

Digitale Edition in Österreich

---

Digital Scholarly Edition in Austria

**Schriften des  
Instituts für Dokumentologie und Editorik**

---

herausgegeben von:

Bernhard Assmann	Roman Bleier
Alexander Czmiel	Stefan Dumont
Oliver Duntze	Franz Fischer
Christiane Fritze	Ulrike Henny-Krahmer
Frederike Neuber	Christopher Pollin
Malte Rehbein	Torsten Roeder
Patrick Sahle	Torsten Schaßan
Gerlinde Schneider	Markus Schnöpf
Martina Scholger	Philipp Steinkrüger
Nadine Sutor	Georg Vogeler

**Band 16**

Schriften des Instituts für Dokumentologie und Editorik – Band 16

# **Digitale Edition in Österreich Digital Scholarly Edition in Austria**

---

herausgegeben von | edited by

Roman Bleier, Helmut W. Klug

2023

BoD, Norderstedt

**Bibliografische Information der Deutschen Nationalbibliothek:**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

**Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 29. April 2023.**

2023

Herstellung und Verlag: Books on Demand GmbH, Norderstedt

ISBN: 978-3-743-102-842

Einbandgestaltung: Stefan Dumont; Coverbild: wurde von Roman Bleier und Helmut Klug für ein KONDE-Poster (DHa 2017) erstellt

Satz: Roman Bleier und LuaTeX

# Contents

Preface . . . . .	I
-------------------	---

## Einleitung – Introduction

Georg Vogeler Einleitung: Gibt es eine österreichische Editionskultur? . . . . .	V
---	---

## Methodische Aufsätze – Methodological essays

Tara L. Andrews Where are the Tools? The Landscape of Semi-Automated Text Edition . . . . .	3
Peter Hinkelmanns Editionen und Graphentechnologie: Vorteile und Hürden digitaler Editionstechniken abseits von TEI-XML . . . . .	19
Claudia Resch Digitale Editionen aus der Perspektive von Expert:innen und User:innen – Rezensionen der Zeitschrift RIDE im Meta-Review . . . . .	37
Markus Ender, Joseph Wang Experience with a Workflow using MS Word and a DOCX to TEI Converter	55
Bernhard Oberreither A Linked Data Vocabulary for Intertextuality in Literary Studies, with some Considerations Regarding Digital Editions . . . . .	69
Daniel Schopper, Thomas Wallnig, Victor Wang “Don’t Worry, We’re also Doing a Book!” — A Hybrid Edition of the Correspondence of Bernhard and Hieronymus Pez OSB . . . . .	89
Sarah Lang Digital Scholarly Editions of Alchemical Texts as Tools for Interpretation . . . . .	111
Martina Bürgermeister Versionierung von digitalen Editionen in der Praxis . . . . .	133

## **Projektberichte – Project reports**

Tara L. Andrews Die Chronik des Matthäus von Edessa (Matt'ēos Ur̄hayec'i) Online . . . . .	153
Roman Bleier, Eva Ortlib, Florian Zeilinger Der Regensburger Reichstag 1576 – digital . . . . .	155
Astrid Böhm, Julia Eibinger, Helmut W. Klug, Christian Steiner CoReMA – Cooking Recipes of the Middle Ages . . . . .	159
Artur R. Boelderl MUSIL ONLINE – interdiskursiver Kommentar . . . . .	161
Ingo Börner, Vanessa Hannesschläger, Isabel Langkabel, Katharina Prager Intertextuality in the Legal Papers of Karl Kraus: A Scholarly Digital Edition	163
Ulrike Czeitschner travel!digital . . . . .	166
Ursula Doleschal, Lisa Rieger Zweisprachiger Spracherwerb: Longitudinalstudie anhand schriftlicher Texte der Hermagoras-Volksschule – Digitalisierung, Auszeichnung, Auswertung	168
Claudia Dürr, Wolfgang Straub Kommentierte Werkausgabe Werner Kofler (Prosa, Lyrik, Radio, Film, Theater)	171
Markus Ender Kommentierte Online-Edition des Gesamtbrieftauswchsels Ludwig von Ficker	173
Konstanze Fliedl, Ingo Börner, Anna Lindner, Marina Rauchenbacher, Isabella Schwentner Arthur Schnitzler – Kritische Edition (Frühwerk) III . . . . .	175
Desiree Hebenstreit, Laura Tezarek, Christiane Fritze, Christoph Steindl Andreas Okopenko: Tagebücher aus dem Nachlass (Hybridedition) . . . . .	177
Peter Hinkelmanns, Katharina Zeppezauer-Wachauer Mittelhochdeutsche Begriffsdatenbank (MHDBDB) . . . . .	179
Mario Klarer, Aaron Tratter, Hubert Alisade Ambrascher Heldenbuch: Transcription and Scientific Dataset . . . . .	182

