides the input of general metadata details. The status of the publication has to be selected from the drop-down list (e.g. Published, Accepted, Forthcoming), as well as the review status of the publication (e.g. no, yes). In addition, the year of publication has to be provided. Finally, the type of publication hast to be chosen from the selective list. Depending on the selected publication type (e.g. Article, Book, Book Section, Event Paper), the subsequent, second part of the metadata input form will be adjusted. Consecutively, the metadata input forms for the various publication types will be described.

The tab `Specific (Publication)-Article` (Figure 4-69) gives an overview of the input form of specific metadata details for publication type `article`. All metadata details that are required for the accurate citation of an article are mandatory. Consequently, the user has to insert the name of the publication source, for example the name of journal or magazine (e.g. International Journal of Digital Curation). The publishing house has to be supplemented with optional city and country details (e.g. UKOLN). Finally, the page range of the article has to be specified (e.g. 68 – 80). In addition, optional metadata details can be added. For example, the type of article (e.g. Journal, Magazine, Newspaper), the number of pages covered by the article (e.g. 13),

the website of the publication source or journal (e.g. http://ijdc.net), as well as if available, the corresponding volume (e.g. 2) and issue (e.g. 7) can be provided.

Figure 4-68: Metadata input wizard tab `Specific (Picture)` with exemplary values

Figure 4-69: Metadata input wizard tab `Specific (Publication)-Article` with exemplary values

The tab `Specific (Publication)-Book` (Figure 4-70) provides an overview of the metadata details that can be added by the user for the publication type book. Again, mandatory metadata details, required for the complete citation of a book have to be provided. This include the name of the publishing house (e.g. Elsevier Science, University of Cologne, Institute of Geography) and the corresponding city and country, where the publisher is located, separated by a comma (e.g. Amsterdam, The Netherlands, Cologne, Germany). In addition, optional metadata details can be complemented. These include the title of the book series (e.g. Kölner Geographische Arbeiten) and the associated editor of the series (e.g. G. Bareth, H. Besler, B. Braun, …). Furthermore, the volume (e.g. 90) and edition (e.g. 1) of the book can be inserted, as well as the number of pages (e.g. 154) and the page range of the book (e.g. 1 – 154). All metadata details have to be inserted into a text box.

Figure 4-70: Metadata input wizard tab `Specific (Publication)-Book` with exemplary values

The tab `Specific (Publication)-Book Section` (Figure 4-71) offers on overview of metadata details, which can be specified by a user for the publication type `Book Section`. To accurately cite a book section, some metadata details have to be provided by the user. Mandatory properties are: the title of the book (e.g. Proceedings of the Data Management Workshop), the editors of the book separated by comma (e.g. Curdt, C., Bareth, G.), as well as the publishing house (e.g. Elsevier Science, University of Cologne, Institute of Geography) and the corresponding city and country, where the publisher is located, separated by a comma (e.g. Amsterdam, The Netherlands, Cologne, Germany. Furthermore, the range of the pages has to be defined (e.g. 27 – 33).

Figure 4-71: Metadata input wizard tab `Specific (Publication)-Book Section` with exemplary values

In addition, optional metadata properties can be offered by the user. These include the title of the series (e.g. Kölner Geographische Arbeiten), the corresponding editor of the series (e.g. G. Bareth, H. Besler, B. Braun, …), the volume of the book (e.g. 90), as well as an available chapter number of the book sections and the number of pages (e.g. 7). All metadata properties are provided by the user in a text box.

The tab `Specific (Publication)` (Figure 4-72) gives an overview of the metadata details, which can be provided by the user for the publication type `Event Paper`. Some details have to be provided to ensure that the publication can be cited accurately. The metadata input form for event paper is subdivided in two parts. The first part focuses on the metadata details of the corresponding event.

Figure 4-72: Metadata input wizard tab `Specific (Publication)`-`Event Paper` with exemplary values

The user has to provide the name of the event, where the paper was presented (e.g. XXII ISPRS Congress), and the corresponding location of the event, including the city and country, separated by a comma (e.g. Melbourne, Australia). Moreover, the duration of the event has to be provided including a start and end date by using the calendar fields (e.g. 2012-08-25 – 2012-09-01). Besides the mandatory event metadata, also optional metadata can be complemented. These are the website of the event (e.g. http//www.isprs2012.org) and the type of the event, which can be chosen from a selective list (e.g. Conference, Workshop, Project Meeting). The second part of the metadata input form facilitates the supplement of information about the available proceedings. The user can provide the editor(s) (e.g. Shortis, M., Wagner, W., Hyyppä, J.), the title (e.g. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences), and the website of the proceedings (e.g. http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net). In addition, information about the publishing house, volume (e.g. 1 – 7), issue, and the number of pages (e.g. 6) can be inserted into a text box. Finally, the range of pages (e.g. 7 – 12) of the paper is mandatory and has to be complemented in a text box.

Finally, after providing all required metadata details to the several wizard tabs and input fields, the user can complete the metadata input process by using the `submit` button. By submitting the dataset, the metadata input fields will be checked for completeness. In case, the submission was successful, the user will be

forwarded to the metadata overview webpage of the appropriate dataset. Otherwise, by submission of an incomplete or invalid input form, an error message will be displayed above the metadata wizard, as shown in Figure 4-73. In addition, all required mandatory input fields, which have to be completed, are highlighted in red.

Insert metadata for dataset 727

To submit your dataset successful to the TR32DB, please complete the invalid highlighted input fields and choice boxes.

file name: 2012_Curdt_ISPRS.pdf data type: Publication uploaded: 2013-12-02 13:59:42 size: 714 kB

| Template | Identification | Responsible Party | Topic | File Details | Constraints | Geographic | Specific (Publication) |

1. title *:   Main Title ▼ delete ⓘ

2. title: add ⓘ

Figure 4-73: Metadata wizard error message

A successful submission of the data also executes the automatic storage of metadata on the metadata (section 4.3 and 6D.3.1). This includes, for example, the storage of the web-interface version number used to upload the metadata details (e.g. Version 4.0) and the metadata language (e.g. English). In addition, the metadata creator and the date of the metadata creation and last update will be stored automatically.

After a successful submission of the metadata for a dataset, the creator of the metadata is able to modify these metadata. For this purpose, the user has to navigate to the metadata overview of the desired dataset and use the feature `Edit Metadata`. There are several options to access this feature. By navigating to the `User Home`, and selecting the function `View Own Metadata`, the user will receive an overview list of his datasets with corresponding metadata, which are similarly designed like in Figure 4-75. He is now able to select the desired dataset. In addition, the user can also use the various search features of the web-interface (section 4.5.2.2) to search for his desired dataset and access the features of the dataset via the metadata overview. By accessing the `Edit Metadata` feature (Figure 4-53), the user will be forwarded to the metadata input wizard, as previously described. In contrast to the metadata entry, the wizard will be completed with all available metadata of the dataset. The user is now able to modify, complement or delete metadata properties of the dataset. By re-submitting the dataset, the user will be forwarded to the updated metadata overview of the dataset. The user is able to repeat the modification of the dataset as often as he likes. By submitting the metadata to the database, all changes, made to the metadata, will be sustained as a precaution in the database. In addition, a successful submission of the metadata also updates some metadata details on metadata automatically. For example, the web-interface version number will be updated (e.g. Version 4.0), as well as the date of the metadata's last update will be modified.

DOI application

The application of a TR32DB DOI is available for datasets in the TR32DB, which have been successfully uploaded and described with metadata. The application of a DOI is only available for the metadata creator of the dataset or a person in charge of it. The general application process of a DOI was adapted from Version 3.1, modified, updated, and extended according to the requirements of the DataCite Metadata Schema Version 2.2 (section 2.2.2.2). As already described in detail section 4.5.1.3, the application of a TR32DB DOI has to follow six general steps (Figure 4-74). They will be described subsequently.
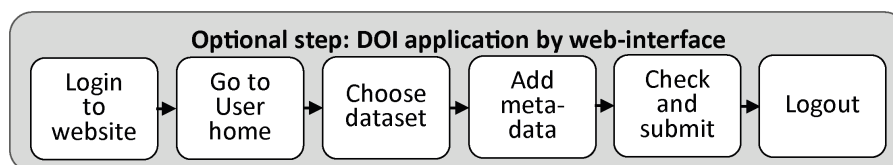
Figure 4-74: DOI application workflow

At first, the user has to navigate to the metadata overview of his desired dataset to access the dataset feature `Apply DOI`. Again, the user can search for his dataset in the TR32DB by using different search functions. In addition, the user can navigate to the `User Home` and access the `Apply DOI` function. The user will receive an overview listing of all datasets in the TR32DB without a DOI, of which the user currently logged in is the metadata creator (Figure 4-75). This list can be sorted as described in detail in section 4.5.2.2
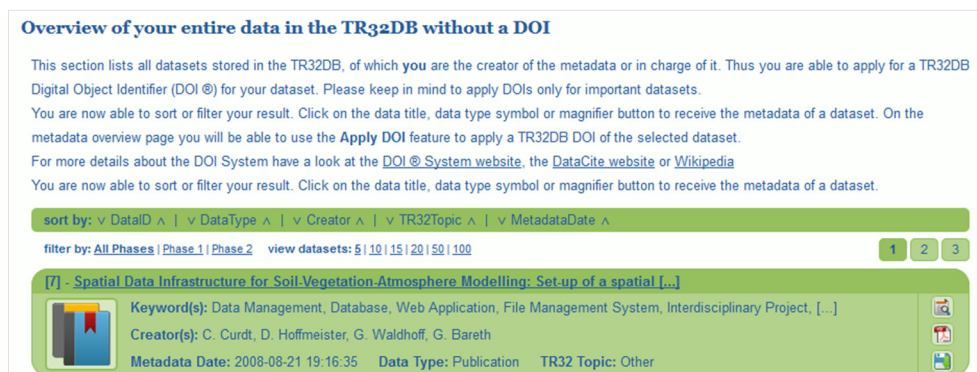
Figure 4-75: Overview of dataset without DOI of a current user

In addition, further information about the process of DOI application is provided. By selecting the desired dataset from the list, the user will initially receive the metadata overview and can select the feature `Apply DOI`. The user will be directly forwarded to a pre-filled metadata application wizard (Figure 4-76 to Figure 4-81) that already contains all available metadata of the dataset. The user is able to review, modify, and delete the available metadata in the metadata input fields. The DOI input wizard is designed in a similar structure and contains similar metadata properties like the general metadata input wizard described in the previous section. The DOI metadata input wizard is simplified and reduced to the metadata properties required by the DataCite Metadata Schema Version 2.2 (section 2.2.2.2). Besides metadata properties, also the obligation of the metadata properties was adjusted as well as some labelling of the metadata property names. The single wizard tabs will be described subsequently.

Initially, some general conditions and guidelines have to be accepted in the tab `Conditions` (Figure 4-76) to enable the DOI application with regard to the data applicant and TR32DB staff responsibility. This includes, for example, the responsibility of the data and metadata quality, administration, and provision.
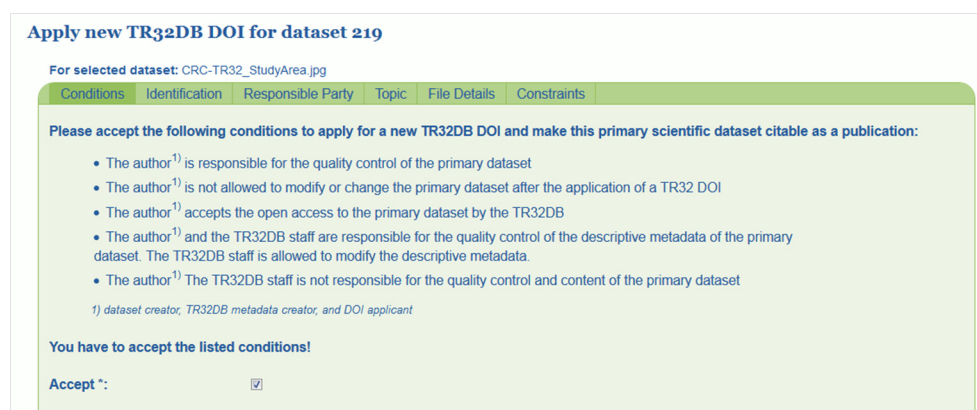
Figure 4-76: DOI application wizard tab `Conditions`

The second tab `Identification` (Figure 4-77) enables the specification of metadata details to identify the dataset. This was reduced to the mandatory property title and associated title type (e.g. CRC-TR32 Study Area Map|Main Title). In addition, optional metadata details can be provided. For example, a description and appropriate description type (e.g. Map of CRC-TR32 study area including sub-regions and major

measurement sites …|Abstract) and an alternate identifier with appropriate identifier type (e.g. 978-3-16-148410-0|ISSN) can be specified. Finally, a related identifier (respectively to relation in metadata wizard) may be supplemented, containing the name of the identifier, identifier type, and appropriate relation type (e.g. 10.5880/TR32DB.8|DOI|Is New Version Of). Multiple inputs of values are possible for all metadata properties in this tab.

Figure 4-77: DOI application wizard tab `Identification`

The third tab `Responsible Party` (Figure 4-78) enables the provision of metadata details related to the responsibility of the dataset. The mandatory properties creator (e.g. Hoffmeister, Dirk) has to be selected from the drop-down list or created using the `new` button, if the desired creator is not available in the listing. In addition, the publisher of the dataset has to be entered into the text box. The default value for the publisher is 'CRC/TR32 Database (TR32DB)'. Furthermore, an optional contributor with associated contributor type can be provided (e.g. German Research Foundation|Funder). Again, multiple inputs of values are possible for the properties `creator` and `contributor`.

Figure 4-78: DOI application wizard tab `Responsible Party`

The fourth tab `Topic` (Figure 4-79) enables the description of the dataset with optional subjects. They can be selected from various drop-down lists. The following listings are available: the CRC/TR32 keyword list (e.g. Catchment, Climate Station, EC), the Gemet-Thesaurus list (e.g. map, catchment area), the ISO/INSPIRE topic category list (e.g. GeoScientificInformation), and the DDC list (e.g. 550 Earth sciences). For the properties `CRC/TR32 keyword`, `Gemet-Thesaurus` and `DDC` multiple inputs of values are enabled.

Figure 4-79: DOI application wizard tab `Topic`

The sixth tab `File Details` (Figure 4-80) enables the specification of metadata details related to the file details and format of the dataset. All metadata properties are optional. A resource type (respectively data type in metadata wizard) can be selected form the drop-down list (e.g. Image) and can be complemented with a value (e.g. Map) in a text box. In addition, the amount of data and the corresponding data size type can be provided using a text box and a selective list (e.g. 1|Datasets, 4038|Kilobytes). A relevant date related to the dataset can be added including date type (e.g. 2010-09-06|Created). Finally, the primary language of the dataset can be selected from a drop-down list (e.g. English). Multiple inputs of values are enabled for the properties `size` and `date`.

Figure 4-80: DOI application wizard tab `File Details`

The tab `Constraints` (Figure 4-81) enables the provision of metadata details related to constraints of the dataset. Initially, the download permission of the dataset has to be selected from the drop-down list (e.g. Only TR32, Free). Recommended values for datasets with a DOI is 'Free'. Furthermore, an optional license value can be selected from the drop-down list (e.g. TR32DB - Data policy agreement).
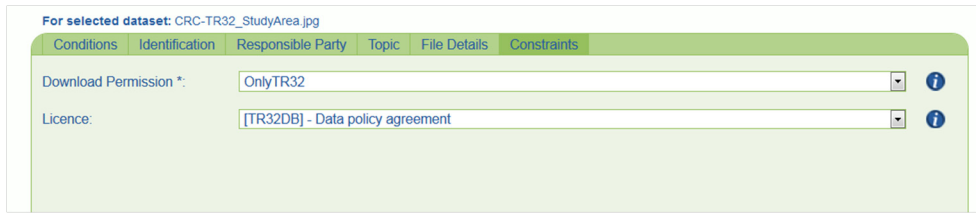
Figure 4-81: DOI application wizard tab `Constraints`

Finally, after completing all required metadata details in the metadata wizard, the user can finish the DOI application process by using the `submit` button. The provided metadata in the input will be checked for completeness. Incomplete or invalid details in the metadata wizard will execute an error message (Figure 4-73). All invalid input fields and selective lists will be highlighted in red (Figure 4-73) and have to be completed to successfully finish the DOI application. A successful application of the dataset will forward the user to the data metadata overview description of the applied DOI (Figure 4-54).

In addition, a successful submission of the metadata executes the automatic storage of metadata details. This includes, for example, the required metadata property publication year, the optional property format of the dataset represented as MIME type, as well as the applicant of the DOI. The DOI name will be assigned automatically (e.g. 10.5880/TR32DB.8) and can be manually modified later. Furthermore, the TR32DB admin will be informed about the application of a new DOI by an automatically generated email.

Finally, the user can view all datasets with a TR32DB DOI by using the function `View Own DOI`. This function is available by navigating to the `User Home` (previous section). An overview will be provided of all datasets stored in the TR32DB, of which the current user logged in has applied a TR32DB DOI. The overview is designed in a similar listing like Figure 4-75. Single datasets can be selected, sorted, and filtered, as well as the metadata overview can be accessed.

Temporary share

A temporary exchange of project data is provided by the TR32DB Share, already described in detail in section 4.5.1.3. The general structure and operating mode of the share sections was adopted from Version 3.1, modified, and adjusted to the new web-interface layout. To access the share, the user has to select the `Share` item from the `data menu bar` and will be forwarded to the overview website of the temporary share (Figure 4-82). This contains general information about the share and provides the download access to all data currently stored in the TR32DB data storage `share` directory (section 3.34).
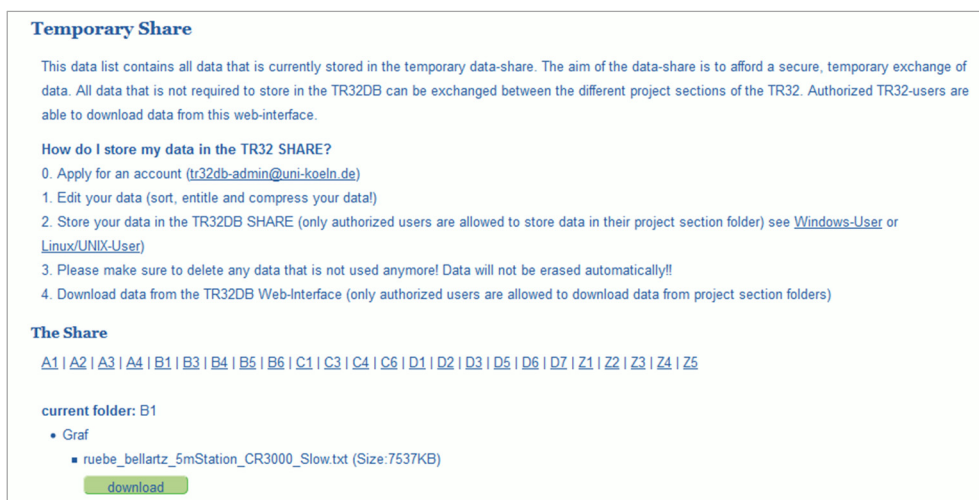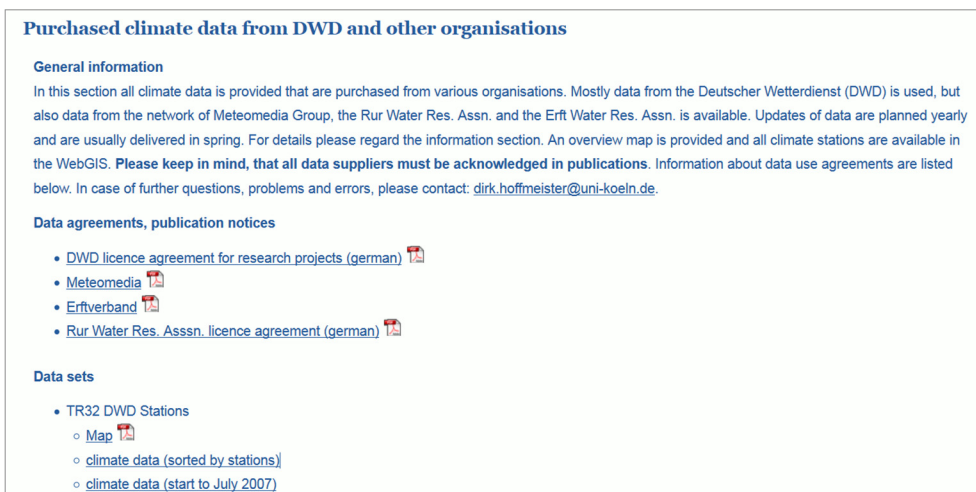
Figure 4-82: Access to project stored in the TR32DB data storage `share` directory

To download a dataset from the share, the user has to select the `download` button of the desired file. A download window will pop up, which enables the storage and download of the data. By accessing the download button, the download will be logged automatically in the database. Download details like the TR32DB user identifier, the IP address (Internet Protocol address), the storage path of the data in the TR32DB data storage including file details, and the download time will be saved.

Climate data section

The climate data section, as already described in detail in section 4.5.1.3 provides an overview of all available climate data from various institutions, which are available for CRC/TR 32 participants. The general structure and operating mode of the climate data section was adopted from Version 3.1. Just the layout of the entire features of the web-interface was adjusted and modified to the new layout of the web-interface Version 4.0 to access the climate data section, the user has to select the `Climate Data` menu item form the `data menu bar`. The user will be forwarded to the overview website of the climate data section (Figure 4-83). General information about the data is provided as well as download links to the purchased climate data.

Figure 4-83: Climate data section overview (partical view)

In addition, the operating mode of the climate data features `query climate stations data` and `query climate charts` provided for purchased DWD data were adopted form Version 3.1 and modified to fit the new layout of the web-interface (Figure 4-84 to Figure 4-85). For example, the layout of the feature `query climate station data` (Figure 4-84) was adapted to the other input forms of the web-interface Version 4.0, for example the combined search for data (section 4.5.2.2), the metadata input wizard (section 4.5.2.3) or the DOI metadata input wizard (section 4.5.2.3).

Figure 4-84: Climate data query with exemplary values

In addition, the structure and operating mode of the output website displaying the result of the feature `query climate station data` was adopted. The layout of the website was modified and adapted to the new layout (Figure 4-85).

Climate Data Query Result

This table displays the result of your query on climate data from DWD according to selected stations and time extent.

Back to Search   Abbreviatons   Export Data

Station	JA	MO	TA	HH	iX	h	VV	N	dd	ff	TT	TD	P0	1RRR	1RRRtR	ww	RR1	3SD24	TM	FG24	UU	VC	RR24	FD24	FL24
Aachen (WEWA)-10498	2010	6	18	0	2	250	3000	2	36	12	135	105	9916	0	1	-1	0	-99	-999	-999	82	-9	-999	-999	-999
Aachen (WEWA)-10498	2010	6	18	1	2	250	2500	2	31	13	128	107	9914	-999	9	-1	0	-99	-999	-999	87	-9	-999	-999	-999
Aachen (WEWA)-10498	2010	6	18	2	2	250	2500	1	31	13	119	105	9911	-999	9	-1	0	-99	-999	-999	91	-9	-999	-999	-999

Figure 4-85: Climate data query result with exemplary values

WebGIS Version 3

The WebGIS can be accessed by authorized TR32DB users form the `data menu bar`. By accessing the `WebGIS` item from the menu bar, the user will be forwarded to the TR32 Viewer (Figure 4-86: The TR32 Viewer Version 3Figure 4-86). The current version of the TR32 Viewer is using the latest Version 3.6 of ESRI's ArcGIS Viewer for Flex (section 3.6). The WebGIS Version 2, described in detail in section 4.5.1.3, was updated, further developed, and adjusted with new features. The WebGIS Version 3 still enables the visualization of basic geodata purchased from various institutions, as well as data provided by project CRC/TR32 participants (e.g. land use data). In addition, the representation of DWD data was extended.
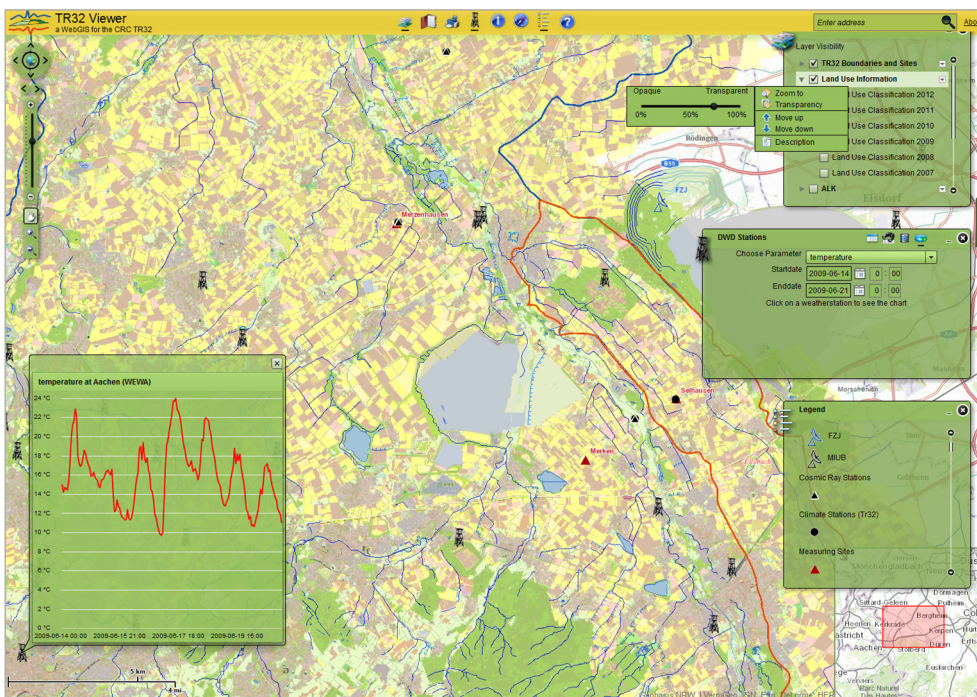
Figure 4-86: The TR32 Viewer Version 3

An example of the current WebGIS is presented in Figure 4-86, presenting the land use classification of the cur catchment from 2012 (LUSSEM & WALDHOFF, 2013) and the digital landscape model in the background. In addition, the climate stations are visible on the map, several measurement stations and sites, as well as various boundaries of the Rur catchment and sub-catchment. For the climate station Aachen (WEWA), a corresponding curve chart is represented of the parameter temperature from 2009-06-14 to 2009-06-21. The

presented parameter and the time extent are presented in the DWD Stations menu. In addition, the overview map is displayed in the background in the bottom right corner, the table of contents with layer features is displayed in the top, and a corresponding legend of current active layer symbols of the table of contents is visible.

The TR32 Viewer is presented color coordinated with the CRC/TR32 logo and TR32DB web-interface. The header controller of the TR32 Viewer represents the logo of the CRC/TR32, the application title, the widget tray, as well as a search feature and an about link. The `about` link provides information about the TR32 Viewer Version and the development team in a pop-up-window. The `search` provides the address search and the representation of the result on the map. Subsequently, the general navigation features and menu bar features (widget tray) will be described.

The mouse wheel enables zooming in and out of the map. In addition, the navigation on the map is provided by several navigation features. They will be described subsequently:

Zoom Slider The map zoom slider (Figure 4-86), presented on the left side of the map, enables zooming in and out. By holding and dragging up and down the circle, as well as by clicking on the plus or minus symbol, the scale will be changed. By clicking and holding the circle, the current scale will be displayed.

Pan Wheel The navigation pan wheel tool (Figure 4-86) enables the panning of the map to the right, left, up, and down by clicking on the appropriate arrow. By clicking on the globe in the center of the pan wheel, the map will be reset to the initial zoom state. The arrows below the pan wheel enable the zoom to previous and next extent.

Zoom Tools The navigation zoom tool (Figure 4-86) enables three features. The pan feature enables the moving of the map by clicking, holding on the map and subsequently moving the map to the desired map extent. In addition, the zoom in and out is enabled by clicking on the plus and minus symbol.

Scale The current scale of the map (Figure 4-86)is displayed at the bottom on the left side of the map in kilometers/meters and in miles/feet respectively.

Scroll and Close Sidebar This toll enables the movement of the single sidebar boxes and is located on the right side. They can be moved up and down, as well as closed by clicking on the appropriate triangle.

Overview Map The overview map (Figure 4-86) is located at the bottom right corner and can be expanded and minimized by using the appropriate arrow. The coherence between the current spatial extent displayed on the map (red rectangle) and the entire map is shown. By moving the red rectangle on the overview map, the map display will be moved to the desired extent.

The menu bar features (widget tray) are described subsequently:

Table of Contents The table of contents (Figure 4-86) lists all available layers. Different features are available for the layers. The visibility of the layer on the map can be changed by modifying the activity of the layer checkbox. By using the `Zoom To` feature, the map will navigate to the favored layer. The transparency of each layer can be changed by selecting the transparency feature of a layer. By moving the dot on the transparency slider to the favored percentage, the transparency of the layer will be modified. In addition, the location of the layer in the table of contents can be moved up and down in the listing. Finally, to receive further information about a layer, the `Description` feature can be used. By selecting this tool, the user will be forwarded to the associated metadata overview website of the TR32DB.

Bookmarks This tool displays a list of predefined map view extents. By selecting a bookmark, the map will show the bookmarked place in the chosen zoom level. The user is able to add new bookmarks, as well as delete unnecessary bookmarks.

Print This tool enables the printing of current map extent. A title and subtitle can be added for the output document.

DWD Stations The climate stations tool (Figure 4-86) enables initially the presentation of all DWD stations on the map. Different features are available. By moving with the mouse pointer over a climate station on the map, the corresponding details of the station will be displayed (e.g. station name, parameter, measurement time extent and gaps). Climate stations can be filtered by name. A result list will be displayed. By selecting a climate station from the list, the map extent moves to the favored climate station. In addition, a chart tool enables the drawing of a curve chart according to a chosen parameter and time extent. By clicking on a climate station, a corresponding chart will be displayed presenting the result.

Identify The identify tool enables the identification of layer characteristics. By clicking on the map, the attributes of each single visible layer in the table of contents will be displayed in a result list.

Mouse Coordinates This tool provides the latitude and longitude coordinates in of the current position of the mouse pointer.

Legend The legend (Figure 4-86) displays the used map symbols of the currently visible layers and an explanatory text. This dynamic legend will be updated automatically if the visibility of a layer changes.

Help The help feature provides a detailed description of all features and symbols used in the WebGIS (CURDT, 2014a).

Address Search This feature provides an address search. As a result, the map will zoom to the favored search location.

About Information about the TR32 Viewer like development team and used Version of ArcGIS Viewer for Flex are provided.

4.5.2.4 Administrative web-interface features

The administrative web-interface Version 4.0 is partially still in development and implementation phase. Consequently, only some functions are currently executable. The main task of the administrative web-interface is the management and maintenance of the TR32DB. The basic principles of the administrative web-interface were adapted from Version 3.1 (section 4.5.1.4). In contrast to the previous TR32DB Admin Tool of Version 3.1, the administrative features of Version 4.0 are directly incorporated in the internal `general menu bar` of the web-interface (Figure 4-87). The menu item `Admin Home` is only visible and accessible for users with administrative rights for the TR32DB. Three main feature groups are provided:

General Administration Administration and maintenance of general properties of the TR32DB including: user and news administration, as well as provision of information about current visitors of the web-interface.

Data Administration Administration and maintenance of data related properties of the TR32DB including: application of TR32DB DOIs, presentation of download details, administration of TR32DB data and geodata.

List Administration Administration and maintenance of selective lists used within the TR32DB (e.g. institution, instrument, keyword, location, and sub-project).
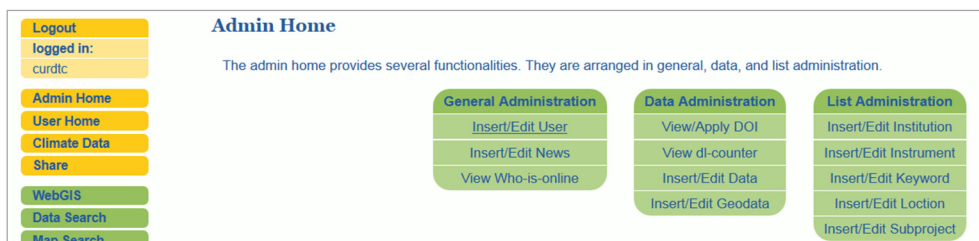
Figure 4-87: Incorporated Admin Home of the web-interface Version 4.0

Subsequently, the individual feature groups will be described.

General administration

The `General Administration` feature group includes three features, which will be described consecutively.

Insert/Edit User The administration of TR32DB users is managed by the `Insert/Edit User` features. A new TR32BD user can be set up with corresponding details, as already described in section 4.5.1.4. In addition, the details of a user can be modified by using the edit operation (Figure 4-88). User details can be modified like the RRZK user account, first name, family name, email-address, address data, as well as the membership in a CRC/TR32 project section and the corresponding funding phase.

Insert/Edit News This feature will administrate the news section displayed on the start page of the TR32DB. The general structure and operating mode of the previous from Version 3.1 will be adopted, modified, and adjusted to the new web-interface layout. This feature will be incorporated in the admin area soon.

View Who-is-online This feature gives an overview of current visitors of the web-interface. This feature will be incorporated in the admin area soon.

Figure 4-88: Modification of exemplary TR32DB user details

Data administration

The `Data Administration` feature group includes four features, which will be described consecutively.

View/Apply DOI The application of DOIs will enable a TR32DB administrator to manage the application of TR32DB DOIs, as already described in detail in section 4.5.1.4. The general structure and operating mode of the DOI application will be adopted from Version 3.1, modified, and adjusted to the new web-interface layout. This feature will be incorporated in the admin area soon.

View dl-counter This feature will enable the listing of all available details collected by the downloading a dataset from the web-interface. For example, available user details, the TR32DB

data identifier, downloaded file name and storage path, as well as the Internet Protocol address (IP address) of the user and the download date and time. All details will be accessed from the database table `dlcounter` (section 3.5.2). This feature will be incorporated in the admin area soon.

Insert/Edit Data This feature enables the administration of all TR32DB data provided by the TR32DB users. For example, all data stored in the TR32DB will be listed. The administrator user will be able to modify existing and insert new metadata for dataset. This feature will be incorporated in the admin area soon.

Insert/Edit Geodata The administration of data form the type geodata will be possible with this feature. The general idea of this section was adopted from Version 3.1 (section 4.5.1.4). The implementation of the geodata administration was set up likewise to features provided in the `User Home` for the input and viewing metadata of dataset, described in detail in section 4.5.2.3. The feature `Insert Metadata` enables the entry of metadata for data from the data type `Geodata`. The admin user will be forwarded to a metadata input wizard (Figure 4-89) that is arranged according to the general metadata input wizard described in detail in section 4.5.2.3. The wizard comprises the general metadata input tabs `Template`, `Identification`, `Responsible Party`, `Topic`, `File Details`, `Constraints`, `Geographic`, and is extended with a `Specific` tab referring to `Geodata`. The metadata input form is arranged according to the described metadata properties for the data type `Geodata` in the TR32DB Metadata Schema (section 4.3). Generals input fields are used in the wizard: text box, text area, select menu, and calendar.

Figure 4-89: Metadata input wizard tab `Specific (Geodata)` with exemplary values

The tab `Specific (Geodata)` (Figure 4-89) enables the administrator user to add specific metadata details with focus on the data type `geodata`. All metadata input fields are optional. Initially, the period of time covered by the content of the dataset (e.g. recording date of data, creation period of data) can be inserted.

This includes a start and end date, e.g. '2010-01-01'- '2010-12-31', which can be entered using a calendar field. A vertical extent can be added for the dataset. This includes an optional lowest and highest vertical extent in the dataset (e.g. 500 – 1000), as well as a corresponding vertical coordinate reference system. The later can be selected from a selective menu, for example 'DHHN92 height (5783)'. Consecutively, a statement can be provided in the lineage text area about the processing history or the overall quality of the dataset. For example, details about the quality assurance or validity of the dataset, legal validity, as well as details about versioning of the dataset can be described. The spatial resolution of the dataset can be added by adding a scale factor or ground distance. The scale can be added by expressing the scale dominator, for example 1: '5000'. The ground distance can be provided by expressing the distance value of the ground distance and selecting a corresponding measure unit of the ground distance (e.g. 20|cm). Subsequently, a spatial representation type can be selected from the drop-down menu (e.g. grid), which describes the method used to spatially represent geographic details in the dataset. A form of presentation can be chosen form the select menu, e.g. 'documentDigital', describing the mode in which the dataset is represented. In addition, the used spatial reference system in the dataset can be added by selecting a value form the drop-down list (e.g. ETRS89 / UTM zone 32N (25832)). Information about the maintenance of the dataset can be provided, describing the scope and frequency of dataset update. A value representing the frequency of maintenance and update can be selected (e.g. annualy, weekly). Furthermore, complementing notes on the maintenance can be added. Finally, a scope, for example 'dataset', can be chosen to describe the information class covered by the specified resource. The initiative type can be selected from the drop-down list (e.g. experiment, campaign, study), describing the framework in which the dataset was created. Multiple inputs of values are enabled for the properties `vertical extent`, `spatial resolution`, `spatial representation type`, `presentation form`, `reference system`, and `maintenance information`.

List administration

The `List Administration` feature group includes five features, which will be described consecutively.

Insert/Edit Institution The administration of the controlled list institution referenced to the database table `listInstitution` (section 3.5.2) will be provided. The administrator user will be able to insert new institutions and modify existing institution. This new feature will be incorporated in the admin area soon.

Insert/Edit Instrument The administration of the controlled lists instrument, parameter, and parameter unit will be possible. They are referenced to the database tables' `listInstrument`, `listInstrumentParameter`, `listInstrumentParameterUnit` (section 3.5.2). The general idea of this section was adopted from Version 3.1 (section 4.5.1.4), expanded, and updated. The feature will be available soon.

Insert/Edit Keyword The administration of the CRC/TR32 keyword list will be available. The database table `listKw` (section 3.5.2) will modifiable by using this feature. The general concept of was adopted from Version 3.1, modified, and extended (section 4.5.1.4). The described feature will be available soon.

Insert/Edit Location The administration of available values in the controlled list location will be facilitated. The database table `listMeasureLocation` (section 3.5.2) is the corresponding table. This new feature for location administration will be forthcoming soon.

Insert/Edit Subproject The administration of project sections details includes the input of new sub-projects with corresponding details, as well as the modification of available details. This feature will be provided soon and is based on a general idea from Version 3.1 (section 4.5.1.4), will be adjusted, and updated.

4.5.2.5   Links to the official TR32 web-interface

The TR32DB web-interface Version 4.0 and the official TR32 web-interface (http://tr32.uni-koeln.de/) have increased the amount of links since the launch of the new official TR32 web-interface in March 2014. This will be subsequently described using the examples of two applications.

The presentation of TR32 Publications at the TR32 web-site (CRC/TR32, 2014a) is automatically generated by accessing metadata details of the TR32DB (Figure 4-90).

Figure 4-90: Presentation at TR32 web-interface of TR32 publication references, which are stored in TR32DB

Figure 4-91: Presentation at TR32 web-interface of project sections with linkage to TR32DB. Exemplary of project section Z1: (A) linking to publications and (B) all data, which are stored in the TR32DB. For project section Z1, a further linking to the TR32DB in provided (C)

All publications stored in the TR32DB are presented by their reference sorted according to the publication year (e.g. 2014, 2013), TR32DB publication type (e.g. article, book section), and the family name of the first author. Basically, the structure of the citation advice is currently following the reference style of Elsevier (e.g. ELSEVIER, 2014). This can be modified according to the requirements of the CRC/TR32. By selecting a

project section at the left site of the website, the visitor will be forwarded to the TR32DB. A search query will be displayed showing all publications of the selected project section. For example, by selecting project Z1 in the TR32 website, the result on the TR32DB website will present all Publication(s) of Project Section Z1 (Figure 4-91).

The overview of a project section at the TR32 web-interface will display general information about the project and members as well as link to the corresponding publications and project data of the selected project. For example, Figure 4-91 presents the linkage of the project section Z1 to their corresponding publications stored in the TR32DB as well as all project data stored in the TR32DB. Since Z1 is the project responsible for the TR32DB, an additional link is provided to the general TR32DB web-interface.

.

5   Discussion

Research data management (RDM) comprises all relevant processes and measures which ensure that research data, created in the context of a research project or institution, are well-organised, preserved, backed up, documented, accessible, and re-usable (CORTI et al., 2014; ENKE & FRITZSCH, 2013). The management of the research data should ideally take place directly during the entire research data life cycle of a research project (section 1), instead of an archiving or management of the collected data at the end of the project or research process (WEICHSELGARTNER et al., 2011). The importance of sharing and managing, which is a part of the research data life cycle, as well as the open access to research, data is mandated by national and global research funders, policy makers, and initiatives (section 2.1.2). For this purpose, RDM systems or repositories form the technical basis to archive research data, enrich them with further details (e.g. documentation with metadata), and, finally, provide them for (re-) use (RAZUM, 2011).

In order to manage all data, created within the framework of the CRC/TR32, the RDM system TR32DB was designed and implemented according to predefined demands and requirements (section 4.1). This was described in detail in the previous chapters. This section will discuss the system. Initially, a critical evaluation of the TR32DB is provided (section 5.1). Subsequently, a comparison of several RDM systems with the TR32DB approach is presented in section 5.2. Finally, the transferability of the TR32DB approach to another interdisciplinary research project is described (section 5.3).

5.1   Evaluation of the TR32DB

The TR32DB is a project database or RDM system which is developed to collect, document, store, backup, and provide all data created in the research life cycle of the CRC/TR32. The TR32DB was designed to cover all demands and requirements of the CRC/TR32 (e.g. heterogeneous data from different research disciplines with different file size) and of the DFG (e.g. cooperation with a computing centre, storage of data for at least 10 years) (section 4.1). In addition, demands on the metadata specifications (section 4.1), as well as general challenges of RDM with a technical, socio-cultural, and ethical nature (section 2.1.1) have to be considered. The TR32DB enables the accurate description of all data with accurate metadata, supports a variety of data formats, facilitates sharing of data, provides the registration of DOIs, and visualizes geodata in a WebGIS. This also corresponds to the general statement of RAZUM (2011) that RDM systems have to be designed according to arising requirements which result from the field of application and the data that have to be managed. In addition, general demands have to be considered, like the linkage of data and metadata, variety of data formats, registration of data identifiers, access management, as well as demands of the funding organisation. Corresponding features should be implemented to meet the demands (RAZUM, 2011).

For the technical implementation of the TR32DB system, a self-designed system was set up, using a mixed form of several components organized in a three-tier architecture (client-server) (section 3.2). Besides the design of a self-developed metadata schema (section 3.3 and 4.3), a file-based data storage was set up (section 3.4 and 4.4) for storing the data, as well as a database-based data storage for managing the corresponding metadata and the user management (section 3.5). Finally, a web-based graphical user interface (section 3.6 and 4.5) was developed to enable browsing, searching, as well as downloading of the data according to the given access rights. Likewise, in other interdisciplinary research projects also self-designed systems are established for RDM which are based on a client-server architecture (LOTZ et al., 2012; NADROWSKI et al., 2013; ZANDER et al., 2013). In addition, RAZUM (2011) determines that in the practical use of RDM systems, usually mixed forms occur. For example, file-based storage (e.g. provided by a digital object repository) is used in combination with relational databases, while the database stores the metadata and the data files are stored in the file system.

As described in section 2.4, several *out-of-the-box* software solutions are available to set up a RDM system. Usually, these solutions are applied by institutional repositories like the Dryad Digital Repository, Datorium (GESIS), DataShare (University of Edinburg), DataBank (University of Oxford) or ePrints Soton (University of Southampton) (section 2.5). Furthermore, it is possible to combine existing RDM solutions with self-developed components. For instance, WILLMES et al. (2013) present an approach to combine various components including the *out-of-the-box* RDM software tool CKAN (CKAN, 2014b). As stated by CORTI et al. (2014), some software solutions, like the open source repository software DSpace or Fedora, are used by single collaborative institutions. However, the implementation of these systems demands significant maintenance and set-up costs. Moreover, CORTI et al. (2014) determine that such repositories are more suitable for research data storage after the end of a research project. Consequently, the use of such a solution was not adequate for the TR32DB, since the system has to operate during the project term.

The demand and requirements on a RDM system change in the course of the project. Consequently, the RDM system has to be adapted according to the needs (MÜCKSCHEL et al., 2005). Hence, also the TR32DB had to be modified in the last years. Initially, the next section will focus on the evaluation and comparison of the previous TR32DB Version 3.1 and the current Version 4.0. Subsequently, the established main components of the TR32DB will be discussed.

5.1.1   TR32DB Version 3.1 versus TR32DB Version 4.0

It is an experience of many interdisciplinary projects that in the course of the project a modification of the RDM system is required, depending on altered demands and requirements arising from the project or funding organisation, as well as technical developments (MÜCKSCHEL et al., 2005). Consequently, MÜCKSCHEL et al. (2007) determine that it is only feasible to implement a prototype of the RDM system on the basis of the present, existing requirements at the beginning of a long-term research project. This prototype will be further developed and adjusted with the specifications of the user and project demands during the project duration. In terms of the implementation of the TR32DB, the same experiences were gained. Initially, the TR32DB was set-up in 2008, further developed, and complemented with new features up to now. Thus, this section provides a comparative overview about main system components and features of the TR32DB Version 3.1 and the further developed TR32DB Version 4.0, described in detail in the previous sections. An outline of the main characteristics of both system versions is presented in Table 5-1.

A significant change from Version 3.1 to Version 4.0 was conducted in terms of the metadata schema. Version 3.1 did only support compliance to basic properties of Dublin Core and DataCite, as well as support of a few necessary elements of ISO 19115 and INSPIRE. In contrast, the complete metadata schema was updated and extended in Version 4.0, as well as documented in detail (section 4.3). In the context of the data storage, no changes were made. With the significant changes of the metadata schema, consequently also modifications and adjustments of the database had to be conducted (section 3.5). The historically grown database structure of Version 3.1 was replaced and reorganized in a simple, extendable structure including linkage of the database tables with foreign keys. Moreover, the labelling of the database tables was revised and optimized. With regard to the web-interface, some changes were conducted (section 4.5). The programming of the web-interface Version 3.1 was based on the Smarty PHP Template Engine. Since the introduction to Smarty for new TR32DB admins was very time-consuming, in the course of the web-site update and redesign, a solution was chosen for Version 4.0 with common web programming, markup, script, and query languages. Besides the layout change from Version 3.1 to Version 4.0, some features were also updated and complemented. For example, a map search was supplemented on the basis of the Google Maps API. Another major modification was conducted with regard to the representation of the research result and the presentation of the metadata. While the metadata details of a single dataset in Version 3.1 were illustrated in a pop-up-window, in Version 4.0 metadata of a single dataset are presented in an additional webpage with a well-arranged metadata view, an optional map, and a permanent URL. In Version 4.0 associated features of a

dataset are directly available on the metadata webpage as well as accessible at the user home. Only the latter was supported in the Version 3.1. Another significant change from Version 3.1 to Version 4.0 is the metadata input. This was based on a long listing form in Version 3.1 and was rearranged to an intuitive, easy-to use, user-friendly metadata input wizard in the new version. An input wizard was also implemented for the DOI application. Moreover, features in the WebGIS were added and updated from Version 3.1 to Version 4.0. Finally, another relevant change was conducted with regard to the TR32DB administration area, which was located on a separate website in Version 3.1 and integrated in the TR32DB web-interface in Version 4.0. Thus, the TR32DB was updated due to many expansions in the previous versions, caused by new demands and requirements of the CRC/TR32, as well as modifications of technical nature (e.g. metadata standard enhancements) (section 4.1). Having discussed the main changes between the two versions, the main components of the system will subsequently be discussed.

Table 5-1: Main characteristics of TR32DB Version 3.1 versus Version 4.0

	TR32DB Version 3.1	TR32DB Version 4.0
Metadata Schema	- Support of Dublin Core, DataCite Version 2.0, elements of ISO 19115 & INSPIRE	- Compliant with Dublin Core, DataCite Version 2.2, ISO 19115, INSPIRE - Metadata Schema is documented
Data Storage	- Andrew File System - Organized folder structure	- Andrew File System - Organized folder structure
Database	- 46 tables, without foreign keys - Historically grown structure	- 115 tables, related with foreign keys - Simple, extendable structure - Grouping of tables (md, data, list-tables)
Web-Interface	- Based on Smarty PHP Template Engine and common web programming and markup languages (e.g. PHP, HTML, CSS) - Search for data (predefined and advanced) - Presentation of data in a result table with basic metadata information of a dataset - Metadata view in pop-up window - Access to dataset features only through user home - Metadata input/edit in a list form - DOI application in a list form - WebGIS 2 with basic features - Admin interface accessible via a separate website	- Usage of common web programming and markup languages (e.g. PHP, HTML, CSS) - Search for data (predefined and complemented advanced search) - Map Search - Presentation of data in a result table with basic metadata information of datasets - Metadata view on a single website with an optional map, and a permanent URL - Access to dataset features through user home and metadata overview website - Metadata input/edit in a wizard form - DOI application in a wizard form - WebGIS 3 with extended features - Admin interface integrated in the main website

5.1.2 Discussion of the TR32DB components

In the previous section, the differences between the TR32DB Version 3.1 and Version 4.0 were discussed. This section will now focus on the evaluation of the single TR32DB components of Version 4.0. After a short overview of the main advantages and disadvantages of the single components, they will all be discussed in detail. At first, the TR32DB Metadata Schema will be described. Afterwards, the focus will be on the data storage based on a file system, followed by a discussion of the database. Finally, the web-interface will be evaluated.

Initially, a short overview is provided with regard to the main advantages and disadvantages of the single TR32DB components (Table 5-2). They will be discussed in detail.