Carina Koch, Lisa Brunner, Anna Huemer, Christoph Würflinger Digitale Edition und Analyse der Medialität diplomatischer Kommunikation: Kaiserliche Gesandte in Konstantinopel in der Mitte des 17. Jahrhunderts . . . . .	184
Philipp Koncar, Roman Bleier InCritApp – Interactive Critical Apparatus . . . . .	187
Sarah Lang, Ursula Gärtner Grazer Repatorium antiker Fabeln (GRaF, 2017–2019) . . . . .	190
Verena Lorber, Joseph Wang-Kathrein Franz und Franziska Jägerstätter Edition . . . . .	193
Oliver Matuschek, Christopher Pollin, Lina Maria Zangerl Stefan Zweig digital . . . . .	196
Frederike Neuber Stefan George Digital . . . . .	198
Helmut Neundlinger, Selina Galka Karl Wiesinger: Digitale Edition der Tagebücher (1961–1973) . . . . .	200
Werner Petermandl, Elisabeth Steiner Celtic Divine Names in the Inscriptions of the Roman Province Germania Inferior . . . . .	203
Claudia Resch ABaC:us – Austrian Baroque Corpus . . . . .	205
Claudia Resch, Nora Fischer, Dario Kampkaspar, Daniel Schopper DIGITARIUM – Das Wien[n]erische Diarium digital . . . . .	209
Sabine Seelbach Virtuelle Benediktinerbibliothek Millstatt . . . . .	213
Thomas Wallnig Die gelehrte Korrespondenz der Brüder Pez (Hybridedition) . . . . .	217
<b>Appendix</b>	
Biographical Notes . . . . .	221

Publications of the Institute for Documentology and Scholarly Editing /  
Schriftenreihe des Instituts für Dokumentologie und Editorik . . . . . 225

## Preface

The present volume, *Digital Scholarly Edition in Austria*, is a collection of essays that originated in the context of the federally funded project “KONDE - Kompetenznetzwerk Digitale Edition” that was conducted at the University of Graz in the years 2016–2020 and included partners from 8 Austrian organizations ranging from literary archives to universities. The overall focus of the project was on digital scholarly editing, but the goals of the project were manifold originating from the partner organisations diverse backgrounds. One major goal was the development of best-practice solutions for publication platforms where digital scholarly editions are closely integrated into repositories for long-term preservation. Another was the development of a strategic concept for the bundling of competences to establish a national infrastructure for digital scholarly editions which meets the constantly changing requirements of modern scholarly editing and research. This includes the development and maintenance of tools for the processing of digital material and their systematic evaluation. One major output of the project was the *Weißbuch Digitale Edition*, which can be seen as an introductory tool for scholars and students who want to venture into the world of digital scholarly editing.

With the present volume the editors provided a space where researchers and editors from Austrian institutions could theorize on their work and present their editing projects. The collection creates a snapshot of the interests and main research areas regarding digital scholarly editing in Austria at the time of the project.

We are grateful to our colleagues in the scholarly network KONDE. Their interests and expertise provide the context of this publication. Special thanks are due to the staff of the Centre for Information Modelling at the University of Graz and the members of the IDE who supported the publication and, in many cases, contributed to it as peer-reviewers or authors. Furthermore we want to thank all authors and peer-reviewers for their professional cooperation during the publication process; it is solely the editors who are to blame for any delay in the publication process. We also want to thank the many people involved in creating the present volume: Karin Kranich and Nicholas Martin for language corrections and formal suggestions, Bernhard Assmann and Patrick Sahle for support and advises during the typesetting and publication process, Stefan Dumont for creating the cover, Elisabeth Raunig for verifying and archiving this volume’s referenced URLs, if possible, in December 2022 on the Internet Archive Wayback Machine ([archive.org](https://archive.org)).

Graz, January 2023, the editors



# **Einleitung**

---

# **Introduction**



# **Einleitung: Gibt es eine österreichische Editionskultur?**

Georg Vogeler

Die internationale Forschung zum Thema „digitale Edition“ ist aktiv. Erst jüngst hat Christopher Ohge (2021) einen Band vorgelegt, der das Thema als Problem der Publikationstechnologien zu beschreiben versucht. Andreas Oberhoff (2021) macht Vorschläge, wie das spezifische Verhältnis zwischen Veränderung und Referenzierbarkeit, das einer digitalen Edition inhärent ist, technologisch realisiert werden könnte. Eine Schweizer Bestandsaufnahme ist von der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (2021) vorgelegt worden. Die Rolle von XML/TEI als Standard digitalen Edierens ist durch die verschiedene Beiträge zur Einbindung von digitalen Editionen in das Semantic Web und die Verwendung von Graphentechnologien als Werkzeuge digitaler Editionen (Spadini, Tomasi und Vogeler 2021) herausgefördert worden. Verschiedene technische Lösungen verbreiten sich als standardisierte Werkzeuge zur Verbreitung von digitalen Editionen (z.B. *EVT*<sup>1</sup> oder *TEI-Publisher*<sup>2</sup>). Automatische Transkriptionsverfahren (HTR) dringen zunehmend in die Planung und Durchführung digitaler Editionen ein, ohne dass sie methodisch schon abschließend positioniert worden sind (z.B. Beloborova, Dillen und Schäuble 2018, 9). Die Forschung diskutiert das Thema immer auch in Auseinandersetzung mit den editionswissenschaftlichen Debatten der Philologien, wofür die Einführung in die digitale Edition von Pierazzo und Mancinelli (2020) ein gutes Beispiel ist, während andere Wissenschaften, wie z.B. die Geschichtswissenschaften das Problem erst jüngst aufgreifen (Vogeler, Pollin und Bleier 2022). Pierazzo und Mancinelli diskutierten explizit die Frage einer nationalen Editionskultur im Kontrast zur internationalen Forschung zur digitalen Edition (17). Gibt es also eine nationale österreichische Editionskultur?