Table 5-2: Main advantages and disadvantages of the single TR32DB components

	Advantages	Disadvantages
Metadata Schema	- Conformity to Dublin Core, DataCite Version 2.2, ISO 19115 core, INSPIRE - Multi-level approach, simplified to general and specific metadata - Support of mandatory, optional, and automatic metadata properties - Design according to project specific demands - Appropriate metadata for different data type - Application of controlled vocabulary lists - Easily extendable with new metadata properties (e.g. new data type, controlled list or values)	- Complexity - Need for Metadata unclear to new users
Data Storage	- Hosted and administrated in RRZK - File-based data storage in AFS - Organized folder structure, which is sustainable, scalable, secure, backed up	- RRZK user account is mandatory to access AFS for data storage - Application of a further software (e.g. WinSCP) to upload data - Metadata input at the earliest possible one day after data upload
Database	- Hosted and administrated in RRZK - MySQL is common in combination with web-interfaces - Simple, extendable database structure - Metadata storage and web-interface administration details	- Only metadata storage, no storage of *real* data possible
Web-Interface	- Hosted and administrated in RRZK - Self-developed web-interface with common web programming, markup, and script languages (e.g. PHP, HTML, CSS, JavaScript) - Short training period for website programming and independent of software updates and security vulnerability - Individually adaptable according to demand - Support of common RDM system features: data search, metadata view, data download, permission management - User-friendly metadata input in a wizard form inclusive template feature, additional description with PDF - Simple DOI application via wizard form including prefilled available metadata - Integrated WebGIS	- Not as stylish as a CMS - No quick integration of out-of-the-box features like in a CMS - No upload of data by web-interface - Metadata input for associated datasets at the earliest possible one day after data upload

TR32DB Metadata Schema

The documentation of research data is an essential duty in the context of RDM. To ensure that the research data can be (re-)used and interpreted, the data has to be well documented (WEICHSELGARTNER et al., 2011). Consequently, sufficient information in form of metadata has to be provided, which also supports the analysis of the data irrespective of the data producer (BÜTTNER et al., 2011). As stated by RAJABIFARD et al. (2009), metadata are essential for data with a spatial reference. They provide the user with information about its accuracy, purpose, actuality, or quality. Moreover, HARVEY (2011) highlights that it is much cheaper to

provide accurate metadata of the data immediately. The subsequent improvement of incomplete metadata afterwards is much more expensive. Consequently, a well-designed metadata schema is essential.

In the context of the TR32DB RDM system, a self-developed metadata schema was designed. Likewise, other interdisciplinary research projects or institutions, which store cross-disciplinary research data, have also created a metadata schema (GREENBERG et al., 2009; LINNE, 2013; TAKEDA et al., 2010). A detailed documentation of the TR32DB Metadata Schema (appendix D) provides supplementary notes about the included single metadata properties and sub-properties. The style of the documentation follows recent metadata schema documentations introduced by VIERKANT et al. (2012), DATACITE METADATA WORKING GROUP (2013) or HAUSSTEIN et al. (2013).

The purpose of the TR32DB Metadata Schema is to describe all research data stored in the TR32DB with accurate and interoperable metadata according to the demands and background of the CRC/TR32. All data stored in the TR32DB, are assigned to different data types (e.g. data, publications, reports). Thus, the TR32DB Metadata Schema is designed in a multi-level approach (section 3.3 and 4.3) and covers (i) general metadata properties (e.g. title, creator, description) to describe all data with basic information, (ii) metadata properties with focus on the CRC/TR32 project background (e.g. specific keywords or topic), as well as (iii) specific metadata properties according to the chosen TR32DB data types (e.g. publication articles can be described with their corresponding journal name, issue, and volume). A multi-level metadata approach, for instance, also supported by the institutional repository of the University of Southampton. Their developed metadata model supports the three levels: core metadata, discipline metadata, and project metadata (TAKEDA et al., 2010).

This TR32DB metadata approach combines metadata properties or elements of different approved metadata schemas and standards, like the DCMI Metadata Terms (chosen as a basis), the DataCite Metadata Schema, as well as core elements of ISO19115:2003 Metadata Standard and INSPIRE. As stated by BACA (2008: 72), "[…] there is no 'one-size-fits-all' metadata schema […]" available. Furthermore, SIMONS & RICHARDSON (2013) determine that in a particular RDM system or data repository which supports cross-disciplinary research data have to support a multiplicity of metadata standards and schemas to cover all different data types. In addition, due to a great variety of research data, it is not possible to apply a single metadata schema for all types of data. Consequently, it is necessary to modify existing metadata schemas and standards according to the specific needs (QIN et al., 2012). Prominent examples, which support and combine multiple metadata schemas, include the Dryad Metadata Application Profile (DRYAD DEVELOPMENT TEAM, 2010; GREENBERG et al., 2009; WHITE et al., 2008) and the Datorium metadata schema (LINNE, 2013; WIRA-ALAM et al., 2012). Both also apply the Dublin Core as a basis for their metadata profile or schema, similar to the TR32DB Metadata Schema.

Moreover, the TR32DB Metadata Schema comprises optional, mandatory, and automatically generated metadata properties. The support of optional and mandatory metadata properties is often used in metadata schemas, for example the DataCite Metadata Schema (DATACITE METADATA WORKING GROUP, 2011) or ISO 19115:2003 (ISO, 2003). Automatically generated properties are not very often documented in metadata schemas, but, for example, the Dryad Metadata Application Profile (DRYAD DEVELOPMENT TEAM, 2010; GREENBERG et al., 2009; WHITE et al., 2008) includes automatic and semi-automatic metadata properties. In the TR32DB Metadata Schema only metadata properties which are of major importance and requested for an accurate description or citation of the data are mandatory. This should reduce the overburdening of the TR32DB users. The previously described, multi-level approach has the advantage that the majority of the metadata properties are the same for all data types and only some properties have to be complemented according to the specific data type. Furthermore, both general and specific metadata properties can easily be extended with new metadata properties according to the demands of the single data types or the CRC/TR32 project participants. Likewise, it is very simple to add a new data type to the TR32DB Metadata Schema, such as data services or additional publication types. In this case, only a new specific data type has to be added to the schema. In contrast, the Dryad Metadata Application Profile consists of modules. The single modules

describe the metadata properties of the supported data types (e.g. publication, data package, data file) (DRYAD DEVELOPMENT TEAM, 2010).

The TR32DB Metadata Schema supports self-established controlled vocabulary lists. Some are based on supported metadata schemas (e.g. DataCite, ISO 19115:2003) and are mandatory for the conformity to a specific metadata schema. In addition, self-created controlled vocabulary lists are set up, according to the needs and demands of the CRC/TR32 and data types. Self-established controlled lists (e.g. supplementation of new measure locations or keywords) can easily be complemented, when required. In addition, the application of controlled lists has the advantage to avoid spelling mistakes as well as the unification of provided values. The selection of predefined values is also much easier and comfortable for the user, instead of inventing new values. For instance, controlled vocabulary lists are available in the DataCite Metadata Schema (DATACITE METADATA WORKING GROUP, 2011, 2013), Re3Data Metadata Schema (VIERKANT et al., 2013) or ISO 19115:2003 (ISO, 2003).

The TR32DB Metadata Schema also provides a mapping of the different metadata schemas (appendix D.4). This enables the linkage to the used metadata standards and provides interoperability. For example, the integrated mapping to the DataCite Metadata Schema facilitates the simple application of DOIs for a dataset. In addition, it is possible to complement further metadata schemas to the TR32DB Metadata Schema, which could be of interest to the CRC/TR32. Metadata mappings are also provided by other metadata schemas. For example, a mapping of Datorium metadata elements is described by WIRA-ALAM et al. (2012).

During the course of the project, the TR32DB Metadata Schema was adjusted according to the demands and requirements of the project, as well as further developments of applied metadata schemas or standards, for example the DataCite Metadata Schema was updated in 2011 and 2013 (DATACITE METADATA WORKING GROUP, 2011, 2013). It is also an experience of other research projects that a defined metadata schema has to be modified, improved, expanded or adapted in the course of the project according to changing requirements of the project members or project data (DRYAD DEVELOPMENT TEAM, 2010; GREENBERG et al., 2009; MÜCKSCHEL et al., 2004; MÜCKSCHEL et al., 2007). Overall, the TR32DB Metadata Schema allows a user-friendly, data type specific, interoperable, easily expandable metadata description of all data stored in the TR32DB RDM system.

TR32DB data storage based on file system

The sustainable storage and backup of research data is one important component of a RDM system. According to RAZUM (2011), a file-based data storage is mainly used for long-term archiving, in contrast to a database-based approach, which supports interactive processing of the data. In the framework of the TR32DB system, a file-based data storage was set up in cooperation and also hosted at the Regional Computing Centre of the University of Cologne. The collaboration with a computing center or library is recommended by the DFG (DFG, 2013d). Likewise, WEICHSELGARTNER et al. (2011) determines with regard to research data that the data storage in a data or computing center offers significant advantages, such as a possible larger amount of data can be managed in contrast to a local server.

The TR32DB uses the Andrew File System (AFS) (section 3.4), hosted and administrated by the RRZK, which is used as a data storage for all TR32DB data. The AFS has already served as an archive storage area for several other research projects in the RRZK (RRZK, 2013a). For example, the AFS is used in the CRC806-Database RDM system as the file-system, also hosted at the RRZK (WILLMES et al., 2014). With the experiences gained in the previous projects in the AFS, a comprehensive folder structure was set up in cooperation with the RRZK according to the demands of the CRC/TR32. This structure enables the storage and backup of all data formats and supports single file sizes up to 8 GB, which is the limit of one folder defined as a default value by the implementation of the folder structure. The storage of large, singles files (few kilobytes to several gigabytes) is a demand of the CRC/TR32 (section 4.1). Furthermore, this well-engineered storage structure provides an organisation of the data and can also enable the access to the data by a manual

recursive search. Furthermore, the data is accessible by a web-interface. Moreover, the folder structure is highly scalable and enables an easy administration by Access Control Lists (ACL). As stated by RAZUM (2011), the main advantages of a file-based RDM system is the support of a number of file formats, the scalability, and security. These benefits are also highlighted by CAMPBELL (1998) and MILICCHIO & GEHRKE (2007). In addition, the TR32DB file-based storage structure is worldwide accessible by various operation systems, like Microsoft Windows or Linux, also described by MILICCHIO & GEHRKE (2007) and CAMPBELL (1998).

The TR32DB folder structure provides a temporary data storage of the data files (Upload-folder) and a final data storage (Phase-folder). This separation of the folders prevents an erasure of the data, which is only permitted to the TR32DB administrator. The application of a script to move the data from the temporary upload-folder to the final data storage folder enables a structured, consistent filling of the single sub-folders (e.g. 000, 001). Moreover, an automatic generated linkage to the TR32DB database is created.

Nevertheless, the application of the AFS has also disadvantages. To access and upload the TR32DB structure, a user account of the RRZK is necessary. Consequently, all CRC/TR32 participants who are not employed at the University of Cologne have to apply for a RRZK University of Cologne account. As described in section 5.3, this can be avoided by integrating the user management into an existing infrastructure and restricting the file upload limit. Furthermore, another disadvantage of the introduced TR32DB data storage is the temporal postponement of the data upload. The input of corresponding metadata is possible one day after the data upload at the earliest, due to the moving of the data files from the temporal to the final data storage. Furthermore, Microsoft Windows users have to use an additional application which supports the Secure File Transfer Protocol (SFTP), like FileZilla or WinSCP (section 4.4), to upload their data to the TR32DB data storage. As described in section 5.3, this can be avoided by integrating the file upload into the web-interface, which causes a restriction of an upload file size of 50 MB (PHP restriction of RRZK server).

Overall, the designed TR32DB file-based data storage structure could have also be implemented in another file system (section 3.4) which offers similar characteristics, like scalability, cross-platform access, backup, and administration of access authorization. However, the AFS is for the TR32DB approach the preferable solution that is provided by the RRZK.

TR32DB data storage based on database

The core component of the TR32DB is the database (section 3.5). The main duty of the database is the storage of the metadata associated to the data files, as well as the handling of the administrative details of the TR32DB system. For this purpose, a database was set up in the relational database management system (RDBMS) MySQL, which is hosted at and administrated by the RRZK. The application of relational databases to store metadata is often used in RDM systems in combination with file systems or digital object repositories (RAZUM, 2011). In addition, the use of MySQL in combination with web-applications is very popular. As stated in the MySQL Whitepaper published by Oracle (ORACLE, 2013c), MySQL is optimized and designed for web applications in contrast to other relational databases. Consequently, it is used very frequently in web-based applications. For example, MySQL is deployed by web-sites with the most traffic in the Internet, such as Facebook, Google, Twitter, Wikipedia, and YouTube (ORACLE, 2013a). According to VASWANI (2010) and DUBOIS (2013) MySQL, as an open source software offers many unique characteristics, like speediness, scalability, simple and easy usability, support of SQL, multi-threading, security, full connectivity via Internet, and platform portability.

So far, the TR32DB database only manages the metadata of the associate data stored in the AFS, and administrative details. Except for some climate data purchased by the DWD, which are stored in a climate database based on MySQL (section 3.5), no *real* data are stored in the TR32DB database. In contrast, other RDM systems store their data in database systems, like PostgreSQL or Sybase (DIEPENBROEK & GROBE,

2007; KUNKEL et al., 2013; ZANDER et al., 2013). This is also very common in biodiversity (KATTGE et al., 2011; LOTZ et al., 2012; NADROWSKI et al., 2013). Overall, the designed TR32DB database structure could have also been implemented in another RDBMS like PostgreSQL or DB2.

TR32DB web-interface

The access to a RDM system is nowadays mostly realized by a web-based graphical user interface (LINNE, 2013; LOTZ et al., 2012; MÄRKER et al., 2011; NADROWSKI et al., 2013; WILLMES et al., 2014; ZANDER et al., 2013). In the context of the TR32DB, a user-friendly web-based graphical user interface was implemented accessible since 2008 at http://www.tr32db.uni-koeln.de or http://www.tr32db.de, and since then continuously adjusted and expanded with new features. The web-interface is connected to the file-based data storage (section 3.4 and 4.4) and the database (section 3.5). The latter is responsible for the majority of dynamic generated website content (e.g. view of metadata details, news), as well as the management of administrative details (e.g. user information and their allocation to a project section). Subsequently, the technical implementation, the web design, as well as the features of the web-interface Version 4.0 (section 4.5.2) will be discussed.

The TR32DB web-interface Version 4.0 is individually self-developed by means of common languages, like PHP, HTML, CSS, as well JavaScript (section 3.6). This ensures a short training period for new TR32DB administrators, instead of a long training period, which is often necessary for websites that are based on common comprehensive CMS systems, like Typo3 or Joomla!. In addition, since the self-developed web-site is based on a code especially written for a specific purpose, a suitable adaptability according to new specific demands is feasible. In addition, a self-developed website is independent of software updates or security vulnerability, which often affects common CMS systems (e.g. HEISE, 2014). In contrast, it is not possible to quickly use predefined, out-of-the-box features or components, such as a calendar or image gallery. Furthermore, self-developed websites provide very often less functionality and are not as stylish as CMS-based websites.

The layout of the TR32DB web-interface is set-up in several components, as described by BEAIRD (2011) (section 3.6.). The static container includes the identify block with CRC/TR32 logo and website title on each page to facilitate the identification of the single webpages. The simple designed navigation on the top and left side ensures the guidance of the users with regard to specific details, like contact information or background information. The content is located in the center. The footer at the bottom of the webpage includes contact details and logos of CRC/TR32 member institutions and DFG. Finally, also the whitespace component is considered to loosen the different containers of the webpage. The colour scheme of the website is arranged in mainly yellow and green colours according to the CRC/TR32 logo.

Recommendations in terms of barrier-free access to the website, as well as usability and accessibility (section 3.6), were also considered while designing and developing the web-interface. For example, the TR32DB web-interface supports a significant title, meaningful headlines organized in a hierarchy, as well as external links on the website are marked with an appropriate hint. The web-interface was designed and implemented according to the target group. The input forms were arranged in a user-friendly and appealing manner. This follows the recommendations of POOKER (2009). Furthermore, suggestion of KLUG (2006) were incorporated, like the "Save me steps wherever you can." (KLUG, 2006:166) rule or the set-up of Frequently Asked Questions (FAQs). In addition, further guidance to use the web-interface is provided by a comprehensive web-interface tutorial (CURDT, 2013c). Many hints including examples are represented by an " i "-symbol, in particular for the input of metadata in the wizard.

The TR32DB web-interface provides several features. Some will subsequently be discussed. General features, available to all visitors of the TR32DB, include the search for data. Different search options are available. For instance, the search according to predefined search (e.g. Figure 4-43) queries (e.g. search for all data of the type `publication`). This search provides an easy and quick overview of all available data of the selected

specific category. In addition, an advanced search (Figure 4-47) enables the combination of various search categories such as title, creator, and data type. This search provides a more qualified and limited search results. This kind of search is very common and often used by other websites. Furthermore, a search for data on a map (Figure 4-48) is offered for all datasets which own geographic coordinates in their metadata details. This search is based on the Google Maps API, which is very common and consequently also user-friendly. So far, the data are only located by their centre coordinates. This will be adjusted in the future by a bounding box. Similar approaches for map-based data search are provided by the CRC806-Database system (WILLMES et al., 2014) and the BrahmaRBIS system (ZANDER et al., 2011).

In addition, all visitors of the TR32DB web-interface are able to view the result of the search, as well as sort and filter it. All search results, with the exception of the DOI-section, will display a list of datasets. For each dataset, main details, like the title or author are presented, as well as a logo to easily identify the data type. Moreover, the main features of the dataset can be accessed immediately. This corresponds to KLUG (2006) in terms of *save steps*. These features include the access to the metadata view, the download of the data (when access permitted), and the download of the additional descriptive PDF. By downloading data from the TR32DB, the CRC/TR32 Project Database: Data Policy Agreement (CRC/TR32, 2011) has to be considered, as well as an optional further selected licence (e.g. Creative Commons Licence). As a negative aspect, it should be mentioned that currently datasets can only be assigned to one project section. By searching for data according to a specific project section, they will accordingly be only presented in one category, although the dataset might be generated in a collaborative work of two project sections. This disadvantage is known and will be solved in the next TR32DB system update.

The presentation of the metadata view is accessible to every visitor of the TR32DB. The specific webpages have a persistent URL for each dataset. This enables the citation of the data, as well as easy linking to this webpage (e.g. when referring to a certain dataset in an e-mail). The webpage of each metadata view is dynamically generated and arranged in a printer-friendly view. By designing this metadata view page, several solutions of other RDM systems were observed (e.g. Dryad, PANGAEA). Initially, the overview page provides the available features of the dataset according to the permission of the web-interface visitor or authorized user. Currently, a maximum of four features (Figure 4-50) are available: download data, download additional file (PDF), edit metadata, and apply DOI. These features can be directly accessed. Subsequently, an automatically generated citation advice will be displayed depending on the data type to simplify the citation of a dataset. This feature is very common for RDM web-interfaces (e.g. zenodo portal, Dryad). The metadata view is arranged according to the TR32DB Metadata Schema (section 4.3) and displays all metadata of the selected dataset. By means of hyperlinks it is possible to link to related datasets of the TR32DB. In addition, a linkage to related external identifiers is supported. For example, a dataset can be related to a publication by providing the identifier (e.g. 10.1117/12.813171), the type of identifier (e.g. DOI), and the type of relation (e.g. cites). By using the hyperlink in the metadata view, the user will be linked directly to the corresponding webpage. Furthermore, other linkages are available, for example to the creator and metadata creator details overview webpage. The metadata view page of DOIs, so-called landing page, is arranged according to a similar principle, but with another layout as well as associated to the metadata of the respective DOI, following the recommendations of DATACITE (2012).

All authorized TR32DB users can also use internal features (4.5.2.3). These will be presented in the left menu bar after login only. Some main features are grouped in the `user home` (Figure 4-17), like the initial metadata input for a specific dataset. As stated by FOULONNEAU & RILEY (2008), metadata input user interfaces have to be well-designed to streamline the process of metadata creation. Since there is no overall design, the interface should be designed according to the specific needs. Furthermore, the authors suggest the usage of radio buttons, drop-down lists for controlled vocabulary lists, automatic completion of value fields, as well as tabbing between input fields. Finally, FOULONNEAU & RILEY (2008) suggest the creation of an easy and error-free metadata input solution. This also corresponds to the findings of CALLAHAN et al. (1996). He outlines that metadata creation interfaces need to be easy to use, should support selective lists and

templates with the aim to minimize the effort and time for the user. Moreover, the solution should be flexible to changes, according to changing requirements. Consequently, a user-friendly input of the metadata in a comprehensive wizard (Figure 4-58 - Figure 4-73) was designed according to the TR32DB Metadata Schema following some of the previous mentioned suggestions. This wizard enables the metadata input grouped in several tabs (e.g. identification, topic). Depending on the data type, the last tab will be modified to the specific input fields. This has the advantage that the user always sees the same input form and only one tab will change. The general design of the wizard is aligned to other metadata wizard, such as the INSPIRE Metadata Editor (EUROPEAN COMMISION, 2014b). The wizard supports the use of drop down list for controlled vocabulary lists, as well as text fields, text area, calendar fields and a map feature, supporting the provision of geographic coordinates. The application of drop down lists is reasonable for short controlled vocabulary lists. Since some controlled lists, such as the list of creators or institutions, are growing with the amount of data in the TR32DB, another solutions has to be implemented (e.g. automatic completion of value fields).

Furthermore, the wizard provides single and multiple entries of data values according to the metadata profile. To prevent the users from being overburdened with many input fields, only some input fields are mandatory, marked with an asterisk. Further metadata can be provided to optional fields. In addition, were possible, some metadata will be also automatically generated by the system (e.g. metadata upload date). Assistance to the single input fields is provided by the corresponding " i "-symbol. Furthermore, the suggested `template` feature was implemented, which supports the re-use of previous submitted metadata according to a specific data type. Overall, as stated by CALLAHAN et al. (1996), it is a key challenge to motivate users to document their data, since the documentation is otherwise left undone. Consequently, the creation of a user-friendly metadata user interface is essential.

The input of metadata for the application of TR32DB DOIs (Figure 4-76 - Figure 4-81) follows the same approach as the previously described general input of metadata. The input form was adapted according to the DataCite Metadata Schema (DATACITE METADATA WORKING GROUP, 2011). To design the application process user-friendly, the input form will be prefilled with already available metadata of the specific dataset. This can be modified and complemented. The application of a DOI is restricted to the metadata creator of the dataset.

The Share feature (Figure 4-82) enables a simple and easy temporary exchange of data, for example of previous research results. Furthermore, the climate data section (Figure 4-83 - Figure 4-85) provides also an easy access to data, purchased from several organisations, like the German Weather forecast service. In addition, visualisation, for a first simple overview of measured weather data is provided. Similar features are also provided in other RDM systems, like the TEODOOR system (KUNKEL et al., 2013).

Furthermore, the TR32DB Viewer (Figure 4-86) allows an easy and fast access to the most important purchased geodata of the whole project. For instance, users can easily search for the land use of a certain area and a certain time, as the integrated land use data is displayed and can be queried. In this case, an open access, based on standards from the Open Geospatial Consortium (OGC), is prohibited to most of the services, as the data itself is restricted due to license issues. An access can be granted for services, which are established on own data analysis or measurements, such as the land use data. This allows the easy integration of the data by other users.

The integrated admin area (Figure 4-85) provides some features to support the administration of the TR32DB. For example, the administration of TR32DB user details and their affiliation to project section can be executed. In addition, the metadata entry of geodata is available by a metadata wizard. Further features are still in development.

The TR32DB web-interface is linked to the official TR32 web-interface (Figure 4-88 and Figure 4-89). Publications, which are stored in the TR32DB, will be automatically displayed in the publication list of the TR32 web-interface. This has the advantage that every TR32DB user can administrate his publications, as well as a redundant storage of this information is avoided. A similar approach is used in the CRC806-database

system, which also has a link to their official project webpage. Furthermore, another link is provided from the project sections in the TR32 web-interface to their data and publications stored in the TR32DB.

5.1.3 Acceptance of the TR32DB

It is well documented in literature that most scientists agree to share their data (MARX, 2012; NELSON, 2009), as well as that sharing of data varies in the various disciplines and communities. For instance, data sharing is very common in astronomy, oceanography, or genetics (REICHMAN et al., 2011; VAN NOORDEN, 2013). In addition, reasons for declining data sharing are well known, like lack of time or experience, technical and social problems, standards or loss of data (GRIFFITHS, 2009; NELSON, 2009; TENOPIR et al., 2011). In addition, sharing of unpublished or not fully completed data is declined (REICHMAN et al., 2011). Furthermore, challenges and problems are identified by MÜCKSCHEL et al. (2007) with regard to the RDM system. This includes the lack of acceptance of a RDM system on the side of the users, as well as a low level of motivation to use it. This also concerns to the documentation of data, which is disregarded by the researchers (WEICHSELGARTNER et al., 2011) (section 2.1).

Similar experiences were gained in the last years in the context of data sharing, data exchange, data disposal, and data documentation by means of the TR32DB. Currently (as of May 19, 2014) ~350 GB of data is stored in the AFS data storage and 773 metadata records (datasets with a complete metadata description) are stored in the MySQL database.

The intensity of the handling of data using the TR32DB varies in the different CRC/TR32 project sections and disciplines. Some scientists and projects use the TR32DB very continuously, while others do avoid it. The common use of the TR32DB is well-tracked by some features, like a download counter or who-is-online feature, which operate in the background of the system. Exemplary, the usage of the provided land use data (WALDHOFF et al., 2012) according to the single project sections is presented in Figure 5-1. Only project sections, which have downloaded the data, are presented. It is clearly visible that the demand of the data varies significantly in the different project sections and disciplines, since not all project sections require the data for their research questions. Consequently, some projects sections generally do not need to use the data.
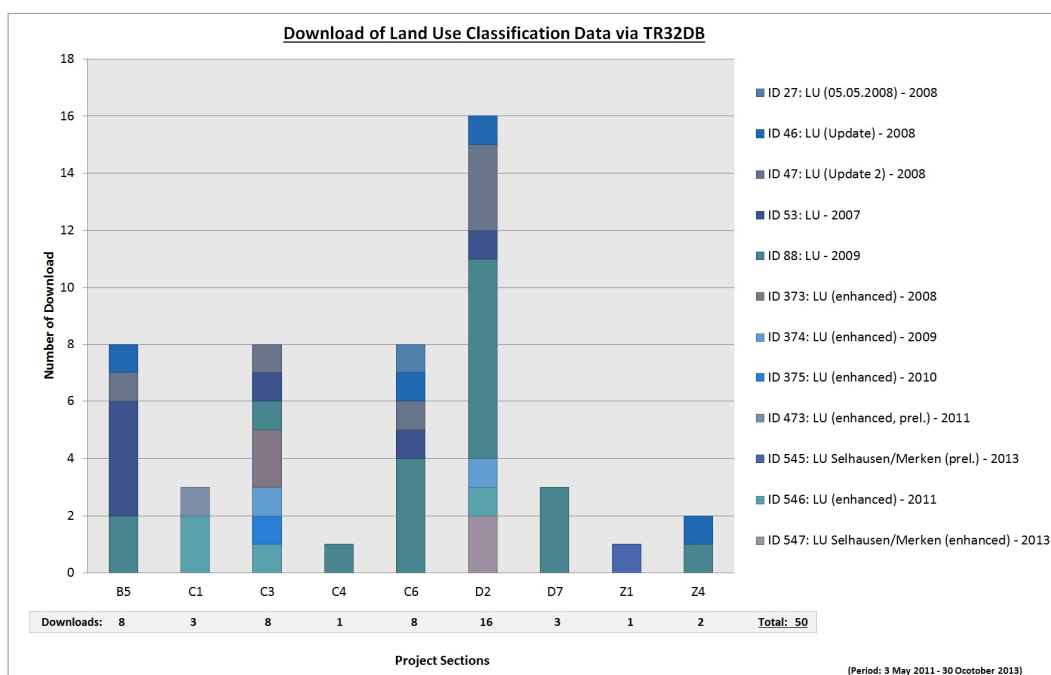
Figure 5-1: Download statistic of the land use data, stored in the TR32DB according to project sections

Furthermore, the additional benefit of a DOI for a dataset, which is cited in a publication, is in evidence by the example of SEIBERT et al. (2011). This dataset is cited in FIENER et al. (2011) as supplement material to the publication (section 2.1.6). In contrast to other datasets, this dataset has been downloaded many times since 2011 from international, interested parties. The total download as of May 19, 2014 is 62 times.

5.2   The TR32DB versus selected RDM systems

This chapter gives a comparative overview of the main system components and features of several selected RDM systems that manage different types of research data, already described with general details in section 2.5. An outline is presented in Table 5-3 and Table 5-4. The selected RDM systems are institutional or project-related systems and focus on various disciplines. With the exception of the BExIS system, all other RDM systems are accessible online or in case of the RBIS system usable with a guest login.

An analysis of the selected RDM systems shows that most of them are in-house developments combining several available components. They are all developed with regard to the specific discipline and requirements of the particular research project or focus of the institution. Particularly, RDM systems have to be designed and consider the project demands (BÜTTNER et al., 2011; LOTZ et al., 2012; MÜCKSCHEL et al., 2007). Correspondingly, the TR32DB is also designed with regard to the CRC/TR32 demands and framework requirements of the DFG (section 4.1). In case of the institutional repositories Datorium and Dryad, the common out-of-the-box repository software DSpace (section 2.4) is used. In addition, the CRC806-Database uses the relatively newly established repository software CKAN (section 2.4) in combination with further components.

All considered RDM systems are accessible by a web-interface. To establish the web-interface, several solutions (section 2.3.4) are used by the RDM systems. Besides established content management systems (CMS), such as Plone or TYPO3, also different web application frameworks are implemented, like the .NET-framework or Ruby on Rails. In addition, RDM web-interfaces are established by combining common web programming, mark-up, scripting, and query languages, like PHP, HTML, CSS, JavaScript or SQL (section 2.3.4). All of these are used for the programming of the TR32DB. In addition, the majority of the selected RDM systems incorporate a WebGIS or web mapping application (section 2.3.5) using various implementations. For instance, the PANAGEA system supports a web mapping application established by the Google Maps API to visualize the location of a specific dataset. A similar approach is used in the TR32DB to visualize the location of a specific dataset, as shown in Figure 4-53. Furthermore, several RDM systems have integrated WebGIS components. For example, the CRC806-Database (WILLMES et al., 2014) and RBIS system (ZANDER et al., 2013) provide a full established WebGIS with several features implemented by the UMN MapServer. In addition, the TEODOOR, BExIS, and FOR816dw system use the GeoServer to support WebGIS functionalities. Likewise, the TR32DB uses ESRI's ArcGIS for Server and the Flex Viewer web application to provide a WebGIS for the users of the TR32DB (section 4.5.1.3 and 4.5.2.3).

For the storage of the data, several solutions are implemented by the various considered projects. File-based data storage in combination with a database (e.g. MySQL, PostgreSQL) is, for example, supported by the CRC806-DB, Datorium, as well as the TR32DB. The project RDM systems BEFdata, BExIS, and FOR816dw, all related to biodiversity data, as well as the TEODOOR system and RBIS system support the storage of the project data in a database, like PostgreSQL, DB2 with pure XML or MySQL. Project data, created and stored in these RDM systems, are all characterized to be homogeneous, which enables the implementation of a database. For instance, the BEFdata system supports the upload and storage of data through a formatted Microsoft Excel Workbook which stores the metadata and the raw data in single sheets (NADROWSKI et al., 2013). Consequently, the application of a database is not feasible for research projects with very heterogeneous research data. Furthermore, some RDM systems, like the BExIS and FOR816dw, both implemented in a follow-up project, are influenced and benefit from experiences gained in previous RDM projects, which applied a file-based data storage approach (LOTZ et al., 2012; NAUSS et al., 2007;

STIETENROTH et al., 2005). With regard to the cooperation with a computing centre or established data centre, several options are available for the different project-related RDM systems. For example, the data storage of the TR32DB and CRC806-Database is established and backed-up in the environment of the RRZK (section 3.4), as well as the FOR816dw stores and back-ups their data in a central infrastructure of the University of Marburg (LOTZ et al., 2012). With regard to the type of data that are stored in the different RDM systems, various approaches are chosen. For instance, the RBIS system supports time series data, geodata, simulation models, and documents (KRALISCH et al., 2009). The CRC806-Database incorporates data, services, literature, presentations, multimedia, and external repository resources. The BEFdata systems and the For816dw maintain data and publications. According to the other systems, the TR32DB uses a similar, extended selection of data types including data, basic geodata, reports, presentations, publications, and pictures (section 3.4).

In terms of supported metadata schemas, several solutions are provided by the considered RDM systems. The BExIS system uses an own metadata schema, only related to the project. All other systems support and use approved metadata standards and schemas related to the specific research area. For instance, the Ecological Metadata Language (EML) is supported by BEFdata and FOR816dw. Repositories related to geoscientific data use the metadata standard ISO 19115 (section 2.2.2.3) or are established in compliance with INSPIRE (section 2.2.2.4), like the CRC806-Database, RBIS, TEODOOR, PANGAEA or the TR32DB. In addition, some repositories combine and support a combination of several metadata standards. All systems use the metadata standard Dublin Core (section 2.2.2.1) as a basis and extend this with other metadata standards. For instance, Dryad combines Dublin Core with elements of Darwin Core, and the bibliographic Ontology (section 2.2.2.5) (DRYAD DEVELOPMENT TEAM, 2010). Likewise, the TR32DB Metadata Schema (section 3.3) was designed and implemented combining the Dublin Core, ISO 19115, INSPIRE, and DataCite Metadata Schema.

Nearly all considered RDM repositories support basic features, like search for data, access for their metadata, and download of the data according to particular user and data permissions (e.g. download of data only for project members). Moreover, unique, permanent identifiers are provided for the datasets (e.g. URL, URN or DOI). Some repositories, like the FOR816dw and the TR32DB support DOIs for datasets in addition to the unique identifier for special dataset (e.g. section 4.5.2.3). Finally, some RDM systems have approved a data policy for the handling of the data in the particular system such as the TEODOOR portal (TERENO, 2014), the FOR816dw (RU816, 2007), and PANGAEA (GROBE, 2005). The CRC/TR32 has also adopted a TR32DB data policy agreement in 2011 (CRC/TR32, 2011).

Table 5-3: Selection of project and institutional RDM system with topic and features

	TR32DB	BEFdata	BExIS	CRC806-Database	Datorium
Website	tr32db.uni-koeln.de	china.befdata.biow.uni-leipzig.de	exploratories.bgc-jena.mpg.de	crc806db.uni-koeln.de	datorium.gesis.org
Discipline	Geosciences	Biodiversity	Biodiversity	Archaeology, Paleoenvironmental Sciences	Social Sciences
Repository type	Project	Project	Project	Project	Institutional
Repository open access	+	+	–	+	+
Repository launch	2008	*	2007	2012	2014
Repository Technology, Software, Components					
out-of-the-box repository software	–	–	–	CKAN	DSpace
Web-interface components	PHP, HTML, JavaScript	Ruby on Rails	.NET framework	TYPO3	*
WebGIS component	ArcGIS for Server, Google Maps API	*	GeoServer, Google Maps API	UMN MapServer, GeoEXT, MapProxy	–
data storage	Andrew File System	PostgreSQL	DB2 with pure XML	Andrew File System	File-based, GESIS Data Archive
database technology	MySQL	PostgreSQL	DB2 with pure XML	MySQL, PostgreSQL	*
metadata schema support	Dublin Core, ISO 19115, INSPIRE, DataCite,	Ecological Metadata Language (EML)	project internal XML-schema, export to DarwinCore	Dublin Core, ISO 19115/ 19119	Dublin Core, GESIS Data Catalogue (based on da|ra & DataCite)
Repository features					
data search	+	+	+	+	+
data download restriction	+	+	+	+	+
data identifier	Unique, permanent URL	Unique, permanent URN	*	Unique, permanent URL	Unique, permanent DOI
support of DOI	+	*	*	–	+
GIS /Mapping features	+	*	*	+	–
metadata open access	+	+	–	+	+
data policy	+	*	+	–	*

+ = available; – = not available; * = unknown

Table 5-4: Selection of project and institutional RDM system with topic and features

	Dryad	**FOR816dw**	**RBIS**	**TEODOOR**	**PANGAEA**
Website	datadryad.org	tropicalmountainforest.org	leutra.geogr.uni-jena.de/brahmaRBIS	teodoor.icg.kfa-juelich.de	pangaea.de
Discipline	Biosciences	Biodiversity	Earth Sciences	Earth Sciences	Earth & Environmental Sciences
Repository type	Institutional	Project	Project	Project	Institutional
Repository open access	+	+	Guest login	+	+
Repository launch	2008	2007	*	2008	1992
Repository Technology, Software, Components					
out-of-the-box repository software	DSpace	–	–	–	–
Web-interface components	*	Struts, J2EE	PHP, HTML, JavaScript	Plone	*
WebGIS component	–	GeoServer, OpenLayers	UMN MapServer, OpenLayers, Google Maps API	GeoServer, OpenLayers, Google Maps API	Google Maps API
data storage	*	MySQL, file system	file system PostgreSQL	file system PostgreSQL	Sybase IQ
database technology	*	MySQL	PostgreSQL /PostGIS	PostgreSQL /PostGIS	Sybase IQ
metadata schema support	Dublin Core, Darwin Core, Bibliographic Ontology	Ecological Metadata Language (EML)	ISO 19115	Dublin Core, INSPIRE SensorML	ISO 19115, DIF, Dublin Core, DataCite, DarwinCore
Repository features					
data search	+	+	+	+	+
data download restriction	+	+	+	+	+
data identifier	Unique, permanent DOI	Unique, permanent URL	Unique, permanent URL	Unique, permanent URN	Unique, permanent DOI
support of DOI	+	+	*	–	+
GIS /Mapping features	+	+	+	+	–
metadata open access	+	+	+	+	+
data policy	+	+	*	+	+

+ = available; – = not available; * = unknown

5.3   Transferability of the TR32DB approach

In the previous sections, the TR32DB approach was discussed in detail in comparison to other RDM system solutions and the single components of the TR32DB were discussed. This section will focus on a feasibility study on the transferability of the designed and implemented TR32DB approach to other interdisciplinary research projects.

The transferability of a self-developed, project-related RDM system was already successfully conducted in other projects. The River Basin Information System (RBIS) (section 2.5), a web-based platform for data management and sharing of environmental data, has been successfully implemented for several national and European funded research projects. For instance, the BrahmaRBIS and the DanubeRBIS platform are based on the RBIS system, which were set-up within the European Commission funded BRAHMATWINN-project. Likewise, the RBIS was used as a basis for the OBIS-platform of the BMBF-funded "The Future Okavango" project and the VuGiaThuBonRBIS portal of the BMBF-funded LUCCI project (ZANDER et al., 2011). In addition, the BEFdata open source web platform (section 2.5) is successfully implemented for the DFG-funded BEF-China project and the FunDivEUROPE project funded by the European Commission (NADROWSKI et al., 2013). Furthermore, the Biodiversity-Exploratories Information System (BExIS) is applied in the DFG-funded Biodiversity Exploratories project (HEIMANN et al., 2010a), where it was originally developed and is used in the DFG-funded Jena Experiment project (JENA EXPERIMENT, 2014). Moreover, BExIS will be implemented in the DFG-funded CRC990 EFForTS (EFFORTS, 2014). In particular, in terms of RDM in CRCs, the DFG demands the re-use of existing technologies and tools (DFG, 2013d).

To successfully set-up or transfer a RDM system or repository for a research project, continuous adjustments have to be made, depending on the changing requirements arising from the research projects, technical developments and altered standards, as well as changing demands of the funding organisations (JENSEN et al., 2011; LUDWIG, 2012; MÜCKSCHEL et al., 2005). In order to successfully transmit the TR32DB approach to another research project, several adjustments of the single system components have to be conducted, to set-up a RDM system according to the demands and requirements of the specific project. In the following, the transmission of the TR32DB approach will be exemplary presented by the implementation of a project database for the DFG-funded Research Unit 1501 "Resilience, Collapse and Reorganisation in Social Ecological Systems of Africa's Savannahs" (RCR, 2014). In the context of this project, the established RCR-Database, accessible online since mid-2012 at http://www.fg1501db.uni-koeln.de, was set-up with the basic features of the TR32DB approach.

The RCR is an interdisciplinary joint research project of the German Universities of Bonn and Cologne, funded since 2010 (currently funded until 2016). The scientists involved work in the field of soil sciences, ethnology, crop and grassland sciences, ecology anthropology, economics, geography, and history. They conduct research on the resilience, collapse, and reorganization in complex coupled social-ecological systems (SES) in Africa. The focus of this research is on selected study sites, located in South Africa (e.g. Kuruman, Thaba Nchu Region) and Kenya (e.g. Lake Naivasha, Lake Baringo, Laikipia Plateau). As a result of their research, the RCR scientists create a lot of different heterogeneous research data, with small file size. For example, these data are acquired during field campaigns, by numerical simulations or empirical analysis, or questionnaires (e.g. GREINER & SAKDAPOLRAK, 2013; KECK & SAKDAPOLRAK, 2013; NAUMANN, 2014). Most of the created data are related to a spatial domain. In addition, the scientists present their research results. Consequently, they create publications, presentations, as well as images or maps which are related to the RCR. All these data should be managed by a RDM system. Moreover, requirements of the DFG (section 2.1.2) have to be considered. For instance, the principles of the *Good Scientific Practise* (DFG, 1998, 2013e), as well as the recommendations and guidelines provided in the *Proposal Preparation Instructions-Project Proposals* (DFG, 2013c, 2014b), which provide instructions for handling of data created in Research Units, have to be considered for the realisation of the RCR-Database, as similarly described for the TR32DB (section 4.1).

The RCR-Database was set-up in order to support the sharing and re-use of the collected research data of the scientists as well as to archive and backup these data during the project funding and beyond. The project database was set up according to the demand of the project participants as well as according to the demands of the DFG.

Initially, the TR32DB Metadata Schema (section 3.3 and 4.3) had to be adjusted and reduced according to the demands and characteristics of the RCR. Since the TR32DB Metadata Schema has been designed in a multi-level approach and supports general, data type specific as well as project specific metadata details, only some minor changes had to be conducted with regard to the project specific metadata properties. The metadata properties associated to the `general level` (appendix D), which describe the basic, ubiquitous information (e.g. title, description or creator of a dataset) could be fully applied. Only some associated CRC/TR32 specific controlled vocabulary lists, had to be adapted according to the RCR-needs. The controlled vocabulary lists `MeasureLocationTR32` (Table D-33), `MeasureRegionTR32` (Table D-34), `TR32Subproject` (Table D-56), `TR32Topic` (Table D-57), and `WebsiteVersion` (Table D-58) had to be adjusted according to the RCR research focus. For example, the measurement region was modified to the RCR study areas (e.g. Kenya, South Africa) and the topics were accordingly adjusted (social, bio-physical, cultural, coupling, other). The metadata properties of the data type category `Data` (appendix D.3.2) were reduced to the RCR-demands. Consequently, only the metadata properties `TemporalExtent` (appendix D.3.2, ID 31) and the property `Lineage` (appendix D.3.2, ID 32) were applied. The metadata category `Geodata` (appendix D.3.3) was not implemented in the RCR-Database. Moreover, in terms of the `specific level` of the TR32DB Metadata Schema, the complete metadata properties and controlled vocabulary lists of the data categories `Picture` (appendix D.3.5), `Presentation` (appendix D.3.6), `Publication` (appendix D.3.7), and `Report` (appendix D.3.4) were completely adopted.

In the next step, the file-based data storage structure in the AFS (section 3.4, Figure 3-7) had to modified to the project structure of the RCR, which is like the CRC/TR32 subdivided in clusters and sub-projects. In contrast to the TR32DB data storage, the single data type folders (e.g. data, picture, publication) were combined to one single data type folder. For instance, all data uploaded to project section Z2, are stored in the temporary folder directory `/afs/rrz/project/fg1501db/Upload_1/Z/Z2/data`. Moreover, the script, which moves the data from the temporary data storage to the final data storage, had to be adjusted. This script operates over night and and moves the data from to the corresponding folder in the final data storage, for example, `/afs/rrz/project/fg1501db/Phase_1/Z/Z2/data/000`. In contrast to the TR32DB moving script, this script updates the storage path of the file in the database, when the file is moved. In addition, the database-based storage (section 3.5), was implemented according to the RCR demands and the modified metadata schema of the RCR.

In a last step, the web-interface for the RCR-Database (Figure 5-2, Figure 5-3) was designed and realized according to the RCR demands and the modified RCR metadata schema. The structure of the RCR web-interface is designed in a similar structure as the TR32DB web-interface (e.g. Figure 4-41). The header and footer of the website were adjusted. Again, two menu bars dominate the website, one general menu bar (top menu) and one data menu bar (side menu A and B). Both menu bars were modified to the requirements of the RCR. For example, the menu items topic and cluster were adjusted. In terms of the web-interface features, two main changes were performed. The menu bars provide the same features just as the TR32DB menu bars. For instance, the search for data is also provided according a predefined search by topic or type, as well as by an advanced search and a map search, as described in detail in section 4.5.2.2 for the TR32DB web-interface Version 4.0.

There are two main differences between the TR32DB and the RCR-Database. The first is related to the management of the users. In contrast to the TR32DB user management, which is performed by RRZK university accounts (e.g. section 3.4.2, 4.2, 4.5.1.3), the RCR-Database user management is connected to the user management of the official RCR-website (RCR, 2014). This web-site, based on a Joomla CMS, had

already been implemented, when the RCR-Database was set-up. Thus, a double user administration was avoided. Consequently, the users just need one account for the two RCR websites. The second main difference between the TR32DB and RCR-Database is the upload of the data to the data storage. Since the RCR scientists only create and provide data files with a small file size (< 50 MB), the upload of the data is directly possible by the web-interface. Consequently, the upload of the data file to the AFS data storage was implemented as a first step for the data provision. In addition, the user has to provide the associated project section (e.g. B2, C3) and data type (e.g. data, report, publication). As a second step, the user will be directly forwarded to a metadata input wizard, which is arranged according to the metadata input wizard of the TR32DB (Figure 4-58 to Figure 4-72) in a simplified design, presented in Figure 5-2. Depending on the selected data type of the data upload, the metadata input wizard will be adjusted in the final, specific tab.

Figure 5-2: RCR-Database metadata input

The RCR-Database wizard enables the input of general and data type specific metadata. In contrast to the TR32DB metadata wizard, the RCR metadata wizard is simplified and limited in terms of multiple entries for some metadata properties, like title, description or date. In addition, the provision of complementing property types, like title type, description type or date type, is reduced to the minimal demand of the supported metadata standards. The RCR wizard supports error detection for missing or invalid metadata values, like the TR32DB wizard. Incorrect input fields will be highlighted in red, as shown in Figure 4-73 for the TR32DB wizard.

Finally, after a successful provision of the metadata, the user will be forwarded to the metadata overview webpage of the dataset, presented in Figure 5-3. This overview is also arranged according to the metadata wizard tabs, but in a simplified layout, in contrast to the TR32DB metadata overview of Version 4.0 (Figure 4-53).

Overall, the TR32DB approach was successfully transferred to another interdisciplinary research project. The RCR-Database was set-up, partly simplified, and extended successfully with new features according to the demands of the RCR participants and project background. By applying the TR32DB approach to the RCR,

time and personal resources could be reduced in contrast to the implementation of a complete new RDM system. Consequently, the developed TR32DB approach can be also easily adapted to the demands of other interdisciplinary research projects in the future. The features of the TR32DB can be limited, simplified, as well as the extension with new features is possible.

Figure 5-3: RCR-Database metadata view

6 Conclusion and perspectives

As previously stated in this thesis, the importance of research data management (RDM) and the establishment of corresponding RDM infrastructures and systems within a research project or institution is demanded by many policy makers (DFG, 2014b; EUROPEAN COMMISSION, 2007a; HOCHSCHULREKTORENKONFERENZ, 2014; NATIONAL SCIENCE FOUNDATION, 2011b), journals (AGU, 2013; PLOS, 2014) or other organisations. The management of research data concerns particularly publicly financed data and has consequently led to the demand of data management plans, which have to be provided with the application of the grant (DFG, 2014b; NATIONAL SCIENCE FOUNDATION, 2011b; NERC, 2014a; WELLCOME TRUST, 2014). In addition, challenges of RDM are identified with a technical, socio-cultural, and ethical nature (GRIFFITHS, 2009; HARTTER et al., 2013; NELSON, 2009; TENOPIR et al., 2011). So far, RDM has been strongly influenced and organized by single disciplines due to different demands on technology, archiving, and metadata. The increasing requirements on crosslinking with researchers in other disciplines demands the evolution of specialists to set up service infrastructures for RDM and interfaces for the entire research process (BÜTTNER et al., 2011a). With regard to interdisciplinary, long-term research projects, like the CRC/TR32, INF-projects assume this task. KLAR & ENKE (2013) state that INF-projects are an appropriate measure to improve the data management in the group domain (section 2.1.4). Furthermore, they outline that INF-projects should focus on the context of the respective project to achieve and make the research results accessible for future re-use. In this context, data documentation is essential. As outlined by WEICHSELGARTNER et al. (2011), there is currently a theoretical trend to already document the data within the creation process of the data or research life cycle (section 1). Correspondingly, data should be already archived and published within the conception and execution phase of the project.

With regard to the TR32DB, the aim is to collect, document, store, backup, share, provide, and publish the data, created by the CRC/TR32 scientists during the project term. To achieve this goal and the already mentioned general challenges of RDM (section 2.1), several solution approaches were conducted. Initially, the TR32DB system was developed in consultation with the project participants and set-up as user-friendly as possible. Though, not all recommendations and demands of the scientists could be considered within the development of the system. Most suggestions for modification were proposed on the TR32DB data management trainings, which introduced the features and handling of the system. They were conducted at irregular intervals upon demand of the CRC/TR32 participants over the last years. The provision of guidance as well as the provision of comprehensive TR32DB tutorials is consequently the second approach to solve the general challenges of RDM. To increase the motivation of RDM in the CRC/TR32, further TR32DB training sessions are planned. In addition, further general information about RDM will be provided by the web-interface. This should include advice to file formats, licensing, data publishing, data sharing or data protection, as well as hints to available online courses to practice the handling of data. A well-designed example is the online portal MANTRA (EDINA, 2014), developed at the University of Edinburgh.

As stated by WINKLER-NEES (2011), the acceptance of a system is also affected by the functionalities and reliability, as well as the quality and available services. Consequently, the TR32DB will be enhanced further with regard to technical developments and demands of the CRC/TR32 participants. This also corresponds to HICKS et al. (2007) who determines that digital repositories or RDM systems have developed from simple collections with simple metadata schemas to complex infrastructures with complex metadata structures to accommodate all demands on the systems.