Dieser Band versammelt Forschungsergebnisse aus einem vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung finanzierten Projekt. Das „Kompetenznetzwerk ‚Digitale Edition‘“ (KONDE) hatte sich zum Ziel gesetzt, die österreichischen Akteure im Feld digitaler Edition so mit einander zu vernetzen, dass Synergien sichtbar werden, die Partner sich im Austausch gegenseitig befürchten sollten und nationale Infrastrukturen entstehen. Der vorliegende Band ist ein Ergebnis des Projektes. Weitere sind unter <https://digitale-edition.at> zugänglich, wozu insbesondere das „Weißbuch ‚Digitale Edition‘“ (Klug 2021) zu zählen ist, das in 219 für die digitale Edition relevante Begriffe einführt, sie untereinander vernetzt und in

---

<sup>1</sup> <http://evt.labcd.unipi.it>.

<sup>2</sup> <https://teipublisher.com>.

25 Projektbeschreibungen exemplifiziert. Der hier vorgelegte Band versammelt nun in einem methodischen Teil wissenschaftliche Beiträge von Kolleginnen und Kollegen aus den Partnerprojekten und überführt die Projektbeschreibungen der digitalen Ressource in die Gutenberg-Galaxis.

Im methodischen Teil werden grundsätzliche Fragen verhandelt. Ich selbst hatte 2017 einen Überblick zu geben versucht, aus welchen Komponenten digitales Editieren besteht und wie weit die technischen Entwicklungen in den jeweiligen Feldern sind (Vogeler 2019). Ich hatte das 2017 in einem Vortrag in folgende Übersicht zu bringen versucht (Abbildung 1):

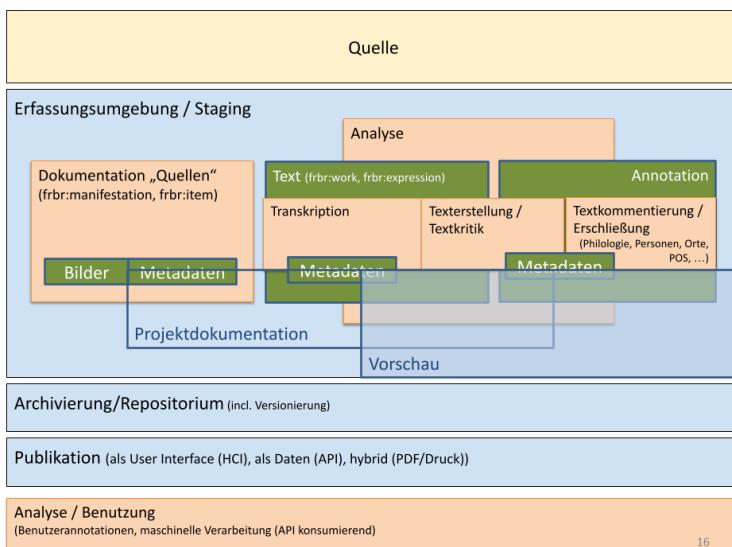


Abbildung 1: Schematische Darstellung von Komponenten einer digitalen Edition.

Digitale Edition wird darin mit Patrick Sahle als „Erschließende Wiedergabe historischer Dokumente unter einem digitalen Paradigma“ verstanden (Sahle 2016). Die historischen Dokumente sind die Quelle der digitalen Edition, die erarbeitet, archiviert und publiziert wird, um dann für weitere Analyse und Benutzung zur Verfügung zu stehen. Die blauen Kästen des Diagramms stehen also für technische Systeme, die bestimmte Stadien in der Erstellung der digitalen Edition abbilden. Die dabei entstandenen Daten werden von einer dieser Komponenten in die nächste übertragen, wenn sie nicht in integrierten Systemen zusammenfallen. Die orangen Blöcke bezeichnen Aktivitäten, in denen Daten (grün) entstehen oder verändert werden: die digitale

Dokumentation der Quellen in Bilddigitalisaten und Metadaten zu den Objekten, die Ergebnisse der Transkription und Textkritik in den Textdaten mit einschlägigen Metadaten die Annotation und Ergebnisse der Textkritik in Annotationsdaten und einschlägigen Metadaten. In der Erfassungsumgebung wird die eigentliche Editionsarbeit dokumentiert und in einer „Staging“-Umgebung werden erste Ansichten und Benutzerinterfaces erzeugt, die die Editionsarbeit unterstützen sollen. Diese müssen nicht identisch sein mit der Form, in der die digitale Edition für die Öffentlichkeit zugänglich gemacht wird, wenn die Vorschau z.B. organisatorische Funktionalitäten wie Statusnotizen oder Validierungshinweise enthält. Das Diagramm ließe sich an vielen Editionsprojekten und technischen Lösungen exemplifizieren: *ediarum*<sup>3</sup> unterstützt z.B. die Transkription und Annotation und enthält eine Vorschau, die auch als Publikationsumgebung genutzt werden kann. Das integrierte System *GAMS* (Stigler und Steiner 2018)<sup>4</sup> verbindet Publikation und Archivierung. Workflow-Management-Systeme wie *Kitodo* legen ihren Schwerpunkt auf die Dokumentation der Quelle. Publikationsframeworks wie das *Edition Visualisation Toolkit* (EVT) (Rosselli del Turco und di Pietro 2019) oder der *TEI-Publisher* konzentrieren sich auf die Unterstützung der Benutzung durch flexible und mächtige graphische Interfaces. Die Beiträge dieses Bandes diskutieren nun verschiedene Aspekte des Diagramms.