Further developments of the TR32DB will comprise, for example, the enhancement of the TR32DB Metadata Schema with regard to map this to an XML structure. The aim is to store all metadata records also in an XML structure in case, the MySQL database will not be accessible in the future. In addition, the import and export of metadata will be possible, as well as an approach to automatically generated metadata. Furthermore, statistical analysis of the use of single datasets in the TR32DB will be taken into account.

Finally, the biggest task for the future is (for an optionally third funding phase of the CRC/TR32), the setting-up of a so-called "TR32DB-Box", which will be hosted in the RRZK environment. The aim is that the TR32DB can operate in the RRZK with little or no maintenance to ensure the availability of all CRC/TR32 project data beyond the funding of the project, as demanded in the *Good Scientific Practice* of the DFG (DFG, 1998, 2013e).

References

AGU, 2013. AGU Publications Data Policy.
http://publications.agu.org/author-resource-center/publication-policies/data-policy/, 2014-04-14.

AILAMAKI, A., KANTERE, V., DASH, D., 2010. Managing Scientific Data. Communications of the ACM 53 (6), 68-78.

ALBERTZ, J., 2013. Einführung in die Fernerkundung : Grundlagen der Interpretation von Luft- und Satellitenbildern. WBG, Darmstadt, Germany.

ALLAN, R., 2011. Management and analysis of large scientific data sets. Technical Report, Science and Technology Facilities Council Daresbury Laboratory, Daresbury, Warrington, UK.

ALLIANCE OF GERMAN SCIENCE ORGANISATIONS, 2010. Principles for the Handling of Research Data.
http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze/, 2012-07-17.

ALONSO, G., CASATI, F., KUNO, H., MACHIRAJU, V., 2004. Web Services: Concepts, Architectures and Applications. Springer, Berlin Heidelberg, Germany. doi: 10.1007/978-3-662-10876-5.

ARACHNE, 2013. Arachne - Objektdatenbank und kulturelle Archive des Archäologischen Instituts der Universität zu Köln und des Deutschen Archäologischen Instituts. http://arachne.uni-koeln.de, 2013-02-25.

ARCHAEOLOGY DATA SERVICE, 2014. Archaeology Data Service Deposit Licence.
http://archaeologydataservice.ac.uk/attach/guidelinesForDepositors/ads_licence_form.pdf, 2014-05-11.

ASCHENBRENNER, A., ENKE, H., FISCHER, T., LUDWIG, J., 2011. Diversity and Interoperability of Repositories in a Grid Curation Environment. Journal of Digital Information 12 (2).

ASCHENBRENNER, A., NEUROTH, H., 2011. Forschungsdaten-Repositorien. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 101-114.

BAASER, U., 2010. Das CampusGIS der Universität zu Köln – webgestützte Geodatendienste für raumbezogene Anwendungen. Ph.D. Thesis, University of Cologne, Germany.

BACA, M., 2008. Introduction to Metadata. Getty Research Institute, Los Angeles, CA, USA.

BACKES, M., DÖRSCHLAG, D., PLÜMER, L., 2005. Landwirtschaftliche Geodaten - Nachhaltige Datenhaltung und -nutzung durch ISO Standards. Zeitschrift für Agrarinformatik 2005 (1), 18-23.

BALL, A., 2009. Scientific Data Application Profile Scoping Study Report.
http://www.ukoln.ac.uk/projects/sdapss/papers/ball2009sda-v11.pdf.

BALL, A., 2013a. The DCC Disciplinary Metadata Catalogue. Proccedings of the International Conference on Dublin Core and Metadata Applications, DC-2013, September 2-6, 2013, Lisbon, Portugal.

BALL, A., 2013b. How to License Research Data. DCC How-to Guides, http://www.dcc.ac.uk/resources/how-guides/license-research-data, 2014-05-19.

BALZERT, H., 2004. Webdesign & Web-Ergonomie : Websites professionell gestalten. W3L, Herdecke, Germany.

BARTELME, N., 2005. Geoinformatik : Modelle, Strukturen, Funktionen. Springer, Berlin, Germany.

BDJ, 2014. Biodiversity Data Journal: Making your data count! http://biodiversitydatajournal.com/, 2014-05-01.

BEAIRD, J., 2011. Gelungenes Webdesign: Die Prinzipien der Webseitengestaltung - Ein Leitfaden für Webprogrammierer. dpunkt.verlag, Heidelberg, Germany.

BEAZLEY, M., 2010. Eprints Institutional Repository Software: A Review. Partnership: the Canadian Journal of Library and Information Practice and Research 5 (2).

BEKAVAC, B., 2013. B 1 Web-Technologien. Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. De Gruyter Saur, Berlin, Germany.
doi: 10.1515/9783110258264.144.

BENN, W., GRINGER, I., 1998. Zugriff auf Datenbanken über das World Wide Web. Informatik-Spektrum 21 (1), 1-8.
doi: 10.1007/s002870050083.

BERLIN DECLARATION, 2014. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, Berlin, Germany.
http://openaccess.mpg.de/Berliner-Erklaerung, 2014-04-19.

BERMAN, F., 2008. Got data? a guide to data preservation in the information age. Communications of the ACM 51 (12), 50-56. doi: 10.1145/1409360.1409376.

BERNERS-LEE, T., 1989. Information Management: A Proposal. http://www.w3.org/History/1989/proposal.html, 2014-05-01.

BEZIRKSREGIERUNG KÖLN, 2014a. ATKIS® - Basis-DLM. http://www.bezreg-koeln.nrw.de/brk_internet/organisation/abteilung07/produkte/landschaftsmodelle/atkis_basis_dlm/index.html, 2014-05-15.

BEZIRKSREGIERUNG KÖLN, 2014b. Digitale (automatisierte) Liegenschaftskarte (ALK). http://www.bezreg-koeln.nrw.de/brk_internet/organisation/abteilung07/produkte/liegenschaftsinformation/katasterinformation/alk/index.html, 2014-05-15.

BEZIRKSREGIERUNG KÖLN, 2014c. Digitale Geländemodelle (DGM). http://www.bezreg-koeln.nrw.de/brk_internet/organisation/abteilung07/produkte/reliefinformationen/gelaendemodelle/index.html, 2014-05-15.

BEZIRKSREGIERUNG KÖLN, 2014d. Digitale Orthophotos (DOP). http://www.bezreg-koeln.nrw.de/brk_internet/organisation/abteilung07/produkte/bildinformationen/orthophoto_dop/index.html, 2014-05-15.

BIBOTOOLS, 2014. Ontology: bibo. http://bibotools.googlecode.com/svn/bibo-ontology/trunk/doc/index.html, 2014-05-20.

BILL, R., 2010. Grundlagen der Geo-Informationssysteme. Wichmann, Heidelberg, Germany.

BLANSIT, B.D., 2006. The Basics of Relational Databases Using MySQL. Journal of Electronic Resources in Medical Libraries 3 (3), 135-148. doi: 10.1300/J383v03n03_10.

BLUE RIBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS (BRTF-SDPA), 2010. Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. Blue Ribbon Task Force on Sustainable Digital Preservation and Access, La Jolla, CA, USA.

BORENSTEIN, N.S., 1996. CMU's Andrew project: a retrospective. Communications of the ACM 39 (12es), 298. doi: 10.1145/272682.272717.

BORGHOFF, U.M., 2005. Vergleich bestehender Archivierungssysteme. Nestor Materialien 3, nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland, Frankfurt, Germany.

BORGMAN, C.L., 2010. Research Data: Who will share what, with whom, when, and why?, RatSWD Working Paper Series. Rat für Sozial- und Wirtschaftsdaten, Berlin, Germany, 23.

BORGMAN, C.L., 2012. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology 63 (6), 1059-1078. doi: 10.1002/asi.22634.

BOSE, R., FREW, J., 2005. Lineage retrieval for scientific data processing: a survey. ACM Computing Surveys 37 (1), 1-28. doi: 10.1145/1057977.1057978.

BOUGHIDA, K., 2005. CDWA lite for Cataloguing Cultural Objects (CCO): A new XML schema for the cultural heritage community. Proccedings of the XVI International Conference of the Association for History and Computing, September 14-17, 2005, Amsterdam, The Netherlands, 49-56.

BRASE, J., 2004. Using Digital Library Techniques – Registration of Scientific Primary Data. In: HEERY, R., LYON, L. (Eds.), Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, 488-494. doi: 10.1007/978-3-540-30230-8_44.

BRASE, J., 2010. DataCite - A global registration agency for research data, RatSWD Working Paper Series. Rat für Sozial- und Wirtschaftsdaten, Berlin, Germany, 7.

BRASE, J., 2013. DataCite and linked data. Italian Journal of Library and Information Science 4 (1), 365-373. doi: 10.4403/jlis.it-5493.

BRASE, J., FARQUHAR, A., 2011. Access to Research Data. D-Lib Magazine 17 (1/2). doi: 10.1045/january2011-brase

BRASE, J., SOCHA, Y., CALLAGHAN, S., BORGMAN, C., UHLIR, P.F., CARROLL, B., 2014. Data Citation: Principles and Practice. In: RAY, J. (Ed.), Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, IN, USA, 167-186.

BREEDE, K., KEMNA, A., ESSER, O., ZIMMERMANN, E., VEREECKEN, H., HUISMAN, J.A., 2011. Joint Measurement Setup for Determining Spectral Induced Polarization and Soil Hydraulic Properties. Vadose Zone Journal 10 (2), 716-726. doi: 10.2136/vzj2010.0110.

BRINKHOFF, T., 2013. Geodatenbanksysteme in Theorie und Praxis : Einführung in objektrelationale Geodatenbanken unter besonderer Berücksichtigung von Oracle Spatial. Wichmann, Heidelberg, Germany.

BROWN, M.L., WHITE, W., 2013. A partnership approach to research data management. In: PRYOR, G., JONES, S., WHITE, A. (Eds.), Delivering Research Data Management Services: Fundamentals of Good Practice. Facet, London, UK.

BROY, M., 1998. Informatik. Eine grundlegende Einführung: Band 2: Systemstrukturen und Theoretische Informatik. Springer, Berlin, Germany.

BUNDESMINISTERIUM DER JUSTIZ UND VERBRAUCHERSCHUTZ, 2002. Gesetz zur Gleichstellung behinderter Menschen (Behindertengleichstellungsgesetz - BGG). http://www.gesetze-im-internet.de/bgg/BJNR146800002.html, 2014-05-16.

BUNDESMINISTERIUM DER JUSTIZ UND VERBRAUCHERSCHUTZ, 2011. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0). http://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html, 2014-05-16.

BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG, 2014. Die Bedeutung von IT-Forschungsinfrastrukturen und der Nationalen Roadmap für die Wettbewerbsfähigkeit der Wissenschaft in Deutschland, Keynote-Speech des Staatssekretärs im Bundesministerium für Bildung und Forschung, Herrn Dr. Georg Schütte, anlässlich des dritten hochschulstrategischen

Werkstattgesprächs der VolkswagenStiftung am 7. Februar 2014 in Hannover. http://www.bmbf.de/pub/reden/Rede_StS_Werkstattspraech_VW_Stiftung.pdf, 2014-04-10.

BUNDESREGIERUNG, 2013. Deutschlands Zukunft gestalten, Koalitionsvertrag zwischen CDU, CSU und SPD, 18. Legislaturperiode. Bundesregierung, Berlin, Germany. https://www.cdu.de/sites/default/files/media/dokumente/koalitionsvertrag.pdf, 2014-04-20.

BURTON, A., GROENEWEGEN, D., LOVE, C., TRELOAR, A., WILKINSON, R., 2012. Making Research Data Available in Australia. IEEE Intelligent Systems 27 (3), 40-43. doi: 10.1109/MIS.2012.57.

BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L., 2011. Research Data Management. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 13-24.

CABINET OFFICE UK, 2013. Policy paper: Open Data Charter. https://www.gov.uk/government/publications/open-data-charter, 2014-04-15.

CALLAHAN, S., JOHNSON, J.K., SHELLEY, P., 1996. Dataset Publishing - a Means to Motivate Metadata Entry. Proccedings of the First IEEE Metadata Conference, April 16-18, 1996, Silver Spring, Maryland.

CAMPBELL, D.G., 2005. Metadata, Metaphor, and Metonymy. Cataloging & Classification Quarterly 40 (3-4), 57-73. doi: 10.1300/J104v40n03_04.

CAMPBELL, R., 1998. Managing AFS : the Andrew file system. Prentice Hall, Upper Saddle River, NJ, USA.

CAPLAN, P., 1995. You Call It Corn, We Call It Syntax-Independent Metadata for Document-Like Objects. The Public-Access Computer Systems Review 6 (4), 19-23.

CAPLAN, P., 2003. Metadata fundamentals for all librarians. American Library Association, Chicago, USA.

CARLSON, J., 2014. The Use of Life Cycle models in Developing and Supporting Data Services. In: RAY, J. (Ed.), Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, IN, USA, 63-86.

CEEC, 2014. CEEC - Codices Electronici Ecclesiae Coloniensis. http://www.ceec.uni-koeln.de/, 2014-03-15.

CEN, 2014a. Geographic information - Metadata - Part 1: Fundamentals (ISO 19115-1:2014). http://esearch.cen.eu/esearch/Details.aspx?id=19217296, 2014-05-20.

CEN, 2014b. Geographic information - Temporal schema (ISO 19108:2002). http://esearch.cen.eu/esearch/Details.aspx?id=19225023, 2014-05-20.

CERN DOC SERVER, 2014. CERN Document Server: Access articles, reports and multimedia content in HEP. http://cds.cern.ch/, 2014-04-10.

CHANDLER, A., FOLEY, D., HAFEZ, A.M., 2000. Mapping and Converting Essential Federal Geographic Data Committee (FGDC) Metadata into MARC21 and Dublin Core. D-Lib Magazine 6 (1).

CHANTELAU, K., BROTHUHN, R., 2010. Multimediale Client-Server-Systeme. Springer, Berlin Heidelberg, Germany. doi: 10.1007/978-3-540-79749-4.

CHRISTIAN-ALBRECHTS-UNIVERSITY OF KIEL, 2014. Data Management at the Christian-Alberts-University Kiel. http://www.uni-kiel.de/forschung/de/data-management, 2014-04-10.

CKAN, 2014a. CKAN: Case Studies. http://ckan.org/instances, 2014-04-10.

CKAN, 2014b. CKAN: The open source data portal software. http://ckan.org/, 2014-04-10.

CLUSTRMAPS, 2014. ClustrMaps Website. http://www.clustrmaps.com, 2014-05-15.

CODA, 2014. What is Coda? http://www.coda.cs.cmu.edu/, 2014-05-01.

CODATA-ICSTI TASK GROUP ON DATA CITATION STANDARDS AND PRACTICES, 2013. Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. Data Science Journal 12 1-75. doi: 10.2481/dsj.OSOM13-043.

CODATA, 2014. CODATA: Who are we? http://www.codata.org/about/who.html, 2014-04-15.

CODD, E.F., 1970. A relational model of data for large shared data banks. Communications of the ACM 13 (6), 377-387. doi: 10.1145/362384.362685.

COLLINS, E., 2011. The national data centres. In: PRYOR, G. (Ed.), Managing Research Data Facet Publishing, London, UK, 151-172.

CONNOLLY, T., BEGG, C., 2005. Database Solutions, A step by step guide to building databases. Person, Edinburgh, Harlow, Essex, UK.

CORTI, L., VAN DEN EYNDEN, V., BISSELL, A., WOOLLARD, M., 2014. Managing and Sharing Research Data: A Guide to Good Practice SAGE Publications Ltd, Los Angeles, CA, USA.

COYLE, K., 2010. Chapter 1: Library Data in a Modern Context. Library Technology Reports 46 (1), 5-13.

CRC/TR32, 2007. Funding Proposal Phase 1: Planned Transregional Collaborative Research Centre 32 Patterns in Soil-Vegetation-Atmosphere Systems: monitoring, modelling and data assimilation. http://tr32.uni-koeln.de/images/stories/download/intranet/proposal/phase_1/sfbtr_060723_01.pdf, 2012-06-24.

CRC/TR32, 2010. Funding Proposal Phase 2: Transregional Collaborative Research Centre 32 Patterns in Soil-Vegetation-Atmosphere Systems: monitoring, modelling and data assimilation.

http://tr32.uni-koeln.de/images/stories/proposals2010/2nd_phase/tr32_proposal_2nd_phase.pdf, 2012-06-24.

CRC/TR32, 2011. CRC/TR32 Project Database: Data Policy Agreement
http://www.tr32db.uni-koeln.de/datapolicy, 2014-03-31.

CRC/TR32, 2014a. TR32 Publications. http://tr32new.uni-koeln.de/index.php/publications-overview, 2014-04-04.

CRC/TR32, 2014b. TR32 Website. http://tr32new.uni-koeln.de/, 2014-04-04.

CREATIVE COMMONS, 2014. Creative Commons: About the Licenses. http://creativecommons.org/licenses/, 2014-03-31.

CURDT, C., 2009. TR32 Viewer Help Version 2. http://www.tr32db.uni-koeln.de/webgis2/help.html, 2014-04-04.

CURDT, C., 2011. TR32 database (TR32DB) Web-Interface user guide (Vers. 3.1).
http://www.tr32db.uni-koeln.de/site/faq.php, 2014-04-04.

CURDT, C., 2013a. Instructions for storing data in the TR32DB using Linux.
http://www.tr32db.uni-koeln.de/site/faq.php, 2014-04-04.

CURDT, C., 2013b. Instructions for storing data in the TR32DB using MS Windows.
http://www.tr32db.uni-koeln.de/site/faq.php, 2014-04-04.

CURDT, C., 2013c. TR32-Database (TR32DB) Web-Interface User Guide (Vers. 4.0).
http://www.tr32db.uni-koeln.de/site/faq.php, 2014-04-04.

CURDT, C., 2014a. TR32 Viewer Help Version 3. http://www.tr32db.uni-koeln.de/webgis/help.html, 2014-04-04.

CURDT, C., 2014b. TR32DB Metadata Schema for the Description of Research Data in the TR32DB. Transregional Collaborative
Research Centre 32, Project Section Z1/INF, Institute of Geography, University of Cologne, Cologne, Germany. doi:
10.5880/TR32DB.10.

D'ARCUS, B., GIASSON, F., 2009. Bibliographic Ontology Specification, Specification Document, 4 November 2009.
http://bibliontology.com/specification, 2014-02-12.

DALLMEIER-TIESSEN, S., 2011. Strategien bei der Veröffentlichung von Forschungsdaten. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER,
L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 157-168.

DALLMEIER-TIESSEN, S., PFEIFFENBERGER, H., 2010. Peer Reviewed Data Publication in Earth System Sciences: Earth-System-
Science-Data, In: PUSCHMANN, C., STEIN, D.A. (Eds.), Towards Open Access Scholarship. Düsseldorf University Press,
Düsseldorf, Germany.

DAMM, T., GÖTZE, H.-J., 2009. Modern Geodata Management: Application of Interdisciplinary Interpretation and Visualization in
Central America. International Journal of Geophysics 2009 13. doi: 10.1155/2009/878324.

DATA.GOV.UK, 2014. data.gov.uk portal. http://data.gov.uk/, 2014-04-10.

DATABIB, 2014. Databib Portal. http://databib.org/, 2014-04-10.

DATACITE, 2012. Business Models Principles v 1.0. DataCite, Hannover, Germany. doi: 10.5438/0007.

DATACITE, 2014a. DataCite Webpage. http://www.datacite.org/, 2014-05-15.

DATACITE, 2014b. Metadata Store. https://mds.datacite.org/, 2014-04-04.

DATACITE METADATA WORKING GROUP, 2011. DataCite Metadata Schema for the Publication and Citation of Research Data,
Version 2.2. DataCite. doi: 10.5438/0005.

DATACITE METADATA WORKING GROUP, 2013. DataCite Metadata Schema for the Publication and Citation of Research Data,
Version 3.0. DataCite. doi: 10.5438/0008.

DAVIS, H., VICKERY, J., 2007. Datasets, a Shift in the Currency of Scholarly Communication: Implications for Library Collections and
Acquisitions. Serials Review 33 (1), 26-32.

DAY, M., 2002. Metadata: Mapping between metadata formats. http://www.ukoln.ac.uk/metadata/interoperability/,
2014-02-04.

DAY, M., 2005. Metadata, In: ROSS, S., DAY, M. (Eds.), DCC Digital Curation Manual.

DCC, 2014a. DCC Checklist for a Data Management Plan. http://www.dcc.ac.uk/resources/data-management-plans/checklist, 2014-
04-20.

DCC, 2014b. The DCC Curation Lifecycle Model. Digital Curation Centre.

DCC, 2014c. Disciplinary Metadata. http://www.dcc.ac.uk/resources/metadata-standards, 2014-04-18.

DCC, 2014d. Overview of funders' data policies. http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies,
2014-04-20.

DCC, 2014e. UK Institutional data policies. http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-
institutional-data-policies, 2014-04-20.

DCC, 2014f. What is digital curation? http://www.dcc.ac.uk/digital-curation/what-digital-curation, 2014-04-18.

DCMI, 2011. User Guide. http://wiki.dublincore.org/index.php/User_Guide, 2013-02-21.

DCMI, 2012a. DCMI Metadata Terms. http://dublincore.org/documents/dcmi-terms/, 2013-02-12.

DCMI, 2012b. Dublin Core Metadata Element Set, Version 1.1. http://dublincore.org/documents/dces/, 2013-02-12.

DCMI, 2014. DCMI Mission and Principles. http://dublincore.org/about-us/, 2014-05-15.

DDI, 2014. Data Documentation Initiative Lifecycle. http://www.ddialliance.org/, 2014-04-14.

DE LANGE, N., 2013. Geoinformatik in Theorie und Praxis. Springer, Berlin Heidelberg, Germany. doi: 10.1007/978-3-642-34807-5.

DE WITT, S., SINCLAIR, R., SANSUM, A., WILSON, M., 2012. Managing Large Data Volumes from Scientific Facilities. ERCIM NEWS 89 (Special theme: Big Data), 15.

DEELMAN, E., BERRIMAN, B., CHERVENAK, A., CORCHO, O., GROTH, P., MOREAU, L., 2010. Metadata and Provenance Management. In: SHOSHANI, A., ROTEM, D. (Eds.), Scientific Data Management: Challenges, Technology, and Deployment CRC Press/Chapman & Hall, Boca Raton, FL, USA, 433-466.

DELISLE, M., 2012. Mastering Phpmyadmin 3.4 for Effective MySQL Management. PACKT Publishing, Birmingham, UK.

DETER, A., 2013. Claas präsentiert Landtechnikneuheiten. top agrar online. http://www.topagrar.com/news/Technik-Techniknews-Claas-praesentiert-Landtechnikneuheiten-1196205.html, 2014-04-30.

DEUTSCHER BUNDESTAG, 2009. Gesetz über den Zugang zu digitalen Geodaten (Geodatenzugangsgesetz - GeoZG). Bundesgesetzblatt Jahrgang 2009 Teil I Nr. 8, Berlin, Germany. http://www.gesetze-im-internet.de/bundesrecht/geozg/gesamt.pdf, 2014-02-10.

DFG, 1998. Proposals for Safeguarding Good Scientific Practice. Recommendations of the Commission on Professional Self Regulation in Science. DFG, Bonn, Germany. http://www.dfg.de/download/pdf/dfg_im_profil/reden_ stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf, 2009-04-01.

DFG, 2006. Position paper "Scientific Library Services & Information Systems – Funding priorities through 2015". DFG, Bonn, Germany. http://www.dfg.de/download/pdf/foerderung/programme/lis/pos_papier_funding_priorities _2015_en.pdf, 2014-04-14.

DFG, 2009a. Merkblatt - Service-Projekte zu Informationsmanagement und Informationsinfrastruktur in Sonderforschungsbereichen INF, bulletin 60.06 - 08/09. DFG, Bonn, Germany.

DFG, 2009b. Recommendations for Secure Storage and Availability of Digital Primary Research Data. DFG, Bonn, Germany. http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901_en.pdf, 2014-04-11.

DFG, 2011. Aufbau von Informationsinfrastrukturen für Forschungsdaten: DFG bewilligt 27 Projekte. http://dfg.de/foerderung/info_wissenschaft/archiv/2011/info_wissenschaft_11_18/index.html, 2014-04-10.

DFG, 2013a. Guidelines Information Infrastructures for Research Data, DFG form 12.14 – 03/13. DFG, Bonn, Germany. http://www.dfg.de/formulare/12_14/12_14_en.pdf, 2014-04-10.

DFG, 2013b. Guidelines Virtual Research Environments, DFG form 12.12 – 03/13. DFG, Bonn, Germany. http://www.dfg.de/formulare/12_12/12_12_en.pdf, 2014-05-15.

DFG, 2013c. Leitfaden für die Antragstellung - Projektanträge, DFG-Vordruck 54.01 - 04/13 DFG, Bonn, Germany. http://www.dfg.de/formulare/54_01/54_01_en.pdf, 2013-07-06.

DFG, 2013d. Merkblatt Sonderforschungsbereiche, DFG-Vordruck 50.06 - 06/13. DFG, Bonn, Germany. http://www.dfg.de/formulare/50_06/50_06_en.pdf, 2013-07-06.

DFG, 2013e. Sicherung guter wissenschaftlicher Praxis. Recommendations of the Commission on Professional Self Regulation in Science. DFG, Bonn, Germany. http://www.dfg.de/download/pdf/dfg_im_profil/reden_ stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf, 2013-07-06.

DFG, 2014a. Koordinierte Programme: Sonderforschungsbereiche. DFG, Bonn, Germany. http://www.dfg.de/foerderung/programme/koordinierte_programme/sfb/index.html, 2014-04-15.

DFG, 2014b. Leitfaden für die Antragstellung - Projektanträge, DFG-Vordruck 54.01 - 04/14. DFG, Bonn, Germany. http://www.dfg.de/formulare/54_01/54_01_de.pdf, 2014-04-19.

DICKMANN, F., ZEHNER, K., 2001. Computerkartographie und GIS. Westermann, Braunschweig, Germany.

DIEPENBROEK, M., GROBE, H., 2007. PANGAEA® als vernetztes Verlags- und Bibliothekssystem für wissenschaftliche Daten. In: BALL, R. (Ed.), WissKom 2007: Wissenschaftskommunikation der Zukunft. Forschungszentrum, Zentralbibliothek, Jülich, 147-158.

DIEPENBROEK, M., GROBE, H., REINKE, M., SCHINDLER, U., SCHLITZER, R., SIEGER, R., WEFER, G., 2002. PANGAEA-an information system for environmental sciences. Computers & Geosciences 28 (10), 1201-1210. doi: 10.1016/S0098-3004(02)00039-0.

DIETRICH, D., ADAMUS, T., MINER, A., STEINHART, G., 2012. De-Mystifying the Data Management Requirements of Research Funders. Issues in Science and Technology Librarianship 70 (1). doi: 10.5062/F44M92G2.

DIRECTORY OF OPEN ACCESS REPOSITORY, 2014. The Directory of Open Access Repositories - OpenDOAR. http://www.opendoar.org, 2014-04-10.

DMPONLINE, 2014. DMPonline: Data Management Plan Online Tool. https://dmponline.dcc.ac.uk/, 2014-04-20.

DMPTOOL, 2014. DMPTool: Guidance and Resources for your Data Management Plan. https://dmp.cdlib.org/, 2014-04-20.

DOBRATZ, S., 2007. Open-Source-Software zur Realisierung von Institutionellen Repositories - Überblick. Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) 54 (4/5), 199-206.

DONNELLY, M., 2011. Data management plans and planning. In: PRYOR, G. (Ed.), Managing Research Data Facet Publishing, London, UK, 83-104.

DOWNS, R.R., CHEN, R.S., LENHARDT, W.C., BOURNE, W., MILLMAN, D., 2007. Cooperative Management of a Long-Term Archive of Heterogeneous Scientific Data. Proceedings of the Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data (PV 2007), Munich, Germany.

DRAFTING TEAM METADATA AND EUROPEAN COMMISSION JOINT RESEARCH CENTRE, 2010. INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. European Commission Joint Research Centre, Ispra, Italy.

DREYER, M., TSCHIDA, U., 2009. eSciDoc – Das Repository-Konzept der Max Planck Digital Library. cms-journal: Wissenschaftliches Publizieren im digitalen Zeitalter 32 56-59.

DRYAD, 2014a. Dryad Digital Repository. http://datadryad.org/, 2014-05-01.

DRYAD, 2014b. Integrated Journals and Costs for Submitters. http://datadryad.org/pages/integratedJournals, 2014-05-04.

DRYAD DEVELOPMENT TEAM, 2010. Dryad Metadata Application Profile, Version 3.0. http://wiki.datadryad.org/wg/dryad/images/8/8b/Dryad3.0.pdf, 2014-04-15.

DSPACE, 2014a. DSpace Repository Software. http://dspace.org/introducing, 2014-04-10.

DSPACE, 2014b. DSpace Repository Software Use Cases. http://dspace.org/use-cases, 2014-04-10.

DUBOIS, P., 2013. MySQL: Developers Library. Addison-Wesley, Upper Saddle River, NJ, USA.

DUVAL, E., HODGINS, W., SUTTON, S., WEIBEL, S.L., 2002. Metadata Principles and Practicalities. D-Lib Magazine 8 (4).

DWD, 2014. Climatological Data. http://www.dwd.de/klimadaten, 2014-05-20.

EARTHSERVER, 2014. The EarthServer Initiative. http://www.earthserver.eu/, 2014-05-01.

ECONOMIST, T., 2010. The data deluge, The Economist, 25.02.2010 ed. The Economist Newspaper, London, UK.

EDINA, 2014. MANTRA Research Data Management Training. http://datalib.edina.ac.uk/mantra/, 2014-05-18.

EFFERTZ, E., 2010. The funder's perspective: Data management in coordinated programmes of the German Research Foundation (DFG). In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29.-30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 35-38. doi: 10.5880/TR32DB.KGA90.7.

EFFORTS, 2014. INF - Web-GIS Based Information System and Research Data Management. http://www.uni-goettingen.de/en/inf---web-gis-based-information-system-and-research-data-management/412103.html, 2014-05-05.

EHLERS, M., SCHIEWE, J., 2010. Geoinformatik. Wissenschaftliche Buchgesellschaft, Darmstadt, Germany.

ELINGS, M.W., WAIBEL, G., 2007. Metadata for all: Descriptive standards and metadata sharing across libraries, archives and museums. First Monday 12 (3). doi: 10.5210/fm.v12i3.1628.

ELMASRI, R.A., NAVATHE, S.B., 2014. Fundamentals of Database Systmes, Sixth Edition. Pearson Education Limited, Edinburgh, Harlow, Essex, UK.

ELSEVIER, 2010. Elsevier and PANGAEA Link Contents for easier access to full earth system research. http://www.elsevier.com/about/press-releases/science-and-technology/elsevier-and-pangaea-link-contents-for-easier-access-to-full-earth-system-research, 2014-05-04.

ELSEVIER, 2014. Guide for Authors. http://www.elsevier.com/journals/geomorphology/0169-555X/guide-for-authors#18000, 2014-03-28.

EMC, 2014. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. http://www.emc.com/leadership/digital-universe/2014iview/index.htm, 2014-04-28.

ENGELHARDT, C., 2013. Forschungsdatenmanagement in DFG-SFBs - Teilprojekte Informationsinfrastruktur (INF-Projekte). LIBREAS. Library Ideas, 23. http://edoc.hu-berlin.de/docviews/abstract.php?id=40316, 2013-10-14.

ENKE, H., FRITZSCH, B., 2013. Erfahrungen aus dem Projekt „WissGrid — Grid für die Wissenschaft". In: NEUROTH, H., LOSSAU, N., RAPP, A. (Eds.), Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft. vwh-Verlag, Universitätsverlag Göttingen, Glückstadt, Göttingen, Germany, 180-193.

ENSOM, T., WOLTON, A., 2012. RDE Metadata Profile for EPrints. UK Data Archive, University of Essex, Essex, UK.

EPRINTS, 2014a. EPrints Flexible Repository Software. http://www.eprints.org/software/, 2014-04-10.

EPRINTS, 2014b. EPrints Repository Examples. http://www.eprints.org/exemplar.php, 2014-04-10.

ERASMUS UNIVERSITEIT ROTTERDAM, 2014. Research life cycle. http://www.eur.nl/ub_informatievaardigheden/ul_instruction/osc/why_collaborate_online/when_to_collaborate_online/research_life_cycle/, 2014-04-14.

ERCIM NEWS, 2012. Special theme: Big Data. http://ercim-news.ercim.eu/images/stories/EN89/EN89-web.pdf, 2014-04-10.

ESF & EUROHORCs, 2008. The EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build It, European Science Foundation, Strasbourg, France. http://www.esf.org/fileadmin/Public_documents/Publications/SPB33_ERARoadMap.pdf, 2014-04-22.

ESFRI, 2014. European Strategy Forum on Research Infrastructures Roadmap. http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap, 2014-04-01.

ESRI, 2012. Was ist eine File-Geodatabase? http://help.arcgis.com/de/arcgisdesktop/10.0/help/index.html#//003n0000007m000000, 2014-05-02.

ESRI, 2014. ArcGIS for Server. http://www.esri.com/software/arcgis/arcgisserver, 2014-05-01.

ESSD, 2014. Earth System Science Data (ESSD). http://www.earth-system-science-data.net/, 2014-05-01.

EUROPEAN COMMISION, 2014a. Digital Agenda for Europe - Action 3: Open up public data resources for re-use. http://ec.europa.eu/digital-agenda/en/content/action-3-open-public-data-resources-re-use, 2014-04-01.

EUROPEAN COMMISION, 2014b. INSPIRE GEOPORTAL - Metadata Editor. http://inspire-geoportal.ec.europa.eu/editor/, 2014-05-15.

EUROPEAN COMMISSION, 2007a. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union L 108 (1), 1-14.

EUROPEAN COMMISSION, 2008. Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata. Official Journal of the European Union L 326 (12), 1-19.

EUROPEAN COMMISSION, 2012a. Commission Recommendation - on access to and preservation of scientific information. European Commission, Brussels, Belguim.

EUROPEAN COMMISSION, 2012b. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Towards better access to scientific information: Boosting the benefits of public investments in research. European Commission, Brussels, Belgium.

EUROPEAN COMMISSION, 2013a. Guidelines on Data Management in Horizon 2020. European Commision, Brussels, Belgium.

EUROPEAN COMMISSION, 2013b. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. European Commision, Brussels, Belgium.

EUROPEAN ENVIRONMENT INFORMATION AND OBSERVATION NETWORK (EIONET), 2012a. GEneral Multilingual Environmental Thesaurus (GEMET), Alphabetic listing of terms, GEMET - Concepts, version 3.1, 2012-07-20. http://www.eionet.europa.eu/gemet/alphabetic?langcode=en, 2014-02-24.

EUROPEAN ENVIRONMENT INFORMATION AND OBSERVATION NETWORK (EIONET), 2012b. INSPIRE Spatial Data Themes, GEMET - INSPIRE themes, version 1.0, 2008-06-01. http://www.eionet.europa.eu/gemet/inspire_themes?langcode=en, 2014-02-24.

FAO, 2014. GeoNetwork opensource: FAO Portal. http://www.fao.org/geonetwork, 2014-04-14.

FAUNDEEN, J.L., BURLEY, T.E., CARLINO, J.A., GOVONI, D.L., HENKEL, H.S., HOLL, S.L., HUTCHISON, V.B., MARTÍN, E., MONTGOMERY, E.T., LADINO, C.C., TESSLER, S., ZOLLY, L.S., 2014. The United States Geological Survey Science Data Lifecycle Model. USGS Open-File Report, U.S. Geological Survey, Reston, VA.

FEDORA, 2014a. Fedora Repository Project. http://fedorarepository.org/about, 2014-04-10.

FEDORA, 2014b. Fedora User Registry. http://registry.duraspace.org/registry/fedora, 2014-04-10.

FEHILY, C., 2008. Virtual Quickstart Guide: SQL. Peachpit Press, Berkeley, CA, USA.

FIENER, P., SEIBERT, S.P., AUERSWALD, K., 2011. A compilation and meta-analysis of rainfall simulation data on arable soils. Journal of Hydrology 409 (1–2), 395-406. doi: 10.1016/j.jhydrol.2011.08.034.

FIGSHARE, 2014. About figshare. http://figshare.com/, 2014-04-01.

FILEZILLA, 2014. FileZilla Overview. https://filezilla-project.org/, 2015-05-16.

FISCHER, P., HOFER, P., 2011. Lexikon der Informatik. Springer Berlin, Heidelberg, Germany. doi: 10.1007/978-3-642-15126-2.

FORCE11, 2014. Joint Declaration of Data Citation Principles. https://www.force11.org/datacitation, 2014-05-01.

FOREIGN & COMMONWEALTH OFFICE UK, 2013. News story: G8 Science Ministers Statement. https://www.gov.uk/government/news/g8-science-ministers-statement, 2014-04-14.

FOULONNEAU, M., RILEY, J., 2008. Metadata for Digital Resources: Implementation, Systems Design and Interoperability. Chandos Publishing, Oxford, UK.

FREYTAG, J.-C., 2014. Grundlagen und Visionen großer Forschungsfragen im Bereich Big Data. Informatik-Spektrum 37 (2), 97-104. doi: 10.1007/s00287-014-0771-y.

FRIEDHOFF, S., MEIER ZU VERL, C., PIETSCH, C., MEYER, C., VOMPRAS, J., LIEBIG, S., 2013a. Replicability and Comprehensibility of Social Research and its Technical Implementation. RatSWD Working Paper Series, SCIVERO Press, Berlin, Germany.

FRIEDHOFF, S., MEIER ZU VERL, C., PIETSCH, C., MEYER, C., VOMPRAS, J., LIEBIG, S., 2013b. Social Research Data: Documentation, Management, and Technical Implementation within the SFB 882. SFB 882 Working Paper Series, DFG Research Center (SFB) 882 From Heterogeneities to Inequalities, Bielefeld, Germany.

GEOLOGISCHER DIENST NRW, 2014. Informationssystem Bodenkarte von NRW 1 : 50 000. http://www.gd.nrw.de/g_bk50d.htm, 2014-05-15.

GEOSCIENCE DATA JOURNAL, 2014. Geoscience Data Journal. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049-6060, 2014-05-01.

GEOSERVER, 2014. What is GeoServer. http://geoserver.org/display/GEOS/What+is+GeoServer, 2014-05-01.

GFZ, 2014. DOIDB Metadata Store. https://doidb.wdc-terra.org/mds/, 2014-04-04.

GIBBS, H., 2009. DataShare Metadata Schema for ePrints Soton (ePrints 3.1) University of Southampton, Southampton, UK.

GIBNEY, E., VAN NOORDEN, R., 2013. Scientists losing data at a rapid rate. Nature News. doi: 10.1038/nature.2013.14416.

GIGASCIENCE, 2014. About GigaScience. http://www.gigasciencejournal.com/about, 2014-05-01.

GILCHRIST, P., 1986. Research Data Management. In: MCGAGHIE, W., FREY, J. (Eds.), Handbook for the Academic Physician. Springer, New York, USA, 251-276. doi: 10.1007/978-1-4684-6328-6_15.

GILLILAND, A.J., 2008. Setting the Stage In: BACA, M. (Ed.), Introduction to Metadata. Gregory M. Britton, Los Angeles, CA, 1-19.

GODBOLE, A.S., KAHATE, A., 2006. Web Technologies. Tata McGraw-Hill, New Delhi, India.

GODBY, C.J., YOUNG, J.A., CHILDRESS, E., 2004. A Repository of Metadata Crosswalks. D-Lib Magazine 10 (12). doi: 10.1045/december2004-godby.

GOOGLE, 2014. Google Maps API. https://developers.google.com/maps/.

GÖTTLICHER, D., BENDIX, J., 2004. Eine modulare Multi-User Datenbank für eine ökologische Forschergruppe mit heterogenem Datenbestand. Zeitschrift für Agrarinformatik 4 95-103.

GÖTTLICHER, D., DOBBERMANN, M., NAUSS, T., BENDIX, J., 2010. Central data services in multidisciplinary environmental research projects - the data-management of the DFG research unit 816. In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29. - 30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 59-64. doi: 10.5880/TR32DB.KGA90.11.

GRAF, A., PROLINGHEUER, N., SCHICKLING, A., SCHMIDT, M., SCHNEIDER, K., SCHUTTEMEYER, D., HERBST, M., HUISMAN, J.A., WEIHERMULLER, L., SCHARNAGL, B., STEENPASS, C., HARMS, R., VEREECKEN, H., 2011. Temporal Downscaling of Soil Carbon Dioxide Efflux Measurements Based on Time-Stable Spatial Patterns. Vadose Zone Journal 10 (1), 239-251. doi: 10.2136/vzj2009.0152.

GRANELL, C., GOULD, M., MANSO, M.Á., BERNABÉ, M.Á., 2009. Spatial Data Infrastructures. In: KARIMI, H.A.E. (Ed.), Handbook of Research on Geoinformatics. Information Science Reference / IGI Global Hershey, USA, 36-41. doi: 10.4018/978-1-59140-995-3.

GRAY, J., LIU, D.T., NIETO-SANTISTEBAN, M., SZALAY, A., DEWITT, D.J., HEBER, G., 2005. Scientific data management in the coming decade. SIGMOD Rec. 34 (4), 34-41. doi: 10.1145/1107499.1107503.

GREENBERG, J., 2003a. Metadata and the World Wide Web. Encyclopedia of Library and Information Science, Second Edition. Taylor & Francis, 1876-1888. doi: 10.1081/E-ELIS-120008663.

GREENBERG, J., 2003b. Metadata Generation: Processes, People and Tools. Bulletin of the American Society for Information Science and Technology 29 (2), 16-19. doi: 10.1002/bult.269.

GREENBERG, J., 2004. Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications. Journal of Internet Cataloging 6 (4), 59-82.

GREENBERG, J., 2005. Understanding Metadata and Metadata Schemes. Cataloging & Classification Quarterly 40 (3-4), 17-36. doi: 10.1300/J104v40n03_02.

GREENBERG, J., SWAUGER, S., FEINSTEIN, E., 2013. Metadata Capital in a Data Repository. Proccedings of the International Conference on Dublin Core and Metadata Applications, DC-2013, September 2-6, 2013, Lisbon, Portugal.

GREENBERG, J., WHITE, H.C., CARRIER, S., SCHERLE, R., 2009. A Metadata Best Practice for a Scientific Data Repository. Journal of Library Metadata 9 (3-4), 194-212. doi: 10.1080/19386380903405090.

GREENSTONE, 2014. Greenstone digital library software. http://www.greenstone.org/, 2014-04-14.

GREINER, C., SAKDAPOLRAK, P., 2013. Rural-urban migration, agrarian change, and the environment in Kenya: a critical review of the literature. Population and Environment 34 (4), 524-553. doi: 10.1007/s11111-012-0178-0.

GRIFFITHS, A., 2009. The Publication of Research Data: Researcher Attitudes and Behaviour. International Journal of Digital Curation 4 (1), 46-56.

GROBE, H., 2005. Data policy of the information system PANGAEA. doi: 10.1594/PANGAEA.327791.

GROBE, H., DIEPENBROEK, M., DITTERT, N., REINKE, M., SIEGER, R., 2006. Archiving and Distributing Earth-Science Data with the PANGAEA Information System. In: FÜTTERER, D., DAMASKE, D., KLEINSCHMIDT, G., MILLER, H., TESSENSOHN, F. (Eds.), Antarctica. Springer, Berlin Heidelberg, Germany, 403-406. doi: 10.1007/3-540-32934-X_51.

GROENEWEGEN, D., TRELOAR, A., 2013. Adding Value by Taking a National and Institutional Approach to Research Data: The ANDS Experience. International Journal of Digital Curation 8 (2), 89-98. doi: 10.2218/ijdc.v8i2.274.

GWK, 2013. Pressemitteilung: Ergebnisse der Sitzung der Gemeinsamen Wissenschaftskonferenz (GWK) am 22. November 2013. http://www.gwk-bonn.de/fileadmin/Pressemitteilungen/pm2013-16.pdf, 2014-04-15.

HAAK MARCIAL, L., HEMMINGER, B.M., 2010. Scientific data repositories on the Web: An initial survey. Journal of the American Society for Information Science and Technology 61 (10), 2029-2048. doi: 10.1002/asi.21339.

HALPIN, T., MORGAN, T., 2008. Information Modeling and Relational Databases. Morgan Kaufmann, Burlington, MA, USA.

HANSON, B., SUGDEN, A., ALBERTS, B., 2011. Making Data Maximally Available. Science 331 (6018), 649. doi: 10.1126/science.1203354.

HARTTER, J., RYAN, S.J., MACKENZIE, C.A., PARKER, J.N., STRASSER, C.A., 2013. Spatially Explicit Data: Stewardship and Ethical Challenges in Science. PLoS Biology 11 (9), e1001634. doi: 10.1371/journal.pbio.1001634.

HARVEY, R., 2011. Preserving Digital Materials. Walter de Gruyter, Berlin, Germany. doi: 10.1515/9783110253696.

HAUSSTEIN, B., QUITZSCH, N., JEUDE, K., SCHLEINSTEIN, N., ZENK-MÖLTGEN, W., 2013. da|ra Metadatenschema Version: 2.1.1. TECHNICAL Reports, GESIS – Leibniz-Institut für Sozialwissenschaften, Berlin, Köln, Germany. doi: 10.4232/10.mdsdoc.2.2.1.

HAYNES, D., 2004. Metadata for Information Management and Retrieval. Facet Publishing, London, UK.

HEERY, R., ANDERSON, S., 2005. Digital Repositories Review. http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf, 2014-04-26.

HEERY, R., PATEL, M., 2000. Application Profiles: Mixing and Matching Metadata Schemas. Ariadne 25.

HEIDERICH, M., 2009. Sichere Webanwendungen: das Praxishandbuch. Galileo Computing Bonn, Germany.

HEIMANN, D., KÖNIG-RIES, B., NIESCHULZE, J., 2010a. The Biodiversity-Exploratory Information System - Towards a service-oriented framework for knowledge-based data and toll integration. In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29. - 30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 65-73. doi: 10.5880/TR32DB.KGA90.12.

HEIMANN, D., NIESCHULZE, J., KÖNIG-RIES, B., 2010b. A flexible statistics web processing service - Added value for information systems for experiment data. Journal of Integrative Bioinformatics - JIB 7 (1). doi: 10.2390/biecoll-jib-2010-140.

HEISE, 2014. Open-Source-CMS: Sicherheitsupdate für Joomla. http://www.heise.de/open/meldung/Open-Source-CMS-Sicherheitsupdate-fuer-Joomla-2138003.html, 2014-05-20.

HELLY, J., STAUDIGEL, H., KOPPERS, A., 2003. Scalable models of data sharing in Earth sciences. Geochemistry, Geophysics, Geosystems 4 (1), 1010. doi: 10.1029/2002GC000318.

HELMHOLTZ-GEMEINSCHAFT, 2014a. Good scientific practice. http://www.helmholtz.de/en/research/good_scientific_practice/, 2014-04-14.

HELMHOLTZ-GEMEINSCHAFT, 2014b. Open Access in der Helmholtz-Gemeinschaft. http://oa.helmholtz.de/index.php?id=55, 2014-04-14.

HEUER, K., SIPPEL, R., 2004. UNIX-Systemadministration. Springer, Berlin, Heidelberg, Germany. doi: 10.1007/978-3-642-18945-6.

HEY, A.J.G., TREFETHEN, A., 2003. The Data Deluge: An e-Science Perspective In: BERMAN, F., FOX, G., HEY, A.J.G. (Eds.), Grid Computing - Making the Global Infrastructure a Reality. Wiley and Sons, West Sussex, UK, 809-824.

HEY, T., TANSLEY, S., TOLLE, K., 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, Washington, USA.

HEY, T., TREFETHEN, A.E., 2002. The UK e-Science Core Programme and the Grid. Future Generation Computer Systems 18 (8), 1017-1031. doi: 10.1016/S0167-739X(02)00082-1.

HICKS, E.A., PERKINS, J., MAURER, M.B., 2007. Application Profile Development for Consortial Digital Libraries. Library Resources & Technical Services 51 (2), 116-131.

HIDER, P., 2012. Information resource description : creating and managing metadata. Facet Publishing, London, UK.

HIGGINS, S., 2007. What are Metadata Standards. http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards, 2013-08-13.

HIGGINS, S., 2011. The lifecycle of data management. In: PRYOR, G. (Ed.), Managing Research Data Facet Publishing, London, UK, 17-46.

HIGH LEVEL EXPERT GROUP ON SCIENTIFIC DATA, 2010. Riding the wave - How Europe can gain from the rising tide of scientific data. European Union, Bruessles, Belgium.

HILBERT, M., LOPEZ, P., 2011. The world's technological capacity to store, communicate, and compute information. Science 332 (6025), 60-65. doi: 10.1126/science.1200970.

HIRWADE, M.A., 2011. A study of metadata standards. Library Hi Tech News 28 (7), 18 - 25. doi: 10.1108/07419051111184052.

HOBBS, J.R., PAN, F., 2006. Time Ontology in OWL. http://www.w3.org/TR/owl-time/, 2014-02-12.

HOCHSCHULREKTORENKONFERENZ, 2014. Management von Forschungsdaten - eine zentrale strategische Herausforderung für Hochschulleitungen, Empfehlung der 16. HRK-Mitgliederversammlung am 13.5.2014 n Frankfurt am Main. http://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf, 2014-05-19.

HODSON, S., MOLLOY, L., 2013. Case Study 5: development of institutional RDM services by projects in the Jisc Managing Research Data programmes. In: PRYOR, G., JONES, S., WHITE, A. (Eds.), Delivering Research Data Management Services: Fundamentals of Good Practice. Facet, London, UK, 205-237.

HOERNES, U., ROWECK, H., 2007. Informations- und Datenmanagement am Beispiel einer Langzeitstudie zum Ökologischen Landbau ("Hof Ritzerau", Norddeutschland). Proccedings of the Zwischen Tradition und Globalisierung - 9. Wissenschaftstagung Ökologischer Landbau, March 20-23, 2007, Universität Hohenheim, Stuttgart, Deutschland.

HOFFMEISTER, D., BOLTEN, A., CURDT, C., WALDHOFF, G., BARETH, G., 2008. High-resolution Crop Surface Models (CSM) and Crop Volume Models (CVM) on field level by terrestrial laser scanning, In: GUO, H., WANG, C. (Eds.), Sixth International Symposium on Digital Earth: Models, Algorithms, and Virtual Reality. SPIE, Beijing, China, 78400E. doi: 10.1117/12.872315.

HOFFMEISTER, D., WALDHOFF, G., CURDT, C., TILLY, N., BENDIG, J., BARETH, G., 2013. Spatial variability detection of crop height in a single field by terrestrial laser scanning. In: STAFFORD, J. (Ed.), Precision agriculture '13. Wageningen Academic Publishers, 267-274. doi: 10.3920/978-90-8686-778-3_31.

HOLDREN, J., 2013. Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research The White House, Executive Office of the President, Office of Science and Technology Policy, Washington, DC, USA.

HOWARD, J.H., 1988. An Overview of the Andrew File System. Proccedings of the USENIX Winter Conference, Dallas, Texas, USA.

HOX, J.J., BOEIJE, H.R., 2005. Data collection, primary versus secondary. In: KEMPF-LEONARD, K. (Ed.), Encyclopedia of social measurement. Elsevier, Amsterdam, The Netherlands, 593-599.

HUDGINS, J., AGNEW, G., BROWN, E., 1999. Getting mileage out of metadata: applications in the library. American Library Association, Chicago, IL, USA.

HUMBOLDT UNIVERSITY OF BERLIN, 2014. Forschungsdatenmanagement an der Humboldt-Universität zu Berlin. https://www.cms.hu-berlin.de/ueberblick/projekte/dataman, 2014-04-10.

HUMPHREY, C., 2008. E-Science and the Life Cycle of Research. http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc, 2013-09-27.

HUSCHKA, D., OELLERS, C., OTT, N., WAGNER, G.G., 2011. Datenmanagement und Data Sharing: Erfahrungen in den Sozial- und Wirtschaftswissenschaften, RatSWD Working Paper Series. Rat für Sozial- und Wirtschaftsdaten, Berlin, Germany, 18.

IBM, 2000. OpenAFS User Guide. http://docs.openafs.org/UserGuide, 2014-05-01.

IBM, 2014. Network attached storage (NAS): IBM NAS for your storage infrastructure. http://www-03.ibm.com/systems/storage/network/, 2015-05-05.

IDF, 2014. The DOI® System. http://www.doi.org/index.html, 2014-05-01.

ILLIK, J.A., 2007. Verteilte Systeme : Architekturen und Software-Technologien. expert verlag, Renningen, Germany.

IMPACTSTORY, 2014. About Impactstory. https://impactstory.org/about, 2014-05-01.

IMPETUS, 2014. IMPETUS Portal. http://geonetwork.impetus.uni-koeln.de, 2014-04-13.

INSPIRE, 2014. HEP Search: High-Energy Physics Literature Database. http://inspirehep.net/, 2014-04-10.

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2012. Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.). Ann Arbor, MI, USA.

INVENIO, 2014a. About INVENIO. http://invenio-software.org/, 2014-04-10.

INVENIO, 2014b. INVENIO Demo and Example Sites. http://invenio-software.org/wiki/General/Demo, 2014-04-10.

ISO, 2003. ISO 19115:2003 Geographic information - Metadata. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020, 2014-02-05.

ISO, 2005. ISO 19115:2005 Geographic information - Metadata. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=39890, 2014-02-05.

ISO, 2006. ISO 19115:2003/Cor 1:2006 Geographic information - Metadata, Technical Corrigendum 1 International Organization for Standardization (ISO), Geneva, Switzerland.

ISO, 2010. ISO 639-2:1998 Codes for the representation of names of languages - Part 2: Alpha-3 code. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=4767, 2014-02-05.

ISO, 2013. ISO 8601:2004 Data elements and interchange formats - Information interchange - Representation of dates and times. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=40874, 2014-02-05.

ISO, 2014a. ISO 15836:2009: Information and documentation-The Dublin Core metadata element set. http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on, 2014-05-15.

ISO, 2014b. ISO/TS 19139:2007 Geographic information - Metadata - XML schema implementation http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32557, 2014-05-20.

ITNEWS, 2013. Inside eBay's 90PB data warehouse. http://www.itnews.com.au/News/342615,inside-ebay8217s-90pb-data-warehouse.aspx, 2014-04-20.

JANßEN, T., SCHMIDT, M., DRESSLER, S., HAHN, K., HIEN, M., KONATÉ, S., LYKKE, A.M., MAHAMANE, A., SAMBOU, B., SINSIN, B., THIOMBIANO, A., WITTIG, R., ZIZKA, G., 2011. Addressing data property rights concerns and providing incentives for collaborative data pooling: the West African Vegetation Database approach. Journal of Vegetation Science 22 (4), 614-620. doi: 10.1111/j.1654-1103.2011.01271.x.

JENA EXPERIMENT, 2014. The Jena Experiment - Data. http://www.the-jena-experiment.de/Data.html, 2014-05-05.

JENSEN, J.R., 2014. Remote sensing of the environment: an earth resource perspective. Pearson Education, Harlow, UK.

JENSEN, U., 2011. Datenmanagementpläne. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 71-82.

JENSEN, U., KATSANIDOU, A., ZENK-MÖLTGEN, W., 2011. Metadaten und Standards. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 83-100.

JESCHKE, S., BURR, B., HAHN, J.U., HELMES, L., KRIHA, W., KRÜGER, M., LIEHR, A.W., OSTEN, W., PFEIFFER, O., RICHTER, T., SCHNEIDER, G., STEPHAN, W., WEBER, K.-H., 2009. Networking Resources for Research and Scientific Education in BW-eLabs. Proccedings of the 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, Catholic University of Daegu, Korea, 47-52.

JISC, 2014a. Digital infrastructure: Research management programme. http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement.aspx, 2014-04-10.

JISC, 2014b. Research Lifecycle diagram. www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx, 2014-04-14.

JOHNS HOPKINS UNIVERSITY, 2008. Johns Hopkins University Policy on Access and Retention of Research Data and Materials. http://jhuresearch.jhu.edu/Data_Management_Policy.pdf, 2014-04-20.

JOHN HOPKINS UNIVERSITY, 2014. Research Data Management Services at John Hopkins University. http://dmp.data.jhu.edu/, 2014-04-10.

JOHNSTON, L., 2014. A Workflow Model for Curating Research Data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot. University of Minnesota Digital Conservancy, Minneapolis, MN, USA.

JONES, S., 2011. Research data policies: principles, requirements and trends. In: PRYOR, G. (Ed.), Managing Research Data Facet Publishing, London, UK, 47-66.

JONES, S., 2012. Developments in Research Funder Data Policy. International Journal of Digital Curation 7 (1). doi: 10.2218/ijdc.v7i1.219.

JONES, S., 2013. The range and components of RDM infrastrucutre and services. In: PRYOR, G., JONES, S., WHITE, A. (Eds.), Delivering Research Data Management Services: Fundamentals of Good Practice. Facet, London, UK, 89-114.

JOSHI, M., KRAG, S., 2010. Issues in Data Management. Science and Engineering Ethics 16 (4), 743-748. doi: 10.1007/s11948-010-9223-5.

JUSER, 2014. Jülich Shared Electronic Resources (JuSER). http://juser.fz-juelich.de/, 2014-04-10.

KALLE, C., 2005. Datenhaltung am ZAIK/RRZK. Kolloquium: Ausgewählte Themen der Datenverarbeitung, RRZK, Cologne.

KALLE, C., DIEKMANN, L., 2010. Kooperation von der Datenhaltung im AFS bis zur private Cloud – prometheus Bilddatenbanken im zeitlichen Wandel, IT an der UNIVERSITÄT ZU KÖLN. Der Rektor der Universität zu Köln, Köln.

KALLE, C., TIGGEMANN, D., 2007. Parallel beschleunigt - HPC-Fileserving mit Panasas, iX. Heise Zeitschriften Verlag, Hannover, 96-99.

KAMAL, R., 2002. Internet and Web Technologies. Tata McGraw-Hill, New Delhi, India.

KANNENGIESSER, M., KANNENGIESSER, C., 2007. PHP5 / MySQL 5 Master Edition. Franzis, Poing, Germany.

KAPPAS, M., 2012. Geographische Informationssysteme. Westermann, Braunschweig, Germany.

KARASTI, H., BAKER, K.S., MILLERAND, F., 2010. Infrastructure Time: Long-term Matters in Collaborative Development. Comput. Supported Coop. Work 19 (3-4), 377-415. doi: 10.1007/s10606-010-9113-z.

KATTGE, J., OGLE, K., BÖNISCH, G., DÍAZ, S., LAVOREL, S., MADIN, J., NADROWSKI, K., NÖLLERT, S., SARTOR, K., WIRTH, C., 2011. A generic structure for plant trait databases. Methods in Ecology and Evolution 2 (2), 202-213. doi: 10.1111/j.2041-210X.2010.00067.x.

KDMT, 2014. Data Management of Kiel Marine Sciences. https://portal.geomar.de/about-us, 2014-04-10.

KECK, M., SAKDAPOLRAK, P., 2013. What is social resilience? Lessons learned and ways forward. Erdkunde 67 (1), 5-19. doi: 10.3112/erdkunde.2013.01.02.

KEMPER, A., EICKLER, A., 2009. Datenbanksysteme - Eine Einführung. Oldenbourg Wissenschaftsverlag GmbH München, Germany.

KENNEDY, M.R., 2008. Nine questions to guide you in choosing a metadata schema. Journal of Digital Information 9 (1).

References

KIT, 2014. Forschungsdatenmanagement am KIT. http://www.bibliothek.kit.edu/cms/forschungsdatenmanagement.php, 2014-04-10.

KLAR, J., ENKE, H., 2013. Forschungsdaten in der Gruppendomäne – Zwischen individuellen Anforderungen und übergreifenden Infrastrukturen. Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) 60 (6), 316-324.
doi: 10.3196/186429501360649.

KLEINSCHMIDT, P., RANK, C., 2005. Relationale Datenbanksysteme - Eine praktische Einführung. Springer, Berlin Heidelberg, Germany. doi: 10.1007/b138052.

KLUG, S., 2006. Don't Make Me Think! A common sense approach to web usability. New Riders, Berkeley, CA, USA.

KLUMP, J., 2009. Registrierung von Datenpublikationen im GFZ Potsdam über DOIDB, Technical Report. GFZ, Potsdam, Germany.

KLUMP, J., 2010. Digitale Forschungsdaten. In: NEUROTH, H., OßWALD, A., SCHEFFEL, R., STRATHMANN, S., HUTH, K. (Eds.), nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Verlag Werner Hülsbusch, Boizenburg, Germany.

KLUMP, J., BERTELMANN, R., 2013. Forschungsdaten. In: KUHLEN, R., SEMAR, W., STRAUCH, D. (Eds.), Grundlagen der praktischen Information und Dokumentation. De Gruyter Saur, Munich, Germany, 575-583.
doi: 10.1515/9783110258264.575.

KLUMP, J., BERTELMANN, R., BRASE, J., DIEPENBROEK, M., GROBE, H., HÖCK, H., LAUTENSCHLAGER, M., SCHINDLER, U., SENS, I., WÄCHTER, J., 2006. Data publication in the open access initiative. Data Science Journal 5 79-83.

KLUMP, J., LUDWIG, J., 2013. Forschungsdaten-Management. In: NEUROTH, H., LOSSAU, N., RAPP, A. (Eds.), Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft. vwh-Verlag, Universitätsverlag Göttingen, Glückstadt, Göttingen, Germany, 257-275.

KNOWLEDGE EXCHANGE, 2014a. Knowledge Exchange - Research Data Activities.
http://www.knowledge-exchange.info/Default.aspx?ID=284, 2014-04-15.

KNOWLEDGE EXCHANGE, 2014b. Knowledge Exchange in brief.
http://www.knowledge-exchange.info/Default.aspx?ID=68, 2014-04-15.

KOFLER, M., ÖGGL, B., 2010. PHP 5.3 & MySQL 5.4: Programmierung, Administration, Praxisprojekte. Addison-Wesley, Munich, Germany.

KOHL, A., 2011. ActionScript 2. Springer, Berlin Heidelberg, Germany. doi: 10.1007/978-3-540-35182-5.

KOMMISSION ZUKUNFT DER INFORMATIONSINFRASTRUKTUR, 2011. Gesamtkonzept für die Informationsinfrastruktur in Deutschland.

KORRES, W., KOYAMA, C.N., FIENER, P., SCHNEIDER, K., 2010. Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions. Hydrology and Earth System Sciences 14 (5), 751-764.
doi: 10.5194/hess-14-751-2010.

KORRES, W., REICHENAU, T.G., SCHNEIDER, K., 2013. Patterns and scaling properties of surface soil moisture in an agricultural landscape: An ecohydrological modeling study. Journal of Hydrology 498 (0), 89-102.
doi: 10.1016/j.jhydrol.2013.05.050.

KOUDELA, D., KÖHLER, K., MÜLLER-PFEFFERKORN, R., 2013. Report "Technik". GFZ Potsdam, Potsdam, Germany. doi: 10.2312/RADIESCHEN_004.

KRALISCH, S., ZANDER, F., 2010. Environmental Data Management with the River Basin Information System (RBIS). In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29. - 30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 83-91.
doi: 10.5880/TR32DB.KGA90.14.

KRALISCH, S., ZANDER, F., KRAUSE, P., 2009. Coupling the RBIS environmental information system and the JAMS modelling framework In: ANDERSSEN, R.S., BRADDOCK, R.D., NEWHAM, L.T.H. (Eds.), 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia, 902-908.

KRAMER, B., 2014. Umgang mit Daten: Der gläserne Forscher. duz Magazin 3 (14).

KRATZ, J., STRASSER, C.A., 2014. Data publication consensus and controversies [v1; ref status: approved with reservations 1]. F1000Research 2014 3 (94). doi: 10.12688/f1000research.4264.

KRESSE, W., FADAIE, K., 2004. ISO Standards for Geographic Information. Springer, Heidelberg, Berlin, Germany.

KROES, N., 2011. Data is the new gold. Opening Remarks given at Press Conference on Open Data Strategy, http://europa.eu/rapid/press-release_SPEECH-11-872_en.htm, 2013-12-17.

KUBEREK, M., 2012. Organisatorisch-technisches Konzept für eine Forschungsdaten-Infrastruktur in der TU Berlin. TU Berlin, Berlin, Germany.

KUBEREK, M., 2013. Die Forschungsdaten-Infrastruktur der TU Berlin. Bibliotheksdienst 47 (11), 833-846.
doi: 10.1515/bd-2013-0096.

KUNKEL, R., SORG, J., ECKARDT, R., KOLDITZ, O., RINK, K., VEREECKEN, H., 2013. TEODOOR: a distributed geodata infrastructure for terrestrial observation data. Environmental Earth Sciences 69 (2), 507-521.
doi: 10.1007/s12665-013-2370-7.

KUNKEL, R., SORG, J., GASCHE, R., KLUMP, J., KOLDITZ, O., FRENZEL, M., NEIDL, F., 2012. TEODOOR: Geodateninfrastruktur zur Verwaltung und Veröffentlichung von terrestrischen Beobachtungsdaten der HGF Infrastrukturmaßnahme TERENO aus

verteilten Quellen. In: BALL, R. (Ed.), WissKom 2012: Vernetztes Wissen – Daten, Menschen, Systeme. Forschungszentrum, Zentralbibliothek, Jülich, 75-92.

LAGOZE, C., PAYETTE, S., SHIN, E., WILPER, C., 2005. Fedora: an architecture for complex objects and their relationships. International Journal on Digital Libraries 6 (2), 124-138. doi: 10.1007/s00799-005-0130-3.

LANDESREGIERUNG NRW, 2004. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz Nordrhein-Westfalen (Barrierefreie Informationstechnik-Verordnung Nordrhein-Westfalen - BITV NRW). http://www.nw3.de/rechtsarchiv/4oerecht/1bgg/BGG/Ver_BGG_NRW.html, 2014-05-16.

LANDESREGIERUNG NRW, 2009. Gesetz über den Zugang zu digitalen Geodaten Nordrhein-Westfalen (Geodatenzugangsgesetz - GeoZG NRW). Düsseldorf, Germany.

LAUDATIO, 2014. LAUDATIO - Long-term Access and Usage of Deeply Annotated Information. http://www.laudatio-repository.org, 2014-04-10.

LAWRENCE, B., JONES, C., MATTHEWS, B., PEPLER, S., CALLAGHAN, S., 2011. Citation and Peer Review of Data: Moving Towards Formal Data Publication. International Journal of Digital Curation 6 (2), 4-37.

LAWRENCE, B.N., LOWRY, R., MILLER, P., SNAITH, H., WOOLF, A., 2009. Information in environmental data grids. Philos Trans A Math Phys Eng Sci 367 (1890), 1003-1014. doi: 10.1098/rsta.2008.0237.

LEIBNIZ-GEMEINSCHAFT, 2014. Forschungsdatenzentren der Leibniz-Gemeinschaft. http://www.leibniz-gemeinschaft.de/infrastrukturen/forschungsdaten/forschungsdatenzentren, 2014-04-15.

LEIBNIZ UNIVERSITY OF HANOVER, 2014. Forschungsdatenmanagement at the Leibniz University of Hannover. http://www.dezernat4.uni-hannover.de/forschungsdaten.html, 2014-04-10.

LENK, A., KLEMS, M., NIMIS, J., TAI, S., SANDHOLM, T., 2009. What's inside the Cloud? An architectural map of the Cloud landscape. Proccedings of the Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing, 23-31.

LEONELLI, S., 2013. Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology. Bulletin of Science, Technology & Society. doi: 10.1177/0270467613496768.

LEONELLI, S., SMIRNOFF, N., MOORE, J., COOK, C., BASTOW, R., 2013. Making open data work for plant scientists. Journal of Experimental Botany 64 (14), 4109-4117. doi: 10.1093/jxb/ert273.

LEONG, J., 1994. Project Andrew. In: KHANNA, R. (Ed.), Distributed Computing: Implementation and Management Strtegies. PTR Prentice-Hall, Eaglewood Cliffs, NJ, USA, 203-220.

LEVI, P., REMBOLD, U., 2003. Einführung in die Informatik für Naturwissenschaftler und Ingenieure. Carl Hanser, Munich, Germany.

LEVINE, M., 2014. Copyright, Open Data, and the Availability-Usability Gap: Challenges, Opportunities, and Approaches for Libraries. In: RAY, J. (Ed.), Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, IN, USA, 129-147.

LIBRARY CONNECT, 2013. Theme: Research Data Management. Library Connect Newsletter 11 (1).

LIBRARY OF CONGRESS, 2014. Web Archiving FAQs. http://www.loc.gov/webarchiving/faq.html#5, 2014-04-20.

LILLESAND, T.M., KIEFER, R.W., CHIPMAN, J.W., 2008. Remote sensing and image interpretation. Wiley, New York, NY, USA.

LINNE, M., 2013. Sustainable Data Preservation using datorium – facilitating the Scientific Ideal of Data Sharing in the Social Sciences. In: BORBINHA, J., MICHAEL NELSON, KNIGHT, S. (Eds.), Proceedings of the 10th International Conference on Preservation of Digital Objects (iPres2013). Biblioteca Nacional de Portugal, Lisboa, Portugal.

LITWIN, L., ROSSA, M., 2011. Geoinformation Metadata in INSPIRE and SDI: Understanding. Editing. Publishing. Springer, Berlin Heidelberg, Germany. doi: 10.1007/978-3-642-15862-9.

LOESER, H., 1998. Techniken für Web-basierte Datenbankanwendungen: Anforderungen, Ansätze, Architekturen. Informatik Forschung und Entwicklung 13 (4), 196-216. doi: 10.1007/s004500050113.

LOTZ, T., NIESCHULZE, J., BENDIX, J., DOBBERMANN, M., KÖNIG-RIES, B., 2012. Diverse or uniform? — Intercomparison of two major German project databases for interdisciplinary collaborative functional biodiversity research. Ecological Informatics 8 10-19. doi: 10.1016/j.ecoinf.2011.11.004.

LUBAS, R.L., JACKSON, A.S., SCHNEIDER, I., 2013. The metadata manual: a practical workbook. Chandos Publishing, Oxford, UK.

LUDWIG, J., 2012. Zusammenfassung und Interpretation. In: NEUROTH, H., STRATHMANN, S., OßWALD, A., SCHEFFEL, R., KLUMP, J., LUDWIG, J. (Eds.), Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. Verlag Werner Hülsbusch, Boizenburg, 295-310.

LUDWIG, J., ENKE, H., 2013. Leitfaden zum Forschungsdaten-Management: Handreichungen aus dem WissGrid-Projekt. Verlag Werner Hülsbusch, Glückstadt, Germany.

LUSSEM, U., WALDHOFF, G., 2013. Land use classification 2012 of the Rur Catchment, CRC/TR32 Database (TR32DB). doi: 10.5880/TR32DB.9.

LÜTZENKIRCHEN, F., 2002. MyCoRe - Ein Open-Source-System zum Aufbau digitaler Bibliotheken. Datenbank Spektrum 2 (4), 23-27.

LYNCH, C., 2008. Big data: How do your data grow? Nature 455 (7209), 28-29.

MA, J., 2006. Managing metadata for digital projects. Library Collections, Acquisitions, and Technical Services 30 (1–2), 3-17. doi: 10.1016/j.lcats.2006.07.001.

MANDL, P., 2013. Grundkurs Betriebssysteme: Architekturen, Betriebsmittelverwaltung, Synchronisation, Prozesskommunikation. Springer Fachmedien Wiesbaden, Wiesbaden, Germany.

MAPSERVER, 2014. Welcome to MapServer. http://mapserver.org/, 2014-05-01.

MARILL, J.L., LUCZAK, E.C., 2009. Evaluation of Digital Repository Software at the National Library of Medicine. D-Lib Magazine 15 (5/6).

MARIOS, D.D., GEORGE, P., DIMITRIOS, K., PANKAJ, M., ATHENA, V., 2009. Cloud Computing: Distributed Internet Computing for IT and Scientific Research, 10-13.

MÄRKER, M., BOLUS, M., KANAEVA, Z., 2011. Spatial assessment of early human expansions using GIS and Database Techniques: Examples from Southern Africa and Eurasia, In: MACCHIARELLI, R., WENIGER, G.-C. (Eds.), Pleistocene Databases: Acquisition, Storing, Sharing. Neanderthal Museum, Mettmann, Germany, 69-75.

MÄRKER, M., KANAEVA, Z., 2010. ROAD: The role of culture in early expansion of humans data base. In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29. - 30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 93-97. doi: 10.5880/TR32DB.KGA90.15.

MARX, V., 2012. My data are your data. Nature Biotechnology 30 (6), 509-511. doi: 10.1038/nbt.2243.

MATTHEWS, B., SUFI, S., FLANNERY, D., LERUSSE, L., GRIFFIN, T., GLEAVES, M., KLEESE, K., 2010. Using a Core Scientific Metadata Model in Large-Scale Facilities. International Journal of Digital Curation 5 (1), 106-118.

MCLAUGHLIN, M., 2010. MySQL workbench: data modeling & development. McGraw Hill, New York, Chicago, San Francisco, USA.

MCNALLY, R., MACKENZIE, A., HUI, A., TOMOMITSU, J., 2012. Understanding the 'Intensive' in 'Data Intensive Research': Data Flows in Next Generation Sequencing and Environmental Networked Sensors. International Journal of Digital Curation 7 (1). doi: 10.2218/ijdc.v7i1.216.

MICHENER, W.K., 2006. Meta-information concepts for ecological data management. Ecological Informatics 1 (1), 3-7.

MICHENER, W.K., BRUNT, J.W., HELLY, J.J., KIRCHNER, T.B., STAFFORD, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications 7 (1), 330-342. doi: 10.2307/2269427.

MICHENER, W.K., VIEGLAIS, D., VISION, T., KUNZE, J., CRUSE, P., JANÉE, G., 2011. DataONE: Data Observation Network for Earth — Preserving Data and Enabling Innovation in the Biological and Environmental Sciences. D-Lib Magazine 17 (1/2). doi: 10.1045/january2011-michener.

MILICCHIO, F., GEHRKE, W.A., 2007. Distributed services with OpenAFS: for enterprise and education. Springer, Berlin, Germany.

MILLER, S.J., 2011. Metadata for digital collections : a how-to-do-it manual. Neal-Schuman Publishers, Inc., New York, USA.

MIT, 2014. Data Management and Publishing at the MIT. http://libraries.mit.edu/guides/subjects/data-management/index.html, 2014-04-10.

MITCHELL, T., EMDE, A., 2008. Web-Mapping mit Open Source-GIS-Tools. O'Reilly, Beijing, China.

MONASH UNIVERSITY, 2014. Research data management at the Monash University. http://monash.edu/library/researchdata/index.html, 2014-04-10.

MONASTERIUM, 2014. Monasterium. http://www.mom-ca.uni-koeln.de/mom/home, 2014-03-15.

MORRIS, J.H., SATYANARAYANAN, M., CONNER, M.H., HOWARD, J.H., ROSENTHAL, D.S., SMITH, F.D., 1986. Andrew: a distributed personal computing environment. Communications of the ACM 29 (3), 184-201. doi: 10.1145/5666.5671.

MPDL, 2014. Max Planck Digital Library. http://www.mpdl.mpg.de, 2014-04-10.

MPG, 2009. Regeln zur Sicherung guter wissenschaftlicher Praxis beschlossen vom Senat der Max-Planck-Gesellschaft am 24. November 2000, geändert am 20. März 2009. Max-Planck-Gesellschaft, Munich, Germany.

MÜCKSCHEL, C., NIESCHULZE, J., SCHACHTEL, G.A., LI, S., SLOBODA, B., KÖHLER, W., 2004. Web-basierte Informationssysteme in interdisziplinären Umwelt-Forschungsprojekten – am Beispiel der beiden DFG-Sonderforschungsbereiche 299 (Giessen) und 552 (Göttingen/Kassel). Zeitschrift für Agrarinformatik 3 46-56.

MÜCKSCHEL, C., NIESCHULZE, J., WEIST, C., SLOBODA, B., KÖHLER, W., 2007. Herausforderungen, Probleme und Lösungsansätze im Datenmanagement von Sonderforschungsbereichen. eZAI 2 1-16.

MÜCKSCHEL, C., SCHACHTEL, G.A., KÖHLER, W., 2005. Management of metadata in an interdisciplinary environmental research project, EFITA. EFITA, Vila Real, Portugal.

MÜCKSCHEL, C., SCHACHTEL, G.A., WEHRUM, A., NIESCHULZE, J., KÖHLER, W., 2006. Grundlegende Anforderungen an das Datenmanagement in interdisziplinären Forschungsprojekten. Proccedings of the 26. Gil Jahrestagung, Potsdam, Germany, 185-188.

MÜCKSCHEL, C., WEIST, C., KÖHLER, W., 2008. Central data management in environmental research projects - selected problems and solutions. Proccedings of the 28. Gil Jahrestagung, Kiel, Germany, 101-104.

MULLINS, J., 2014. The Policy and Institutional Framework. In: RAY, J. (Ed.), Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, IN, USA, 25-44.

MYCORE, 2014. MyCoRe Applications: Alle Standorte und Projekte. http://www.mycore.de/applications/list.html, 2014-04-10.

MYSQL, 2014a. Chapter 11 Data Types. http://dev.mysql.com/doc/refman/5.7/en/data-types.html, 2014-03-03.

MYSQL, 2014b. MySQL - The world's most popular open source database. http://www.mysql.com/, 2014-03-04.

MYSQL, 2014c. MySQL Workbench. http://www.mysql.com/products/workbench/, 2014-03-04.

MYSQL, 2014d. What is MySQL? http://dev.mysql.com/doc/refman/5.7/en/what-is-mysql.html, 2014-03-04.

NADROWSKI, K., RATCLIFFE, S., BÖNISCH, G., BRUELHEIDE, H., KATTGE, J., LIU, X., MAICHER, L., MI, X., PRILOP, M., SEIFARTH, D., WELTER, K., WINDISCH, S., WIRTH, C., 2013. Harmonizing, annotating and sharing data in biodiversity–ecosystem functioning research. Methods in Ecology and Evolution 4 (2), 201-205. doi: 10.1111/2041-210x.12009.

NADROWSKI, K., SEIFARTH, D., RATCLIFFE, S., WIRTH, C., MAICHER, L., 2012. Identifiers in e-Science platforms for the ecological sciences, In: KÖHLER, T., KAHNWALD, N. (Eds.), Communities in New Media: Virtual Enterprises, Research Communities & Social Media Networks. Proceedings of GeNeMe 2012. TUDpress, Dresden, Germany, 259-272.

NASA, 2014. Global Change Master Directory: Discover Earth science data and services. http://gcmd.gsfc.nasa.gov/index.html, 2014-04-10.

NATIONAL ARCHIVES, 2014. Open Government Licence for public sector information. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/, 2014-05-19.

NATIONAL SCIENCE BOARD, 2005. Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf, 2014-05-01.

NATIONAL SCIENCE BOARD, 2011. Digital Research Data Sharing and Management. https://www.nsf.gov/nsb/publications/2011/nsb01211.pdf, 2014-05-20.

NATIONAL SCIENCE FOUNDATION, 2007. Cyberinfrastructure Vision for 21st Century Discovery. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf, 2013-09-02.

NATIONAL SCIENCE FOUNDATION, 2011a. Chapter II - Proposal Preparation Instructions. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp, 2014-04-15.

NATIONAL SCIENCE FOUNDATION, 2011b. NSF Data Management Plan Requirements. http://www.nsf.gov/eng/general/dmp.jsp, 2013-10-01.

NATIONAL SCIENCE FOUNDATION, 2013. Chapter VI - Other Post Award Requirements and Considerations. http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4, 2014-04-15.

NATURE, 2008. Big Data: Science in the Petabyte Era. http://www.nature.com/nature/journal/v455/n7209/index.html#ed, 2014-04-20.

NATURE, 2009. Data - What data? http://www.nature.com/nature/journal/v461/n7261/index.html#ed, 2014-04-20.

NAUMANN, C., 2014. Stability and Transformation in a South African Landscape: Rural Livelihoods, Governmental Interventions and Agro-Economic Change in Thaba Nchu. Journal of Southern African Studies 40 (1), 41-57. doi: 10.1080/03057070.2014.877653.

NAUSS, T., GÖTTLICHER, D., DOBBERMANN, M., BENDIX, J., 2007. Central Data Services in Multidisciplinary Environmental Research Projects. eZAI 2 1-13.

NEHER, G., RITSCHEL, B., 2011. Semantische Vernetzung von Forschungsdaten. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 169-190.

NELSON, B., 2009. Empty archives. Nature 461 (7261), 160-163. doi: 10.1038/461160a.

NERC, 2014a. Data management planning. http://www.nerc.ac.uk/research/sites/data/dmp/, 2014-04-20.

NERC, 2014b. NERC Data centres. http://www.nerc.ac.uk/research/sites/data/, 2014-04-10.

NESTOR, 2014. Nestor - ein deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung. http://www.langzeitarchivierung.de, 2014-04-15.

NETAPP, 2014. Network-Attached Storage (NAS). http://www.netapp.com/de/products/protocols/nas/index.aspx, 2015-05-05.

NETWORK WORKING GROUP, 2007. The Dublin Core Metadata Element Set. http://www.ietf.org/rfc/rfc5013.txt, 2014-05-15.

NEUROTH, H., 2013. Die wissenschaftliche Bibliothek im Kontext von Forschungsinfrastrukturen. In: NEUROTH, H., LOSSAU, N., RAPP, A. (Eds.), Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft. vwh-Verlag, Universitätsverlag Göttingen, Glückstadt, Göttingen, Germany, 325-344.

NEUROTH, H., OßWALD, A., SCHEFFEL, R., STRATHMANN, S., HUTH, K., 2010. nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Verlag Werner Hülsbusch, Boizenburg, Germany.

NHMRC, 2007. Australian Code for the Responsible Conduct of Research, Australian Government, National Health and Medical Research Council, Canberra, ACT, Australia. http://www.nhmrc.gov.au/guidelines/publications/r39.

NIELSEN, J., 1997. How Users Read on the Web. http://www.nngroup.com/articles/how-users-read-on-the-web/, 2014-05-05.

NIH, 2014. NIH Data Sharing Policy. http://grants.nih.gov/grants/policy/data_sharing/, 2014-04-15.

NISO, 2004. Understanding Metadata. Bethesda, MD, USA. http://www.niso.org/standards/resources/UnderstandingMetadata.pdf, 2012-08-20.

NISO, 2008. Building a metadata schema – where to start. Baltimore, Maryland, USA. http://www.niso.org/apps/group_public/download.php/5271/N800R1_Where_to_start_advice_on_creating_a_metadata_schema.pdf, 2013-05-16.

NISO, 2012. The Dublin Core Metadata Element Set. http://www.niso.org/apps/group_public/download.php/10256/Z39-85-2012_dublin_core.pdf, 2013-05-16.

NOGUERAS-ISO, J., ZARAZAGA-SORIA, F.J., LACASTA, J., BÉJAR, R., MURO-MEDRANO, P.R., 2004. Metadata standard interoperability: application in the geographic information domain. Computers, Environment and Urban Systems 28 (6), 611-634.

NOGUERAS-ISO, J., ZARAZAGA-SORIA, F.J., MURO-MEDRANO, P.R., 2005. Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval. Springer, Berlin Heidelberg, Germany. doi: 10.1007/3-540-27508-8.

NPG, 2014a. About Scientific Data. http://www.nature.com/scientificdata/about/, 2014-05-01.

NPG, 2014b. Availability of data and materials. http://www.nature.com/authors/policies/availability.html, 2014-04-10.

O'REILLY, T., 2005. What Is Web 2.0. http://oreilly.com/web2/archive/what-is-web-20.html, 2014-05-01.

ODE PROJECT, 2011. Ten Tales of Drivers & Barriers in Data Sharing. The Hague, The Netherlands. http://www.alliance permanentaccess.org/wp-content/uploads/downloads/2011/10/7836_ODE_brochure_final.pdf, 2014-05-20.

OECD, 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. http://www.oecd.org/science/sci-tech/38500813.pdf, 2014-04-10. doi: 10.1787/9789264034020-en-fr.

OELTJEN, W., 2007. Metadaten-Management mit MyCoRe. In: BALL, R. (Ed.), WissKom 2007: Wissenschaftskommunikation der Zukunft. Forschungszentrum, Zentralbibliothek, Jülich, Germany, 185-192.

OELTJEN, W., 2012. Vernetzung mit MyCoRe - Eine Repository-Software vernetzt Systeme, Daten und Menschen. In: BALL, R. (Ed.), WissKom 2012: Vernetztes Wissen – Daten, Menschen, Systeme. Forschungszentrum, Zentralbibliothek, Jülich, Germany, 225-235.

ONLINE COMPUTER LIBRARY CENTER (OCLC), 2014. Dewey Decimal Classification (DDC). http://dewey.info/, 2014-02-24.

ONSRUD, H., CAMPBELL, J., 2007. Big Opportunities in Access to "Small Science" Data. Data Science Journal 6 (Open Data Issue), OD59-OD66. doi: 10.2481/dsj.6.OD58.

OPEN DATA COMMONS, 2014. Open Data Commons Licenses. http://opendatacommons.org/licenses/, 2014-03-31.

OPENAFS, 2013. Welcome to the home of OpenAFS. http://openafs.org, 2013-02-23.

OPPEL, A., 2009. Databases: A Beginner's Guide. McGraw Hill, New York, Chicago, San Francisco, USA.

OPPEL, A., SHELDON, R., 2009. SQL: A Beginner's Guide. McGraw Hill, New York, Chicago, San Francisco, USA.

ORACLE, 2013a. MySQL Reference Architectures for Massively Scalable Web Infrastructure. A MySQL® Strategy White Paper, http://www.mysql.com/why-mysql/white-papers/mysql-reference-architectures-for-scalable-web-infrastructure/, 2014-03-10.

ORACLE, 2013b. MySQL Workbench: Database Design. Development. Administration. Migration. A MySQL® White Paper, http://www.mysql.com/why-mysql/white-papers/mysql-workbench-database-design-development-administration/, 2014-03-04.

ORACLE, 2013c. Top 10 Reasons to Choose MySQL for Next Generation Web Applications. A MySQL Whitepaper, http://www.mysql.com/why-mysql/white-papers/top-10-reasons-to-choose-mysql-for-next-generation-web-applications/, 2014-03-15.

ORACLE, 2014. MySQL Workbench. Oracle, Redwood City, CA, USA. http://downloads.mysql.com/docs/workbench-en.pdf, 2015-03-15.

OREGON STATE UNIVERSITY LIBRARIES, 2014. Data Management Druing the Reserach Lifecycle. http://guides.library.oregonstate.edu/lifecycle, 2014-04-10.

OSI (OPEN SOCIETY INSTITUTE), 2004. A Guide to Institutional Repository Software. Chain Bridge Group, New York, NY, USA. http://www.budapestopenaccessinitiative.org/pdf/OSI_Guide_to_IR_Software_v3.pdf, 2014-04-14.

OSSWALD, A., SCHEFFEL, R., NEUROTH, H., 2012. Langzeitarchivierung von Forschungsdaten Einführende Überlegungen. In: NEUROTH, H., STRATHMANN, S., OßWALD, A., SCHEFFEL, R., KLUMP, J., LUDWIG, J. (Eds.), Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. Verlag Werner Hülsbusch, Boizenburg, Germany, 13-23.

OVERPECK, J.T., MEEHL, G.A., BONY, S., EASTERLING, D.R., 2011. Climate Data Challenges in the 21st Century. Science 331 (6018), 700-702. doi: 10.1126/science.1197869.

PAJARES, G., 2011. Advances in Sensors Applied to Agriculture and Forestry. Sensors 11 (9), 8930-8932. doi: 10.3390/s110908930.

PALAIOLOGK, A., ECONOMIDES, A., TJALSMA, H., SESINK, L., 2012. An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS. International Journal on Digital Libraries12 (4), 195-214. doi: 10.1007/s00799-012-0092-1.

PAMPEL, H., BERTELMANN, R., 2011. "Data Policies" im Spannungsfeld zwischen Empfehlung und Verpflichtung. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 49-61.

PAMPEL, H., GOEBELBECKER, H.-J., VIERKANT, P., 2012. re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositorien. Ein Werkstattbericht. In: BALL, R. (Ed.), WissKom 2012: Vernetztes Wissen – Daten, Menschen, Systeme. Forschungszentrum, Zentralbibliothek, Jülich, Germany, 61-74.

PAMPEL, H., VIERKANT, P., SCHOLZE, F., BERTELMANN, R., KINDLING, M., KLUMP, J., GOEBELBECKER, H.-J., GUNDLACH, J., SCHIRMBACHER, P., DIEROLF, U., 2013. Making research data repositories visible: the re3data.org registry. PeerJ PrePrints 1 (e21v1). doi: 10.7287/peerj.preprints.21v1.

PANGAEA, 2014. PANGAEA - Data Publisher for Earth & Environmental Science. http://www.pangaea.de/, 2014-05-01.

PARK, J.R., TOSAKA, Y., 2010. Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability. Information Technology and Libraries 29 (3), 104-116.

PERONI, S., SHOTTON, D., 2012. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Web Semantics: Science, Services and Agents on the World Wide Web 17 33-43. doi: 10.1016/j.websem.2012.08.001.

PHILLIPS, S., GREEN, C., MASLOV, A., MIKEAL, A., LEGGETT, J., 2007. Manakin: A New Face for DSpace D-Lib Magazine 13 (9/10).

PHPMYADMIN, 2014. Bringing MySQL to the web. http://www.phpmyadmin.net/home_page/, 2014-04-10.

PINDE, F., SUN, J., 2011. Web GIS: Principles and Applications. ESRI Press, Redlands, CA, USA.

PIWOWAR, H.A., DAY, R.S., FRIDSMA, D.B., 2007. Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2 (3), e308.

PLEWE, B., 1997. GIS online : information retrieval, mapping, and the Internet. OnWord Press, Santa Fe, NM, USA.

PLOS, 2014. PLOS Editorial and Publishing Policies. http://www.plosone.org/static/policies.action, 2014-04-10.

POMASKA, G., 2012. Webseiten-Programmierung. Springer, Wiesbaden, Germany. doi: 10.1007/978-3-8348-2485-1.

POOKER, N., 2009. Der erfolgreiche Webdesigner, Der Praxisleitfaden für Selbstständige. Galileo Press, Bonn, Germany.

POPULAR SCIENCE, 2011. Data Is Power. http://www.popsci.com/announcements/article/2011-10/november-2011-data-power, 2014-04-20.

PRÖLL, S., ZANGERLE, E., GASSLER, W., 2013. MySQL 5.6: Das umfassende Handbuch. Galileo Computing, Bonn, Germany.

PROMETHEUS, 2014. prometheus - Das verteilte digitale Bildarchiv für Forschung & Lehre. http://www.prometheus-bildarchiv.de/, 2014-03-15.

PRUDLO, M., 2005. E-Archiving: An Overview of Some Repository Management Software Tools. Ariadne(43).

PRYOR, G., 2011a. Preface. In: PRYOR, G. (Ed.), Managing Research Data Facet Publishing, London, UK, vii.

PRYOR, G., 2011b. Why manage research data? In: PRYOR, G. (Ed.), Managing Research Data Facet Publishing, London, UK, 1-16.

PRYOR, G., JONES, S., WHYTE, A., 2013. Delivering Research Data Management Services: Fundamentals of good practice. Facet Publishing, London, UK.

PURDUE UNIVERSITY, 2014. Research and Data Services at Purdue University. https://www.lib.purdue.edu/research, 2014-04-10.

QIN, J., BALL, A., GREENBERG, J., 2012. Functional and Architectural Requirements for Metadata: Supporting Discovery and Management of Scientific Data. DCMI International Conference on Dublin Core and Metadata Applications DC-2012 The Kuching Proceedings 62-71.

RADAR, 2014. RADAR - Research Data Repositorium. http://www.radar-projekt.org/display/RD/Home, 2014-04-04.

RADIESCHEN, 2013. Report "Synthese". GFZ Potsdam, Potsdam, Germany. doi: 10.2312/RADIESCHEN_007.

RAIMOND, Y., ABDALLAH, S., 2007a. The Event Ontology. http://www.motools.sourceforge.net/event/event.html, 2014-02-12.

RAIMOND, Y., ABDALLAH, S., 2007b. The Timeline Ontology. http://motools.sourceforge.net/timeline/timeline.html, 2014-02-12.

RAINER, R.K., CEGIELSKI, C., 2011. Introduction to Information Systems: Enabling and Transforming Business. Wiley, Hoboken, NJ, USA.

RAJABIFARD, A., KALANTARI, M., BINNS, A., 2009. SDI and metadata entry and updating tools. In: VAN LOENEN, B., BESEMER, J.W.J., ZEVENBERGEN, J.A. (Eds.), SDI Convergence. Research, Emerging Trends, and Critical Assessment. NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission, Delft, 121-136.

RASDAMAN, 2014. The ultimate Raster Server. http://www.rasdaman.com/, 2014-05-01.

Ray, J.M., 2014. Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, USA.

Razum, M., 2011. Systeme und Systemarchitekturen für das Datenmanagement. In: Büttner, S., Hobohm, H.-C., Müller, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 123-138.

Razum, M., Schwichtenberg, F., Fridman, R., 2007. Versioning of Digital Objects in a Fedora-based Repository. Procceedings of the German e-Science Conference, Baden Baden, Germany.

Razum, M., Schwichtenberg, F., Wagner, S., Hoppe, M., 2009. eSciDoc Infrastructure: A Fedora-Based e-Research Framework. In: Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C., Tsakonas, G. (Eds.), Research and Advanced Technology for Digital Libraries. Springer, Berlin, Heidelberg, Germany, 227-238. doi: 10.1007/978-3-642-04346-8_23.

RCR, 2014. Resilience, Collapse and Reorganisation in Social Ecological Systems of Africa's Savannahs. http://www.fg1501.uni-koeln.de/, 2014-05-05.

RCUK, 2013. RCUK Policy and Guidelines on Governance of Good Research Conduct. http://www.rcuk.ac.uk/Publications/researchers/grc/, 2014-04-15.

RDA, 2014. About RDA. https://www.rd-alliance.org/about.html, 2014-05-15.

Registry of Open Access, 2014. Registry of Open Access Repositories. http://roar.eprints.org, 2014-04-10.

Registry of Research Data Repositories, 2014. Registry of Research Data Repositories. http://www.re3data.org/, 2014-04-10.

Reibold, H., 2013. phpMyAdmin 4.0 kompakt. Brain-Media, Saarbrücken, Germany.

Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and Opportunities of Open Data in Ecology. Science 331 (6018), 703-705. doi: 10.1126/science.1197962.

Reilly, S., Schallier, W., Schrimpf, S., Smith, E., Wilkinson, M., 2011. Report on integration of data and publications.

Research Information Network (RIN), 2008. Stewardship of digital research data: a framework of principles and guidelines. A Research Information Network Report, Research Information Network.

Rice, R., 2007. DISC-UK DataShare Project: Building Exemplars for Institutional Data Repositories in the UK. IASSIST Quarterly 2007 (Fall/Winter).

Rice, R., 2014. Edinburgh datashare – reflections from a data repository manager. Bulletin of the American Society for Information Science and Technology 40 (2), 39-40. doi: 10.1002/bult.2014.1720400212.

Rice, R., Haywood, J., 2011. Research Data Management Initiatives at University of Edinburgh. International Journal of Digital Curation 6 (2), 232-244.

Rice, R., Macdonald, S., George, H., 2008. DSpace Metadata Schema for Edinburgh DataShare - Ver. 1. EDINA National Data Centre, Edinburgh University Data Library Edinburgh, UK. http://www.disc-uk.org/docs/Edinburgh_DataShare_DC-schema1.pdf, 2014-02-17.

Riley, J., 2009a. Glossary of Metadata Standards. http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf, 2013-05-14.

Riley, J., 2009b. Seeing Standards: A Visualization of the Metadata Universe. http://www.dlib.indiana.edu/~jenlrile/metadatamap/, 2013-05-14.

RIN, 2008. To share or not to share: publication and quality assurance of research data outputs. A Research Information Network Report, http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs, 2014-05-01.

Ross, R., 2011. File Systems. In: Padua, D. (Ed.), Encyclopedia of Parallel Computing. Springer US, 671-679. doi: 10.1007/978-0-387-09766-4_188.

RRZK, 2013a. AFS - Andrew File System. http://rrzk.uni-koeln.de/afs.html, 2013-02-23.

RRZK, 2013b. Backup-System TSM. http://rrzk.uni-koeln.de/tsm.html, 2013-02-23.

RRZK, 2013c. Schreibrechte für Webserver in Projekten. http://rrzk.uni-koeln.de/schreibrechte-webserver.html?&L=2, 2013-02-23.

RRZK, 2013d. Uni-Account. http://rrzk.uni-koeln.de/mitarbeiteraccount.html, 2013-02-23.

RRZK, 2014a. Das RRZK. http://rrzk.uni-koeln.de/das-rrzk.html, 2014-05-05.

RRZK, 2014b. Data Storage. http://rrzk.uni-koeln.de/datenhaltung.html?&L=1, 2014-05-01.

RRZK, 2014c. HPC Systeme. http://rrzk.uni-koeln.de/ressourcen-hpc.html?&L=1, 2014-05-01.

RRZK, 2014d. Regulations for publishing web sites at the RRZK. http://rrzk.uni-koeln.de/regeln-webseiten.html?&L=1, 2014-05-05.

RU816, 2007. Data User Agreement DFG Research Unit 816 Biodiversity and Sustainable Management of a Megadiverse Mountain Ecosystem in South Ecuador. http://www.bergregenwald.de/files/generalinformation/for816_datenvereinbarung.pdf, 2014-05-01.

RUMSEY, S., JEFFERIES, N., 2013a. Challenges in Building an Institutional Research Data Catalogue. International Journal of Digital Curation 8 (2), 205-214. doi: 10.2218/ijdc.v8i2.284.

RUMSEY, S., JEFFERIES, N., 2013b. DataFinder: A Research Data Catalogue for Oxford. Ariadne(71).

RWTH AACHEN, 2014. Forschungsdaten an der RWTH Aachen. http://www.bth.rwth-aachen.de/index.php?id=70, 2014-04-10.

SALLANS, A., LAKE, S., 2014. Data Management Assessment and Planning Tools. In: RAY, J. (Ed.), Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, IN, USA, 87-107.

SATYANARAYANAN, M., 1989. Integrating security in a large distributed system. ACM Trans. Comput. Syst. 7 (3), 247-280. doi: 10.1145/65000.65002.

SCHARNHORST, A., TEN BOSCH, O., DOORN, P., 2012. Looking at a digital research data archive - Visual interfaces to EASY. Digital Libraries.

SCHICKLING, A., GRAF, A., PIERUSCHKA, R., PLÜCKERS, C., GEIß, H., LAI, I.-L., SCHWEEN, J.H., ERENTOK, K., SCHMIDT, M., WAHNER, A., CREWELL, S., RASCHER, U., 2010. The influence of leaf photosynthetic efficiency and stomatal closure on canopy carbon uptake and evapotranspiration – a model study in wheat and sugar beet. Biogeosciences Discuss. 7, 7131-7172. doi: 10.5194/bgd-7-7131-2010.

SCHILL, A., SPRINGER, T., 2012. Verteilte Systeme. Springer, Berlin Heidelberg, Germany. doi: 10.1007/978-3-642-25796-4.

SCHMIDT, M., JANßEN, T., DRESSLER, S., HAHN, K., HIEN, M., KONATÉ, S., LYKKE, A.M., MAHAMANE, A., SAMBOU, B., SINSIN, B., THIOMBIANO, A., WITTIG, R., ZIZKA, G., 2012. The West African Vegetation Database. Biodiversity & Ecology 4, 105-110. doi: 10.7809/b-e.00065.

SCHMIDT, P., MÄRKER, M., 2010. Spatial assessment of early human expansions using ROAD: an example from Southern Africa. In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29. - 30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 113-117. doi: 10.5880/TR32DB.KGA90.18.

SCHOMBURG, A., VENEMA, V., LINDAU, R., AMENT, F., SIMMER, C., 2010. A downscaling scheme for atmospheric variables to drive soil-vegetation-atmosphere transfer models. Tellus Series B-Chemical and Physical Meteorology 62 (4), 242-258. doi: 10.1111/j.1600-0889.2010.00466.x.