Die Komponenten des Schemas realisieren sich in verschiedenen technischen Werkzeugen, die kontinuierlich in Entwicklung sind, aber an sich das gesamte Feld abzudecken scheinen (Vogeler 2019).<sup>5</sup> TARA ANDREWS greift diesen praktischen Zugang zur digitalen Edition in ihrem Beitrag auf und versucht auf dem aktuellen Stand Orientierung zu geben, welche in Österreich entwickelten Werkzeuge in einem typischen Prozess editorischen Arbeitens zur Anwendung kommen könnten und damit die Planung einer digitalen Edition zu unterstützen. Sie konzentriert sich auf in Österreich entwickelte Tools, die gute Beispiele für allgemeine Probleme sind, insbesondere die Abhängigkeit der Werkzeugwahl von den in der jeweiligen Edition angewendeten Konzepten. In Österreich sind Werkzeuge wie der *Classical Text Editor*, *Transkribus*, *Recogito* oder die *DSE-Baseapp* entstanden, die bei unterschiedlichen Aufgaben die Editorinnen und Editoren unterstützen. Sie bringen aber alle Eigenschaften mit sich, die zu Entscheidungen zwingen. Diese Entscheidungen liegen bei den Editorinnen und Editoren und ergeben sich aus den Spezifika der jeweiligen Edition.

In den letzten Jahren werden zunehmend Verfahren gestestet, Graphentechnologien in digitalen Editionen einzusetzen. Erst jüngst haben Elena Spadini, Francesca Tomasi und Georg Vogeler (2021) eine Sammlung von Beiträgen zum Thema vorgelegt. PETER

<sup>3</sup> <https://www.ediarum.org>.

<sup>4</sup> <https://gams.uni-graz.at/doku>.

<sup>5</sup> Vgl. auch die Tools-Rubrik im KONDE-Weißbuch; oder die Übersicht über Tools zur Textbearbeitung in der Lehrressource *forText* (Literatur digital erforschen, Hamburg und Darmstadt 2016-, <https://fortext.net>).

HINKELMANNS gibt einen Überblick über diese Technologien im Feld digitaler Editionen, der einen ersten Leitfaden für ihren Einsatz bietet. Graphentechnologien sind insbesondere für überlappende Strukturen, Variantengraphen und semantische Annotationen im Bereich der digitalen Edition im Einsatz und können, so Hinkelmanns, in Kombination mit XML/TEI eine gute Wahl für digitale Editionen sein.

Aus Benutzersicht nähert sich CLAUDIA RESCH digitalen Editionen über Rezensionen, die sie korpuslinguistisch auswertet. Mit diesem Verfahren kann sie in den 35 von ihr untersuchten *Texten im Review Journal of the IDE* (RIDE) Muster von Unzufriedenheit und Verbesserungsvorschlägen ausmachen. Sechs Kritikpunkte werden immer wieder aufgegriffen: mangelnde theoretische Einordnung, zu wenig Transparenz der editorischen Verfahren, Schwächen im Interface, unzureichende Suchfunktionen, zu geringe Auswahl an Darstellungsformen und fehlender Zugang zu den Rohdaten. Claudia Resch schlägt vor, die Nutzerinnen und Nutzer frühzeitig in die Entscheidungen über die Präsentation der Edition einzubinden, weist aber auch auf einen langsamem Bewusstseinswandel unter den Editorinnen und Editoren hin in ihrer Bereitschaft, Quelldaten zugänglich zu machen. Dass hier schon die Rückwirkung aus den sechs Jahren RIDE-Rezensionen vorliegt, ist für sie anzunehmen.

In einem österreichischen Projekt ist als Editionsumgebung eine Pipeline von XSLT-Transformationen entstanden, mit der JOSEPH WANG und MARKUS ENDER XML/TEI aus DOCX-Dokumenten extrahieren: „DOCX2TEI“. Die Methode weicht vom in der TEI-Community verbreiteten Verfahren in Oxgarage<sup>6</sup> insofern ab, als sie projektspezifische Textstrukturen und semantische Annotationen in Form von Word-Kommentaren verarbeiten kann. Mit DOCX2TEI machen sie damit MS Word zu einem projektspezifischen XML-Editor, der eine Brücke zwischen fachlichen Projektmitarbeiterinnen und Projektmitarbeitern und technischer Infrastruktur einer digitalen Edition baut.

Der Umstand, dass digitale Editionspraxis von den verwendeten Datenmodellen abhängig ist, verleiht dem Vorschlag für eine formale Beschreibung von intertextuellen Bezügen von BERNHARD OBERREITHER besondere Bedeutung. Intertextualität ist der Kern der Kommentierungsleistung, die „Quellen“ eines Textes zu identifizieren sucht. INTRO – *the Intertextual Relationships Ontology* versucht, eine pragmatische Balance zwischen einfachen Grundkonzepten und komplexen Abstraktionen zu finden – und dabei insbesondere dem Umstand Rechnung zu tragen, dass die Feststellung intertextueller Bezüge eine Interpretation der Editorin und des Editors ist.

Die Definition digitaler Edition von Patrick Sahle (2013, 2016), grenzt Editionen im digitalen Paradigma unter anderem dadurch ab, dass sie Eigenschaften besitzen, die sich nicht im Druck realisieren lassen würden. DANIEL SCHOPPER, THOMAS WALLNIG und VICTOR WANG machen einen Vorschlag, wie man beide Medienwelten in hybriden Editionen verbinden kann. Am Beispiel der Kooperation zwischen der Edition der

---

<sup>6</sup> <https://github.com/TEIC/oxgarage>.