SCHUBERT, M., 2007. Datenbanken: Theorie, Entwurf und Programmierung relationaler Datenbanken. B. G. Teubner Verlag, Wiesbaden, Germany. doi: 10.1007/978-3-8351-9108-2.

SCIENCE, 2011. Special Online Collection: Dealing with Data. http://www.sciencemag.org/site/special/data/, 2014-04-20.

SCIENCE EUROPE, 2013a. About Science Europe. http://www.scienceeurope.org/about-us/about-us-full/, 2014-04-10.

SCIENCE EUROPE, 2013b. Science Europe - Roadmap. Science Europe, Brussels, Belgium.

SEIBERT, S., AUERSWALD, K., FIENER, P., DISSE, M., MARTIN, W., HAIDER, J., MICHAEL, A., GERLINGER, K., 2011. Surface runoff from arable land - a homogenized data base of 726 rainfall simulation experiments, CRC/TR32 Database (TR32DB). doi: 10.1594/GFZ.TR32.2.

SELBACH, C., SCHWEEN, J., CREWELL, S., 2010. Field Report: TR32 2007-2010 - Description of field measurements performed in the Ellebach Catchment. Institute of Meteorology, University of Cologne, Cologne, Germany.

SEN, A., 2004. Metadata management: past, present and future. Decision Support Systems 37 (1), 151-173. doi: 10.1016/s0167-9236(02)00208-7.

SENCKENBERG, 2014. West African Vegetation – A vegetation data network for West Africa. http://www.westafricanvegetation.senckenberg.de/, 2014-04-10.

SHUMILOV, S., ROGMANN, A., LAUBACH, J., 2008. GLOWA Volta GeoPortal: an interactive geodata repository and communication system, In: EHLERS, M., BEHNCKE, K., GERSTENGARBE, F.-W., HILLEN, F., KOPPERS, L., STROINK, L., WÄCHTER, J. (Eds.), Digital Earth Summit on Geoinformatics 2008: Tools for Global Change Research. Wichmann, Heidelberg, Germany, 363-368.

SIDEBOTHAM, B., 1986. Volumes: The Andrew File System Data Structuring Primitive. Proccedings of the European Unix User Group Conference (EUUG), Manchester, UK.

SIMONS, N., RICHARDSON, J., 2013. New Content in Digital Repositories: The changing research landscape Chandos Publishing Oxford, UK.

SMARTY, 2014. The Philosophy of Smarty. http://www.smarty.net/about_smarty, 2014-05-15.

SMITH, M., BARTON, M., BASS, M., BRANSCHOFSKY, M., MCCLELLAN, G., STUVE, D., TANSLEY, R., WALKER, J.H., 2003. DSpace: An Open Source Dynamic Digital Repository. D-Lib Magazine 9 (1). doi: 10.1045/january2003-smith.

SORG, J., TIMON KIRSCHKE, KUNKEL, R., 2013. Konzeption und Implementierung eines OGC-konformen Sensor Observation Service für den standardisierten Zugriff auf Rasterzeitreihendaten. gis.SCIENCE 2013 (2), 39-50.

SPRINGER, P., MEHRTENS, H., FLEISCHER, D., SCHIRNICK, C., CZERNIAK, A., 2013. Kiel Data Management Infrastructure for Marine Sciences. Proccedings of the IMDIS 2013 International Conference on Matine Data and Information Systems, Lucca, Italy, 257-258.

STAIR, R., REYNOLDS, G., 2010. Principles of Information Systems: A Managerial Approach. Course Technology, Cengage Learning, Boston, MA, USA.

STANFORD UNIVERSITY LIBRARIES, 2014. Data Management Services - Best Practices. http://library.stanford.edu/research/data-management-services/data-best-practices, 2014-04-20.

STARR, J., GASTL, A., 2011. isCitedBy: A Metadata Scheme for DataCite. D-Lib Magazine 17 (1/2). doi: 10.1045/january2011-starr

STIETENROTH, D., NIESCHULZE, J., AREND, K., 2005. Rechtliche Aspekte und Umsetzung des Datenmanagement in internationalen interdisziplinären Forschungsprojekten. Zeitschrift für Agrarinformatik 3 64-75.

STOCKHAUSE, M., HÖCK, H., 2010. Integrated Climate Data Center (ICDC) at the Cluster of Excellence at the University of Hamburg. In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29.-30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 127-135. doi: 10.5880/TR32DB.KGA90.20.

STRASSER, C.A., COOK, R.B., MICHENER, W.K., BUDDEN, A., KOSKELA, R., 2011. DataONE: Promoting Data Stewardship Through Best Practices. Proccedings of the Environmental Information Management Conference 2011, 126-131.

STRATHMANN, S., 2013. Status of Discussion and Current Activities: The International Perspective In: NEUROTH, H., STRATHMANN, S., OßWALD, A., LUDWIG, J. (Eds.), Digital Curation of Research Data - Experiences of a Baseline Study in Germany. Verlag Werner Hülsbusch, Glückstadt, Germany, 18-36.

STROBEL, C., 2004. Web-Technologien: in E-Commerce-Systemen. Oldenbourg, Munich, Germany.

SUMATHI, S., ESAKKIRAJAN, S., 2007. Fundamentals of Relational Database Management Systems. Springer, Berlin, Heidelberg, Germany.

SYNOLOGY, 2014. Synology Products. http://www.synology.com/en-uk/products/index, 2014-05-05.

TAKEDA, K., BROWN, M., COLES, S., CARR, L., EARL, G., FREY, J., HANCOCK, P., WHITE, W., NICHOLS, F., WHITTON, M., GIBBS, H., FOWLER, C., WAKE, P., PATTERSON, S., 2010. Data management for all: the Institutional Data Management Blueprint Project. Proccedings of the 6th International Digital Curation Conference.

TANENBAUM, A.S., 2009. Moderne Betriebssysteme. Pearson Studium München, Germany.

TANSLEY, R., BASS, M., STUVE, D., BRANSCHOFSKY, M., CHUDNOV, D., MCCLELLAN, G., SMITH, M., 2003. The DSpace institutional digital repository system: current functionality. Proccedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries, Houston, Texas, USA, 87-97.

TECHNICAL UNIVERSITY OF BERLIN, 2014. Servicezentrum Forschungsdaten und –publikationen an der TU Berlin. http://www.szf.tu-berlin.de/, 2014-04-10.

TENOPIR, C., ALLARD, S., DOUGLASS, K., AYDINOGLU, A.U., WU, L., READ, E., MANOFF, M., FRAME, M., 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6 (6), e21101.

TEOREY, T., LIGHTSTONE, S., NADEAU, T., JAGADISH, H.V., 2011. Database Modeling and Desing. Elsevier, Morgan Kaufmann, Burlington, USA.

TERENO, 2014. TERENO Data Policy. http://teodoor.icg.kfa-juelich.de/downloads-de/TERENO%20Data%20policy.pdf, 2014-05-01.

THALER, A.D., ZELNIO, K.A., FREITAG, A., MacPHERSON, R., SHIFFMAN, D., BIK, H., GOLDSTEIN, M.C., McCLAIN, C., 2012. Digital environmentalism: tools and strategies for the evolving online ecosystem. In: GALLAGHER, D. (Ed.), Environmental Leadership: A Reference Handbook. SAGE, London, UK.

THALHEIM, B., 2000. Entity-Relationship Modeling: Foundations of Database Technology. Springer, Berlin, Heidelberg, Germany. doi: 10.1007/978-3-662-04058-4.

THE ECONOMIST, 2010. Data, data everywhere. The Economist Special report: Managing information 16.

THE ROYAL SOCIETY, 2012. Science as an open enterprise. The Royal Society, London, UK.

THIESING, F., KORTEMEYER, S., 2008. Entwicklung moderner Web-Anwendungen mit Open-Source-Bausteinen. Informatik-Spektrum 31 (2), 115-132. doi: 10.1007/s00287-008-0226-4.

THOMSON REUTERS, 2014. The Data Citation Index. http://wokinfo.com/products_tools/multidisciplinary/dci/, 2014-05-01.

TIB, 2014a. GetInfo - Das Portal für Technik und Naturwissenschaften. http://www.tib-hannover.de/de/getinfo/, 2014-04-10.

TIB, 2014b. TIB DOI Service. http://www.tib-hannover.de/en/services/doi-service/, 2014-05-01.

TOUSSAINT, F., 1999. Wissenschaftliches Datenmanagement: Das CERA-2 Daten- und Metadatenmodell, In: GROß, H., RICHTER, D. (Eds.), Einsatz von Datenbanktechnologie in der Metrologie, Berlin, Germany.

TOUSSAINT, F., LAUTENSCHLAGER, M., REINKE, M., 1999. CERA-2: Ein Raumbezogenes Daten- und Metadatenmodell. In: KRAMER, R., HOSENFELD, F. (Eds.), Praxis der Umweltinformatik. Metropolis-Verlag, Marburg, Germany.

TRAMBOO, S., HUMMA, S.M.S., GUL, S., 2012. A study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone. International Journal of Computer Applications 59 (16), 1-9.

TRELOAR, A., 2009. Design and Implementation of the Australian National Data Service. International Journal of Digital Curation 4 (1), 122-137.

TRELOAR, A., HARBOE-REE, C., 2008. Data management and the curation continuum: how the Monash experience is informing repository relationships. Proccedings of the VALA 2008, Melbourne, Australia.

TRELOAR, A., WILKINSON, R., 2008. Access to Data for eResearch: Designing the Australian National Data Service Discovery Serivces. International Journal of Digital Curation 2 (2), 151-158.

TROPPENS, U., 2009. Storage networks explained : basics and application of Fibre Channel SAN, NAS, iSCSI, InfiniBand, and FCoE. Wiley, Chichester, UK.

UK DATA ARCHIVE, 2014. The reserch data lifecycle. http://www.data-archive.ac.uk/create-manage/life-cycle, 2014-04-20.

UK DATA SERVICE, 2014a. Terms and conditions of access. http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx, 2014-05-11.

UK DATA SERVICE, 2014b. UK Data Service Data Management Plan Checklist. http://ukdataservice.ac.uk/manage-data/plan/checklist.aspx, 2014-04-20.

UNIVERSITY OF BIELEFELD, 2014a. Forschungsdatenmanagement an der Universität Bielefeld. http://data.uni-bielefeld.de/de, 2014-04-10.

UNIVERSITY OF CENTRAL FLORIDA LIBRARIES, 2014. The research lifecycle at UCF. library.ucf.edu/ScholarlyCommunication/ResearchLifecycleUCF.php, 2014-04-20.

UNIVERSITY OF EDINBURGH, 2011. Research Data Management Policy. www.ed.ac.uk/is/research-data-policy, 2014-04-20.

UNIVERSITY OF GÖTTINGEN, 2014. Datenmanagement und -archivierung at the University of Göttingen. http://www.uni-goettingen.de/de/datenmanagement/221506.html, 2014-04-10.

UNIVERSITY OF HEIDELBERG, 2014. Forschungsdatenmanagement at the University of Heidelberg. http://www.ub.uni-heidelberg.de/service/forschung/Welcome.html, 2014-04-10.

UNIVERSITY OF LONGHBOROUGHT LIBRARY, 2014. Research support - Research Lifecycle. http://www.lboro.ac.uk/services/library/research/, 2014-04-14.

UNIVERSITY OF MANNHEIM, 2014. Das Forschungsdaten Service Center of the University of Mannheim. http://service.informatik.uni-mannheim.de/de/das-center/, 2014-04-10.

UNIVERSITY OF MELBOURNE, 2005. Policy on the Management of Research Data and Records. http://www.unimelb.edu.au/records/pdf/research.pdf, 2014-04-20.

UNIVERSITY OF NEWCASTLE, 2008. Research Data and Materials Management Policy. http://www.newcastle.edu.au/policy/000869.html, 2014-04-20.

UNIVERSITY OF OREGON LIBRARIES, 2014. Research Data Management - Best Practices. https://library.uoregon.edu/datamanagement/guidelines.html, 2014-04-20.

UNIVERSITY OF OXFORD, 2014. Policy on the Management of Research Data and Records. http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf, 2014-04-20.

UNIVERSITY OF TENNESSEE, 2009. Research Data Policy. http://research.utk.edu/files/2013/05/policy_research-data1.pdf, 2014-04-14.

UNIVERSITY OF VIRGINIA LIBRARY, 2014. Steps in the Research Life Cycle. http://dmconsult.library.virginia.edu/lifecycle/, 2014-04-14.

UNTERSTEIN, M., MATTHIESSEN, G., 2012. Relationale Datenbanken und SQL in Theorie und Praxis. Springer, Berlin, Heidelberg, Germany.

UoB, 2014b. Resolution zum Forschungsdatenmanagement. http://data.uni-bielefeld.de/de/resolution, 2014-04-10.

VAN DEN EYNDEN, V., CORTI, L., WOOLLARD, M., BISHOP, L., HORTON, L., 2011. Managing and Sharing Data: Best Practice for Researchers. UK Data Archive, Colchester, UK.

VAN DER GRAAF, M., WAAIJERS, L., 2012. A Surfboard for Riding the Wave: Towards a four country action programme on research data. Knowledge Exchange, Copenhagen, Denmark.

VAN NOORDEN, R., 2013. Data-sharing: Everything on display. Nature 500 (7461), 243-245. doi: 10.1038/nj7461-243a.

VASWANI, V., 2010. MySQL Database Usage & Administration. McGraw Hill, New York, Chicago, San Francisco, USA.

VAUGHAN, P.J., 2001. System Implementation Success Factors; It's not just the technology. Proccedings of the CUMREC 2001, May 13-16, 2001, Phoenix, Arizona, USA.

VBA-GEOPORTAL, 2014. Volta Basin Authority Geoportal. http://131.220.109.2/geonetwork/srv/en/about, 2014-04-13.

VELLUCCI, S., 2000. Metadata and Authority Control. LRTS 44 (1), 33-43.

VETTERLI, T., VADUVA, A., STAUDT, M., 2000. Metadata standards for data warehousing: open information model vs. common warehouse metadata. SIGMOD Rec. 29 (3), 68-75. doi: 10.1145/362084.362138.

References

VIERKANT, P., SPIER, S., RUECKNAGEL, J., GUNDLACH, J., FICHTMUELLER, D., PAMPEL, H., KINDLING, M., KIRCHHOFF, A., GOEBELBECKER, H.-J., KLUMP, J., BERTELMANN, R., SCHIRMBACHER, P., SCHOLZE, F., 2012. Vocabulary for the Registration and Description of Research Data Repositories. re3data.org, Potsdam, Germany. doi: 10.2312/RE3.002.

VIERKANT, P., SPIER, S., RUECKNAGEL, J., PAMPEL, H., GUNDLACH, J., FICHTMUELLER, D., KINDLING, M., KIRCHHOFF, A., GOEBELBECKER, H.-J., KLUMP, J., KLOSKA, G., REUTER, E., SEMRAU, A., SCHNEPF, M., SKARUPIANSKI, M., BERTELMANN, R., SCHIRMBACHER, P., SCHOLZE, F., KRAMER, C., 2013. Schema for the Description of Research Data Repositories, Version 2.1. re3data.org, Potsdam, Germany. doi: 10.2312/re3.004.

VISION, T., 2010. The Dryad Digital Repository: Published evolutionary data as part of the greater data ecosystem. iEvoBio 2010-Informatics for Phylogenetics, Evolution, and Biodiversity, http://hdl.handle.net/10101/npre.2010.4595.1, 2014-05-01.

VOGEL, O., ARNOLD, I., CHUGHTAI, A., IHLER, E., KEHRER, T., MEHLIG, U., ZDUN, U., 2009. Software-Architektur: Grundlagen - Konzepte - Praxis. Spektrum Akademischer Verlag, Heidelberg, Germany.

VOGEL, O., ARNOLD, I., CHUGHTAI, A., KEHRER, T., 2011. Software Architecture: A Comprehensive Framework and Guide for Practitioners. Springer, Heidelberg Berlin, Germany.

W3C, 2008. Web Content Accessibility Guidelines (WCAG) 2.0. http://www.w3.org/TR/WCAG20/, 2014-05-15.

W3C, 2014. W3C Standards. http://www.w3.org/standards/, 2014-05-15.

W3DTF, 2014. Date and Time Formats. http://www.w3.org/TR/NOTE-datetime, 2014-05-01.

WALDHOFF, G., 2012. Enhanced Land Use Classification of 2009 for the Rur catchment, CRC/TR32 Database (TR32DB). doi: 10.5880/TR32DB.2.

WALDHOFF, G., BARETH, G., 2009. GIS- and RS-based land use and land cover analysis - case study Rur-Watershed, Germany. Geoinformatics 2008 and Joint Conference on Gis and Built Environment: Advanced Spatial Data Models and Analyses 7146 (26). doi: 10.1117/12.813171.

WALDHOFF, G., CURDT, C., HOFFMEISTER, D., BARETH, G., 2012. Analysis of Multitemporal and Multisensor Remote Sensing Data for Crop Rotation Mapping. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. I-7 177-182. doi: 10.5194/isprsannals-I-7-177-2012.

WANG, T.J., SAEMANN, G., DU, H., 2007. A Systems Development Life Cycle Project for the AIS Class. Journal of College Teaching & Learning (TLC) 4 (1), 67-94.

WEIBEL, S.L., 1995. Metadata: The Foundations of Resource Description. D-Lib Magazine 1 (1).

WEICHSELGARTNER, E., GÜNTHER, A., DEHNHARD, I., 2011. Archivierung von Forschungsdaten. In: BÜTTNER, S., HOBOHM, H.-C., MÜLLER, L. (Eds.), Handbuch Forschungsdatenmanagement. Bock u. Herchen, Bad Honnef, Germany, 191-202.

WELLCOME TRUST, 2010. Policy on data management and sharing. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm, 2014-04-15.

WELLCOME TRUST, 2014. Guidance for researchers: Developing a data management and sharing plan. http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm, 2014-04-20.

WENZ, C., HAUSER, T., MAURICE, F., 2011. Das Website-Handbuch: Programmierung und Design. Markt + Technik, Munich, Germany.

WHITE, H.C., CARRIER, S., THOMPSON, A., GREENBERG, J., SCHERLE, R., 2008. The Dryad data repository: a Singapore framework metadata architecture in a DSpace environment. Proccedings of the International Conference on Dublin Core and Metadata Applications, DC-2008, Berlin, Germany, 157-162.

WHYTE, A., TEDDS, J., 2011. Making the Case for Research Data Management. DCC Briefing Papers, Digital Curation Centre, Edinburgh, UK.

WIKIBON BLOG, 2012. A Comprehensive List of Big Data Statistics. http://wikibon.org/blog/big-data-statistics/, 2014-04-20.

WILLIS, C., GREENBERG, J., WHITE, H., 2012. Analysis and synthesis of metadata goals for scientific data. Journal of the American Society for Information Science and Technology 63 (8), 1505-1520. doi: 10.1002/asi.22683.

WILLMES, C., BROCKS, S., HOFFMEISTER, D., HÜTT, C., KÜRNER, D., VOLLAND, K., BARETH, G., 2012. Facilitating integrated spatio-temporal visualization and analysis of heterogeneous archaeological and palaeonvironmental research data. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. I-2 223-228. doi: 10.5194/isprsannals-I-2-223-2012.

WILLMES, C., KÜRNER, D., BARETH, G., 2014. Building Research Data Management Infrastructure using Open Source Software. Transactions in GIS 18 (4), 496-509. doi: 10.1111/tgis.12060.

WILSON, A., JEFFREYS, P., 2013. Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford. International Journal of Digital Curation 8 (2), 235-246. doi: 10.2218/ijdc.v8i2.287.

WILSON, J.A.J., MARTINEZ-URIBE, L., FRASER, M.A., JEFFREYS, P., 2011. An Institutional Approach to Developing Research Data Management Infrastructure. International Journal of Digital Curation 6 (2), 274-287.

WILSON, M., 2014. Metadata: Describing geospatial data. In: NEBERT, D. (Ed.), Developing Spatial Data Infrastrucutres: The SDI Cookbook. GSDI, 24-38.

WINKLER-NEES, 2013. Status of Discussion and Current Activities: National Developments In: NEUROTH, H., STRATHMANN, S., OßWALD, A., LUDWIG, J. (Eds.), Digital Curation of Research Data - Experiences of a Baseline Study in Germany. Verlag Werner Hülsbusch, Glückstadt, Germany, 18-36.

WINKLER-NEES, S., 2011. Anforderungen an wissenschaftliche Informationsstrukturen, RatSWD Working Paper Series. Rat für Sozial- und Wirtschaftsdaten, Berlin, Germany, 21.

WINKLER-NEES, S., 2012. Stand der Diskussion und Aktivitäten - National. In: NEUROTH, H., STRATHMANN, S., OßWALD, A., SCHEFFEL, R., KLUMP, J., LUDWIG, J. (Eds.), Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. Verlag Werner Hülsbusch, Boizenburg, Germany, 23-40.

WINSCP, 2014. Introducing WinSCP. http://winscp.net/eng/docs/introduction, 2015-05-16.

WIRA-ALAM, A., DIMITROV, D., ZENK-MÖLTGEN, W., 2012. Extending Basic Dublin Core Elements for an Open Research Data Archive. DCMI International Conference on Dublin Core and Metadata Applications DC-2012--The Kuching Proceedings 56-61.

WIRED, 2014. Amazon's Cloud Keeps Growing Despite Fears of NSA Spying. http://www.wired.com/2014/02/amazon-cloud-size/, 2014-04-20.

WISSENSCHAFTSRAT, 2011a. Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen. Wissenschaftsrat, Berlin, Germany. http://www.wissenschaftsrat.de/download/archiv/10464-11.pdf, 2012-07-13.

WISSENSCHAFTSRAT, 2011b. Übergreifende Empfehlungen zu Informationsinfrastrukturen. Wissenschaftsrat, Berlin, Germany. http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf, 2012-07-20.

WISSENSCHAFTSRAT, 2012. Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. Wissenschaftsrat, Berlin, Germany. http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf, 2012-07-20.

WONCZAK, S., 2014. Personal Communication on 2014-05-06 with Dipl. Chem. Dr. Stephan Wonczak of the Regional Computing Centre of University of Cologne, Germany.

WOODLEY, M.S.-. 2008. Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information. In: BACA, M. (Ed.), Introduction to Metadata. Gregory M. Britton, Los Angeles, CA, USA, 38-62.

YEUNG, A., HALL, G.B., 2007a. Spatial Database Systems: Design, Implementation and Project Management. Springer, Dordrecht, The Netherlands.

YEUNG, A.W., HALL, G.B., 2007b. Spatial Data Standards and Metadata. In: YEUNG, A.W., HALL, G.B. (Eds.), Spatial Database Systems. Springer Netherlands, 129-173. doi: 10.1007/1-4020-5392-4_5.

ZACHARIAS, S., REYERS, M., PINTO, J., SCHWEEN, J., CREWELL, S., KERSCHGENS, M., 2012. Heat and moisture budgets from airborne measurements and high-resolution model simulations. Meteorology and Atmospheric Physics 117 (1), 47-61. doi: 10.1007/s00703-012-0188-6.

ZANDER, F., KRALISCH, S., BUSCH, C., FLÜGEL, W.-A., 2013. Data and information management for integrated research - requirements, experiences and solutions. Proccedings of the 20th International Congress on Modelling and Simulation, December 1-6, 2013, Adelaide, Australia, 2201-2206.

ZANDER, F., KRALISCH, S., FLÜGEL, W.-A., 2011. Environmental data management with the River Basin Information System Proccedings of the 19th International Congress on Modelling and Simulation, December 12-16, 2011, Perth, Australia, 3191-3197.

ZENG, M.L., QIN, J., 2008. Metadata. Neal-Schuman Publishers, New York, NY, USA.

ZENK-MÖLTGEN, W., HABBEL, N., 2012. Der GESIS Datenbestandskatalog und sein Metadatenschema, Version 1.8. GESIS Technical Report http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-01.pdf, 2014-05-15.

ZENODO, 2014. About ZENODO. http://zenodo.org/about, 2014-04-01.

A   Appendix: Results of the survey
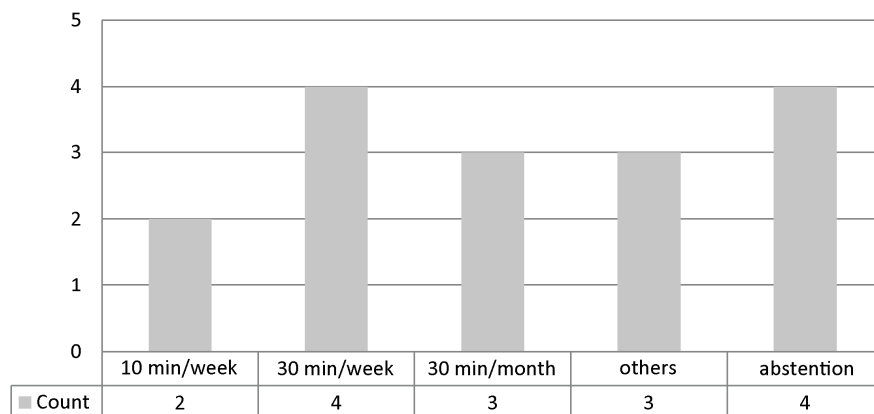
Figure A-1: How much time is available for data preparation and maintenance?

	10 min/week	30 min/week	30 min/month	others	abstention
Count	2	4	3	3	4
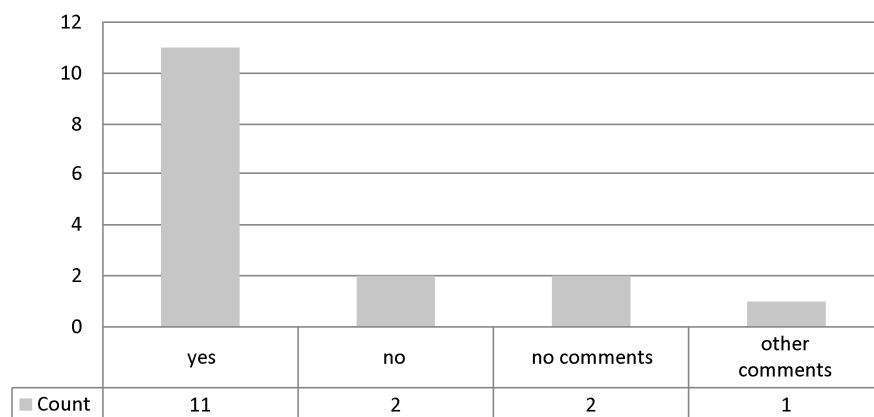
Figure A-2: Do you support the development of a project database in the Internet that is password protected?

	yes	no	no comments	other comments
Count	11	2	2	1
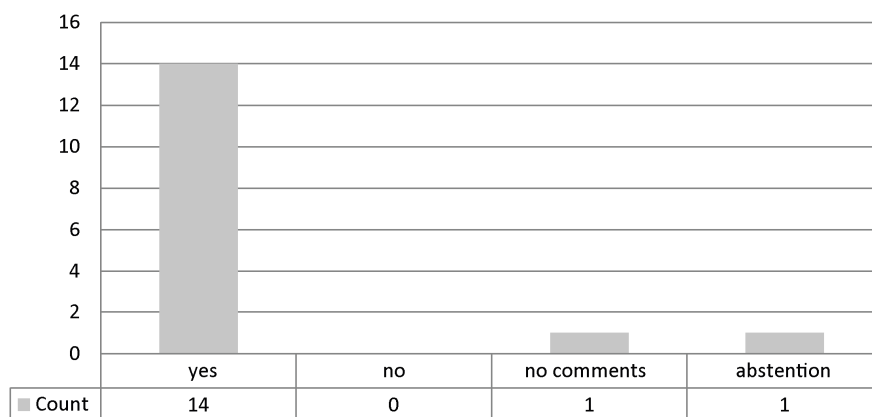
Figure A-3 Do you agree that all participants use one database?

	yes	no	no comments	abstention
Count	14	0	1	1
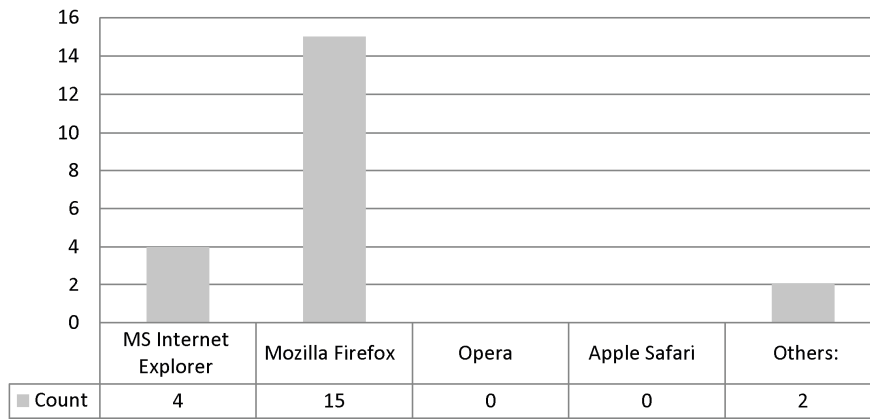
Figure A-4: Which web browser do you use?
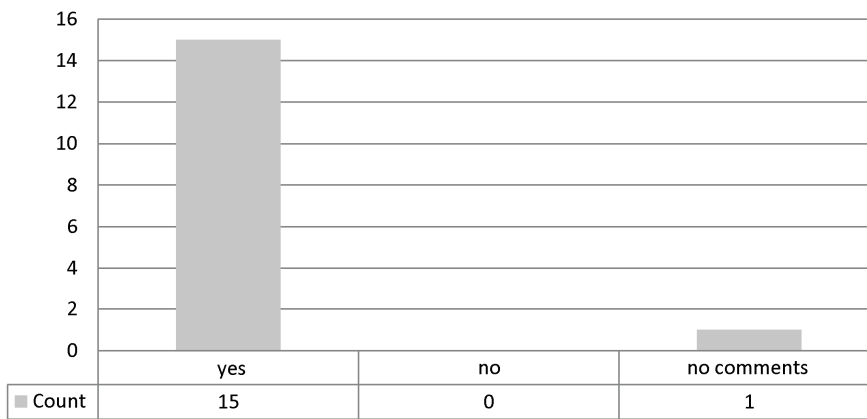
Figure A-5: Should additional data (e.g. publications, dissertations, reports, …) be stored in the project database?
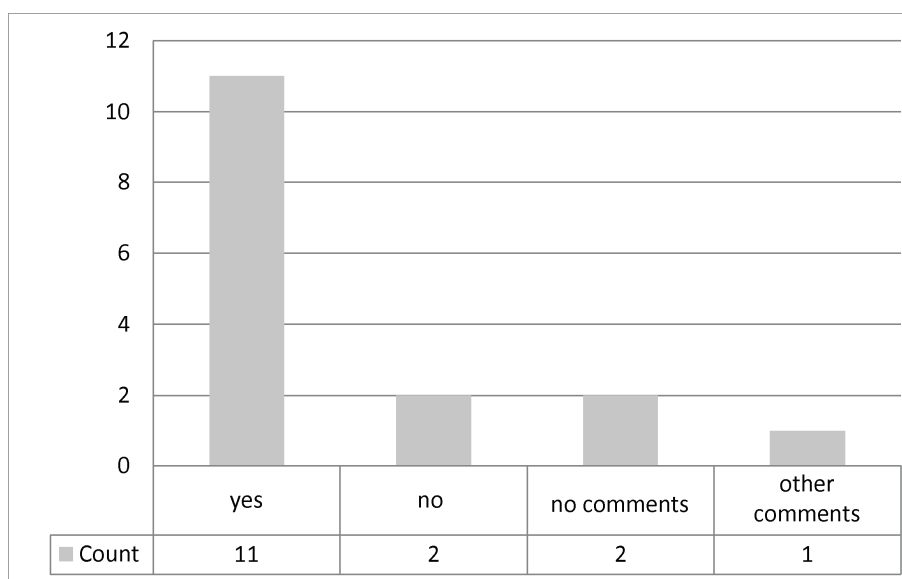
Figure A-6: Do you support the development of a project database in the Internet that is password protected?

Count	MS Windows	Apple Mac OS	Linux	Others:
	13	0	10	1
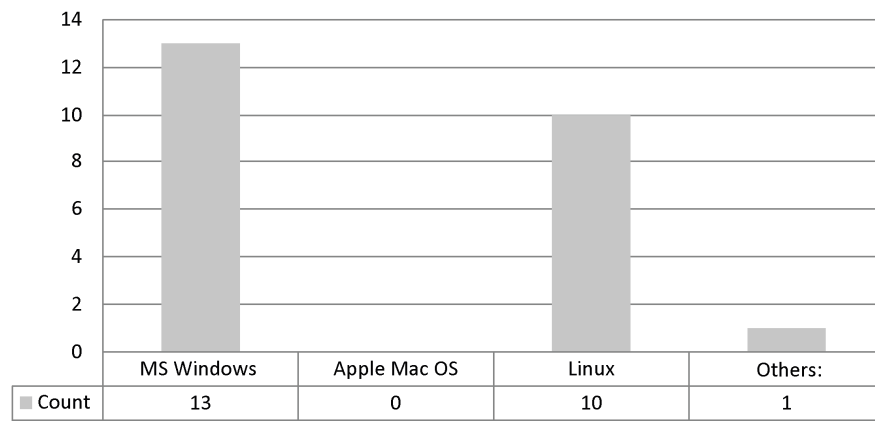
Figure A-7: Which operating system do you use?

B   Appendix: TR32DB web-interface Version 3.1

Figure B-1: Metadata input form for general properties

Figure B-2: Metadata input form for `specific properties` of data type `data`

Metadata for geodata

* required fields

Inserted Metadata

(Choose...) — Metadata inserted by your subproject.
You can choose this if the metadata you want to insert, is
similar or equal to older metadata used in your subproject.

selected dataset as guideline:
Digital Orthophotos Rur catchment 2010

General Properties

Please insert your metadata in english!

user: *	Hoffmeister, Dirk	User of subproject Z1 responsible for this dataset.
dataset name: *	DigitalOrthophotosRurcatchment2010	Dataset name, eg. ASTER-LU_050508
extension: *	.TIFF	Dataset extension, eg. zip
title: *	Digital Orthophotos Rur catchment 2010	Maintitle of the dataset.
description: *	Orthophotos of Rur Catchment recorded in 2010 An orthophoto is a geometrically corrected aerial photograph. The very diverse uses of orthophotos include: - Urban and rural planning - Environmental and agricultural studies Due to terrain shape, movements of the aircraft	Short abstract of the dataset.

1. creator-organization * Bezirksregierung Koeln [www.lverma.nrw.de/]

role: *
distributor
add organization

1. creator-person * name:*
Hoffmeister, Dirk (dirk.hoffmeister@uni-koeln.de) — Dataset creator (person responsible for content of the resource).
role: * point of contact — Role of creator.
1. creator-institution: * University of Cologne, Institute of Geography — e.g. University of Cologne, Institute of Geography
add creator

contributor: — Dataset contributor (person and/or institution responsible for making contributions to dataset content; in addition to creator). e.g. John Doe, University of Cologne, Institute of Geography

publisher: University of Cologne, Institute of Geography, TR32DB — Dataset publisher (organization responsible for making the resource available).

citation advice: — Advices of creator for use and citation of dataset e.g. Schmitz, Lars; Motani, Ryosuke (2011): Nocturnality in Dinosaurs Inferred from Scleral Ring and Orbit Morphology. Science, 332 (6030), pp.705-708. DOI: 10.1126/science.1200043

identifier(s): add identifier — Identifier of the resource, if available.
relation: add relation — A reference to a related resource including type of relation, if available e.g. dataset [5] is version of.

creation date: *	2007-04-05	Date of creation of the dataset.
publication date: *	2013-04-05	Date of publication of dataset.
file size:	0 data unit: (Choose...)	Amount of data, number of pages, etc. e.g. 5 Datasets.
dataset type: *	image	Type of dataset.
format-version:	TIFF - * TIFF	Format and version of dataset, e.g. MS Word - 2007.
format *	type: image format: tiff	You can find the whole list of mime types here.
dataset language:*	german	Primary language of the dataset.
TR32 - topic: *	land use	TR32 specific topic of the dataset.
keywords: *	(Choose...)	Select min. 1 and max. 6 keywords (please contact Z1 for further new keywords!).

Catchment, Geography,

Reset Keywords

status: *	completed	Dataset processing status.
permission: *	only TR32	Permission of dataset.

Data Specific Properties

measurement / model region: *	NRW	Region of measurement or model domain.
measurement / model location:*	OTHER	Location of measurement or model position.
temporal extent:	begin: 0000-00-00 00:0 end: 0000-00-00 00:0	Time period covered by the content of the dataset.
spatial resolution-scale:	1 :	Equivalent scale e.g. 1 : 10000.
spatial resolution - distance:	(Choose...)	Resolution distance/unit of length (e.g. 1mm, 1cm, 1dm).
measurement/model frequency:	(Choose...)	Frequency of measuremet/modeling (eg. 30 seconds, 10 minutes).
additional description	Durchsu	Upload a pdf for additional description (with filesize max. 5 MB).

Geodata Specific Properties

INSPIRE topic category: imageryBaseMapsEarthCover

geographic bounding box: — Extent of the dataset in geographic space. Values in decimal degrees, with a precision of at least 2 decimals. E.g. 49.88

north bound latitude
51.3405

west bound longitude   east bound longitude
5.20137   7.39863

south bound latitude
50.2994

reference system identifier:	25832	Dataset reference system, eg. 31462.
reference system name:	ETRS89	Name of reference system.
lineage:		Description of process history and/or overall quality and validity of the data set.
representation:	(Choose...)	-

choose a button to continue: Reset Submit

Figure B-3: Metadata input form for data type `geodata`; general and data specific properties(complement to Figure 4-39)

Figure B-4: DOI application form

```xml
<?xml version="1.0" encoding="utf-8"?>
<resource xmlns="http://datacite.org/schema/kernel-2.2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://datacite.org/schema/kernel-2.2
http://schema.datacite.org/meta/kernel-2.2/metadata.xsd">
  <identifier identifierType="DOI">10.5880/TR32DB.9</identifier>
  <creators>
    <creator>
      <creatorName>Lussem, Ulrike</creatorName>
    </creator>
    <creator>
      <creatorName>Waldhoff, Guido</creatorName>
    </creator>
  </creators>
  <titles>
    <title>Land use classification 2012 of the Rur Catchment</title>
  </titles>
  <publisher>CRC/TR32 Database (TR32DB)</publisher>
  <publicationYear>2013</publicationYear>
  <subjects>
    <subject>Agriculture</subject>
    <subject>ATKIS</subject>
    <subject>Land Cover</subject>
    <subject>Land Use</subject>
    <subject>Remote Sensing</subject>
  <contributors>
    <contributor contributorType="ContactPerson">
      <contributorName>Waldhoff, Guido</contributorName>
    </contributor>
    <contributor contributorType="Producer">
      <contributorName>CRC/Transregio 32, Sub-Project Z1/INF</contributorName>
    </contributor>
    <contributor contributorType="Funder">
      <contributorName>German Research Foundation (DFG)</contributorName>
    </contributor>
    <contributor contributorType="HostingInstitution">
      <contributorName>Regional Computing Centre Cologne</contributorName>
    </contributor>
  </contributors>
  <dates>
    <date dateType="Created">2013-12-23</date>
  </dates>
  <language>eng</language>
  <resourceType resourceTypeGeneral="Dataset"></resourceType>
  <sizes>
    <size>1 Datasets</size>
  </sizes>
  <formats>
    <format>application/zip</format>
  </formats>
  <version></version>
  <rights>TR32DB Data policy agreement</rights>
  <descriptions>
    <description descriptionType="Abstract">This data set contains the land use
classification of 2012 for the study area of the CRC/TR 32: "Patterns in Soil-
Vegetation-Atmosphere Systems", which corresponds to the Rur-catchment. The
classification is provided in GeoTIFF and in ASCII format.
Spatial resolution: 15 m; Projection: WGS84, UTM Zone 32N.</description>
  </descriptions>
</resource>
```

Figure B-5: Example XML Code for DOI registration

Figure B-6: Climate data overview at web-interface Version 3.1

C Appendix: Database structure

C.1.1 Database structure Version 3.1

Overview of the database structure Version 3.1
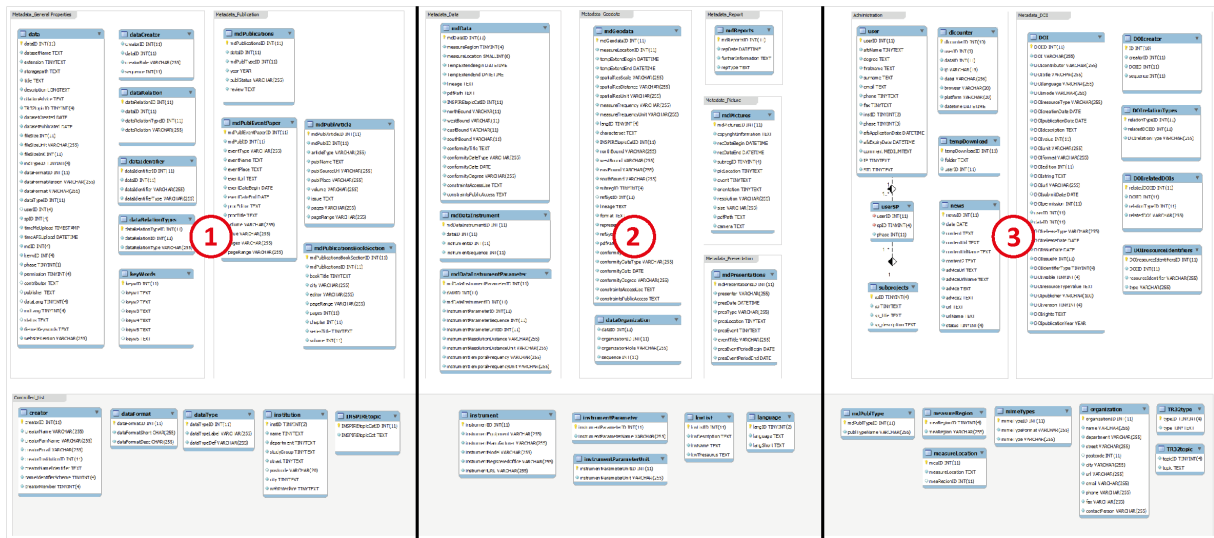
Figure C-1: Overview of database structure Version 3.1

Detailed structure

The detailed structure of the TR32DB Database Version 3.1 is presented on the following pages. The overview presents the arrangement of the single pages of the structure.

Metadata_General Properties

data
- dataID INT(11)
- datasetName TEXT
- extension TINYTEXT
- storagepath TEXT
- title TEXT
- description LONGTEXT
- citationAdvice TEXT
- TR32topicID TINYINT(4)
- datasetCreated DATE
- datasetPublicated DATE
- fileSize INT(11)
- fileSizeUnit VARCHAR(255)
- fileSizeInK INT(11)
- mdTypeID TINYINT(4)
- dataFormatID INT(11)
- dataFormatVersion VARCHAR(255)
- dataFormat VARCHAR(255)
- dataTypeID INT(11)
- userID INT(4)
- spID INT(4)
- timeMdUpload TIMESTAMP
- timeAFSUpload DATETIME
- mdID INT(4)
- keywID INT(4)
- phase TINYINT(1)
- permission TINYINT(4)
- contributor TEXT
- publisher TEXT
- dataLang TINYINT(4)
- mdLang TINYINT(4)
- status TEXT
- GemetKeywords TEXT
- websiteVersion VARCHAR(255)

dataCreator
- creatorID INT(11)
- dataID INT(11)
- creatorRole VARCHAR(255)
- sequence INT(11)

dataRelation
- dataRelationID INT(11)
- dataID INT(11)
- dataRelationTypeID INT(11)
- dataRelation VARCHAR(255)

dataIdentifier
- dataIdentifierID INT(11)
- dataID INT(11)
- dataIdentifier VARCHAR(255)
- dataIdentifierType VARCHAR(255)

dataRelationTypes
- dataRelationTypeID INT(11)
- dataRelationID INT(11)
- dataRelationType VARCHAR(255)

keyWords
- keywID INT(11)
- keyw1 TEXT
- keyw2 TEXT
- keyw3 TEXT
- keyw4 TEXT
- keyw5 TEXT
- keyw6 TEXT

Metadata_Publication

mdPublications
- mdPublicationsID INT(11)
- dataID INT(11)
- mdPublTypeID INT(11)
- year YEAR
- publStatus VARCHAR(255)
- review TEXT

mdPublEventPaper
- mdPublEventPaperID INT(11)
- mdPublID INT(11)
- eventType VARCHAR(255)
- eventName TEXT
- eventPlace TEXT
- eventUrl TEXT
- eventDateBegin DATE
- eventDateEnd DATE
- procEditor TEXT
- procTitle TEXT
- volume VARCHAR(255)
- issue VARCHAR(255)
- pages VARCHAR(255)
- pageRange VARCHAR(255)

mdPublArticle
- mdPublArticleID INT(11)
- mdPublID INT(11)
- articleType VARCHAR(255)
- publName TEXT
- publSourceUrl VARCHAR(255)
- publPlace VARCHAR(255)
- volume VARCHAR(255)
- issue TEXT
- pages VARCHAR(255)
- pageRange VARCHAR(255)

mdPublicationsBookSection
- mdPublicationsBookSectionID INT(11)
- mdPublicationsID INT(11)
- bookTitle TINYTEXT
- city VARCHAR(255)
- editor VARCHAR(255)
- pageRange VARCHAR(255)
- pages INT(11)
- chapter INT(11)
- seriesTitle TINYTEXT
- volume INT(11)

① next page

Controlled_List

creator
- creatorID INT(11)
- creatorName VARCHAR(255)
- creatorFamName VARCHAR(255)
- creatorEmail VARCHAR(255)
- creatorInstitutionID INT(11)
- creatorNameIdentifier TEXT
- nameIdentifierScheme TINYINT(4)
- creatorMember TINYINT(4)

dataFormat
- dataFormatID INT(11)
- dataFormatShort CHAR(255)
- dataFormatDesc CHAR(255)

dataType
- dataTypeID INT(11)
- dataTypeLabel VARCHAR(255)
- dataTypeDef VARCHAR(255)

institution
- instID TINYINT(2)
- name TINYTEXT
- department TINYTEXT
- studyGroup TINYTEXT
- street TINYTEXT
- postcode VARCHAR(20)
- city TINYTEXT
- webInterface TINYTEXT

INSPIREtopic
- INSPIREtopicCatID INT(11)
- INSPIREtopicCat TEXT

Metadata_Data

mdData
- mdDataID INT(11)
- measureRegion TINYINT(4)
- measureLocation SMALLINT(6)
- TempExtendBegin DATETIME
- TempExtendEnd DATETIME
- lineage TEXT
- pdfPath TEXT
- INSPIREtopicCatID INT(11)
- northBound VARCHAR(11)
- westBound VARCHAR(11)
- eastBound VARCHAR(11)
- southBound VARCHAR(11)
- conformityTitle TEXT
- conformityDateType VARCHAR(255)
- conformityDate DATE
- conformityDegree VARCHAR(255)
- constraintsAccessUse TEXT
- constraintsPublicAccess TEXT

mdDataInstrument
- mdDataInstrumentID INT(11)
- dataID INT(11)
- instrumentID INT(11)
- instrumentSequence INT(11)

mdDataInstrumentParameter
- mdDataInstrumentParameterID INT(11)
- dataID INT(11)
- mdDataInstrumentID INT(11)
- instrumentParameterID INT(11)
- instrumentParameterSequence INT(11)
- instrumentParameterUnitID INT(11)
- instrumentResolutionDistance VARCHAR(255)
- instrumentResolutionDistanceUnit VARCHAR(255)
- instrumentTemporalFrequency VARCHAR(255)
- instrumentTemporalFrequencyUnit VARCHAR(255)

Metadata_Geodata

mdGeodata
- mdGeodataID INT(11)
- measureLocationID INT(11)
- tempExtendBegin DATETIME
- tempExtendEnd DATETIME
- spatialResScale VARCHAR(255)
- spatialResDistance VARCHAR(255)
- spatialResUnit VARCHAR(255)
- measureFrequency VARCHAR(255)
- measureFrequencyUnit VARCHAR(255)
- langID TINYINT(4)
- characterset TEXT
- INSPIREtopicCatID INT(11)
- northBound VARCHAR(255)
- westBound VARCHAR(255)
- eastBound VARCHAR(255)
- southBound VARCHAR(255)
- subregID TINYINT(4)
- refSysID INT(11)
- lineage TEXT
- format TEXT
- representation TEXT
- refSysName TEXT
- pdfPath TEXT
- conformityTitle TEXT
- conformityDateType VARCHAR(255)
- conformityDate DATE
- conformityDegree VARCHAR(255)
- constraintsAccessUse TEXT
- constraintsPublicAccess TEXT

dataOrganization
- dataID INT(11)
- organizationID INT(11)
- organizationRole VARCHAR(255)
- sequence INT(11)

Metadata_Report

mdReports
- mdReportsID INT(11)
- repDate DATETIME
- furtherInformation TEXT
- repType TEXT

Metadata_Picture

mdPictures
- mdPicturesID INT(11)
- copyrightInformation TEXT
- recDateBegin DATETIME
- recDateEnd DATETIME
- subregID TINYINT(4)
- picLocation TINYTEXT
- event TINYTEXT
- orientation TINYTEXT
- resolution VARCHAR(255)
- size VARCHAR(255)
- pdfPath TEXT
- camera TEXT

Metadata_Presentation

mdPresentations
- mdPresentationsID INT(11)
- presenter VARCHAR(255)
- presDate DATETIME
- presType VARCHAR(255)
- presLocation TINYTEXT
- presEvent TINYTEXT
- eventTitle VARCHAR(255)
- presEventPeriodBegin DATE
- presEventPeriodEnd DATE

instrument
- instrumentID INT(11)
- instrumentEquipment VARCHAR(255)
- instrumentManufacturer VARCHAR(255)
- instrumentModel VARCHAR(255)
- instrumentRegisteredOffice VARCHAR(255)
- instrumentURL VARCHAR(255)

instrumentParameter
- instrumentParameterID INT(11)
- instrumentParameterName VARCHAR(255)

instrumentParameterUnit
- instrumentParameterUnitID INT(11)
- instrumentParameterUnit VARCHAR(255)

kwList
- kwListID INT(11)
- kwDescription TEXT
- kwName TEXT
- kwThesaurus TEXT

language
- langID TINYINT(2)
- language TEXT
- langShort TEXT

②

previous page

next page

previous page

Administration

user
- userID INT(11)
- afsName TINYTEXT
- degree TEXT
- firstname TEXT
- surname TEXT
- email TEXT
- phone TINYTEXT
- fax TINYTEXT
- instID TINYINT(2)
- phase TINYINT(3)
- afsApplicationDate DATETIME
- afsExpiryDate DATETIME
- comment MEDIUMTEXT
- IP TINYTEXT
- SID TINYTEXT

dlcounter
- dlcounterID INT(10)
- userID INT(3)
- dataID INT(11)
- ip VARCHAR(15)
- datei VARCHAR(256)
- browser VARCHAR(20)
- platform VARCHAR(20)
- datetime DATETIME

tempDownload
- tempDownloadID INT(11)
- folder TEXT
- userID INT(11)

1
1..*

userSP
- userID INT(11)
- spID TINYINT(4)
- phase INT(11)

1..*

1

subprojects
- spID TINYINT(4)
- sp TINYTEXT
- sp_title TEXT
- sp_description TEXT

news
- newsID INT(11)
- date DATE
- content TEXT
- contentUrl TEXT
- contentUrlName TEXT
- content2 TEXT
- adviceUrl TEXT
- adviceUrlName TEXT
- advice TEXT
- advice2 TEXT
- url TEXT
- urlName TEXT
- status TINYINT(4)

Metadata_DOI

DOI
- DOIID INT(11)
- DOI VARCHAR(255)
- DOIcontributor VARCHAR(255)
- DOItitle VARCHAR(255)
- DOIlanguage VARCHAR(255)
- DOImode VARCHAR(255)
- DOIresourceType VARCHAR(255)
- DOIcreationDate DATE
- DOIpublicationDate DATE
- DOIdescription TEXT
- DOIvalue INT(11)
- DOIunit VARCHAR(255)
- DOIformat VARCHAR(255)
- DOIedition INT(11)
- DOIstring TEXT
- DOIurl VARCHAR(255)
- DOIsubmitDate DATE
- DOIpermission INT(11)
- userID INT(11)
- dataID INT(11)
- DOIreleaseType VARCHAR(255)
- DOIreleaseDate DATE
- DOIissueDate DATE
- DOIissueNr INT(11)
- DOIidentifierType TINYINT(4)
- DOIvisible TINYINT(4)
- DOIresourceTypeValue TEXT
- DOIpublisher VARCHAR(100)
- DOIversion TINYINT(4)
- DOIrights TEXT
- DOIpublicationYear YEAR

③

DOIcreator
- ID INT(10)
- creatorID INT(11)
- DOIID INT(11)
- sequence INT(11)

DOIrelationTypes
- relationTypeID INT(11)
- relatedDOIID INT(11)
- DOIrelationType VARCHAR(255)

DOIrelatedDOIs
- relatedDOIID INT(11)
- DOIID INT(11)
- relationTypeID INT(11)
- relatedDOI VARCHAR(255)

DOIresourceIdentifiers
- DOIresourceIdentifiersID INT(11)
- DOIID INT(11)
- resourceIdentifier VARCHAR(255)
- type VARCHAR(255)

mdPublType
- mdPublTypeID INT(11)
- publTypeName VARCHAR(255)

measureRegion
- meaRegionID TINYINT(4)
- meaRegion VARCHAR(255)

measureLocation
- meaID INT(11)
- measureLocation TEXT
- meaRegionID INT(11)

mimeTypes
- mimeTypeID INT(11)
- mimeTypeFormat VARCHAR(255)
- mimeType VARCHAR(255)

organization
- organizationID INT(11)
- name VARCHAR(255)
- department VARCHAR(255)
- street VARCHAR(255)
- postcode INT(11)
- city VARCHAR(255)
- url VARCHAR(255)
- email VARCHAR(255)
- phone VARCHAR(255)
- fax VARCHAR(255)
- contactPerson VARCHAR(255)

TR32type
- typeID TINYINT(4)
- type TINYTEXT

TR32topic
- topicID TINYINT(4)
- topic TEXT

C.1.2 Database structure Version 4.0

Overview of the database structure Version 4.0

Figure C-2: Overview of database structure Version 4.0

Detailed structure

The detailed structure of the TR32DB Database Version 4.0 is presented on the following pages. The overview presents the arrangement of the single pages of the structure.