Korrespondenz der Gebrüder Pez und dem Wiener Böhlau Verlag können sie zeigen, dass eine gedruckte Edition als Derivat einer digitalen Edition durchaus sinnvoll ist. Sie gehen dabei von Benutzungsszenarien aus, in denen das physische Objekt „Buch“ dem digitalen Medium überlegen ist und schlagen ein Szenario vor, in dem diese Eigenschaften in eine komplexe Infrastruktur von digitalem Archiv und graphischer Benutzerschnittstelle integriert sind. Für die Nachnutzbarkeit der Infrastruktur legt sich die Pez-Edition auf ein Datenmodell fest und leitet eine Strategie für die Normalisierung zukünftiger Datensätze daraus ab. So verteilen sich die Aufgaben der Publikation einer Edition auf Verlag und Forschungsdatenrepositorium.

Digitale Editionen können Teil einer automatischen Verarbeitung von Text sein. SARAH LANG diskutiert, mit welchen Textmining-Methoden die Texte des Hieroalchemisten Michael Maier untersucht werden könnten. Sie müssen dabei die Texte semantisch anreichern, da reine Bag-of-Words-Methoden nicht ausreichen, die Verschleierungstaktiken der Alchemisten zu durchschauen. Mit der Verbindung zu formalen Wissensressourcen können die sogenannten „Decknamen“ kontextualisiert und damit verständlich gemacht werden. Es wird damit noch einmal deutlich, wie wichtig die kritische Erschließung aus der oben zitierte Definition von Patrick Sahle für das Verständnis digitalen Edierens ist.

MARTINA BÜRGERMEISTER diskutiert die Folgen der Veränderbarkeit digitaler Edition. Ihre Auswertung von existierenden Praktiken von Versionierung kann vier Strategien identifizieren, die auch in einander greifen können: Zusammenfassende Dokumentation in einer Änderungsbeschreibung, Dokumentation einzelner Änderungsschritte in *Revision Descriptions* der einzelnen Dokumente, technische Versionskontrolle bei der Datenerzeugung und Versionierung des Gesamtsystems. Die von ihr beobachteten Strategien werden aber nur in einem Viertel der untersuchten digitalen Editionen überhaupt angewendet. Das demonstriert die Reichweite von Martina Bürgermeisters Ergebnissen: So sehr die Dokumentation von Änderungen die Nachvollziehbarkeit von Forschungsarbeit und die Zitierbarkeit von digitalen Editionen unterstützt und damit das Vertrauen in digitale Editionsformen stärkt, so sehr ist auch notwendig, Praktiken der Versionierung überhaupt erst zu etablieren.

Die Landschaft digitalen Edierens in Österreich bilden die Kurzbeschreibungen von 25 Projekten ab, die sich im zweiten Teil dieses Bandes versammeln. Dazu gehören Editionen „österreichischer“ Autorinnen und Autoren wie Arthur Schnitzler (FLIEDL, BÖRNER, LINDNER, RAUCHENBACHER und SCHWENTNER), Karl Kraus (BÖRNER, HANNESSCHLÄGER, LANGKABEL, PRAGER), Robert Musil (BOELDERL), Werner Kofler (DÜRR und STRAUB), Ludwig von Ficker (ENDER), Karl Wiesinger (NEUNDLINGER und GALKA), Andreas Okopenko (HEBENSTRIET, TEZAREK und FRITZE), Franz und Franziska Jägerstätter (LOBER und WANG-KATHREIN), oder des *Ambraser Heldenbuchs* (KLARER, TRATTER und ALISADE). Die Gebrüder Pez (WALLNIG) sind Österreicher, aber zeigen in ihrer internationalen Vernetzung ebenso wie die Korrespondenz Habsburgischer

Diplomaten in Konstantinopel (KOCH, BRUNNER, HUEMER und WÜRFLINGER), dass das Konzept „österreichische Autoren“ nicht sehr weit reicht, um eine österreichische Editionslandschaft zu beschreiben. In Österreich werden nämlich auch die Chronik des Matthäus von Edessa (Andrews), die Akten des Reichstags von 1576 (BLEIER, ORTLIEB und ZEILINGER), mittelalterlichen Kochrezepte (BÖHM, EIBINGER und KLUG), Stefan George (NEUBER) oder Inschriften der römischen Provinz Germania Inferior (PETERMANDL und STEINER) ediert, deren Texte kaum noch expliziten Österreich-bezug aufweisen. Schließlich finden sich unter den Projekten auch Aktivitäten, die am Rande dessen bewegen, was man mit der Edition von Patrick Sahle als digitale Edition fassen kann: Linguistisch annotierte Korpora wie das Austrian Baroque Corpus ABACUS (RESCH) oder das Korpus der Baedeker-Reiseführer (CZEITSCHNER) sind literarische und kulturhistorische Quellen ersten Ranges. Das gilt auch für retrodigitalisierte Textsammlungen wie das DIGITARIUM, die digitale Repräsentation der ältesten Ausgaben der Wiener Zeitung (RESCH, FISCHER und KAMPKASPAR). In der Mittelhochdeutschen Begriffsdatenbank werden die Texte als Belege für lexikographische Arbeit zusammengestellt (HINKELMANN und ZEPPEZAUER-WACHER), im Grazer Repository antiker Fabeln die für Unterrichtszwecke kommentiert (LANG und GÄRTNER). Als Vorbereitung für linguistische Arbeit dient das Korpus schriftlicher Texte der Hermagoras-Volksschule (DOLESCHAL und RIEGER). Am Rande digitalen Edierens steht auch die Dokumentation und Digitalisierung von Originalen wie in der virtuellen Benediktinerbibliothek Millstatt (SEELBACH) oder in der Nachlassdokumentation „Stefan Zweig digital“ (MATUSCHEK, POLLIN und ZANGERL). Das Werkzeug zur Visualisierung eines textkritischen Apparats (KONCAR und BLEIER) ist selbst keine digitale Edition, kann aber gut ein Teil einer digitalen Edition werden. All diese Projekte bilden Teile dessen ab, was ich oben in der Graphik der Komponenten digitalen Edierens zusammenzufassen versucht habe, auch wenn sie nicht immer das zum Ziel haben, was Sahles Definition einer digitalen Edition beschreibt. Die Abgrenzung zwischen „Digitalisierung“ und „Digitaler Edition“, wie sie von Kenneth Price (2009) und Patrick Sahle (2007) und jüngst von Woud Dillen (2019) diskutiert worden ist, bleibt weiterhin schwierig und hat z.B. im Rezensionorgan *RIDE*, das sich auf kritische Vorstellung von digitalen Editionen spezialisiert hat,<sup>7</sup> dazu geführt, dass in den Jahren 2017 und 2018 „Textsammlungen“ anhand eines speziellen Kriterienkatalog rezensiert wurden,<sup>8</sup> der sich an den vom IDE für digitale Editionen entwickelten anlehnt (Henny-Krahmer und Neuber 2017). Die Projekte zeigen aber auch, dass sich Abgrenzungsprobleme insbesondere daraus ergeben, dass das Bedürfnis nach einer wissenschaftlich verlässlichen digitalen Textrepräsentation auch von Forschungsinteressen geleitet sein kann, in denen Texte anders aufgefasst werden als in der klassischen Editionsphilologie.

<sup>7</sup> <https://ride.i-d-e.de>.

<sup>8</sup> <https://ride.i-d-e.de/issues/issue-6/>, <https://ride.i-d-e.de/issues/issue-8> und <https://ride.i-d-e.de/issues/issue-9>.

Gibt es nun eine österreichische Editionskultur? Die in Österreich diskutierten methodischen Überlegungen, die im Laufe des Projektes angewendeten Techniken und Methoden, das Weißbuch und die Code-Beispiele im Projektrepository<sup>9</sup> sind keineswegs auf Österreich beschränkt. Die im Laufe des Projektes entstandenen digitalen Editionen und editionsähnlichen Datensätze haben thematisch einen leichten „nationalen“ Bezug, denn Editionen von österreichischen Autorinnen und Autoren sowie Textüberlieferung in Österreich haben selbstverständlich eine nationale Komponente. Diese Komponente produziert aber noch keinen nationalen Diskurs, denn die digitalen Verfahren sind Beiträge zur internationalen Forschung. „Nationale“ Wissenschaft ist aber immer mehr ein soziales als ein epistemologisches Phänomen. Es kann sich also gut noch aus der Zusammenarbeit im Projekt eine nationale Editionskultur entwickeln, die wie im Projekt selbst in fruchtbarem Austausch mit den Kolleginnen und Kollegen außerhalb Österreichs steht. Die hier versammelten Beiträge bilden einen guten Startpunkt, darüber nachzudenken, auf was sich eine solche österreichische Kultur digitalen Edierens konzentrieren könnte: Workflows, Infrastrukturen, Werkzeuge, Methoden, Forschungsinteressen, Abgrenzungen digitalen Edierens stehen zur Debatte.

## Literatur

- Beloborodova, Olga, Wout Dillen und Joshua Schäuble. 2018. „CATCH 2020 project, manuscript genetics, digital scholarly editing, Samuel Beckett (1906–1989)“. Folien zur Präsentation auf der Transkribus User Conference, 2018. Zugriff: 12. Dezember 2021. <https://readcoop.eu/wp-content/uploads/2018/11/BELOBORODOVA-DILLEN-SCHAUBLE.pdf>.
- Dillen, Wout. 2019. „On Edited Archives and Archived Editions.“ *International Journal of Digital Humanities* 1 (2): 263–77. doi:10.1007/s42803-019-00018-4.
- Klug, Helmut W., Hrsg. 2021. *KONDE Weißbuch*. GAMS. 562.50. Graz: Zentrum für Informationsmodellierung. <https://hdl.handle.net/11471/562.50>.
- Mancinelli, Tiziana und Elena Pierazzo. 2020. *Che cos’è un’edizione scientifica digitale*. Roma: Carocci.
- Oberhoff, Andreas. 2021. *Digitale Editionen im Spannungsfeld des Medienwechsels: Analysen und Lösungsstrategien aus Sicht der Informatik*. Bielefeld: transcript. doi:10.14361/9783839459058.
- Ohge, Christopher. 2021. *Publishing Scholarly Editions: Archives, Computing, and Experience*. Cambridge: University Press. doi:10.1017/9781108766739.
- Price, Kenneth M. 2009. „Edition, Project, Database, Archive, Thematic Research Collection: What’s in a Name?“ *Digital Humanities Quarterly* 3 (3). Zugriff: 12. Dezember 2021. <http://digitalhumanities.org/dhq/vol/3/3/000053/000053.html>.
- Rosselli del Turco, Roberto. 2019. „La visualizzazione di edizioni digitali con EVT: una soluzione per edizioni diplomatiche e critiche“. *Ecdotica* 16: 148–73. doi:10.7385/99301.