ControlledVocabularyList_specific

listPublicationStatus
- publicationStatusID TINYINT(3)
- publicationStatus VARCHAR(20)

listArticleType
- articleTypeID TINYINT(3)
- articleType VARCHAR(30)

listPublicationType
- publicationTypeID TINYINT(3)
- publicationType VARCHAR(20)
- publicationTypeVal VARCHAR(20)

1

1

1

1

Metadata_Publication

mdPubBookSection
- dataID INT(10)
- bookTitle VARCHAR(250)
- bookEditor VARCHAR(150)
- city VARCHAR(50)
- volume VARCHAR(50)
- numPages SMALLINT(6)
- pageStart VARCHAR(6)
- pageEnd VARCHAR(6)
- chapter VARCHAR(50)
- seriesTitle VARCHAR(150)
- seriesEditor VARCHAR(150)

mdPubBook
- dataID INT(10)
- seriesEditor VARCHAR(170)
- seriesTitle VARCHAR(150)
- city VARCHAR(50)
- volume VARCHAR(50)
- numPages SMALLINT(6)
- pageStart SMALLINT(6)
- pageEnd SMALLINT(6)
- edition VARCHAR(50)

mdPubEventPaper
- dataID INT(10)
- eventTypeID TINYINT(3)
- event VARCHAR(150)
- eventLocation VARCHAR(150)
- eventUrl VARCHAR(150)
- eventStart DATE
- eventEnd DATE
- procEditor VARCHAR(150)
- procTitle VARCHAR(250)
- procUrl VARCHAR(150)
- issue VARCHAR(150)
- volume VARCHAR(50)
- numPages SMALLINT(6)
- pageStart SMALLINT(6)
- pageEnd SMALLINT(6)

mdPubArticle
- dataID INT(10)
- articleTypeID TINYINT(3)
- source VARCHAR(150)
- sourceUrl VARCHAR(250)
- issue VARCHAR(150)
- volume VARCHAR(50)
- numPages SMALLINT(6)
- pageStart SMALLINT(6)
- pageEnd SMALLINT(6)

mdPublication
- dataID INT(10)
- publicationTypeID TINYINT(3)
- publicationStatusID TINYINT(3)
- year YEAR
- review TINYINT(1)

0..*

1

1

1

1

1

1..*

1..*

1

1

1

1

0..

1

①

Metadata_MetadataDetails

afsData
- dataID INT(10)
- mdTypeID TINYINT(3)
- subprojectID TINYINT(3)
- phaseID TINYINT(3)
- name VARCHAR(150)
- extension VARCHAR(10)
- path VARCHAR(250)
- size INT(11)
- afsUpload DATETIME

1

1..*   1..*   1..*

Administration

news
- newsID SMALLINT(5)
- date DATE
- content TEXT
- contentUrl VARCHAR(150)
- contentUrlName VARCHAR(150)
- content2 TEXT
- advice TEXT
- adviceUrl VARCHAR(150)
- adviceUrlName VARCHAR(150)
- advice2 TEXT
- url VARCHAR(150)
- urlName VARCHAR(150)
- status TINYINT(1)

userAdmin
- userID SMALLINT(5)
- adminTypeID TINYINT(3)

listUser
- userID SMALLINT(5)
- institutionID SMALLINT(5)
- afsName VARCHAR(10)
- degree VARCHAR(30)
- firstname VARCHAR(50)
- surname VARCHAR(75)
- mail VARCHAR(150)
- phone VARCHAR(20)
- fax VARCHAR(20)
- comment TEXT
- ip VARCHAR(39)
- SID TINYTEXT

whoIsOnline
- userIP VARCHAR(39)
- userID SMALLINT(5)
- stamp DATETIME

dlCounter
- dlCounterID INT(10)
- dataID INT(10)
- userID SMALLINT(5)
- ip VARCHAR(39)
- data VARCHAR(250)
- browser VARCHAR(50)
- platform VARCHAR(50)
- dlTime DATETIME

userSubproject
- userID SMALLINT(5)
- subprojectID TINYINT(3)
- phaseID TINYINT(3)

1..*   1..*

1   1

0..*

1..*

1..*

1

1..*

1

1

0..*

0..*

1..*   1..*

next page

page 249

listEventType
- 🔑 eventTypeID TINYINT(3)
- 🔹 eventType VARCHAR(30)

listPresentationType
- 🔑 presentationTypeID TINYINT(3)
- 🔹 presentationType VARCHAR(30)

listReportType
- 🔑 reportTypeID TINYINT(3)
- 🔹 reportType VARCHAR(30)
- 🔹 reportTypeVal VARCHAR(30)

listOrientation
- 🔑 orientationID TINYINT(3)
- 🔹 orientation VARCHAR(20)

Metadata_Picture

mdPicture
- 🔑 dataID INT(10)
- 🔹 orientationID TINYINT(3)
- 🔹 copyright TEXT
- 🔹 recStart DATETIME
- 🔹 recEnd DATETIME
- 🔹 resolution SMALLINT(6)
- 🔹 sizeX SMALLINT(6)
- 🔹 sizeY SMALLINT(6)
- 🔹 recordPlace VARCHAR(150)
- 🔹 recordMethod VARCHAR(150)
- 🔹 eventTypeID TINYINT(3)
- 🔹 event VARCHAR(150)
- 🔹 eventLocation VARCHAR(150)
- 🔹 eventUrl VARCHAR(150)
- 🔹 eventStart DATE
- 🔹 eventEnd DATE

Metadata_Presentation

mdPresentation
- 🔑 dataID INT(10)
- 🔴 eventTypeID TINYINT(3)
- 🔴 presentationTypeID TINYINT(3)
- 🔹 presenter VARCHAR(150)
- 🔹 presDate DATE
- 🔹 eventStart DATE
- 🔹 eventEnd DATE
- 🔹 event VARCHAR(200)
- 🔹 eventLocation VARCHAR(150)
- 🔹 eventUrl VARCHAR(250)

Metadata_Report

mdReport
- 🔑 dataID INT(10)
- 🔴 reportTypeID TINYINT(3)
- 🔹 repDate DATE
- 🔹 furtherInfo TEXT
- 🔹 city VARCHAR(50)
- 🔹 institution VARCHAR(150)
- 🔹 volume VARCHAR(50)
- 🔹 numPages SMALLINT(6)
- 🔹 pageStart SMALLINT(6)
- 🔹 pageEnd SMALLINT(6)

②

Metadata_Identification

mdIdentification
- 🔑 dataID INT(10)
- 🔹 citationAdvice TEXT
- 🔹 pdfPath VARCHAR(100)
- 🔹 graphicFileName VARCHAR(70)

dataTitle
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- 🔴 titleTypeID TINYINT(3)
- 🔹 dataTitle VARCHAR(250)

dataDescription
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- 🔴 descriptionTypeID TINYINT(3)
- 🔹 dataDescription TEXT

dataIdentifier
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- 🔴 identifierTypeID TINYINT(3)
- 🔹 dataIdentifier VARCHAR(300)

dataRelation
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- 🔹 dataRelation VARCHAR(300)
- 🔴 relationTypeID TINYINT(3)
- 🔴 identifierTypeID TINYINT(3)

previous page

next page

page 250

245

listReferenceSystemType ▼
- 🔑 referenceSystemTypeID TINYINT(4)
- ◇ referenceSystemType VARCHAR(20)
- ◇ referenceSystemTypeDesciption TEXT

listReferenceSystem ▼
- 🔑 referenceSystemID TINYINT(3)
- ◇ referenceSystemCode VARCHAR(15)
- ◇ referenceSystemName VARCHAR(50)
- ◇ referenceSystemDescription TEXT
- ● referenceSystemTypeID TINYINT(4)

listPresentationForm ▼
- 🔑 presentationFormID TINYINT(3)
- ◇ presentationForm VARCHAR(20)
- ◇ presentationFormDescription TEXT

listMaintenanceFrequencyUnit ▼
- 🔑 maintenanceFrequencyUnitID TINYINT(3)
- ◇ maintenanceFrequencyUnit VARCHAR(20)
- ◇ maintenanceFrequencyUnitDescription TEXT

Metadata_Geodata

dataReferenceSystem ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ● referenceSystemID TINYINT(3)

dataVerticalExtent ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◇ minimumValue DECIMAL(10,0)
- ◇ maximumValue DECIMAL(10,0)
- ● referenceSystemID TINYINT(3)

dataPresentationForm ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ● presentationFormID TINYINT(3)

dataMaintenanceInformation ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◇ maintenanceInformation TEXT
- ● maintenanceFrequencyUnitID TINYINT(3)

③

mdMetadata ▼
- 🔑 dataID INT(10)
- ● websiteVersionID TINYINT(3)
- ● userID SMALLINT(5)
- ◇ firstUpload DATETIME
- ◇ lastUpload DATETIME
- ● languageID SMALLINT(5)
- ◇ mdVisible TINYINT(2)

Metadata_ResponsibleParty

dataCreator ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ● roleTypeID TINYINT(3)
- ◇ creatorID SMALLINT(5)

dataContributor ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ● contributorTypeID TINYINT(3)
- ◇ dataContributor VARCHAR(150)

dataOrganization ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ● roleTypeID TINYINT(3)
- ● organizationID SMALLINT(5)

mdResponsibleParty ▼
- 🔑 dataID INT(10)
- ◇ publisher VARCHAR(150)

page 251

previous page

next page

246

listSpatialRepresentationType ▼
- 🔑 spatialRepresentationTypeID TINYINT(3)
- ◆ spatialRepresentationType VARCHAR(30)
- ◆ spatialRepresentationTypeDescription TEXT

listMeasureUnit ▼
- 🔑 measureUnitID TINYINT(3)
- ◆ measureUnit VARCHAR(20)
- ◆ measureUnitDescription TEXT

listInitiativeType ▼
- 🔑 initiativeTypeID TINYINT(3)
- ◆ initiativeType VARCHAR(20)
- ◆ initiativeTypeDescription TEXT

listScopeCode ▼
- 🔑 scopeCodeID TINYINT(3)
- ◆ scopeCode VARCHAR(25)
- ◆ scopeCodeVal VARCHAR(20)
- ◆ scopeCodeDescription TEXT

dataSpatialRepresentation ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ spatialRepresentationTypeID TINYINT(3)

dataSpatialResolution ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◇ measureUnitID TINYINT(3)
- ◇ distance SMALLINT(6)
- ◇ equivalentScale DECIMAL(10,0)

mdGeodata ▼
- 🔑 dataID INT(10)
- ◆ initiativeTypeID TINYINT(3)
- ◆ scopeCodeID TINYINT(3)
- ◆ tempExtentStart DATETIME
- ◆ tempExtentEnd DATETIME
- ◇ lineage TEXT

④

dataConformity ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ specification TINYINT(3)
- ◆ dateTypeID TINYINT(3)
- ◇ conformDate DATE
- ◇ conformityDegreeID TINYINT(3)
- ◇ explanation TEXT

Metadata_Topic

mdTopic ▼
- 🔑 dataID INT(10)
- ◆ tr32TopicID TINYINT(3)
- ◆ isoTopicCategoryID TINYINT(3)

dataGemetInspireSpatialDataTheme ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ gemetInspireSpatialDataThemeID TINYINT(3)

dataKw ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ kwID SMALLINT(5)

dataGemet ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ dataGemet VARCHAR(75)

dataDDC ▼
- 🔑 dataID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ DDCID TINYINT(3)

previous page

next page

page 252

247

listInstrumentParameter
- instrumentParameterID SMALLINT(6)
- instrumentParameter VARCHAR(150)

listInstrument
- instrumentID SMALLINT(5)
- instrument VARCHAR(150)
- manufacturer VARCHAR(150)
- model VARCHAR(150)
- registeredOffice VARCHAR(150)
- url VARCHAR(150)

listResolutionDistanceUnit
- resolutionDistanceUnitID SMALLINT(5)
- resolutionDistanceUnit VARCHAR(20)

listInstrumentParameterUnit
- instrumentParameterUnitID SMALLINT(5)
- instrumentParameterUnit VARCHAR(25)

listTemporalFrequencyUnit
- temporalFrequencyUnitID SMALLINT(5)
- temporalFrequencyUnit VARCHAR(20)

Metadata_Data

dataInstrumentParameter
- dataInstrumentID INT(10)
- dataID INT(10)
- sequence TINYINT(3)
- instrumentParameterID SMALLINT(6)
- instrumentParameterUnitID SMALLINT(5)
- resolutionDistanceUnitID SMALLINT(5)
- temporalFrequencyUnitID SMALLINT(5)
- resolutionDistance SMALLINT(6)
- temporalfrequency SMALLINT(6)

dataInstrument
- dataInstrumentID INT(10)
- dataID INT(10)
- instrumentID SMALLINT(5)
- sequence TINYINT(3)

mdData
- dataID INT(10)
- initiativeTypeID TINYINT(3)
- scopeCodeID TINYINT(3)
- tempExtentStart DATETIME
- tempExtentEnd DATETIME
- lineage TEXT

⑤

previous page

Metadata_FileDetails

mdFileDetails
- dataID INT(10)
- languageID SMALLINT(5)
- dataStatusID TINYINT(3)
- dataTypeID TINYINT(3)
- dataFormatID TINYINT(3)
- dataFormatVersion VARCHAR(30)

dataDate
- dataID INT(10)
- sequence TINYINT(3)
- dateTypeID TINYINT(3)
- dataDate DATE

dataSize
- dataID INT(10)
- sequence TINYINT(4)
- sizeTypeID TINYINT(3)
- dataSize VARCHAR(20)

Metadata_Constraints

mdConstraint
- dataID INT(10)
- accessCon TEXT
- accessUse TEXT
- licenceID TINYINT(3)
- dlPermID TINYINT(3)
- dlInfo TEXT

Metadata_Geographic

mdGeographic
- dataID INT(10)
- north VARCHAR(12)
- east VARCHAR(12)
- south VARCHAR(12)
- west VARCHAR(12)
- measureLocationID TINYINT(4)

page 253

248

page 244

ControlledVocabularyList_general

listPhase
🔑	phaseID TINYINT(3)
🔷	phase VARCHAR(5)

listAdminType
🔑	adminTypeID TINYINT(3)
🔷	adminType VARCHAR(50)
🔷	adminTypeDescription TEXT

listSubproject
🔑	subprojectID TINYINT(3)
🔷	subproject VARCHAR(10)
🔷	comment TEXT

listMdType
🔑	mdTypeID TINYINT(3)
🔷	mdType VARCHAR(20)
🔷	mdTypeVal VARCHAR(20)

⑥

Metadata_DOI

page 245

listDescriptionType
- descriptionTypeID TINYINT(3)
- descriptionType VARCHAR(20)
- descriptionTypeDOI VARCHAR(20)
- descriptionTypeVal VARCHAR(20)
- descriptionTypeDOIVal VARCHAR(20)

listRelationType
- relationTypeID TINYINT(3)
- relationType VARCHAR(30)
- relationTypeDOI VARCHAR(30)
- relationTypeDescription TEXT
- relationTypeDescriptionDOI TEXT
- relationTypeVal VARCHAR(30)
- relationTypeDOIVal VARCHAR(30)

listTitleType
- titleTypeID TINYINT(3)
- titleType VARCHAR(20)
- titleTypeDOI VARCHAR(20)
- titleTypeVal VARCHAR(20)
- titleTypeDOIVal VARCHAR(20)

listIdentifierType
- identifierTypeID TINYINT(3)
- identifierType VARCHAR(30)
- identifierTypeDOI VARCHAR(30)
- identifierTypeUrl VARCHAR(150)
- identifierTypeDescription TEXT
- identifierTypeDescriptionShort VARCHAR(50)
- identifierTypeVal VARCHAR(30)
- identifierTypeDOIVal VARCHAR(30)

⑦

previous page

next page

DOItitle
- DOIID INT(10)
- sequence TINYINT(3)
- titleTypeID TINYINT(3)
- DOItitle VARCHAR(250)

DOIcreator
- DOIID INT(10)
- sequence TINYINT(3)
- creatorID SMALLINT(5)

DOIdescription
- DOIID INT(10)
- sequence TINYINT(3)
- descriptionTypeID TINYINT(3)
- DOIdescription TEXT

DOIrelatedIdentifier
- DOIID INT(10)
- sequence TINYINT(3)
- identifierTypeID TINYINT(3)
- relationTypeID TINYINT(3)
- DOIrelatedIdentifier VARCHAR(150)

DOIalternateIdentifier
- DOIID INT(10)
- sequence TINYINT(3)
- identifierTypeID TINYINT(3)
- DOIalternateIdentifier VARCHAR(150)

page 246

listCreatorStatus
- creatorStatusID TINYINT(3)
- creatorStatus VARCHAR(30)
- creatorStatusVal VARCHAR(30)

listContributorType
- contributorTypeID TINYINT(3)
- contributorType VARCHAR(30)
- contributorTypeVal VARCHAR(30)
- contributorTypeDOI VARCHAR(30)
- contributorTypeDOIVal VARCHAR(30)
- contributorTypeDescription TEXT
- contributorTypeDescriptionShort VARCHAR(50)

listRoleType
- roleTypeID TINYINT(3)
- roleType VARCHAR(30)
- roleTypeDOI VARCHAR(30)
- roleTypeVal VARCHAR(30)
- roleTypeDOIVal VARCHAR(30)

listCreator
- creatorID SMALLINT(5)
- creatorStatusID TINYINT(3)
- institutionID SMALLINT(5)
- creatorSchemeID TINYINT(3)
- identifier VARCHAR(150)
- firstname VARCHAR(50)
- surname VARCHAR(75)
- mail VARCHAR(150)

listInstitution
- institutionID SMALLINT(5)
- name VARCHAR(100)
- department VARCHAR(100)
- street VARCHAR(100)
- postcode VARCHAR(20)
- city VARCHAR(50)
- country VARCHAR(50)
- website VARCHAR(150)
- institutionStatusID TINYINT(3)

listWebsiteVersion
- websiteVersionID TINYINT(3)
- websiteVersion VARCHAR(10)

listCreatorScheme
- creatorSchemeID TINYINT(3)
- creatorScheme VARCHAR(15)
- creatorSchemeDOI VARCHAR(15)
- creatorSchemeUrl VARCHAR(150)
- creatorSchemeDescription TEXT

listLanguage
- languageID SMALLINT(5)
- language VARCHAR(100)
- languageShort VARCHAR(3)
- languageShort2 VARCHAR(2)

⑧

listInstitutionStatus
- institutionStatusID TINYINT(3)
- institutionStatus VARCHAR(35)

previous page

next page

DOI
- DOIID INT(10)
- dataID INT(10)
- dlPermID TINYINT(3)
- userID SMALLINT(5)
- mimeTypeID VARCHAR(5)
- dataTypeID TINYINT(3)
- licenceID TINYINT(3)
- languageID SMALLINT(5)
- identifierTypeID TINYINT(3)
- DOIidentifier VARCHAR(150)
- DOIvisible TINYINT(1)
- DOIversion TINYINT(3)
- DOIpublisher VARCHAR(150)
- DOIpublicationYear YEAR
- DOIurl VARCHAR(150)
- DOIstring TEXT
- DOIresourceTypeValue VARCHAR(150)
- DOIsubmitToDatacite TINYINT(1)

DOIcontributor
- DOIID INT(10)
- sequence TINYINT(3)
- contributorTypeID TINYINT(3)
- creatorSchemeID TINYINT(3)
- DOIcontributor VARCHAR(150)
- DOIcontributorIdentifier VARCHAR(150)

DOIsubjectGemet
- DOIID INT(10)
- sequence TINYINT(3)
- DOIsubjectGemet VARCHAR(75)

251

page 247

listKw
- kwID SMALLINT(5)
- kw VARCHAR(75)
- kwDescription TEXT

listConformityDegree
- conformityDegreeID TINYINT(3)
- conformityDegree VARCHAR(15)
- conformityDegreeVal VARCHAR(15)
- conformityDegreeDescription TEXT

listDDC
- DDCID TINYINT(3)
- DDC VARCHAR(50)

listTr32Topic
- tr32TopicID TINYINT(3)
- tr32Topic VARCHAR(30)
- tr32TopicVal VARCHAR(30)

listConformitySpecification
- specificationID TINYINT(3)
- specification VARCHAR(60)
- specificationDescription TEXT
- specificationURL VARCHAR(45)

listGemetInspireSpatialDataTheme
- gemetInspireSpatialDataThemeID TINYINT(3)
- gemetInspireSpatialDataTheme VARCHAR(70)
- gemetInspireSpatialDataThemeDOI VARCHAR(50)
- gemetInspireSpatialDataThemeDescription TEXT

listIsoTopicCategory
- isoTopicCategoryID TINYINT(3)
- isoTopicCategory VARCHAR(35)
- isoTopicCategoryDOI VARCHAR(35)
- isoTopicCategoryDescription TEXT

previous page

next page

⑨

DOIsubjectDDC
- DOIID INT(10)
- sequence TINYINT(3)
- DDCID TINYINT(3)

DOIsubjectGemetInspire
- DOIID INT(10)
- sequence TINYINT(3)
- gemetInspireSpatialDataThemeID TINYINT(3)

DOIsubjectTr32
- DOIID INT(10)
- sequence TINYINT(3)
- kwID SMALLINT(5)

DOIsubjectIso
- DOIID INT(10)
- sequence TINYINT(3)
- isoTopicCategoryID TINYINT(3)

252

page 248

listDataFormat ▼
- 🔑 dataFormatID TINYINT(3)
- ◇ dataFormat VARCHAR(30)
- ◇ dataFormatDescription TEXT
- ◆ dataFormatTypeID TINYINT(3)

listDataStatus ▼
- 🔑 dataStatusID TINYINT(3)
- ◇ dataStatus VARCHAR(30)
- ◇ dataStatusVal VARCHAR(30)

listSizeType ▼
- 🔑 sizeTypeID TINYINT(3)
- ◇ sizeType VARCHAR(30)
- ◇ sizeTypeDOI VARCHAR(30)
- ◇ sizeTypeVal VARCHAR(30)
- ◇ sizeTypeDOIVal VARCHAR(30)

listMeasureLocation ▼
- 🔑 measureLocationID TINYINT(4)
- ◆ measureRegionID TINYINT(3)
- ◇ measureLocation VARCHAR(30)

listDataFormatType ▼
- 🔑 dataFormatTypeID TINYINT(3)
- ◇ dataFormatType VARCHAR(30)

listDlPerm ▼
- 🔑 dlPermID TINYINT(3)
- ◇ dlPerm VARCHAR(30)

listMeasureRegion ▼
- 🔑 measureRegionID TINYINT(3)
- ◇ measureRegion VARCHAR(30)

listDateType ▼
- 🔑 dateTypeID TINYINT(3)
- ◇ dateType VARCHAR(20)
- ◇ dateTypeVal VARCHAR(20)
- ◇ dateTypeDescription TEXT
- ◇ dateTypeDOI VARCHAR(20)
- ◇ dateTypeDOIVal VARCHAR(20)
- ◇ dateTypeDOIDescription TEXT
- ◇ dateTypeIso VARCHAR(20)
- ◇ dateTypeIsoVal VARCHAR(20)
- ◇ dateTypeIsoDescription TEXT

listDataType ▼
- 🔑 dataTypeID TINYINT(3)
- ◇ dataType VARCHAR(30)
- ◇ dataTypeVal VARCHAR(30)
- ◇ dataTypeDOI VARCHAR(30)
- ◇ dataTypeDOIVal VARCHAR(30)
- ◇ dataTypeDescription TEXT
- ◇ dataTypeComment TEXT

listLicence ▼
- 🔑 licenceID TINYINT(3)
- ◇ licenser VARCHAR(30)
- ◇ licence VARCHAR(50)
- ◇ licenceShort VARCHAR(20)
- ◇ licenceText TEXT
- ◇ licenceUrl VARCHAR(150)

listMimeType ▼
- 🔑 mimeTypeID VARCHAR(5)
- ◇ mimeType VARCHAR(100)

⑩

previous page

DOIdate ▼
- 🔑 DOIID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ dateTypeID TINYINT(3)
- ◇ DOIdate DATE

DOIsize ▼
- 🔑 DOIID INT(10)
- 🔑 sequence TINYINT(3)
- ◆ sizeTypeID TINYINT(3)
- ◇ DOIsize INT(5)

D   Appendix: TR32DB Metadata Schema

TR32DB Metadata Schema

for the Description of Research Data in the TR32DB

Version 4.0

April 2014
doi: 10.5880/TR32DB.10

Author:

Constanze Curdt

Collaborative Research Centre/Transregio 32, Project Section Z1/INF

University of Cologne, Institute of Geography

Contact:

tr32db-admin@uni-koeln.de

http://www.tr32db.uni-koeln.de

D.1   Introduction

D.1.1   The TR32DB

Managing and archiving research data in a well-organised framework is an essential task in every interdisciplinary, long-term research project. All data created or collected within the project funding has to be stored and backed up including accurate description with metadata. In DFG-funded Collaborative Research Centres (CRC) is a project section *Information Infrastructure* (so-called *INF*-project) responsible to set up a systematically research data management (RDM) system. The focus of the RDM system should be on long-term usage to manage all relevant research data created within the context of the project. This should fulfil the obligations of the *Good Scientific Practice* to archive the data for at least ten years and also support synergies and communication between the researchers.

Within the framework of the CRC/Transregio 32 "Patterns in Soil-Vegetation-Atmosphere-Systems: Monitoring, Modeling and Data Assimilation" (CRC/TR32, www.tr32.de), a web-based RDM system, a so-called project database, was implemented by the INF-project section to handle all relevant research data. The CRC/TR32-Database (TR32DB, www.tr32db.uni-koeln.de) is operating since early 2008 and is physically located at the Regional Computing Centre of the University of Cologne (RRZK). Secure, sustainable archive and back up is provided within this environment. The developed metadata management system (MMS) is the central element of the system. It was designed, realized, and implemented according to the needs of the interdisciplinary project background, the demands of the stored data types, and to recent metadata standards. The present TR32DB Metadata Schema provides the basis for the MMS.

D.1.2   The TR32D Metadata Schema

The TR32DB Metadata Schema is a structured list of metadata properties chosen to describe all data in the TR32DB with accurate metadata properties and thus to improve their searchability. The entire data provided to the TR32DB can be described with a number of descriptive metadata properties (e.g. creator, title, abstract, keywords, etc.) and administrative or technical properties (e.g. file format, file type, rights statement, etc.). The stored data are organized in six main data type categories: `Data`, `Geodata`, `Report`, `Picture`, `Presentation`, and `Publication`.

The TR32DB Metadata Schema is set up in two levels (Figure D-1) to describe the various types of data collected by the CRC/TR32 participants. The first level is the `general` level. This level includes metadata properties classified in seven categories: Identification, Responsible Party, Topic, File Details, Constraints, Geographic, and automatic generated Metadata Details. The second level is the `specific` level and contains the data type specific metadata properties. Currently, six data types are included: `Data`, `Geodata`, `Report`, `Picture`, `Presentation`, and `Publication`. `Publication` takes a special position and is once again subdivided into the sub-categories: Article, Book, Book Section, and Event Paper.

The TR32DB Metadata Schema comprises a defined number of metadata properties. Some of these properties are part of the core set and consequently mandatory. In addition, optional metadata properties are provided as well as automatically generated. All metadata properties are described in detail including the TR32DB property name, a definition, the maximum occurrence, the obligation, as well as supplementary notes with regard to allowed values, examples or other constraints.

The TR32DB Metadata Schema is designed in compliance with Dublin Core[1] and DataCite[2]. It is expanded with metadata elements from ISO 19115[3], INSPIRE[4], various properties appropriate to the stored data types, as well as specific properties related to the CRC/TR32 project background.

---

[1] Dublin Core Metadata Initiative; http://dublincore.org/documents/dcmi-terms/

[2] DataCite Metadata Schema, Version 2.2, http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf

[3] ISO 19115:2003 Geographic information – Metadata; http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020
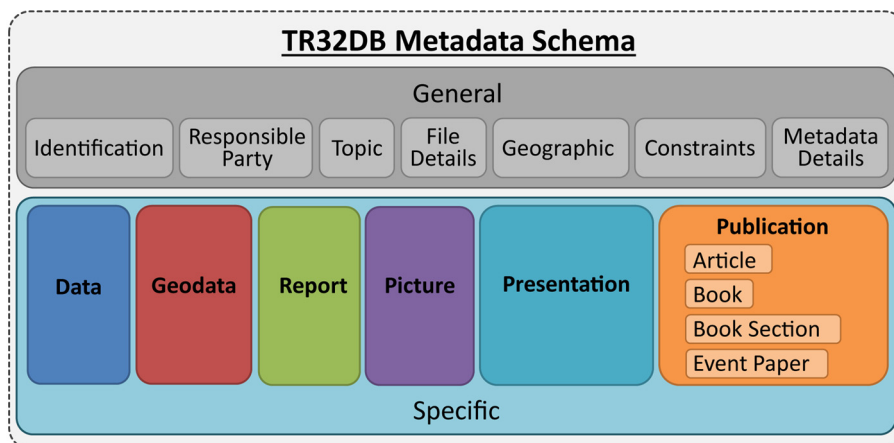
Figure D-1: TR32DB Metadata Schema Structure

D.2 Overview of TR32DB Metadata Schema Properties

An overview of the metadata properties of the TR32DB Metadata Schema is presented in this section. They will be described in detail in the following section. The metadata properties are allocated in three levels of obligation: **Mandatory (M)** properties must be provided, **Optional (O)** properties may be provided, and **Automatic (A)** properties are automatically generated by the TR32DB.

The following tables provide a short form of the metadata properties assorted according to the obligation level. Table D-1 lists all properties with the obligation mandatory (M). These properties must be supplied for a dataset. All properties with the obligation level optional (O) are listed in Table D-1. These may be supplied for a dataset. Table D-1 and Table D-2 are both structured with regard to the general and data type specific metadata properties. All properties listed in Table 3 are automatically generated.

---

4 INSPIRE Implementing Directive; http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:EN:PDF

Table D-1: TR32DB Mandatory Properties

ID	Property
	GENERAL
1	Title (with type)
2	Description (with type)
3	Date (with type)
7	Creator (Creator-Person with role, name, emailAddress, TR32Membership; CreatorInstitution with universityName, instituteName, streetName, postcode, city, country, website; Creator-Organisation with organisationRole, organisationName ,organisationDepartment, streetName, postcode, city, country, website, contactPerson, eMailAddress, phone)
9	Publisher
10	Subject (with TR32topic, TR32keyword)
11	DataType
15	Language
16	DataStatus
17	Download (with DownloadPermission)
19	MeasuringSite (with measuringRegion, measuringLocation)
	DATA
31	TemporalExtent (with startDate, endDate)
	GEODATA
34	TemporalExtent (with startDate, endDate)
	REPORT
44	ReportDate
46	ReportCity
47	ReportInstitution
	PICTURE
52	RecordDate (with startDate, endDate)
53	RecordPlace
	PRESENTATION
61	PresentationDate
63	Event (with eventName, eventLocation, eventPeriod (with startDate, endDate))
	PUBLICATION-Article
64	Status
65	Review
66	Year
67	PublicationType
69	PublicationSource
73	PageRange (with startPage, endPage)
	PUBLICATION-Book
74	Status
75	Review
76	Year
77	PublicationType
80	City
	PUBLICATION-BookSection
85	Status
86	Review
87	Year
88	PublicationType
89	BookTitle
90	BookEditor
91	City
96	PageRange (with startPage, endPage)
	PUBLICATION-EventPaper
97	Status
98	Review
99	Year
100	PublicationType
101	Event (with eventName, eventLocation, eventPeriod)
104	PageRange (with startPage, endPage)

Table D-2: TR32DB Optional Properties

ID	Property
	GENERAL
3	Identifier (with identifierType)
4	Relation (with Identifier, identifierType, relationType)
5	CitationAdvice
6	AdditionDescription (descriptionFileName, graphicFileName)
8	Contributor (with contributorName, contributorType, nameIdentifier, nameIdentifierSchema)
10	Subject (with GemetThesaurus, TopicCategory, InspireTheme, DDC)
12	DataFormat (with dataFormatVersion)
13	DataSize (with dataSizeType)
18	Constraints (with AccessAndUseConstraints, AccessLimitations)
20	GeographicBoundingBox (with westBoundLongitude, eastBoundLongitude, northBoundLatitude, southBoundLatitude)
	DATA
32	Lineage
33	MeasuringInstrumentModelMethod (Instrument with instrumentName, instrumentModel, instrumentManufacturer, manufacturerRegisterdOffice, manufacturerWebsite; InstrumentParameter with ParameterName, parameterUnit, ResolutionDistance, resolutionDistanceUnit, TemporalFrequency, temporalFrequencyUnit)
	GEODATA
35	VerticalExtent (with minimumValue, maximumValue, ReferenceSystem, referenceSystemCode, referenceSystemName)
36	Lineage
37	Spatial Resolution (with scale, distanceNumber, distanceUnit)
38	SpatialRepresentationType
39	PresentationForm
40	ReferenceSystem (with referenceSystemCode, referenceSystemName)
41	MaintenanceInformation (with maintenanceFrequency, maintenanceNote)
42	Scope
43	InitiativeType
	REPORT
45	ReportType
48	ReportVolume
49	NumberOfPages
50	PageRange (with startPage, endPage)
51	FurtherInformation
	PICTURE
54	RecordMethod
55	Orientation
56	Resolution
57	Size (with width, height)
58	CopyrightInformation
59	Event (with eventType, eventName, eventLocation, eventWebsite, eventPeriod (with startDate, endDate))
	PRESENTATION
60	Presenter
62	PresentationType
	PUBLICATION-Article
68	ArticleType
70	Volume
71	Issue
72	NumberOfPages
70	Volume
	PUBLICATION-Book
78	SeriesTitle
79	SeriesEditor
81	Volume
82	Edition
83	NumberOfPages
84	PageRange (with startPage, endPage)
	PUBLICATION-BookSection
92	SeriesTitle
93	SeriesEditor
94	Volume

95	Chapter
96	NumberOfPages
92	SeriesTitle
	PUBLICATION-EventPaper
102	Proceedings (ProceedingsEditor, ProceedingsTitle, ProceedingsURL, Volume, Issue)
103	NumberOfPages

Table D-3: TR32DB Automatic Generated Properties

ID	Property
	METADATA DETAILS (METADATA ON METADATA)
21	TR32DBIdentifier (with TR32DBUrl)
22	TR32MetaDataType
23	MetadataCreator (with creatorFirstName, creatorFamiliyName, academicTitle, eMailAddress, phone, fax, TR32subproject)
24	MetadataCreatorInstitution (with universityName, instituteName, streetName, postcode, city, country, website)
25	mdCreationDate
26	mdLastUpdateDate
27	mdWebVersion
28	mdLanguage
29	fileInformation (with filename, fileExtention, fileFormat, fileSize, fileUploadTime, fileSubproject, fileFundingPhase)
30	Conformity (with specificationTitle, date, dateType, degree, explanation)

D.3 TR32DB Metadata Properties

The following tables provide a detailed description of all mandatory, optional and automatic generated TR32DB metadata properties with corresponding sub-properties. They are clearly arranged as illustrated in Figure 1. First of all, the `general` metadata properties are described in details, following by the data type `specific` metadata properties.

Each property is described with the following information:

- the **identifier (ID)** number of the metadata property
- the **definition** describes the metadata property
- the **occurrence (OCC)** indicates the quantity/cardinality constrains of the property:

    0-n = metadata property is optional and repeatable

    0-1 = metadata property is optional, but not repeatable

    1-n = metadata property is required and repeatable

    1 = metadata property is required, but not repeatable

- the level of **obligation (OB)** for each metadata property is distinguished in:

    Mandatory (M) properties must be provided.

    Optional (O) properties may be provided.

    Automatic (A) properties are automatically generated.

- the **notes** of a metadata property provide information about the allowed values, example, or other constraints, e.g. data value type (free text, date, controlled vocabularies), syntax encoding schema (e.g. encoding for a date)

Throughout the entire document, a naming convention has been used for the properties and sub-properties as follows. All properties begin with a capital letter (e.g. Title, Description, Identifier …). All sub-properties begin with a lower case letter (e.g. titleType, descriptionType, identifierType, …).

D.3.1 TR32DB **General** Metadata Properties

Table D-4: TR32DB General Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Identification					
1	Title	A name given to the dataset. Typically a title will be a name by which the dataset is formally known.	1-n	M	Free Text *Example:* Enhanced Land Use Classification of 2008 for the Rur catchment
1.1	titleType	The type of Title.	1	M	*Controlled List:* **Table D-54: TitleType** See appendix for values and definitions.
2	Description	A textual description of the content of the dataset including additional information.	1-n	M	Free Text
2.1	descriptionType	The type of Description.	1	M	*Controlled List:* **Table D-23: DescriptionType** See appendix for values and definitions.
3	Identifier	A uniquely value (string or number) identifying the dataset. Any alphanumeric string which is unique within its domain of issue.	0-n	O	Free Text *Example:* 10.5880/TR32DB.1
3.1	identiferType	The name of the Identifier type.	0-1	O/ M	If Identifier is used, then identiferType is mandatory. *Controlled List:* **Table D-27: IdentifierType** See appendix for values and definitions.
4	Relation	Relation represents the relationship between the described dataset and another resource or identifier that is related to the described dataset in some way.	0-n	O	
4.1	identifier	A uniquely value (string or number) identifying the dataset. Any alphanumeric string which is unique within its domain of issue.	0-1	O/ M	If relationType and identiferType is used, then identifer is mandatory. Free Text *Example:* 10.5880/TR32DB.1 http://tr32db.uni-koeln.de/data.php?dataID=27
4.2	identifierType	The name of the identifier type.	0-1	O/ M	If relationType and identifier is used, then identiferType is mandatory. *Controlled List:* **Table D-27: IdentifierType** See appendix for values and definitions.
4.3	relationType	The type of the Relation.	0-1	O/ M	If identifierType and identifer is used, then relationType is mandatory. *Controlled List:* **Table D 45: RelationType** See appendix for values and definitions.
5	CitationAdvice	An advice to cite the dataset.	0-1	O	Free Text *Example:* Waldhoff, Guido (2012): Enhanced Land Use Classification of 2008 for the Rur catchment. DOI: 10.5880/TR32DB.1
6	AdditionalDescription	Additional information describing the dataset.			
6.1	descriptionFileN	The name of a text file that	0-1	O	Only text files in PDF format are

	ame	provides additional descriptive information about the dataset. For example detailed explanations can be given about the measured/modelled parameters, site details, or instrument or model method.			permitted with a maximal file size of 5MB. Free Text *Example:* TR32_LU2011_description.pdf
6.2	graphicFileName	The name of a graphic that provides an illustration of the dataset (should include a legend for the graphic).	0-1	O	Only graphic files in JPG or PNG format are permitted with a maximal file size of 2MB. Free Text *Example:* TR32_LU2011_preview.jpg
Responsible Party					
7	Creator	The person(s) and/or organisation(s) associated with the dataset. They can be involved in producing the dataset, responsible for establishment, management, maintenance and distribution of the resource. By default CreatorPerson should be used for research data created within the CRC/TR32.			CreatorOrganisation is currently only applicable for TR32MetaDataType GEODATA. If claiming conformance to INSPIRE and ISO 19115, CreatorOrganisation should be used instead of CreatorPerson.
7.1	CreatorPerson	The main researcher(s) involved in producing the dataset (person responsible for content of the resource) or author(s) of the publication, in priory order. (Creator-Person)	1-n	M	*Controlled List:* Available at **http://www.tr32db.de/creatorList***
7.1.1	creatorRole	The role of the Creator.	1	M	*Controlled List:* **Table D 48: RoleType** See appendix for values and definitions.
7.1.2	creatorFirstName	The first name of the Creator.	1	M	Free Text Example: John
7.1.3	creatorFamilyName	The family name of the Creator.	1	M	Free Text *Example:* Doe
7.1.4	nameIdentifier	The unique identifier of an individual, according to various schemes. Any alphanumeric string which is unique within its domain of issue.	0-1	O	Free Text The format is dependent upon the identifier scheme.
7.1.4.1	nameIdentifierScheme	The name of the name identifier scheme.	0-1	O/ M	If nameIdentifer is used, then nameIdentifierScheme is mandatory. *Controlled List:* **Table D-18: CreatorScheme** See appendix for values and definitions.
7.1.5	emailAddress	The email address of the Creator.	1	M	Free Text *Example:* JohnDoe@gmx.de
7.1.6	TR32Membership	The status of the membership in the CRC/TR32.	1	M	*Controlled List:* **Table D-19: CreatorStatus** See appendix for values and definitions.
7.1.7	CreatorInstitution	The corresponding institution of the dataset creator.	1	M	*Controlled List:* Available at **http://www.tr32db.de/institutionList***
7.1.7.1	universityName	The name of the university or organisation.	1	M	Free Text *Example*: University of Cologne
7.1.7.2	instituteName	The name of the institute.	1	M	Free Text

						Example: Institute of Geography
7.1.7.3	streetName	The name of the postal street and house number as part of the address of the institution.	1	M	Free Text *Example*: Albertus-Magnus-Platz
7.1.7.4	postCode	The postcode as part of the address of the institution.	1	M	Free Text *Example*: 50923
7.1.7.5	city	The name of the city as part of the address of the institution.	1	M	Free Text *Example*: Cologne
7.1.7.6	country	The name of the country as part of the address of the institution.	1	M	Free Text *Example*: Germany
7.1.7.7	website	The website of the institution.	1	M	Free Text *Example*: www.geographie.uni-koeln.de
7.1.7.8	TR32Membership	The status of the membership in the CRC/TR32.	1	M	*Controlled List:* **Table D-28: InstitutionStatus** See appendix for values and definitions.
7.2	CreatorOrganisation	The organisation responsible for the establishment, management, maintenance and distribution of the resource.	0-1	O/M	Mandatory if claiming conformance to INSPIRE and ISO 19115. Therefore, Creator-Organisation should be used instead of Creator-Person. Currently only applicable for TR32MetaDataType GEODATA. *Controlled List:* Available at **http://www.tr32db.de/institutionList***
7.2.1	organisationRole	The role of the organisation.	1	M	Mandatory if claiming conformance to INSPIRE and ISO 19115. *Controlled List:* **Table D 48: RoleType** See appendix for values and definitions.
7.2.2	organisationName	The name of the organisation.	1	M	Mandatory if claiming conformance to INSPIRE and ISO 19115. Free Text *Example*: Bezirksregierung Köln
7.2.3	organisationDepartment	The responsible department of the organisation.	0-1	O	Free Text *Example*: Abteilung 07 - GEObasis.nrw
7.2.4	streetName	The name of the postal street and house number as part of the address of the organisation.	0-1	O	Free Text *Example*: Muffendorfer Str. 19-21
7.2.5	postCode	The postcode as part of the address of the organisation.	0-1	O	Free Text *Example*: 53177
7.2.6	city	The name of the city as part of the address of the organisation.	0-1	O	Free Text *Example*: Bonn
7.2.7	country	The name of the country as part of the address of the organisation.	0-1	O	Free Text *Example*: Germany
7.2.8	website	The website of the organisation.	1	M	Free Text *Example*: http://www.bezreg-koeln.nrw.de
7.2.9	contactPerson	A name of the responsible contact person working at the organisation in order by person surname, given name, and title separated by a	0-1	O	Free Text *Example*: Doe, John

		comma.			
7.2.10	eMailAddress	The e-mail address at which the organisation or an individual person may be contacted.	1	M	Mandatory if claiming conformance to INSPIRE and ISO 19115. Free Text *Example*: JohnDoe@gmx.de
7.2.11	phone	Telephone numbers at which the organisation or an individual person may be contacted.	0-1	O	Free Text *Example*: +49-(0)228-4705555
8	Contributor	The person and/or institution responsible for making contributions to the dataset content, responsible for collecting, managing or distributing the data (in addition to creator).	0-n	O	
8.1	contributorName	The name of the contributor.	0-1	O	Free Text *Example*: Person: John Doe Institution: University of Cologne, Institute of Geography
8.2	contributorType	The type of contributor.	0-1	O/ M	If Contributor is used, then contributorType is mandatory. *Controlled List:* **Table D 17: ContributorType** See appendix for values and definitions.
8.3	nameIdentifier	The unique identifier of an individual, according to various schemes. Any alphanumeric string which is unique within its domain of issue.	0-1	O	Free Text The format is dependent upon the identifier scheme.
8.3.1	nameIdentifierScheme	The name of the name identifier scheme.	0-1	O/ M	If nameIdentifer is used, then nameIdentifierScheme is mandatory. *Controlled List:* **Table D-18: CreatorScheme** See appendix for values and definitions.
9	Publisher	The organization responsible for making the dataset available. In case of publications this is the publishing house of the publication.	1	M	Default value: CRC/TR32 Database (TR32DB) Free Text *Example*: Elsevier Science, Amsterdam, The Netherlands
Topic					
10	Subject	The topic of the dataset. Typically, the subject will be represented using keywords, key phrases, or classification codes.			
10.1	TR32topic	A subject describing the CRC/TR32 specific topic.	1	M	*Controlled List:* **Table D-57: TR32Topic** See appendix for values and definitions.
10.2	TR32keyword	Keywords describing the subject of the dataset with focus on the CRC/TR32 research topic.	1-n	M	*Controlled List:* Available at **http://www.tr32db.de/keywordList***
10.3	GEMETThesaurus	Keywords describing the dataset taken from the General Multi-Lingual Environmental Thesaurus (GEMET).	0-n	O	*Controlled List:* Available at **http://www.eionet.europa.eu/gemet**
10.3.1	subjectScheme	The name and/or information about the used subject.	0-1	O/ M	If GemetThesaurus is used, then subjectScheme is mandatory. Mandatory value: GEMET - Concepts *Controlled List:* **Table D-52: SubjectScheme** See appendix for values and

					definitions.
10.4	topicCategory	Topic category describing the dataset taken from the INSPIRE directive (Infrastructure for Spatial Information in Europe) and ISO 19115.	0-1	O/ M	Mandatory if claiming conformance to INSPIRE and ISO 19115 *Controlled List:* **Table D 30: IsoTopicCategory** Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ. do?uri=OJ:L:2008:326:0012:01:EN:HT ML
10.5	INSPIRETheme	Subject describing the dataset according to the GEMET INSPIRE Spatial Data Themes.	0-n	O/ M	Conditional: Mandatory if claiming conformance to INSPIRE *Controlled List:* **Table D-26: InspireTheme** Available at http://www.eionet.europa.eu/gemet/i nspire_themes
10.5.1	subjectScheme	The name and/or information about the used subject.	0-1	O/ M	If INSPIRETheme is used, then subjectScheme is mandatory. Mandatory value: GEMET – INSPIRE themes *Controlled List:* **Table D-52: SubjectScheme** See appendix for values and definitions.
10.6	DDC	Category describing the dataset taken from the Dewey Decimal Classification System (DDC).	0-n	O	*Controlled List:* Available at http://dewey.info/
10.6.1	subjectScheme	The name and/or information about the used subject.	0-1	O/ M	If DDC is used, then subjectScheme is mandatory. Mandatory value: DDC *Controlled List:* **Table D-52: SubjectScheme** See appendix for values and definitions.
File Details					
11	DataType	The nature or genre of the dataset.	1	M	Recommended and default setting for datasets is [dataset]. [Dataset collection] is recommended for dataset series. *Controlled List:* **Table D-21: DataType** See appendix for values and definitions.
12	DataFormat	The format of the dataset. The format consists of the format name and a corresponding category.	0-1	O/ M	If dataFormatVersion is used, then DataFormat is mandatory. *Controlled List:* Available at **http://www.tr32db.de/dataFormat List**
12.1	dataFormatVersi on	The version of the format.	0-1	O	Free Text *Example*: 2007
13	DataSize	The size of the dataset, represented by a number.	0-n	O	Free Text Example: 255
13.1	dataSizeType	The corresponding data size type.	0-1	O/ M	*Controlled List:* **Table D-50: SizeType** See appendix for values and definitions.
14	Date	A date that refers or is relevant to the dataset.	1-n	M	YYYY-MM-TT or any other format described in W3DTF[5] *Example*: 2012-10-17
14.1	dateType	The type of Date.	1	M	*Controlled List:*