<sup>9</sup> <https://github.com/KONDE-AT>.

- Sahle, Patrick. „What Is a Scholarly Digital Edition?“ 2016. In *Digital Scholarly Editing: Theories and Practices*, hg. v. Matthew James Driscoll und Elena Pierazzo, 19–40. Cambridge: Open Book Publishers. doi:10.11647/OBP.0095.02.
- . 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. Norderstedt: Books on Demand.
- . 2007. „Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung“. In *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, hg. v. Michael Stolz et. al., 64–84. Zürich: germanistik.ch.
- Schweizerische Akademie der Geistes- und Sozialwissenschaften, Hrsg. 2021. „Edieren: Geisteswissenschaften im digitalen Wandel | Éditer: les sciences humaines en mutation“. *Bulletin der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (SAGW-Bulletin)* 27 (3). Bern: SAGW. doi:10.5281/zenodo.5716099.
- Spadini, Elena, Francesca Tomasi und Georg Vogeler, Hrsg. 2021. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*. Norderstedt: Books on Demand.
- Stigler, Johannes Hubert und Elisabeth Steiner. 2018. „GAMS – An Infrastructure for the Long-Term Preservation and Publication of Research Data from the Humanities“. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (1): 207–16. doi:10.31263/voebm.v71i1.1992.
- Vogeler, Georg, Christopher Pollin und Roman Bleier. 2022. „Ich glaube, Fakt ist...: der geschichtswissenschaftliche Zugang zum digitalen Edieren“. In *Digital History. Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*, hg. v. Karoline Dominika Döring, Stefan Haas, Mareike König und Jörg Wetzlaufer. Berlin, Boston: De Gruyter.
- . 2019. „Digitale Editionspraxis. Vom pluralistischen Textbegriff zur pluralistischen Softwarelösung“. In *Textgenese in der digitalen Edition*, hg. v. Anke Bosse und Walter Fanta, 117–36. Berlin, Boston: De Gruyter. doi:10.1515/9783110575996-008.

## **Methodische Aufsätze**

---

**Methodological essays**



# **Where are the Tools? The Landscape of Semi-Automated Text Edition**

Tara L. Andrews

## **Abstract**

The aim of this article is to answer the question: given that there have been so many tools and methods developed to help prepare scholarly critical editions of texts, why do so many scholars have trouble knowing where to start? The article walks the reader through the typical process of creating an edition, mentioning along the way a variety of tools that have been developed or used in the Austrian landscape in particular, and aims thereby to illustrate many of the considerations that the scholar setting out on an edition project must account for.

## **Zusammenfassung**

Ziel dieses Artikels ist es, eine Antwort auf die Frage anzubieten: Wenn so viele Werkzeuge und Methoden entwickelt worden sind, um kritische Editionen von Texten vorzubereiten, warum finden es dann so viele WissenschaftlerInnen schwierig zu wissen, wo sie anfangen sollen? Der Artikel führt den Leser durch den typischen Prozess der Erstellung einer Edition, erwähnt dabei eine Vielzahl von Werkzeugen, die insbesondere in der österreichischen Landschaft entwickelt wurden oder verwendet werden, und versucht damit viele der Überlegungen sichtbar zu machen, die zu Beginn eines Editionsprojekts berücksichtigt werden müssen.

As more or less any professor, teaching fellow, or research assistant in the field of Digital Humanities can attest, there is a great deal of interest from scholars in the literary and historical fields about digital editions of texts and how they might be feasibly done. Yet many of these scholars have little idea where to start or what tools exist that are relevant for the particular work they wish to do, despite the fact that textual criticism has been on an increasingly digital trajectory since well before the availability of the World Wide Web. After so many years of development of the digital edition, this seems like a rather odd state of affairs—there are, after all, a plethora of tools available for use (Klug, Galka and Steiner 2021; see also the discussion of several of these tools and methods in Vogeler 2019). Why, then, can it be so difficult to advise scholars about a way forward?

The purpose of this article is to walk the reader through the process of creating a digital edition, from initial transcription to publication and including various methods

of source analysis. Along the way we will cover a range of tools that have emerged, especially in Austria, to assist with the creation of digital editions. We must, however, stress the following. Although many textual scholars hope for a single full-featured software package designed to take their editions all the way from conception to publication without the need either to learn their way around computer programming or to hire someone who does, this hope is sadly misplaced. There are almost as many possible forms of digital edition as there are texts to be edited. Every editor will have a different set of priorities for her edition, not to mention a text (or a corpus) that differs from other texts in ways that are perhaps small, but certainly crucial, for the purpose of preparing that edition. While the argument has been made elsewhere that there can be no “monolithic” general-purpose tool (van Zundert and Boot 2011), we hope with this walk-through of digital editing processes to illuminate why this is the case.

Given the high degree of specialization that any edition project must reach, a scholar who wishes to produce a digital edition should expect to exercise a significant amount of control over the process. As such, the scholar will need to know the principles, and the limitations, of the data modelling system that is used to render texts into the digital medium, and will need to understand how and where the choices made for her particular project might differ from the assumptions built into the tools that are available for analysis and publication of the result. In many cases, in fact, different tools take a different set of assumptions as their starting point, and so the editor will eventually need to understand the data models and their associated technologies well enough to assess how—or indeed whether—these differences can be bridged. We contend here that, in order to create a digital edition on time and within budget, a scholar cannot hope to rely entirely on “IT experts” hired for the purpose. She will need to gain enough knowledge, not only about how text encoding is done, but also about what is done with the result of that encoding and the parameters of the technologies that are used to do it, to be able to make informed decisions.