---

[5] http://www.w3.org/TR/NOTE-datetime

					Table D-22: DateType See appendix for values and definitions.
15	Language	The language of the dataset.	1	M	Language code according to ISO 639-2[6]. *Controlled List:* Available at **http://www.tr32db.de/LanguageList** Example: eng
16	DataStatus	The status of the dataset.	1	M	*Controlled List:* **Table D-20: DataStatus** See appendix for values and definitions.
Constraints					
17	Download	Information about the file download.			
17.1	downloadPermission	The download permission of the dataset.	1	M	*Controlled List:* **Table D 24: DownloadPermissio** Table D-24: DownloadPermissionSee appendix for values and definitions.
17.2	downloadInformation	Additional information about the download of the dataset.	0-1	O	Free Text
18	Constraints	Information about the constraints for using the dataset.			
18.1	accessAndUseConstraints	The conditions applying to access and use of the dataset.	0-1	O	The value [conditions unknown] is recommended, if the conditions are unknown. If no conditions apply, the value [no conditions apply] is recommended. Free Text
18.2	accessLimitations	The limitations and reasons (on public access) of the dataset.	0-1	O	Free Text The value [no limitations] is recommended, if there are no limitations. *Example*: no limitations
18.3	licence	The licence applying to the dataset.	0-1	O	*Controlled List:* **Table D-31: Licence** See appendix for values and definitions.
Geographic Information					
19	MeasuringSite	The place or spatial region, where the dataset was measured, modelled, collected or about which the dataset is focused on.			
19.1	measuringRegion	The region of the measurement or model domain of the dataset.	1	M	*Controlled List:* **Table D-34: MeasureRegionTR32** See appendix for values and definitions.
19.2	measuringLocation	The location of the measurement or model domain of the dataset.	1	M	*Controlled List:* **Table D-33: MeasureLocationTR32** See appendix for values and definitions.
20	GeographicBoundingBox	The geographic coverage of the dataset expressed in terms of geographic coordinates given as a bounding box.	0-1		Free Text
20.1	westBoundLongitude	Western-most coordinate of the limit of the dataset extent,	0-1	O	Free Text *Example*:

---

[6] http://www.loc.gov/standards/iso639-2/php/code_list.php

		expressed in longitude in decimal degree.			5.2263
20.2	eastBoundLongit ude	Eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degree.	0-1	O	Free Text *Example*: 7.4236
20.3	northboundLatit ude	Northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degree.	0-1	O	Free Text *Example*: 54.5915
20.4	southBoundLatit ude	Southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degree.	0-1	O	Free Text *Example*: 50.2058
Metadata Details (automatic generation)					
21	TR32DBIdentifie r	Unique identifier of the dataset within the TR32DB.	1	A	Free Text *Example*: 50
21.1	TR32DBUrl	Unique identifier of the dataset using a Uniform Resource Locator (URL) address.	1	A	Free Text *Example*: http://tr32db.uni-koeln.de/view?dataID=2
22	TR32MetaDataT ype	The data type of the dataset that defines the corresponding metadata extension.	1	A	*Controlled List*: **Table D-55: TR32MetaDataType** See appendix for values and definitions.
23	MetadataCreator	Party responsible for creation and updating the metadata information.	1	A	
23.1	creatorFirstNam e	The first name of the Metadata Creator.	1	A	Free Text *Example*: John
23.2	creatorFamiliyNa me	The family name of the Metadata Creator.	1	A	Free Text *Example*: Doe
23.3	academicTitle	The academic title of the Metadata Creator.	0-1	A	Free Text *Example*: Prof. Dr.
23.4	eMailAddress	The e-mail address of the Metadata Creator.	1	A	Free Text *Example*: JohnDoe@gmx.de
23.5	phone	The telephone number of the Metadata Creator.	1	A	Free Text *Example*: +49-(0)221-4708839
23.6	fax	The fax number of the Metadata Creator.	1	A	Free Text *Example*: +49-(0)221-4708838
23.7	TR32subproject	The corresponding TR32 project section of the Metadata Creator.	1-n	A	*Controlled List*: **Table D-56: TR32Subproject*** See appendix for values and definitions.
24	MetadataCreatorI nstitution	The corresponding institution of the Metadata Creator.	1	A	
24.1	universityName	The name of the university or organisation.	1	A	Free Text *Example*: University of Cologne
24.2	instituteName	The name of the institute.	1	A	Free Text *Example*: Institute of Geography
24.3	streetName	The postcode as part of the address of the institution.	1	A	Free Text *Example*: Albertus-Magnus-Platz
24.4	postCode	The postcode as part of the address of the institution.	1	A	Free Text *Example*: 50923

24.5	city	The name of the city as part of the address of the institution.	1	A	Free Text *Example*: Cologne
24.6	country	The name of the country as part of the address of the institution.	1	A	Free Text *Example*: Germany
24.7	website	The website of the institution.	1	A	Free Text Exa*m*ple: www.geographie.uni-koeln.de
25	MdCreationDate	The date that the metadata was created.	1	A	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2012-10-17
26	MdLastUpdateDate	The date that the metadata was last updated.	1	A	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2012-10-17
27	MdWebVersion	The version of the TR32DB web site and metadata (profile).	1	A	Free Text *Example*: V41
28	MdLanguage	The language used for documenting metadata.	1	A	Language code according to ISO 639-26. *Controlled List*: Available at **http://www.tr32db.de/LanguageList** *Example*: eng
29	FileInformation	Automatic generated information about the uploaded file.	1	A	
29.1	fileName	The title of the uploaded file.	1	A	Free Text *Example*: LandUse_Classification_2013
29.2	fileExtention	The file extension of the uploaded file.	1	A	Free Text Example: .zip
29.3	fileFormat	The technical format of the uploaded file using MIME type.	1	A	*Controlled List*: Available at **http://www.tr32db.de/mimeTypeList**
29.4	fileSize	The file size of the uploaded file in kb.	1	A	Free Text *Example*: 555
29.5	fileUploadTime	The date that the file was updated to the TR32DB data storage.	1	A	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2012-10-17
29.6	fileSubproject	The TR32 project section, where the file was created.	1	A	*Controlled List*: **Table D-56: TR32Subproject*** See appendix for values and definitions.
29.7	fileFundingPhase	The TR32 project funding phase, when the file was created.	1	A	*Controlled List*: **Table D-37: Phase** See appendix for values and definitions.
30	Conformity	Information about the outcome of evaluating the obtained value (or set of values) against a specified acceptable conformance quality level.	0-n	A	
30.1	specificationTitle	The specification describes the citation of product specification or user requirement against which data is being evaluated.	1	A	Free Text *Controlled List*: **Table D-16: ConformitySpecification** See appendix for values and definitions.
30.2	date	The reference date of the	1	A	YYYY-MM-TT or any other format

ID		Definition			described in W3DTF5 *Example*: 2010-04-26
		specification.			
30.3	dateType	The reference date type.	1	A	*Controlled List:* **Table D-22: DateType** See appendix for values and definitions.
30.3	degree	The degree describes the indication of the conformance result.	1	A	*Controlled List:* **Table D-15: ConformityDegree** See appendix for values and definitions.
30.3	explanation	The explanation describes information of the meaning of conformance for this result.	1	A	Free Text *Example*: Only mandatory items are included.

D.3.2 TR32DB specific `Data` Metadata

Table D-5: TR32DB `Data` Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Data					
31	TemporalExtent	The time period covered by the content of the dataset including start and end date. In case of one day period startDate and endDate are the same date.	1	M	
31.1	startDate	The start date of the time period.	1	M	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
31.2	endDate	The end date of the time period.	1	M	For one day period, startDate is equivalent to endDate YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-02
32	Lineage	A statement on process history and/or overall quality of the dataset. Where appropriate, it may include a statement about the validity and quality assurance of the dataset, information about the version (if multiple versions exist), or the legal validity.	0-1	O	Free Text
33	MeasuringInstrumentModelMethod	The measuring instrument or modeling method used to create or included in the described dataset and corresponding measured or modelled output parameter.	0-n	O	
33.1	Instrument	Details of the measuring instrument or modeling method used to create or included in the described dataset.	0-n	O	*Controlled List:* Available at **http://www.tr32db.de/ instrumentList*** *Example*: Gas Analyser COSMO-model
33.1.2	instrumentName	The name of the measuring instrument or modeling method.	1	M	Free Text *Example*: Gas Analyzer Atmosphere-Model
33.1.2	instrumentModel	The model name of the measuring instrument or modeling method.	1	M	Free Text *Example*: Li-7500 COSMO

33.1.3	instrumentManuf acturer	The manufacturer of the measuring instrument or modeling method.	0-1	O	Free Text *Example*: Jenoptik
33.1.4	manufacturerReg isteredOffice	The location of the registered office of the measuring instrument or modeling method.	0-1	O	Free Text *Example*: Jena, Germany
33.1.5	manufacturerWe bsite	The website of the registered office.	0-1	O	Free Text *Example*: www.jenoptik.com http://www.cosmo-model.org/
33.2	Parameter	Details of the measured or modelled output parameter included in the described dataset.	0-n	O	
33.2.1	ParameterName	The name of the measured or modelled output parameter.	0-1	O	*Controlled List:* Available at **http://www.tr32db.de/ parameterList***
33.2.2	parameterUnit	The corresponding unit of the measured or modelled output parameter.	0-1	O/ M	If ParameterName is used, then parameterUnit is mandatory. *Controlled List:* Available at **http://www.tr32db.de/ parameterUnitList**
33.2.3	ResolutionDistan ce	The resolution distance of the parameter.	0-1	O	Free Text *Example*: 5
33.2.4	resolutionDistanc eUnit	The corresponding unit of the resolution distance.	0-1	O/ M	If ResolutionDistance is used, then resolutionDistanceUnit is mandatory. *Controlled List:* **Table D-47: ResolutionDistanceUnit** See appendix for values and definitions.
33.2.5	TemporalFreque ncy	The frequency of measurement or modelling of the parameter.	0-1	O	Free Text *Example*: 30
33.2.6	temporalFrequen cyUnit	The corresponding unit of the temporal frequency.	0-1	O/ M	If TemporalFrequency is used, then temporalFrequency-Unit is mandatory. *Controlled List:* **Table D-53: TemporalFrequencyUnit** See appendix for values and definitions.

D.3.3 TR32DB specific `Geodata` Metadata

Table D-6: TR32DB `Geodata` Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Geodata					
34	TemporalExtent	The time period covered by the content of the dataset including start and end date. In case of one day period startDate and endDate are the same date.	0-1	O	
34.1	startDate	The start date of the time period.	0-1	O	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
34.2	endDate	The end date of the time period.	0-1	O	For one day period, startDate is equivalent to endDate YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-02

35	VerticalExtent	The vertical domain of the dataset.	0-n	O	
35.1	minimumValue	The lowest vertical extent contained in the dataset.	0-1	O	Free Text *Example*: 5
35.2	maximumValue	The highest vertical extent contained in the dataset.	0-1	O	Free Text *Example*: 10
35.3	referenceSystem	Information about the vertical coordinate reference system (CRS) to which the maximum and minimum elevation values are measured. The CRS identification includes unit of measure.	0-1	O/M	If minimumValue or maximumValue is used, then referenceSystem is mandatory.
35.3.1	referenceSystem Code	The identifier of the reference system represented as EPSG code. For further information about EPSG see http://www.epsg-registry.org or http://www.epsg.org/.	0-1	O/M	If referenceSystemCode is used, then ReferenceSystem-Name is mandatory. *Controlled List:* **Table D-43: ReferenceSystem** See <u>appendix</u> for values and definitions.
35.3.2	referenceSystem Name	The name of the reference system.	0-1	O/M	If referenceSystemName is used, then referenceSystem-Code is mandatory. *Controlled List:* **Table D-43: ReferenceSystem** See <u>appendix</u> for values and definitions.
36	Lineage	A statement on process history and/or overall quality of the dataset. Where appropriate it may include a statement whether the dataset has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity.	0-1	O	Free Text
37	Spatial Resolution	The factor which provides a general understanding of the density of spatial data in the dataset. The level of detail expressed as a scale factor or a ground distance (consisting of the number and a unit of measure)	0-n	O	
37.1	scale	The level of detail expressed as the scale of a comparable hardcopy map or chart typically expressing the scale denominator 1:x.	0-1	O	Free Text *Example*: 5000
37.2	distanceNumber	The number expressing the distance value of the ground distance.	0-1	O	Free Text *Example*: 5
37.3	distanceUnit	The corresponding unit of measure of the distance value of the ground distance.	0-1	O/M	If distanceNumber is used, then distanceUnit is mandatory. *Controlled List:* **Table D-35: MeasureUnit** See <u>appendix</u> for values and definitions.
38	SpatialRepresenta tionType	The spatial representation type describes the method used to spatially represent geographic information in the dataset.	0-n	O	*Controlled List:* See <u>appendix</u> for values and definitions.
39	PresentationForm	The presentation form describes the mode in which the resource is represented.	0-n	O	*Controlled List:* **Table D-38: PresentationForm** See <u>appendix</u> for values and definitions.
40	ReferenceSystem	Description of the spatial reference systems used in the dataset.	0-n	O	
40.1	referenceSystem Code	The identifier of the reference system represented as EPSG code.	0-1	O/M	If referenceSystemCode is used, then referenceSystem-Name is mandatory.

		For further information about the EPSG see http://www.epsg-registry.org or http://www.epsg.org/.			*Controlled List:* Table D-43: ReferenceSystem See <u>appendix</u> for values and definitions.
40.2	referenceSystem Name	The name of the reference system.	0-1	O/ M	If referenceSystemName is used, then referenceSystem-Code is mandatory. *Controlled List:* **Table D-43: ReferenceSystem** See <u>appendix</u> for values and definitions.
41	MaintenanceInfo rmation	Information about the scope and frequency of updating of the dataset.	0-n	O	
41.1	maintenanceFreq uency	The maintenance and update frequency represents the frequency with which changes and additions are made to the resource after the initial resource is completed.	0-1	O	*Controlled List:* **Table D-32: MaintenanceFrequencyUnit** See <u>appendix</u> for values and definitions.
41.2	maintenanceNot e	The maintenance note describes information regarding specific requirements for maintaining the resource.	0-1	O	Free Text
42	Scope	The scope describes the class of information covered by the resource.	0-1	O	*Controlled List:* **Table D-49: ScopeCode** See <u>appendix</u> for values and definitions.
43	InitiativeType	The initiative type describes the type of initiative under which the aggregate dataset was produced.	0-1	O	*Controlled List:* **Table D-29: InitiativeType** See <u>appendix</u> for values and definitions.

D.3.4 TR32DB specific `Report` Metadata

Table D-7: TR32DB `Report` Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Report					
44	ReportDate	The date of the report.	1	M	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
45	ReportType	The type of report.	1	M	*Controlled List:* **Table D-46: ReportType** See <u>appendix</u> for values and definitions.
46	ReportCity	The corresponding city and country, where the report was created.	1	M	Free Text *Example*: Cologne, Germany
47	ReportInstitution	The corresponding institution, where the report was created.	1	M	Free Text *Example*: Insitute of Geography, University of Cologne, Germany
48	ReportVolume	The volume of the report.	0-1	O	Free Text *Example*: 5
49	NumberOfPages	The number of pages of the report.	0-1	O	Free Text. Only numerical values are allowed. *Example*: 10
50	PageRange	The page range of the report.	0-1	O	
50.1	startPage	The start page of the report.	0-1	O/ M	If startPage is used, then endPage is mandatory. Only numerical values are allowed.

					Free Text *Example*: 1
50.2	endPage	The end page of the report.	0-1	O/ M	If endPage is used, then startPage is mandatory. Only numerical values are allowed. Free Text *Example*: 10
51	FurtherInformati on	Additional information about the report.	0-1	O	Free Text

D.3.5 TR32DB specific `Picture` Metadata

Table D-8: TR32DB `Picture` Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Picture					
52	RecordDate	The recording date of the picture including start and end date. In case of one day recording startDate and endDate are the same date.	1	M	
52.1	startDate	The start date of the time period.	1	M	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
52.2	endDate	The end date of the time period.	1	M	For one day events, startDate is equivalent to endDate YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-02
53	RecordPlace	The place, where the picture or image was taken or that is displayed.	1	M	Free Text *Example*: Selhausen CRC/TR32 test field
54	RecordMethod	Details about the recording method that created the picture. For instance, the make and model name of a camera or the used software.	0-1	O	Free Text *Example*: Canon EOS 5D Adobe Photoshop CS4
55	Orientation	The orientation of the picture.	0-1	O	*Controlled List:* **Table D-36: Orientation** See <u>appendix</u> for values and definitions.
56	Resolution	The resolution of the picture in dpi (dots per inch).	0-1	O	Free Text *Example*: 200
57	Size	The size of the picture or image (horizontal and vertical) in pixels.	0-1	O	
57.1	width	The width of the picture in pixels (horizontal or x dimension).	0-1	O/ M	If height is used, then width is mandatory. Free Text *Example*: 10
57.2	height	The height of the picture in pixels (vertical or y dimension).	0-1	O/ M	If width is used, then heigth is mandatory. Free Text *Example*: 1
58	CopyrightInform ation	Information about the copyright of the picture.	0-1	O	Free Text *Example*: Copyright 2013 TR32. All rights reserved

59	Event	Details about the event, where the picture or image was taken.	0-1	O	
59.1	eventType	The type of event.	0-1	O	*Controlled List:* **Table D-25: EventType** See <u>appendix</u> for values and definitions.
59.2	eventName	The name of the event.	0-1	O	Free Text *Example*: CRC/TR32 Field Measurement Campaign
59.3	eventLocation	The location of the event including the city and the country separated by a comma, where the picture or image was taken.	0-1	O	Free Text *Example*: Selhausen, Gemany
59.4	eventWebsite	The website of the event.	0-1	O	Free Text *Example*: http://www.tr32.de/fieldmeasurements
59.5	eventPeriod	The duration of the event. In case of one day events startDate and endDate are the same date.	0-1	O	
59.5.1	startDate	The start date of the event.	0-1	O	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
59.5.2	endDate	The end date of the event.	0-1	M/O	If startDate is used, then endDate is mandatory. For one day events, startDate is equivalent to endDate. YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-10

D.3.6 TR32DB specific `Presentation` Metadata

Table D-9: TR32DB `Presentation` Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Presentation					
60	Presenter	The presenter of the presentation.	0-1	O	Free Text *Example*: John Doe
61	PresentationDate	The date, when the presentation was given.	1	M	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
62	PresentationType	The type of presentation.	1	M	*Controlled List:* **Table D-39: PresentationType** See <u>appendix</u> for values and definitions.
63	Event	Details about the event, where the dataset was presented.	1	M	
63.1	eventType	The type of event.	0-1	O	*Controlled List:* **Table D-25: EventType** See <u>appendix</u> for values and definitions.
63.2	eventName	The name of the event.	1	M	Free Text *Example*: European Geosciences Union General Assembly 2013
63.3	eventLocation	The location of the event including the city and the country separated	1	M	Free Text *Example*:

		by a comma.			Vienna, Austria
63.4	eventWebsite	The website of the event.	0-1	O	Free Text *Example*: http://www.egu.eu
63.5	eventPeriod	The duration of the event. In case of one day events startDate and endDate are the same date.	1	M	
63.5.1	startDate	The start date of the event.	1	M	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-01
63.5.2	endDate	The end date of the event.	1	M	For one day events, startDate is equivalent to endDate. YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2013-07-10

D.3.7   TR32DB specific `Publication` Metadata

Table D-10: TR32DB `Publication` Metadata Properties

ID	TR32DB property name	Definition	Occ	OB	Notes (allowed values, examples, other constraints)
Article					
64	Status	The status of the article.	1	M	*Controlled List:* **Table D 41: PublicationStatus** See <u>appendix</u> for values and definitions.
65	Review	The review status of the article.	1	M	*Controlled List:* **Table D-40: PublicationReview** See <u>appendix</u> for values and definitions.
66	Year	The year of the article.	1	M	Free Text *Example*: 2013
67	PublicationType	The type of publication.	1	M	Mandatory value: article *Controlled List:* **Table D-42: PublicationType** See <u>appendix</u> for values and definitions.
68	ArticleType	The type of article.	0-1	O	*Controlled List:* **Table D-14: ArticleType** See <u>appendix</u> for values and definitions.
69	PublicationSource e	The name of the publication source, e.g. name of the journal or magazine.	1	M	Free Text *Example*: Nature
69.1	PublicationSourc eWebsite	The website of the publication source.	0-1	O	Free Text *Example*: www.nature.com
70	Volume	The volume of the publication source.	0-1	O/ M	Mandatory, if Volume is available. Free Text *Example*: 5
71	Issue	The issue of the publication source.	0-1	O/ M	Mandatory, if Issue is available. Free Text *Example*: 1
72	NumberOfPages	The number of pages of the article.	0-1	O	Free Text. Only numerical values are allowed. *Example*: 10
73	PageRange	The page range of the article.	1	M	

73.1	startPage	The start page of the article.	1	M	Free Text. Only numerical values are allowed. Example: 1
73.2	endPage	The end page of the article.	1	M	Free Text. Only numerical values are allowed. *Example*: 10
Book					
74	Status	The status of the book.	1	M	*Controlled List:* **Table D 41: PublicationStatus** See <u>appendix</u> for values and definitions.
75	Review	The review status of the book.	1	M	*Controlled List:* **Table D-40: PublicationReview** See <u>appendix</u> for values and definitions.
76	Year	The year of the book.	1	M	Free Text *Example*: 2013
77	PublicationType	The type of publication.	1	M	Mandatory value: book *Controlled List:* **Table D-42: PublicationType** See <u>appendix</u> for values and definitions.
78	SeriesTitle	The series title of the book.	0-1	O/M	Mandatory, if SeriesTitle is available. Free Text *Example*: Kölner Geographisches Arbeiten
79	SeriesEditor	The series editor of book.	0-1	O/M	Mandatory, if SeriesEditor is available. Free Text *Example*: Bareth, G., Braun, B., .Schneider, K.
80	City	The city and county, where the publisher is located.	1	M	Free Text *Example*: Cologne, Germany
81	Volume	The volume of the book.	0-1	O/M	Mandatory, if Volume is available. Free Text *Example*: 90
82	Edition	The edition of the book.	0-1	O	Free Text *Example*: 1st ed.
83	NumberOfPages	The number of pages of the book.	0-1	O	Free Text. Only numerical values are allowed. *Example*: 145
84	PageRange	The page range of the book.	1	O	
84.1	startPage	The start page of the book.	0-1	O/M	If startPage is used, then endPage is mandatory. Only numerical values are allowed. Free Text *Example*: 1
84.2	endPage	The end page of the book.	0-1	O/M	If endPage is used, then startPage is mandatory. Only numerical values are allowed. Free Text *Example*: 144
BookSection					
85	Status	The status of the publication.	1	M	*Controlled List:* **Table D 41: PublicationStatus** See <u>appendix</u> for values and definitions.
86	Review	The review status of the publication.	1	M	*Controlled List:* **Table D-40: PublicationReview** See <u>appendix</u> for values and definitions.
87	Year	The year of the publication.	1	M	Free Text *Example*: 2013

88	PublicationType	The type of publication.	1	M	Mandatory value: book *Controlled List:* **Table D-42: PublicationType** See <u>appendix</u> for values and definitions.
89	BookTitle	The title of the book	1	M	Free Text *Example*: Proceedings of the Data Management Workshop
90	BookEditor	The editor of the book.	1	M	Free Text *Example*: Curdt, C. and Bareth, G.
91	City	The city and county separated by a comma, where the publisher is located.	1	M	Free Text *Example*: Cologne, Germany
92	SeriesTitle	The title of the book series.	0-1	O	Free Text *Example*: Kölner Geographisches Arbeiten
93	SeriesEditor	The editor of book series.	0-1	O	Free Text *Example*: G. Bareth, H. Besler, B. Braun, E. Brunotte
94	Volume	The volume of the book.	0-1	O	Free Text *Example*: 90
95	Chapter	The chapter of the book.	0-1	O	Free Text *Example*: 5
96	NumberOfPages	The number of pages of the book section.	0-1	O	Free Text. Only numerical values are allowed. *Example*: 8
97	PageRange	The page range of the book section.	1	M	
97.1	startPage	The start page of the book section.	1	M	Free Text. Only numerical values are allowed. *Example*: 82
97.2	endPage	The end page of the book section.	1	M	Free Text. Only numerical values are allowed. *Example*: 90
EventPaper					
98	Status	The status of the publication.	1	M	*Controlled List:* **Table D 41: PublicationStatus** See <u>appendix</u> for values and definitions.
99	Review	The review status of the publication.	1	M	*Controlled List:* **Table D-40: PublicationReview** See <u>appendix</u> for values and definitions.
100	Year	The year of the publication.	1	M	Free Text *Example*: 2013
101	PublicationType	The type of publication.	1	M	Mandatory value: eventPaper *Controlled List:* **Table D-42: PublicationType** See <u>appendix</u> for values and definitions.
102	Event	Details about the event, where the dataset was presented.	1	M	
102.1	eventType	The type of event.	0-1	O	*Controlled List:* **Table D-25: EventType** See <u>appendix</u> for values and definitions.
101.2	eventName	The name of the event.	1	M	Free Text *Example*: ISPRS Workshop Laserscanning 2009
102.3	eventLocation	The location of the event including the city and the country separated by a comma.	1	M	Free Text *Example*: Paris, France
102.4	eventWebsite	The website of the event.	0-1	O	Free Text *Example*:

					http://laserscanning2009.ign.fr/
102.5	eventPeriod	The duration of the event. In case of one day events startDate and endDate are the same date.	1	M	
102.5.1	startDate	The start date of the event.	1	M	YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2009-07-01
102.5.2	endDate	The end date of the event.	1	M	For one day events, startDate is equivalent to endDate. YYYY-MM-TT or any other format described in W3DTF5 *Example*: 2009-07-10
103	Proceedings	Information about the proceedings of the event.			
103.1	ProceedingsEditor	The editor of the proceedings.	0-1	O	Free Text *Example*: Bretar, F., Pierrot-Deseilligny, M., and Vosselman, G.
103.2	ProceedingsTitle	The title of the proceedings.	0-1	O	Free Text *Example*: ISPRS Workshop Laserscanning 2009
103.3	ProceedingsURL	The website of the proceedings.	0-1	O	Free Text
103.4	Volume	The volume of the publication source.	0-1	O/ M	Mandatory, if Volume is available. Free Text *Example*: 5
103.5	Issue	The issue of the publication source.	0-1	O/ M	Mandatory, if Issue is available. Free Text *Example*: 1
104	NumberOfPages	The number of pages of the event paper.	0-1	O	Free Text. Only numerical values are allowed. *Example*: 8
105	PageRange	The page range of the event paper.	1	M	
105.1	startPage	The start page of the event paper in the proceedings.	1	M	Free Text. Only numerical values are allowed. *Example*. 82
105.2	endPage	The end page of the event paper in the proceedings.	1	M	Free Text. Only numerical values are allowed. *Example*: 90

D.4   TR32DB Metadata Schema Mapping

D.4.1   Overview

This section gives an overview of mappings between the metadata properties of the TR32DB Metadata Schema and elements of four metadata schemas and standards. Mandatory elements or properties of the used metadata standard are highlighted (section D.4.2). The mapping was on one hand used to create this metadata schema and on the other hand represents the interoperability to the used schemas and standards. The following metadata schemas and standards are applied (Table D-11):

Table D-11: Metadata Schemes and Standards for Mapping

Schema	Full Title	Web Access Address
DataCite	DataCite Metadata Schema for the Publication and Citation of Research Data, Version 2.2	http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf
Dublin Core	Dublin Core Metadata Element Set, Version 1.1	http://dublincore.org/documents/dces/
	DCMI Metadata Terms	http://dublincore.org/documents/dcmi-terms/
INSPIRE	INSPIRE Implementing Directive;	http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:EN:PDF
	INSPIRE Implementing Rules for Metadata, Version 1.2	http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf
ISO	ISO 19115:2003 Geographic information – Metadata	http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

For each property of the metadata schema and standard, the identifier number (ID) and the particular obligation (O.) is specified. The identifier number of INSPIRE reverences to Part B of the INSPIRE Implementing Directive. In addition, further metadata schemes are considered with regard to further use of the TR32DB Metadata Schema and presented in `Other Elements`. The following table (Table D-12) provides an overview of the used abbreviations represented in the mapping.

Table D-12: Abbreviation applied for the Mapping

Abbreviation	Full Title	Web Access Address
bibo	The Bibliographic Ontology	http://bibliontology.com/
dc	Dublin Core Element Set, Version 1.1	http://dublincore.org/documents/dces/
dcterms	DCMI Metadata Terms	http://dublincore.org/documents/dcmi-terms/
event	The Event Ontology	http://purl.org/NET/c4dm/event.owl#
tl	The Timeline Ontology	http://motools.sourceforge.net/timeline/timeline.html
time	Time Ontology in OWL	http://www.w3.org/TR/owl-time/; http://www.w3.org/2006/time#

D.4.2 TR32DB Metadata Schema Mapping

Table D-13: TR32DB Metadata Schema Mapping

ID	TR32DB property	Dublin Core	ID	ISO	O.	ID	INSPIRE	O.	ID	DataCite	O.	Other Elements
Identification												
1	Title	dc.title dcterms:title	360	CI_Citation.title	M	1.1	Resource title	M	3	Title	M	
1.1	TitleType	dcterms.alternative	361	CI_Citation.alternateTitle	O					3.1	TitleType	C	
2	Description	dc.description dcterms.description	25	MD_Identification.abstract	M	1.2	Abstract	M	17	Description	O	
2.1	descriptionType	dcterms.abstract dcterms.tableOfContents								17.1	descriptionType	O / M	
3	Date	dc.date dcterms.date	362/ 394	CI_Date.date	M	5.2, 5.3, 5.4	Date	C	8	Date	O	
3.1	dateType	dc.date dcterms.available dcterms.created dcterms.dateAccepted dcterms.dateCopyrighted dcterms.dateSubmitted dcterms.issued dcterms.modified	362/ 395	CI_Date.dateType	M	5.2, 5.3, 5.4	Date of publication, revision and creation	C	8.1	dateType	O / M	
4	Identifier	dc.identifier dcterms.identifier	365	CI_Citation.Identifier	O	1.5	Unique resource identifier Code	M	11	AlternateIdentifier	O	
4.1	identiferType	dc.identifier dcterms.identifier	207	MD_Identifier.Code	M	1.5	Identifier Namespace	O	11.1	alternateIdentifierType	O / M	bibo:issn bibo:isbn bibo:uri bibo:doi
5	Relation											
5.1	identifier	dc.identifier dcterms.identifier								12	RelatedIdentifier	O	
5.2	identifierType	dc.identifier dcterms.identifier								12.1	relatedIdentifierType	O / M	
5.3	relationType	dc.relation dcterms.relation								12.2	relationType	O /	

		dcterms.conformsTo dcterms.isReferencedBy dcterms.references dcterms.isVersionOf dcterms.hasVersion dcterms.isFormatOf dcterms.hasFormat dcterms.isPartOf dcterms.hasPart dcterms.isReplacedBy dcterms.replaces dcterms.source									M	
6	CitationAdvice	dcterms.bibliographic Citation										
7	AdditionalDescription											
7.1	descriptionFileName											
7.2	graphicFileName		49	MD_BrowseGraphic.fileName	M							
Responsible Party												
8	Creator											
8.1	Creator-Person	dc.creator dcterms.creator							2	Creator	M	bibo:authorList
8.1.1	creatorRole		379	CL_ResponsibleParty.role	O	9.2	Responsible party role	M				
8.1.2	creatorFirstName	dc.creator dcterms.creator	375	CI_ResponsibleParty.individualName	M				2.1	CreatorName	M	
8.1.3	creatorFamilyName	dc.creator dcterms.creator	375	CI_ResponsibleParty.individualName	M				2.1	CreatorName	M	
8.1.4	nameIdentifier								2.2	nameIdentifier	C	
8.1.4.1	nameIdentifierScheme								2.2.1	nameIdentifierScheme	C/M	
8.1.5	emailAddress		386	CI_Address.electronicMailAddress	O	9.1	Responsible party	M				
8.1.6	TR32Membership											
8.1.7	CreatorInstitution								2.1	CreatorName	M	
8.1.7.1	universityName		376	CI_ResponsibleParty.organisationName	M				2.1	CreatorName	M	
8.1.7.2	instituteName		376	CI_ResponsibleParty.or	M							

				ganisationName									
8.1.7.3	streetName		381	CI_Address.deliveryPoint	O								
8.1.7.4	postCode		384	CI_Address.postalCode	O								
8.1.7.5	city		382	CI_Address.city	O								
8.1.7.6	country		385	CI_Address.country	O								
8.1.7.7	website		390	CI_Contact.onlineResource	O								
8.1.7.8	TR32Membership												
8.2	Creator-Organisation	dc.creator dcterms.creator											
8.2.1	OrganisationRole		379	CL_ResponsibleParty.role	O	9.2	Responsible party role	M					
8.2.2	OrganisationName		376	CI_ResponsibleParty.organisationName	M					2.1	CreatorName	M	
8.2.3	OrganisationDepartment		376	CI_ResponsibleParty.organisationName	M					2.1	CreatorName	M	
8.2.4	streetName		381	CI_Address.deliveryPoint	O								
8.2.5	postCode		384	CI_Address.postalCode	O								
8.2.6	city		382	CI_Address.city	O								
8.2.7	country		385	CI_Address.country	O								
8.2.8	website		390	CI_Contact.onlineResource	O								
8.2.9	contactPerson		375	CI_ResponsibleParty.individualName	M								
8.2.10	eMailAddress		386	CI_Address.electronicMailAddress	O	9.1	Responsible party	M					
8.2.11	phone												
9	Contributor	dc.contributor dcterms.contributor								7	Contributor	O	bibo:contributor
9.1	contributorName	dc.contributor dcterms.contributor								7.2	contributorName	O / M	
9.2	contributorType									7.1	contributorType	O / M	
9.3	nameIdentifier									2.2	nameIdentifier	O	
9.3.1	nameIdentifierScheme									2.2.1	nameIdentifierScheme	O / M	
10	Publisher	dc.publisher	376/	CI_ResponsibleParty.or	O	9.1	Responsible party	M	4	Publisher	M		

		dcterms.publisher	379	ganisationName CI_ResponsibleParty.role (=publisher)		9.2	Responsible party role (=publisher)				
Topic											
11	Subject	dc.subject dcterms.subject							6	Subject	O
11.1	TR32topic	dc.subject dcterms.subject	53	MD_Keywords.keyword	M	3.1	Keyword value	M	6	Subject	O
11.2	TR32keyword	dc.subject dcterms.subject	53	MD_Keywords.keyword	M	3.1	Keyword value	M	6	Subject	O
11.3	GemetThesaurus	dc.subject dcterms.subject	53	MD_Keywords.keyword	M	3.1	Keyword value	M	6	Subject	O
11.3.1	subjectScheme		55	MD_Keywords.ThesaurusName	C	3.2	Originating controlled vocabulary	M	6.1	subjectScheme	O
11.4	topicCategory	dc.subject dcterms.subject	41	MD_DataIdentification.topicCategory	C	2.1	Topic category	M	6	Subject	O
11.5	INSPIRETheme	dc.subject dcterms.subject	53	MD_Keywords.keyword	M	3.1	Keyword value	M	6	Subject	O
11.5.1	subjectScheme		55	MD_Keywords.ThesaurusName	C	3.2	Originating controlled vocabulary	M	6.1	subjectScheme	O
11.6	DDC	dc.subject dcterms.subject	53	MD_Keywords.keyword	M	3.1	Keyword value	M	6	Subject	O
11.6.1	subjectScheme	dcterms.ddc	55	MD_Keywords.ThesaurusName	C	3.2	Originating controlled vocabulary	M	6.1	subjectScheme	O
File Details											
12	DataType	dc.type dcterms.type	6	MD_Metadata.hierarchyLevel	C	1.3	Resource Type	M	10.1	resourceTypeGeneral	O/M
13	DataSize	dc.format dcterms.extent	276	MD_DigitialTransferOptions.transferSize	O				13	Size	O
13.1	dataSizeType	dc.format dcterms.sizeOrDuration	275	MD_DigitialTransferOptions.unitsOfDistribution	O				13	Size	O
14	DataFormat	dc.format dcterms.mediaType	285	MD_Format.name	M				14	Format	O
14.1	dataFormatVersion	dc.format dcterms.mediaTypeOrExtent	286	MD_Format.version	M						
15	Language	dc.language dcterms.language	39	MD_DataIdentification.language	M	2.2.7	Resource language	C	9	Language	O
16	DataStatus										
Constraints											
17	Download										

17.1	DownloadPermission	dc.rights dcterms.rights										
17.2	DownloadInformation	dc.rights dcterms.rightsStatement							16	Rights	O	
18	Constraints											
18.1	AccessAndUse-Constraints	dc.rights dcterms.rightsStatement	68	MD_Constraints.useLimitation	O	8.1	Condition applying to access and use	M	16	Rights	O	
18.2	AccessLimitations	dc.rights dcterms.accessRights	72	MD_LegalConstraints.otherConstraints	C	8.2	Limitations on public access	M	16	Rights	O	
18.3	Licence	dc.rights dcterms.license							16	Rights	O	cc:license
Geographic Information												
19	MeasuringSite	dc.coverage dcterms.location										
19.1	measuringRegion	dc.coverage dcterms.location										
19.2	measuringLocation	dc.coverage dcterms.location										
20	GeographicBounding-Box	dc.coverage dcterms.spatial dcterms.box	343	EX_GeographicBoundingBox	M							
20.1	westBoundLongitude	dc.coverage dcterms.spatial dcterms.box	344	EX_GeographicBoundingBox.westBoundLongitude	M	4.1	Geographic bounding box	M/C				
20.2	eastBoundLongitude	dc.coverage dcterms.spatial dcterms.box	345	EX_GeographicBoundingBox.eastBoundLongitude	M	4.1	Geographic bounding box	M/C				
20.3	northBoundLatutude	dc.coverage dcterms.spatial dcterms.box	347	EX_GeographicBoundingBox.northBoundLatitude	M	4.1	Geographic bounding box	M/C				
20.4	southBoundLatitude	dc.coverage dcterms.spatial dcterms.box	346	EX_GeographicBoundingBox.southBoundLatitude	M	4.1	Geographic bounding box	M/C				
Metadata on Metadata (automatic generation)												
21	tr32dbIdentifier	dc.identifier dcterms.identifer	2	MD_Metadata.fileIdentifer	O							
21.1	tr32dbUrl		397	CI_OnlineResource.linkage	M	1.4	Resource locator	M				
22	TR32MetaDataType											
23	MetadataCreator											

23.1	creatorFirstName		8	MD_Metadata.contact	M	10.1	Metadata point of contact	M			
23.2	creatorFamiliyName		8	MD_Metadata.contact	M						
23.3	academicTitle										
23.4	eMailAddress		386	CI_Address.electronicMailAddress	O	10.1	Metadata point of contact	M			
23.5	phone		408	CI_Telephone.voice	O						
23.6	fax		409	CI_Telephone.facsimile	O						
23.7	TR32subproject										
24	MetadataCreatorInstitution										
24.1	universityName		376	CI_ResponsibleParty.organisationName	C						
24.2	instituteName		376	CI_ResponsibleParty.organisationName	C						
24.3	streetName		381	CI_Address.deliveryPoint	O						
24.4	postCode		384	CI_Address.postalCode	O						
24.5	city		382	CI_Address.city	O						
24.6	country		385	CI_Address.country	O						
24.7	website										
25	mdCreationDate	dc.date dcterms.date	9	MD_Metadata.dateStamp	M	10.2	Metadata date	M			
26	mdLastUpdateDate	dc.date dcterms.modified									
27	mdWebVersion										
28	mdLanguage	dc.Language dcterms.language	3	MD_Metadata.language	C	10.3	Metadata Language	M			
29	fileInformation										
29.1	fileName										
29.2	fileExtention	dcterms.mediaType									
29.3	fileFormat	dc.format dcterms.fileFormat							14	Format	O
29.4	fileSize	dc.format dcterms.extent	276	MD_DigitialTransferOptions.transferSize	O				13	Size	O
29.5	fileUploadTime	dc.date dcterms.modified									
29.6	fileSubproject										
29.7	fileFundingPhase										
30	Conformity	dcterms.conformsTo	128	DQ_Result							
30.1	specificationTitle		130	DQ_ConformanceResult.specification	M	7.1	Specification	M			
30.2	date	dc.date	362/	CI_Date.date	M	7.1	Specification Date	M			

		dcterms.date	394										
30.3	dateType	dc.date dcterms.created dcterms.issued dcterms.modified	362/ 394	CI_Date.dateType	M	7.1	Specification Code	M					
30.3	degree		132	DQ_ConformanceResult.pass	M	7.2	Degree	M					
30.3	Explanation		131	DQ_ConformanceResult.explanation	M	7.1	Explanation						
Data													
31	TemporalExtent	dc.coverage dcterms.termporal	350	EX_TemporalExtent									
31.1	startDate	dc.date dcterms.date	351	EX_TemporalExtent.extent	M	5.1	Temporal extent	C					
31.2	endDate	dc.date dcterms.date	351	EX_TemporalExtent.extent	M	5.1	Temporal extent	C					
32	Lineage	dc.source dcterms.provenence	83	LI_Lineage.statement	C	6.1	Lineage	M					
33	MeasuringModellingInstrument	dc.source											
33.1	instrumentName												
33.1.2	instrumentModel												
33.1.2	instrumentManufac-turer												
33.1.3	manufacturerRegisterdOffice												
33.1.4	manufacturerWebsite												
33.1.5	InstrumentParameter												
33.2	ParameterName												
33.2.1	parameterUnit												
33.2.2	ResolutionDistance												
33.2.3	resolutionDistanceUnit												
33.2.4	TemporalFrequency												
33.2.5	temporalFrequencyUnit												
Geodata													
34	TemporalExtent	dc.coverage dcterms.extent	350	EX_TemporalExtent									

34.1	startDate	dc.date dcterms.date	351	EX_TemporalExtent.extent	M	5.1	Temporal extent	C					
34.2	endDate	dc.date dcterms.date	351	EX_TemporalExtent.extent	M	5.1	Temporal extent	C					
35	VerticalExtent	dc.coverage dcterms.coverage	354	EX_VerticalExtent									
35.1	minimumValue	dc.coverage dcterms.coverage	355	EX_VerticalExtent.minimumValue	C/M								
35.2	maximumValue	dc.coverage dcterms.coverage	356	EX_VerticalExtent.maximumValue	C/M								
35.3	ReferenceSystem		358	EX_VerticalExtent.verticalCRS	C/M								
35.3.1	referenceSystem Code												
35.3.2	referenceSystem Name												
36	Lineage	dc.source dcterms.provenence	83	LI_Lineage.statement	C	6.1	Lineage	M					
37	Spatial Resolution		59	MD_Resolution	C	6.2	Spatial resolution	M					
37.1	scale		60	MD_Resolution.equivalentScale	C	6.2	Equivalent scale	M					
37.2	distanceNumber		61	MD_Resolution.distance	C	6.2	Ground sample distance	M					
37.3	distanceUnit		61	MD_Resolution.distance	C	6.2	Ground sample distance	M					
38	SpatialRepresentationType	dc.type dcterms.type	37	MD_DataIdentification.spatialRepresentationType	O								
39	PresentationForm		368	CI_Citation.presentationForm	O								
40	ReferenceSystem		186	MD_ReferenceSystem									
40.1	referenceSystem Code		187	MD_ReferenceSystem.referenceSystemIdentifier	C/M								
40.2	referenceSystem Name		196	RS_ReferenceSystem.name	M								
41	MaintenanceInforma-tion	dcterms.frequency	142	MD_MaintenanceInformation									
41.1	maintenanceFrequency	dcterms.frequency	143	MD_MaintenanceInformation. maintenanceAndUpdate Frequency	M								
41.2	maintenanceNote		148	MD_MaintenanceInfor	O								

				mation. maintenanceNote									
42	Scope	dc.type dcterms.type	6	MD_Metadata.hierarchy Level	C	1.3	ResourceType	M					
43	InitiativeType		66.5	MD_AggregateInformat ion.initiativeType	C								
Report													
44	ReportDate	dc.date dcterms.date											
45	ReportType												bibo:report
46	ReportCity	dcterms.location											event:place
47	ReportInstitution												
48	ReportVolume												bibo:volume
49	NumberOfPages												bibo:numPages
50	PageRange												bibo:pages
50.1	startPage												bibo:pageStart
50.2	endPage												bibo:pageEnd
51	FurtherInformati on												
Picture													
52	RecordDate	dc.date											
52.1	startDate	dc.date											
52.2	endDate	dc.date											
53	RecordPlace	dcterms.location											
54	RecordMethod												
55	Orientation												
56	Resolution												
57	Size	dcterms.sizeOrDuratio n											
57.1	width	dcterms.sizeOrDuratio n											
57.2	height	dcterms.sizeOrDuratio n											
58	CopyrightInform ation	dc.rights											
59	Event												bibo:event
59.1	EventType												bibo:conferenc e bibo:workshop
59.2	EventName	dc.title dcterms.title											bibo:event
59.3	EventLocation	dc.coverage dcterms.location											event:place

59.4	EventWebsite	dc.date dcterms.periodOfTime		bibo:website
59.5	EventPeriod	dc.date dcterms.periodOfTime		event:time tl:interval
59.5.1	startDate	dc.date dcterms.periodOfTime		tl:beginsAt
59.5.2	endDate	dc.date dcterms.periodOfTime		tl:endsAt
Presentation				
60	Presenter			bibo:presents
61	PresentationDate	dc.date dcterms.date		
62	PresentationType			bibo:slideshow
63	Event			bibo:event
63.1	EventType			bibo:conference bibo:workshop
63.2	EventName	dc.title dcterms.title		bibo:event bibo:presentedAt
63.3	EventLocation	dc.coverage dcterms.location		event.place
63.4	EventWebsite			bibo:presentedAt bibo:website
63.5	EventPeriod	dc.date dcterms.periodOfTime		event:time tlinterval
63.5.1	startDate	dc.date dcterms.periodOfTime		tl:beginsAt
63.5.2	endDate	dc.date dcterms.periodOfTime		tl:endsAt
Publication				
Article				
64	Status			bibo:status
65	Review			bibo:status
66	Year	dc.date dcterms.issued		
67	PublicationType	dcterms.bibliographicResource		bibo:Article
68	ArticleType			bibo:journal bibo:magazine
69	PublicationSource	dc.title dcterns.title		bibo:journal
69.1	PublicationSourc			bibo:website

	e-Website												
70	Volume												bibo:volume
71	Issue												bibo:issue
72	NumberOfPages												bibo:numPages
73	PageRange												bibo:pages
73.1	startPage												bibo:pageStart
73.2	endPage												bibo:pageEnd
Book													
74	Status												bibo:status
75	Review												bibo:status
76	Year	dc.date dcterms.issued											
77	PublicationType	dcterms.bibliographic Resource											bibo:book
78	SeriesTitle	dc.title dcterms.title											
79	SeriesEditor												bibo:editorList
80	City	dcterms.location											event:place
81	Volume												bibo:volume
82	Edition												bibo:edition
83	NumberOfPages												bibo:numPages
84	PageRange												bibo:pages
84.1	startPage												bibo:pageStart
84.2	endPage												bibo:pageEnd
BookSection													
85	Status												bibo:status
86	Review												bibo:status
87	Year	dc.date dcterms.issued											
88	PublicationType	dcterms.bibliographic Resource											bibo:BookSection
89	BookTitle	dc.title dcterms.title											
90	BookEditor												bibo:editorList
91	City												event:place
92	SeriesTitle	dc.title dcterms.title											
93	SeriesEditor												bibo:editorList
94	Volume												bibo:volume
95	Chapter												bibo:chapter
96	NumberOfPages												bibo:numPages
97	PageRange												bibo:pages

97.1	startPage												bibo:pageStart
97.2	endPage												bibo:pageEnd
EventPaper													
98	Status												bibo:status
99	Review												bibo:status
100	Year	dc.date dcterms.issued											
101	PublicationType	dcterms.bibliographic Resource											bibo:Proceedings
102	Event												
102.1	EventType												bibo:conference bibo:workshop
102.2	EventName	dc.title dcterms.title											bibo:conference
102.3	EventLocation	dc.coverage dcterms.location											event:place
102.4	EventWebsite												bibo:presentedAt bibo:website
102.5	EventPeriod	dc.date dcterms.periodOfTime											event:time tl:interval
102.5.1	startDate	dc.date dcterms.periodOfTime											tl:beginsAt
102.5.2	endDate	dc.date dcterms.periodOfTime											tl:endsAt
103	Proceedings												bibo:reproducedIn
103.1	ProceedingsEditor												bibo:editor
103.2	ProceedingsTitle												bibo:proceedings
103.3	ProceedingsURL												bibo:proceedings
103.4	Volume												bibo:volume
103.5	Issue												bibo:issue
104	NumberOfPages												bibo:numPages
105	PageRange												bibo:pages
105.1	startPage												bibo:pageStart
105.2	endPage												bibo:pageEnd

D.5 Controlled lists and attribute values

An overview of the controlled lists and attribute values of the TR32DB Metadata Schema are presented in this section. The content of the lists and values is composed of various standards and schemas (see section 6D.1.2) used to design this TR32DB Metadata Schema, as well as own supplements. The origin of the value names and definition is indicated in the header line of the particular tables. Controlled lists with an asterisk * will continuously updated according to the CRC/TR32 project participants requirements. In the following sections, the controlled lists and attribute values will be described in detail.

Table D-14: ArticleType

ID	Value name TR32DB	Definition
1.	Journal	Journal article
2.	Magazine	Magazine article
3.	Newspaper	Newspaper article
4.	Electronic	Electronic article

Table D-15: ConformityDegree

ID	Value name TR32DB (ISO, INSPIRE)	Definition
1.	conformant	The dataset is fully conformant with the cited specification. (true)
2.	notConformant	The dataset does not conform to the cited specification. (false)
3.	notEvaluated	Conformance has not been evaluated.

Table D-16: ConformitySpecification

ID	Value name TR32DB	Definition
1.	DCMES	Dublin Core Metadata Element Set, Version 1.1, http://dublincore.org/documents/dces/
2.	DCMI Metadata Terms	Dublin Core Metadata Initiative Metadata Terms http://dublincore.org/documents/dcmi-terms/
3.	ISO 19115	ISO 19115:2003 Geographic information – Metadata; Core Metadata Elements
4.	INSPIRE	EU INSPIRE Directive, Metadata Implementing Rules Technical Guidelines, http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf
5.	DataCite	DataCite Metadata Schema, Version 2.2, http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf

Table D-17: ContributorType

ID	Value name TR32DB (DataCite)	Definition
1.	ContactPerson	Person with knowledge of how to access, troubleshoot, or otherwise field issues related to the resource.
2.	DataCollector	Person/institution responsible for finding, gathering/collecting data under the guidelines of the author(s) or Principal Investigator (PI).
3.	DataManager	Person (or organization with a staff of data managers, such as a data centre) responsible for maintaining the finished resource.
4.	Distributor	Institution tasked with responsibility to generate/disseminate copies of the resource in either electronic or print form.
5.	Editor	A person who oversees the details related to the publication format of the resource.
6.	Funder	Institution that provided financial support for the development of the resource.
7.	HostingInstitution	Typically, the organization allowing the resource to be available on the Internet through the provision of its hardware/software/operating support.
8.	Producer	Typically a person or organization responsible for the artistry and form of a media product.
9.	ProjectLeader	Person officially designated as head of project team or sub-project team instrumental in the work necessary to development of the resource.
10.	ProjectMember	Person on the membership list of a designated project/project team.
11.	RelatedPerson	A person without a specifically defined role in the development of the resource, but who is someone the author wishes to recognize.
12.	Researcher	A person involved in analysing data or the results of an experiment or formal study. May indicate an intern or assistant to one of the authors who helped with research but who was not so "key" as to be listed as an author.
13.	RightsHolder	Person or institution owning or managing property rights, including intellectual property rights over the resource.
14.	Sponsor	Person or organization that issued a contract or under the auspices of which a work has been written, printed, published, developed, etc.
15.	Supervisor	Designated administrator over one or more groups/teams working to produce a resource or over one or more steps of a development process.
16.	WorkPackageLeader	A Work Package is a recognized data product, not all of which is included in publication. The package, instead, may include notes, discarded documents, etc. The Work Package Leader is responsible for ensuring the comprehensive contents, versioning, and availability of the Work Package during the development of the resource.

Table D-18: CreatorScheme

ID	Value name TR32DB	Definition
1.	ISNI	International Standard Name Identifier scheme (http://www.isni.org/).
2.	ORCID	Open Researcher and Contributor Identifier (http://about.orcid.org/).
3.	OpenID	OpenID is an open standard that describes how users can be authenticated in a decentralized manner, eliminating the need for centralized registration services (http://openid.net).
4.	ResearcherID	ResearcherID is an identifying system for scientific authors created and owned by Thomson Reuters (http://www.researcherid.com).
5.	VIAF	The Virtual International Authority File is an international authority file created by merging the national authority files of several national libraries and operated by the Online Computer Library Center (OCLC) (http://viaf.org/).

Table D-19: CreatorStatus

ID	Value name TR32DB	Definition
1.	CreatorIsNoMember	Creator is no member of CRC/TR32.
2.	CreatorIsFullMember	Creator is full member of CRC/TR32.
3.	CreatorWasMember	Creator is former member of CRC/TR32.
4.	CreatorIsCorrespondingMember	Creator is corresponding member of CRC/TR32.
5.	CreatorIsAssociateMember	Creator is associate member of CRC/TR32.

Table D-20: DataStatus

ID	Value name TR32DB	Definition
1.	Completed	Completed dataset.
2.	inProcess	Data in process.
3.	RawData	Raw data.

Table D-21: DataType

ID	Value name TR32DB (dcterms)	Value name (DataCite)	Value name (ISO, INSPIRE)	Definition
1.	Collection	Collection	series	An aggregation of resources. A collection is described as a group; its parts may also be separately described.
2.	Dataset	Dataset	dataset	Data encoded in a defined structure. Examples include lists, tables, and databases. A dataset may be useful for direct machine processing.
3.	Event	Event		A non-persistent, time-based occurrence. Metadata for an event provides descriptive information that is the basis for discovery of the purpose, location, duration, and responsible agents associated with an event. Examples include an exhibition, webcast, conference, workshop, open day, performance, battle, trial, wedding, tea party, and conflagration.
4.	MovingImage	Audiovisual		A series of visual representations imparting an impression of motion when shown in succession. Examples include animations, movies, television programs, videos, zoetrope's, or visual output from a simulation. Instances of the type Moving Image must also be describable as instances of the broader type Image.
5.	Image	Image	dataset	A visual representation other than text. Examples include images and photographs of physical objects, paintings, prints, drawings, other images and graphics, animations and moving pictures, film, diagrams, maps, musical notation. Note that Image may include both electronic and physical representations.
6.	InteractiveResource	InteractiveResource		A resource requiring interaction from the user to be understood, executed, or experienced. Examples include forms on Webpages, applets, multimedia learning objects, chat services, or virtual reality environments.
7.	PhysicalObject	PhysicalObject		An inanimate, three-dimensional object or substance. Note that digital representations of, or surrogates for, these objects should use Image, Text or one of the other types.
8.	Service	Service	service	A system that provides one or more functions. Examples include a photocopying service, a banking service, an authentication service, interlibrary loans, a Z39.50 or Web server.
9.	Software	Software		A computer program in source or compiled form. Examples include a C source file, MS-Windows .exe executable, or Perl script.
10.	Sound	Sound		A resource primarily intended to be heard. Examples include a music playback file format, an audio compact disc, and recorded speech or sounds.
11.	Text	Text		A resource consisting primarily of words for reading. Examples include books, letters, dissertations, poems,

				newspapers, articles, archives of mailing lists. Note that facsimiles or images of texts are still of the genre Text.
12.	StillImage	Image		A static visual representation. Examples include paintings, drawings, graphic designs, plans and maps. Recommended best practice is to assign the type Text to images of textual materials. Instances of the type StillImage must also be describable as instances of the broader type Image.
13		Model		An abstraction of the real thing, i.e. some generalisation and interpretation. A symbolic representation.

Table D-22: DateType

ID	Value name TR32DB (dcterms)	Value name (DataCite)	Value name (ISO/INSPIRE)	Definition
1.	Created	Created	creation	This property refers to a description of the date or range of the creation of a dataset. According to the one-to-one principle this has to be the creation date of the dataset being described and not the creation date of any other dataset from which the described dataset derives (e.g. a former version or a superior dataset). So a dataset is created only once, every other date of creation belongs to another dataset that has to be described on its own.
2.	Issued	Issued	publication	This property refers to a description of the date of the formal issuance resp. publication of a dataset. A dataset is issued only once, every other issuance belongs to another dataset that has to be described on its own. If the issuance of a dataset is not formal the property "available" should be used.
3.	Available	Available		This property refers to a description of the date a dataset did become or will become available. A dataset becomes available only once, every other availability belongs to another dataset that has to be described on its own. If the availability of a dataset starts with the formal issuance resp. publication use "issued".
4.	Modified	Updated	revision	This property refers to a description of the date a dataset was changed. You may record every date a dataset was modified by repeating this property or record only one date (this should be the last one).
5.	Valid	Valid		This property refers to a description of the date or range a dataset is, was or will be valid. This property should be used if a dataset is only valid resp. relevant until a particular date.
6.	DateCopyrighted	Copyrighted		This property refers to a description of the date or range of the copyright of the dataset.
7.	DateSubmitted	Submitted		This property refers to a description of the date a dataset was submitted (e.g. a thesis at a university department, an article at the editorial board of a journal, etc.).
8.	DateAccepted	Accepted		This property refers to a description of the date a dataset was accepted (e.g. a thesis by a university department, an article by the editorial board of a journal, etc.).

Table D-23: DescriptionType

ID	Value name TR32DB (dcterms)	Value name (DataCite)	Definition
1.	Abstract	Abstract	Abstract describing the resource.
2.		SeriesInformation	Series Information describes a dataset that is part of a series.
3.	TableOfContents	TableOfContents	Table of Contents of a resource.
4.		Other	Other type of description.

Table D-24: DownloadPermission

ID	Value name TR32DB	Definition
1.	Free	Download is possible for everybody.
2.	OnlyOwnSubproject	Download of a data is only possible for registered users of the same CRC/TR32 project section.
3.	OnlyTR32	Download of a data is only possible for members of the CRC/TR32 who are registered users of the TR32DB.

Table D-25: EventType

ID	Value name TR32DB	Definition
1.	Other	Other event type
2.	Conference	Conference
3.	CrossGroupMeeting	CRC/TR32 Cross group meeting
4.	ProjectMeeting	Project meeting
5.	Workshop	Workshop
6.	FieldCampaign	Field Measurement Campaign

Table D-26: InspireTheme

ID	Value name TR32DB (ISO, INSPIRE)	Definition
1.	Addresses	Location of properties based on address identifiers, usually by road name, house number, postal code.
2.	Administrative units	Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries.
3.	Agricultural and aquaculture facilities	Farming equipment and production facilities (including irrigation systems, greenhouses and stables).
4.	Area management/restriction/regulation zones and reporting units	Areas managed, regulated or used for reporting at international, European, national, regional and local levels. Includes dumping sites, restricted areas around drinking water sources, nitrate-vulnerable zones, regulated fairways at sea or large inland waters, areas for the dumping of waste, noise restriction zones, prospecting and mining permit areas, river basin districts, relevant reporting units and coastal zone management areas.
5.	Atmospheric conditions	Physical conditions in the atmosphere. Includes spatial data based on measurements, on models or on a combination thereof and includes measurement locations.
6.	Bio-geographical regions	Areas of relatively homogeneous ecological conditions with common characteristics.
7.	Buildings	Geographical location of buildings.
8.	Cadastral parcels	Areas defined by cadastral registers or equivalent.
9.	Coordinate reference systems	Systems for uniquely referencing spatial information in space as a set of coordinates (x, y, z) and/or latitude and longitude and height, based on a geodetic horizontal and vertical datum.
10.	Elevation	Digital elevation models for land, ice and ocean surface. Includes terrestrial elevation, bathymetry and shoreline.
11.	Energy resources	Energy resources including hydrocarbons, hydropower, bio-energy, solar, wind, etc., where relevant including depth/height information on the extent of the resource.
12.	Environmental monitoring facilities	Location and operation of environmental monitoring facilities includes observation and measurement of emissions, of the state of environmental media and of other ecosystem parameters (biodiversity, ecological conditions of vegetation, etc.) by or on behalf of public authorities.
13.	Geographical grid systems	Harmonised multi-resolution grid with a common point of origin and standardised location and size of grid cells.
14.	Geographical names	Names of areas, regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest.
15.	Geology	Geology characterised according to composition and structure. Includes bedrock, aquifers and geomorphology.
16.	Habitats and biotopes	Geographical areas characterised by specific ecological conditions, processes, structure, and (life support) functions that physically support the organisms that live there. Includes terrestrial and aquatic areas distinguished by geographical, abiotic and biotic features, whether entirely natural or semi-natural.
17.	Human health and safety	Geographical distribution of dominance of pathologies (allergies, cancers, respiratory diseases, etc.), information indicating the effect on health (biomarkers, decline of

		fertility, epidemics) or well-being of humans (fatigue, stress, etc.) linked directly (air pollution, chemicals, depletion of the ozone layer, noise, etc.) or indirectly (food, genetically modified organisms, etc.) to the quality of the environment.
18.	Hydrography	Hydrographic elements, including marine areas and all other water bodies and items related to them, including river basins and sub-basins. Where appropriate, according to the definitions set out in Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (2) and in the form of networks.
19.	Land cover	Physical and biological cover of the earth s surface including artificial surfaces, agricultural areas, forests, (semi-)natural areas, wetlands, water bodies.
20.	Land use	Territory characterised according to its current and future planned functional dimension or socio-economic purpose (e.g. residential, industrial, commercial, agricultural, forestry, recreational).
21.	Meteorological geographical features	Weather conditions and their measurements; precipitation, temperature, evapotranspiration, wind speed and direction.
22.	Mineral resources	Mineral resources including metal ores, industrial minerals, etc., where relevant including depth/height information on the extent of the resource.
23.	Natural risk zones	Vulnerable areas characterised according to natural hazards (all atmospheric, hydrologic, seismic, volcanic and wildfire phenomena that, because of their location, severity, and frequency, have the potential to seriously affect society), e.g. floods, landslides and subsidence, avalanches, forest fires, earthquakes, volcanic eruptions.
24.	Oceanographic geographical features	Physical conditions of oceans (currents, salinity, wave heights, etc.).
25.	Orthoimagery	Geo-referenced image data of the Earth's surface, from either satellite or airborne sensors.
26.	Population distribution - demography	Geographical distribution of people, including population characteristics and activity levels, aggregated by grid, region, administrative unit or other analytical unit.
27.	Production and industrial facilities	Industrial production sites, including installations covered by Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control (1) and water abstraction facilities, mining, storage sites.
28.	Protected sites	Area designated or managed within a framework of international, Community and Member States legislation to achieve specific conservation objectives.
29.	Sea regions	Physical conditions of seas and saline water bodies divided into regions and sub-regions with common characteristics.
30.	Soil	Soils and subsoil characterised according to depth, texture, structure and content of particles and organic material, stoniness, erosion, where appropriate mean slope and anticipated water storage capacity.
31.	Species distribution	Geographical distribution of occurrence of animal and plant species aggregated by grid, region, administrative unit or other analytical unit.
32.	Statistical units	Units for dissemination or use of statistical information.
33.	Transport networks	Road, rail, air and water transport networks and related infrastructure. Includes links between different networks. Also includes the trans-European transport network as defined in Decision No 1692/96/EC of the European Parliament and of the Council of 23 July 1996 on Community Guidelines for the development of the trans-European transport network (1) and future revisions of that Decision.
34.	Utility and governmental services	Includes utility facilities such as sewage, waste management, energy supply and water supply, administrative and social governmental services such as public administrations, civil protection sites, schools and hospitals.

Table D-27: IdentifierType

ID	Value name TR32DB (dcterms)	Value name (DataCite)	Definition
1.	ARK	ARK	The ARK (Archival Resource Key) naming scheme is designed to facilitate the high-quality and persistent identification of information objects [ARK].
2.	DOI	DOI	The Digital Object Identifier (DOI) is a persistent identifier of intellectual property entities. [http://www.doi.org/]
3.		EAN13	The EAN-13 barcode (originally European Article Number, now renamed International Article Number) is a 13 digit barcoding standard.
4.		EISSN	Electronic International Standard Serial Number [http://www.issn.org/]
5.	Handle	Handle	The Handle System (HANDLE) is a comprehensive system for assigning, managing, and resolving persistent identifiers, known as "handles," for digital objects and other resources on the Internet. [http://www.handle.net/]
6.	ISBN	ISBN	An International Standard Book Number (ISBN) identifies an edition of a

			monographic work and is defined by the standard NISO/ANSI/ISO 2108:1992.
7.	ISSN	ISSN	The International Standard Serial Number (ISSN) is an eight-digit number which identifies periodical publications, including electronic serials. [http://www.issn.org/]
8.		ISTC	International Standard Text Code [http://www.istc-international.org/html/]
9.		LISSN	Linking International Standard Serial Number [http://www.issn.org/]
10.		LSID	Life Science Identifier [http://www.ipni.org/lsids.html
11.	PURL	PURL	A PURL is a Persistent Uniform Resource Locator. Functionally, a PURL is a URL. However, instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. Because PURLs conform to the URI specification, they can be used unmodified in DC metadata. [http://purl.org/]
12.		UPC	Universal Product Code
13.	URL	URL	Uniform Resource Locator
14.	URN	URN	Uniform Resource Name
15.	SICI		Serial Item and Contribution Identifier

Table D-28: InstitutionStatus

ID	Value name TR32DB	Definition
1.	InstitutionIsFullMember	Institution is full member of CRC/TR32.
	InstitutionIsNoMember	Institution is no member of CRC/TR32.
2.	InstitutionWasMember	Institution is former member of CRC/TR32.
3.	InstitutionIsAssociateMember	Institution is associate member of CRC/TR32.
4.	InstitutionIsCorresponding Member	Institution is corresponding member of CRC/TR32.

Table D-29: InitiativeType

ID	Value name TR32DB (ISO)	Definition
1.	campaign	Series of organized planned actions.
2.	collection	Accumulation of datasets assembled for a specific purpose.
3.	dataDictionary	Element and entity definitions.
4.	exercise	Specific performance of a function or group of functions.
5.	experiment	Process designed to find if something is effective or valid.
6.	investigation	Search or systematic inquiry.
7.	mission	Specific operation of a data collection system.
8.	sensor	Device or piece of equipment which detects or records.
9.	operation	Action that is part of a series of actions.
10.	platform	Vehicle or other support base that holds a sensor.
11.	process	Method of doing something involving a number of steps.
12.	program	Specific planned activity.
13.	project	Organized undertaking, research, or development.
14.	sciencePaper	Document based on an experiment or research.
15.	study	Examination or investigation.
16.	task	Piece of work.
17.	trial	Process of testing to discover or demonstrate something.
18.	userGuide	Operating manual for users.

Table D-30: IsoTopicCategory

ID	Value name TR32DB (ISO, INSPIRE)	Definition
1.	farming	rearing of animals and/or cultivation of plants; Examples: agriculture, irrigation, aquaculture, plantations, herding, pests and diseases affecting crops and livestock
2.	biota	flora and/or fauna in natural environment; Examples: wildlife, vegetation, biological sciences, ecology, wilderness, sea life, wetlands, habitat, biological resources
3.	boundaries	legal land descriptions; Examples: political and administrative boundaries, governmental units, marine boundaries, voting districts, school districts, international boundaries
4.	climatologyMeteorologyAtmosphere	processes and phenomena of the atmosphere; Examples: cloud cover, weather, climate, atmospheric conditions, climate change, precipitation
5.	economy	economic activities, conditions, and employment; Examples: production, labor, revenue, business, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, exploration and exploitation of resources such as minerals, oil and gas
6.	elevation	height above or below sea level; Examples: altitude, bathymetry, digital elevation models, slope, derived products, DEMs, TINs
7.	environment	environmental resources, protection and conservation; Examples: environmental pollution, waste storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape, water quality, air quality, environmental modelling
8.	geoscientificInformation	information pertaining to earth sciences; Examples: geophysical features and processes, geology, minerals, sciences dealing with the composition, structure and origin of the earth's rocks, risks of earthquakes, volcanic activity, landslides, gravity information, soils, permafrost, hydrogeology, groundwater, erosion
9.	health	health, health services, human ecology, and safety; Examples: disease and illness, factors affecting health, hygiene, substance abuse, mental and physical health, health services, health care providers, public health
10.	imageryBaseMapsEarthCover	base maps; Examples: land/earth cover, topographic maps, imagery, unclassified images, annotations, digital ortho imagery
11.	intelligenceMilitary	military bases, structures, activities; Examples: barracks, training grounds, military transportation, information collection
12.	inlandWaters	inland water features, drainage systems and characteristics; Examples: rivers and glaciers, slat lakes, water utilization plans, dams, currents, floods and flood hazards, water quality, hydrographic charts, watersheds, wetlands, hydrography
13.	location	positional information and services; Examples: addresses, geodetic networks, geodetic control points, postal zones and services, place names, geographic names
14.	oceans	features and characteristics of salt water bodies (excluding inland waters) Examples: tides, tidal waves, coastal information, reefs, maritime, outer continental shelf submerged lands, shoreline
15.	planningCadastre	information used for appropriate actions for future use of the land; Examples: land use maps, zoning maps, cadastral surveys, land ownership, parcels, easements, tax maps, federal land ownership status, public land conveyance records
16.	society	characteristics of society and culture; Examples: settlements, housing, anthropology, archaeology, education, traditional beliefs, manners and customs, demographic data, tourism, recreational areas and activities, parks, recreational trails, historical sites, cultural resources, social impact assessments, crime and justice, law enforcement, census information, immigration, ethnicity
17.	structure	man-made construction; Examples: buildings, museums, churches, factories, housing, monuments, shops, towers, building footprints, architectural and structural plans
18.	transportation	means and aids for conveying persons and/or goods; Examples: roads, airports/airstrips, shipping routes, tunnels nautical charts, vehicle or vessel location, aeronautical charts, railways
19.	utilitiesCommunication	energy, water and waste systems and communications infrastructure and services; Examples: hydroelectricity, geothermal, solar and nuclear sources of energy, water purification and distribution, sewage collection and disposal, electricity and gas distribution, data communication, telecommunication, radio, communication networks

Table D-31: Licence

ID	Value name TR32DB	Definition
1.	Creative Commons - Attribution 3.0 Unported [CC BY]	This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials. [http://creativecommons.org/licenses/by/3.0/
2.	Creative Commons - Attribution-NoDerivs 3.0 Unported [CC BY-ND]	This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you. [http://creativecommons.org/licenses/by-nd/3.0/
3.	Creative Commons - Attribution-NonCommercial 3.0 Unported [CC BY-NC]	This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms. [http://creativecommons.org/licenses/by-nc/3.0/
4.	Creative Commons - Attribution-NonCommercial-NoDerivs 3.0 Unported [CC BY-NC-ND]	This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially. [http://creativecommons.org/licenses/by-nc-nd/3.0/
5.	Creative Commons - Attribution-NonCommercial-ShareAlike 3.0 Unported [CC BY-NC-SA]	This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms. [http://creativecommons.org/licenses/by-nc-sa/3.0/
6.	Creative Commons - Attribution-ShareAlike 3.0 Unported [CC BY-SA 3.0]	This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects. [http://creativecommons.org/licenses/by-sa/3.0/]
7.	Open Data Commons - Open Database License [ODC-ODbL]	Attribution Share-Alike for data/databases [http://opendatacommons.org/licenses/odbl/summary/]
8.	Open Data Commons - Public Domain Dedication and License [PDDL]	Public Domain for data/databases [http://opendatacommons.org/licenses/pddl/summary/]
9.	Open Data Commons - Attribution License [ODC-By]	Attribution for data/databases [http://opendatacommons.org/licenses/by/summary/]
10.	TR32DB Data policy agreement	Data Policy Agreement of the CRC/TR32 Project Database [http://www.tr32db.uni-koeln.de/datapolicy/]
11.	Creative Commons - Attribution 3.0 Unported [CC BY]	This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials. [http://creativecommons.org/licenses/by/3.0/]
12.	Creative Commons - Attribution-NoDerivs 3.0 Unported [CC BY-ND]	This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you. [http://creativecommons.org/licenses/by-nd/3.0/]
13.	Creative Commons - Attribution-NonCommercial 3.0 Unported [CC BY-NC]	This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms. [http://creativecommons.org/licenses/by-nc/3.0/]
14.	Creative Commons - Attribution-NonCommercial-NoDerivs 3.0 Unported [CC BY-NC-ND]	This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially. [http://creativecommons.org/licenses/by-nc-nd/3.0/]
15.	Creative Commons - Attribution-NonCommercial-ShareAlike 3.0 Unported [CC BY-NC-SA]	This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms. [http://creativecommons.org/licenses/by-nc-sa/3.0/
16.	Creative Commons - Attribution-ShareAlike 3.0 Unported [CC BY-SA 3.0]	This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any

		derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects. [http://creativecommons.org/licenses/by-sa/3.0/]

Table D-32: MaintenanceFrequencyUnit

ID	Value name TR32DB (ISO)	Definition
1.	continual	Data is repeatedly and frequently updated.
2.	daily	Data is updated each day.
3.	weekly	Data is updated on a weekly basis.
4.	fortnightly	Data is updated every two weeks.
5.	monthly	Data is updated each month.
6.	quarterly	Data is updated every three months.
7.	biannually	Data is updated twice each year.
8.	annually	Data is updated every year.
9.	asNeeded	Data is updated as deemed necessary.
10.	irregular	Data is updated in intervals that are uneven in duration.
11.	notPlanned	There are no plans to update the data.
12.	unknown	Frequency of maintenance for the data is not known.

Table D-33: MeasureLocationTR32*

ID	Value name TR32DB	Definition
1.	Wüstebach	TR32 measure/model location Wüstebach located in Erkensruhr.
2.	Rollesbroich	TR32 measure/model location Rollesbroich located in Kall.
3.	Merken	TR32 measure/model location Merken located in Ellebach.
4.	Merzenhausen	TR32 measure/model location Merzenhausen located in Ellebach.
5.	Other	Other measure location, not listed.
6.	None	No measure/model location available.
7.	Laboratory	TR32 measure/model location located in the laboratory.
8.	Krauthausen	TR32 measure/model location Krauthausen located in Ellebach.
9.	Jülich (reserach centre)	TR32 measure/model location Research Centre Jülich located in Ellebach.
10.	Ruraue	TR32 measure/model location Ruraue located in Ellebach.
11.	Selhausen	TR32 measure/model location Selhausen located in Ellebach.

Table D-34: MeasureRegionTR32*

ID	Value name TR32DB	Definition
1.	Belgium	TR32 measure/model region located in Belgium.
2.	C4ModellingArea	TR32 measure/model region covering the C4 Modelling Area.
3.	Ellebach	TR32 measure/model region located in the sub-catchment Ellebach.
4.	Erkensruhr	TR32 measure/model region located in the sub-catchment Erkensruhr.
5.	Germany	TR32 measure/model region located in Germany.
6.	Kall	TR32 measure/model region located in the sub-catchment Kall.
7.	Laboratory	TR32 measure/model region located in the laboratory.
8.	None	No TR32 measure/model region available.
9.	North Rhine-Westphalia	TR32 measure/model region located in North Rhine Westphalia.
10.	Other	Other TR32 measure/model region not listed.
11.	Rhineland-Palatinate	TR32 measure/model region located in Rhineland-Palatinate.
12.	RurCatchment	TR32 measure/model region located in the catchment of river Rur.
13.	TheNetherlands	TR32 measure/model region located in The Netherlands.

Table D-35: MeasureUnit

ID	Value name TR32DB (ISO)	Definition
1.	mm	millimetres
2.	cm	centimetres
3.	m	meters

Table D-36: Orientation

ID	Value name TR32DB	Definition
1.	Horizontal/Landscape	Horizontal/landscape image orientation.
2.	Vertical/Portrait	Vertical/portrait image orientation.
3.	Other	Other image orientation.

Table D-37: Phase*

ID	Value name TR32DB	Definition
1.	1	CRC/TR32 funding phase 1 (2007-2010).
2.	2	CRC/TR32 funding phase 2 (2011-2014).

Table D-38: PresentationForm

ID	Value name TR32DB (ISO)	Definition
1.	documentDigital	Digital representation of a primarily textual item (can contain illustrations also).
2.	documentHardcopy	Representation of a primarily textual item (can contain illustrations also) on paper, photographic material, or other media.
3.	imageDigital	Likeness of natural or man-made features, objects, and activities acquired through the sensing of visual or any other segment of the electromagnetic spectrum by sensors, such as thermal infrared, and high resolution radar and stored in digital format.
4.	imageHardcopy	Likeness of natural or man-made features, objects, and activities acquired through the sensing of visual or any other segment of the electromagnetic spectrum by sensors, such as thermal infrared, and high resolution radar and reproduced on paper, photographic material, or other media for use directly by the human user.
5.	mapDigital	Map represented in raster or vector form.
6.	mapHardcopy	Map printed on paper, photographic material, or other media for use directly by the human user.
7.	modelDigital	Multi-dimensional digital representation of a feature, process, etc.
8.	modelHardcopy	3-dimensional, physical model.
9.	profileDigital	Vertical cross-section in digital form.
10.	profileHardcopy	Vertical cross-section printed on paper, etc.
11.	tableDigital	Digital representation of facts or figures systematically displayed, especially in columns.
12.	tableHardcopy	Representation of facts or figures systematically displayed, especially in columns, printed on papers, photographic material, or other media.
13.	videoDigital	Digital video recording.
14.	videoHardcopy	Video recording on film.

Table D-39: PresentationType

ID	Value name TR32DB	Definition
1.	Keynote	Type of presentation is keynote.
2.	Poster	Type of presentation is poster.
3.	Talk	Type of presentation is talk.
4.	other	Type of presentation is other.

Table D-40: PublicationReview

ID	Value name TR32DB	Value name (bibo)	Definition
1.	PeerReview	peerReviewed	A review of the publication was conducted.
2.	NoPeerReview	nonPeerReviewed	There was no review of the publication conducted.

Table D-41: PublicationStatus

ID	Value name TR32DB	Value name (bibo)	Definition
1.	InPrint		Document is in print.
2.	InReview		Document is currently in review process.
3.	Published	published	Document is published.
4.	Submitted		Document is submitted to journal.
5.	Unpublished	unpublished	Document is unpublished.
6.	Accepted	accepted	Document is accepted for publication after peer reviewing.
7.	Draft	draft	Document is drafted.
8.	Forthcoming	forthcoming	Document to be published.
9.	Legal	legal	Legal document.
10.	NonPeerReviewed	nonPeerReviewed	A document that is not peer reviewed.
11.	PeerReviewed	peerReviewed	The process by which articles are chosen to be included in a refereed journal. An editorial board consisting of experts in the same field as the author review the article and decide if it is authoritative enough for publication.
12.	Rejected	rejected	Rejected for publication after peer reviewing.

Table D-42: PublicationType

ID	Value name TR32DB	Definition
1.	Article	Type of publication is article.
2.	Book	Type of publication is book.
3.	BookSection	Type of publication is book section or chapter.
4.	EventPaper	Type of publication is event paper (conference paper).

Table D-43: ReferenceSystem

ID	Value name TR32DB	Definition incl. ReferenceSystemType
1.	32632	WGS 84 / UTM zone 32N (World Geodetic System 1984 / UTM zone 32N); Projected CRS
2.	31466	DHDN / 3-degree Gauss-Kruger zone 2 (Deutsches Hauptdreiecksnetz / 3-degree Gauss-Kruger zone 2); Projected CRS
3.	28992	Amersfoort / RD New (Amersfoort / RD New); Projected CRS
4.	4326	WGS 84 (World Geodetic System 1984); Geodetic CRS
5.	4258	ETRS89 (European Terrestrial Reference System 1989); Geodetic CRS
6.	25832	ETRS89 / UTM zone 32N (European Terrestrial Reference System 1989 / UTM zone 32N)
7.	32631	WGS 84 / UTM zone 31N (World Geodetic System 1984 / UTM zone 31N); Projected CRS
8.	5783	DHHN92 height (Deutsches Haupthoehennetz 1992); Vertical CRS
9.	5621	EVRF2007 (European Vertical Reference Frame 2007); Vertical CRS
10.	31467	DHDN / 3-degree Gauss-Krüger zone 3 (Deutsches Hauptdreiecksnetz / 3-degree Gauss-Krüger zone 2); Projected CRS

Table D-44: ReferenceSystemType

ID	Value name TR32DB	Definition
1.	Projected CRS	Projected coordinate reference system.
2.	Geodetic CRS	Geodetic coordinate reference system (geographic 2D).
3.	Geodetic CRS	Geodetic coordinate reference system (geographic 3D).
4.	Geodetic CRS	Geodetic coordinate reference system (geocentric).
5.	Vertical CRS	Vertical coordinate reference system.
6.	Engineering CRS	Engineering coordinate reference system.
7.	Compound CRS	Compound coordinate reference system.

Table D-45: RelationType

ID	Value name TR32DB (dcterms)	Value name (DataCite)	Definition
1.	IsReferencedBy	IsCitedBy	This property describes the relationship between a resource and another resource that points to the described resource by citation, acknowledgement, etc (e.g. the described resource is a book cited in an article, or a play pointed to in an interview, etc.). For the reciprocal statement use references. Indicates that the related resource includes the described resource in a citation.
2.	References	Cites	This property describes the relationship between the described resource and another resource that is cited, referenced, or otherwise pointed to by the described resource (e.g. the described resource is an article citing a book, or an interview pointing to a play). For the reciprocal statement use isReferencedBy. Indicates that the described resource includes the related resource in a citation.
3.		IsSupplementTo	Indicated that the described resource is a supplement to the related resource.
4.		IsSupplementedBy	Indicated that the related resource is a supplement to the described resource.
5.		IsContinuedBy	Indicates that the described resource is continued by the work of the related resource.
6.		Continues	Indicates that the related resource is a continuation of the work of the described resource.
7.	IsVersionOf		This property describes the relationship between the described resource and another resource that is a former version, edition or adaptation of the described resource (e.g. the described resource is the revision of a book, or another recording of a song, etc.). Another version implies changes in the content of a resource. For resources with different formats use isFormatOf. For the reciprocal statement use hasVersion.
8.	IsPartOf		This property describes the relationship between the described resource and another resource of which the described resource is a physical or logical part (e.g. a painting as part of a collection, an article as part of a journal, etc.). The described resource is like a "child" in a hierarchical or "parent/child" relationship. For the reciprocal statement use hasPart.
9.	HasPart		This property describes the relationship between the described resource and another resource which is a physical or logical part of the described resource (e.g. the described resource is a collection of paintings, or a journal with different articles, etc.). The described resource is like the "parent" in a hierarchical or "parent/child" relationship. For the reciprocal statement use isPartOf.
10.		IsDocumentedBy	Indicates that the related resource is a documentation about/explaining the described resource.
11.		Documents	Indicates that the described resource is a documentation about/explaining the related resource.
12.		IsCompiledBy	Indicates that the related resource is used to compile or create the described resource.
13.		Compiles	Indicates that the related resource is the result of a compile or creation event using the described resource.
14.		IsVariantFormOf	Indicates that the described resource is a variant of different form of related resource, e.g. calculated or calibrated form of different packaging.
15.		IsOriginalFormOf	Indicates that the described resource is the original form of the related resource.
16.		IsPreviousVersionOf	Indicates that the described resource is a previous edition of the related resource.
17.	HasVersion		This property describes the relationship between the described property and another property that is a later version, edition or adaptation of the described resource (e.g. the described resource is the older version of a revised book, or of a song, etc.). Another version implies changes in the content of a resource. For resources with different formats use hasFormat. For the reciprocal statement use isVersionOf.
18.	IsFormatOf		This property describes the relationship between the described resource and another resource that is a former version of the described resource with the same intellectual content but presented in another format (e.g. the described resource is the microfilm version of a printed book or the pdf version of a doc document). For intellectual changes between resources use isVersonOf. For the reciprocal statement use hasFormat.
19.	HasFormat		This property describes the relationship between the described resource and another resource that is a later version of the described resource with the same intellectual content but presented in another format (e.g. the described resource is a printed book that is also available as a microfilm, or a doc document that is also available as

			pdf). For intellectual changes between resources use hasVersion. For the reciprocal statement use isFormatOf.
20.	Replaces		This property describes the relationship between the described resource and another resource that has been supplanted, displaced or superseded by the described resource. It is used for the valid version in chain of versions (e.g. the described resource is the last draft of a contract, or the current version of guidelines). For the reciprocal statement use isReplacedBy.
21.	IsReplacedBy		This property describes the relationship between the described resource and another resource, that supplants, displaces or supersedes the described resource. It is used, when in chain of versions only one version is valid (e.g. the described resource is one of the former drafts of a contract, or a former version of guidelines). For the reciprocal statement use replaces.
22.	Requires		This property describes the relationship between the described resource and another resource supporting the function, delivery or coherence of the content of the described resource (e.g. the described resource is an application that can be used only with a particular software, or hardware). For the reciprocal statement use isRequiredBy.
23.	IsRequiredBy		The described resource is necessary for the function, delivery or coherence of the content of the resource the property references to (e.g. the described resource is a software or hardware necessary to use a particular application). For the reciprocal statement use requires.
24.	ConformsTo		This property describes the relationship between a resource and an established standard, to which the described resource conforms (e.g. a metadata record that conforms to the RDA standard, or a pipe that conforms to ISO 3183, etc.).

Table D-46: ReportType

ID	Value name TR32DB	Definition
1.	Booklet	Report type is booklet.
2.	BookReview	Report type is book review.
3.	ConferenceReview	Report type is conference review.
4.	Glossary	Report type is glossary.
5.	Minutes	Report type is minutes.
6.	OnlineTutorial	Report type is online tutorial.
7.	Other	Report type is other.
8.	PhDReport	Report type is PhD report.
9.	Policy	Report type is guideline.
10.	Guideline	Report type is guideline.
11.	Presentation	Report type is presentation.
12.	DiplomaThesis	Report type is an unpublished/unpublished Diploma thesis.
13.	MasterThesis	Report type is an unpublished/unpublished Master's thesis.
14.	BachelorThesis	Report type is an unpublished/unpublished Bachelor's thesis.
15.	PhDThesis	Report type is an unpublished/unpublished PhDs thesis.
16.	WorkshopReview	Report type is workshop review.

Table D-47: ResolutionDistanceUnit

ID	Value name TR32DB	Definition
1.	μm	Micrometres
2.	mm	Millimetres
3.	cm	Centimetres
4.	dm	Decimetres
5.	m	Metres
6.	km	Kilometres
7.	other	Other

Table D-48: RoleType

ID	Value name TR32DB (ISO, INSPIRE)	Definition
1.	ResourceProvider	Party that supplies the resource.
2.	Custodian	Party that accepts accountability and responsibility for the data and ensures appropriate care and maintenance of the resource.
3.	Owner	Party that owns the resource.
4.	User	Party who uses the resource.
5.	Distributor	Party who distributes the resource.
6.	Originator	Party who created the resource.
7.	PointOfContact	Party who can be contacted for acquiring knowledge about or acquisition of the resource.
8.	PrincipalInvestigator	Key party responsible for gathering information and conducting research.
9.	Processor	Party who has processed the data in a manner such that the resource has been modified.
10.	Publisher	Party who published the resource.
11.	Author	Party who authored the resource.

Table D-49: ScopeCode

ID	Value name TR32DB (ISO)	Value name (dcterms)	Definition
1.	attribute	Dataset	Information applies to the attribute class.
2.	attributeType	Dataset	Information applies to the characteristic of a feature.
3.	collectionHardware	Dataset	Information applies to the collection hardware class.
4.	collectionSession	Event	Information applies to the collection session.
5.	dataset	Dataset	Applies to the dataset.
6.	series	Collection	Applies to the series.
7.	nonGeographicDataset	Dataset	Information applies to non-geographic data.
8.	dimensionGroup	Dataset	Information applies to a dimension group.
9.	featureInformation	Dataset	Applies to a feature.
10.	featureType	Dataset	Information applies to a feature type.
11.	propertyType	Dataset	Information applies to a property type.
12.	fieldSession	Dataset	Information applies to a field session.
13.	software	Software	Information applies to a computer program or routine.
14.	service	Service	Information applies to a capability which a service provider entity makes available to a service user entity through a set of interfaces that define a behaviour, such as a use case.
15.	model	Dataset	Information applies to a copy or imitation of an existing or hypothetical object.
16.	tile	Dataset	Information applies to a tile, a spatial subset of geographic data.

Table D-50: SizeType

ID	Value name TR32DB	Definition
1.	DataPoints	The amount of data points within a dataset.
2.	Datasets	The amount of datasets contained in a dataset.
3.	Experiments	The mount of experiments described or contained in the dataset.
4.	Numbers	The amount of numbers describing the dataset.
5.	Pages	The amount of pages contained in the dataset.
6.	Kilobytes	The size of the dataset in kilobytes.

Table D-51: SpatialRepresentationType

ID	Value name TR32DB (ISO)	Definition
1.	vector	Vector data is used to represent geographic data
2.	grid	Grid / raster data is used to represent geographic data
3.	textTable	Textual or tabular data is used to represent geographic data
4.	tin	Triangulated irregular network
5.	stereoModel	Three-dimensional view formed by the intersecting homologous rays of an overlapping pair of images
6.	video	Scene from a video recording

Table D-52: SubjectScheme

ID	Value name TR32DB	Definition
1.	GEMET - Concepts	Title: GEMET - Concepts, version 3.1 publication date: 2012-07-20 URL: http://www.eionet.europa.eu/gemet/alphabetic?langcode=en KeywordTypeCode: theme
2.	GEMET - INSPIRE themes	Title: GEMET - INSPIRE themes, version 1.0 publication date: 2008-06-01 URL: http://www.eionet.europa.eu/gemet/inspire_themes KeywordTypeCode: theme
3.	DDC	Dewey Decimal Classification URL: http://dewey.info/ KeywordTypeCode: theme

Table D-53: TemporalFrequencyUnit

ID	Value name TR32DB	Definition
1.	day/s	Measure frequency per day/s.
2.	hour/s	Measure frequency per hour/s.
3.	Hz	Measure frequency per hertz.
4.	MHz	Measure frequency per megahertz.
5.	minute/s	Measure frequency per minute/s.
6.	month/s	Measure frequency per month/s.
7.	other	Other frequency, not listed.
8.	second/s	Measure frequency per second/s.
9.	week/s	Measure frequency per week/s.
10.	year/s	Measure frequency per year/s.

Table D-54: TitleType

ID	Value name TR32DB	Value name (dcterms)	Value name (DataCite)	Value name (ISO)	Definition
1.	mainTitle				Main title of the dataset.
2.	alternativeTitle	alternative	alternativeTitle	alternateTitle	Alternative title of the dataset.
3.	subtitle		subtitle		Subtitle of the dataset.
4.	translatedTitle		translatedTitle		Translated title of the dataset.

Table D-55: TR32MetaDataType

ID	Value name TR32DB	Definition
1.	data	Data type data.
2.	geodata	Data type geodata.
3.	picture	Data type picture.
4.	presentation	Data type presentation.
5.	publication	Data type publication.
6.	report	Data type report.

Table D-56: TR32Subproject*

ID	Value name TR32DB	Definition
1.	A1	Pore structure, moisture, and water transport in soil by NMR
2.	A2	Characterization of flow and transport properties of unsaturated soils using non-invasive methods and numerical modeling at the μm to dm scale
3.	A3	Inverse modeling of soil hydraulic property patterns from non-invasive electrical measurements
4.	A4	Multiphase flow simulations coupled to NMR and SIP signatures in reconstructed natural porous media using Lattice Boltzmann methods
5.	B1	Characterizing spatio-temporal patterns of water and C-fluxes at field-scale
6.	B3	Spatial and temporal patterns of soil carbon pools - a MIR-spectroscopic approach
7.	B4	Upscaling of local CO_2, water and heat fluxes to the field scale from observed spatio-temporal patterns using an integrative 3-D mechanistic process model
8.	B5	Modeling the spatio-temporal variability of crop and cropping system processes under heterogeneous field conditions
9.	B6	Soil moisture content estimation by inverting surface and horizontal borehole GPR data
10.	C1	Spatio-temporal variability of catchment properties and their effect on water, solute, and CO_2 fluxes from the micro to the mesoscale
11.	C2	Analysis of temporal and spatial structures of soil moisture by integrating remote sensing and process based modeling
12.	C3	Scale–consistent two-way coupling of land-surface and atmospheric models
13.	C4	Process-based modeling of regional water and energy fluxes taking into account multi-sensor and multi-scale observation patterns
14.	C6	Process-based modeling of regional water and energy fluxes taking into account multi-sensor and multi-scale observation patterns
15.	D1	Modeling of time-space patterns of energy and mass fluxes over heterogeneous land surfaces with special consideration of CO_2 and H_2O
16.	D2	Experimental study of spatio-temporal structures in atmosphere-land surface energy, water and CO_2 exchange
17.	D3	CO_2 and water flux estimation by fourdimensional variational assimilation of in situ & RS data
18.	D5	A high-resolution multi-scale space-time precipitation model from direct measurements & RS
19.	D6	Large-Eddy Simulation of Atmosphere and Land Interactions over Heterogeneous Surfaces
20.	D7	Influence of subsurface hydrodynamics on the lower atmosphere at the catchment scale
21.	D8	High resolution parallel simulation of variably saturated flow with adaptive mesh refinement
22.	Z1/INF	Project Database and Data Management
23.	Z2	SFB/TR32 Central Services, Administration and Coordination
24.	Z3	Measurement Support and Data Processing
25.	Z4	Model Development and Maintenance Services
26.	Z5-IRTG	Integrated Research Training Group (IRTG)

Table D-57: TR32Topic

ID	Value name TR32DB	Definition
1.	atmosphere	Data related to the topic atmosphere.
2.	land use	Data related to the topic land use.
3.	other	Other topic, not listed.
4.	remote sensing	Data related to the topic remote sensing.
5.	soil	Data related to the topic soil.
6.	topography	Data related to the topic topography.
7.	vegetation	Data related to the topic vegetation.

Table D-58: WebsiteVersion*

ID	Value name TR32DB	Definition
1.	V25	Website and Metadata Version 2.5
2.	V26	Website and Metadata Version 2.6
3.	V30	Website and Metadata Version 3.0
4.	V31	Website and Metadata Version 3.1
5.	V40	Website and Metadata Version 4.0

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von unten angegebenen Teilpublikationen − noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Georg Bareth betreut worden.

Constanze Curdt

Teilpublikationen:

CURDT, C., 2014. TR32DB Metadata Schema for the Description of Research Data in the TR32DB. Collaborative Research Centre/Transregio 32, Project Section Z1/INF, Institute of Geography, University of Cologne, Cologne, Germany.
doi: 10.5880/TR32DB.10.

CURDT, C., HOFFMEISTER, D., WALDHOFF, G., JEKEL, C., BARETH, G., 2012. Scientific Research Data Management for Soil-Vegetation-Atmosphere Data: The TR32DB. International Journal of Digital Curation 7 (2), 68-80.
doi: 10.2218/ijdc.v7i2.0.

CURDT, C., HOFFMEISTER, D., WALDHOFF, G., JEKEL, C., BARETH, G., 2012. Developement of a metadata management system for an interdisciplinary research project. Proccedings of the XXII ISPRS Congress, 25 August-01 September, 2012, Melbourne, Australia, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 7-12.

CURDT, C., HOFFMEISTER, D., JEKEL, C., BROCKS, S., WALDHOFF, G., BARETH, G., 2011. TR32DB - Management and visualization of heterogeneous scientific data. Proccedings of the 19th International Conference on Geoinformatics June 24-26, 2011, Shanghai, China, 1-6. doi: 10.1109/GeoInformatics.2011.5981092

CURDT, C., HOFFMEISTER, D., JEKEL, C., WALDHOFF, G., BARETH, G., 2011. TR32DB: Implementation of a Scientific Data Management System for the CRC/TR32. Proccedings of the Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data (PV 2011), November 15-17, 2011, Toulouse, France.

CURDT, C., HOFFMEISTER, D., JEKEL, C., UDELHOFEN, K., WALDHOFF, G., BARETH, G., 2010. Implementation of a centralized data management system for the CRC Transregio 32 'Patterns in soil-vegetation-atmosphere-systems'. In: CURDT, C., BARETH, G. (Eds.), Proceedings of the Data Management Workshop. 29. - 30.10.2009, University of Cologne, Germany. Geographisches Institut der Universität zu Köln, Cologne, Germany, 27-33. doi: 10.5880/TR32DB.KGA90.6.

CURDT, C., HOFFMEISTER, D., WALDHOFF, G., BARETH, G., 2009. Implementierung eines Projektdatenmanagements für das interdisziplinäre Forschungsprojekt TR32 „Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation" Proccedings of the 29. Gil Jahrestagung, Rostock, Germany, 37-40.

CURDT, C., HOFFMEISTER, D., WALDHOFF, G., BARETH, G., 2009. Implementation of a metadatabase for a regional SDI for soil-vegetation-atmosphere modelling. Proccedings of the GSDI 11 World Conference: Spatial Data Infrastructure Convergence: Building SDI Bridges to address Global Challenges, June 15-19, 2009, Rotterdam, The Netherlands.

CURDT, C., HOFFMEISTER, D., WALDHOFF, G., BARETH, G., 2009. Management and visualization of interdisciplinary research data with ArcGIS Server. Proccedings of the ESRI Intern. User Conference, July 13-17, 2009, San Diego, CA, USA.

CURDT, C., HOFFMEISTER, D., WALDHOFF, G., BARETH, G., 2008. Spatial Data Infrastructure for Soil-Vegetation-Atmosphere Modelling: Set-up of a spatial database for a research project (SFB/TR32). Proccedings of the XXI ISPRS Congress Beijing 2008, July 3-11, 2008, Beijing, China, 131-136.