## 1 From scholarly work to digital model

Texts can be published into the digital medium by a variety of means. The basic requirement for any online publication is that it must be expressed as one or more documents rendered in HTML—the standard format for web documents—and hosted on a server connected to the Internet, under a publicly reachable URL. In order for this publication to be a critical edition, it is only necessary that those documents, in one way or another, contain a faithful representation of the critical text and any apparatuses or other commentary that the editor felt necessary to include.

There are many possible pathways to this end state. For example, it is easy to envision the preparation of the edition in a word processor, where the document is then saved into HTML format and given to a hosting provider. That would result in an online publication that may be “digital enough” for many purposes, but could not be considered “a digital edition” in the sense proposed by Sahle (2016).

A middle ground between the print and the digital can be found with software such as the *Classical Text Editor* (CTE), developed at the Austrian Academy of Sciences (ÖAW) by Stefan Hagel (1997-). This is a package intended specifically for the creation of critical editions from multiple witness copies. Its user interface is intended to hew as closely as possible to the familiar interface of a word processor, while extending the functionality to provide for the things that an editor will need, such as the definition of sigla to correspond to witnesses, the possibility to record variant readings to the edited text based on those witnesses, the possibility to add other sorts of scholarly apparatus according to the conventions of classical philology.

Although the primary output of CTE is a print-ready document suitable for submission to a book publisher, it also offers the option of export to a TEI-XML format, which could then be transformed to HTML and published online (see below). This feature, added by CTE’s author in the hope of encouraging the proliferation of companion digital publications by CTE users alongside the more usual print publications, has not had widespread use (Hagel 2007, 78). The author attributes this to a lack of institutional interest in digital publications; while this may still have been true in 2007, it is much less true today, and yet those who call themselves digital philologists still do not usually recommend CTE as a means to produce a digital edition. The reasons for this, we would argue, go deeper than institutional interest.

One of the major design decisions of CTE is to allow the scholar to create a text edition that conforms to her exacting scholarly specifications, but without troubling her with “any sort of surfacing tags or other sorts of ‘code’ [which] can be detrimental” to the ability of the editor to “remain devoted to scholarly questions” (Hagel 2007, 79). While this is an admirable goal, it may actually be part of the problem. CTE encourages the scholar to concentrate primarily on how the edition will look once it is printed. Although the software has a reasonably complex conceptual model, the scholar not inclined to study this model or to familiarise herself with the advanced features of CTE will quickly find that she can produce an edition that “looks right” on the page even when the model is violated internally.

For example, CTE provides a mechanism for recording additions, omissions, and transpositions, and for specifying the abbreviations that should appear in the *apparatus criticus* when one of these situations arises. The user can, on the other hand, achieve the same outward effect simply by inputting these abbreviations as though the abbreviation was itself the text of a variant. The distinction is invisible in the resulting print proof, but in the companion XML output, incorrect use of the data

model will instantly undermine any automated attempts to parse the document for the edition text and the variant witness texts. Several potential inconsistencies of this type were encountered in the attempt to write a parser for CTE’s flavour of TEI-XML, for use with the Stemmwaweb service (Andrews n.d.).

It is thus clear that, if a user of CTE wishes to produce a *useful* digital output alongside the print output of an edition, she must take care to understand not only the scholarly features of CTE, but also the data model that underlies what is displayed on the screen. She will soon find that there is a set of technical assumptions embedded in the software about how to represent the data that has been entered by the scholar, and more assumptions about how to translate the text from the CTE model into the TEI double-endpoint-attachment means of encoding variant text. She will then need to draw upon her understanding of the data model as it was expressed in the TEI output, in order to use the tool(s) that she eventually chooses for further analysis of the edited text or for its publication.

## 2 The typical digital edition

Although the XML export functionality of CTE offers a route to HTML-based online presentation, this might not necessarily be considered a true digital edition. Many scholars follow the definition of “digital edition” offered by Sahle (2016), who argues that “scholarly digital editions are scholarly editions that are guided by a digital paradigm in their theory, method and practice”—for Sahle this also means that a digital edition must offer some significant content and functionality that would be lost in an analogous print edition.

What does this mean in practice? Although Bordalejo (2018) argues, with some justification, that textual scholarship has not actually progressed beyond traditional paradigms on a methodological level, a more or less mainstream set of steps have emerged toward creating something that is commonly acknowledged as a “digital scholarly edition”. It begins with the transcription of sources, usually from digitized images, in which not only the sequence of interpreted text but also the features of the textual expression (such as decorated or otherwise highlighted text, authorial or scribal corrections, marginal notes, and the like) are recorded. The ability to pay close attention to these features of the text brings the scholar immediately beyond the capabilities of word processors, or even the CTE.

Thus, when we speak of a “typical digital edition”, we usually envision the transcription of source text from digitized images, the comparison (if necessary) and analysis (as desired) of these source texts, the production of a commentary usually including information that is linked to specific portions of the text or specific words therein, and the presentation of the whole in an online format, so that a reader (or

user) of the edition can view the text and its commentary in whatever form best suits the purpose that the editor had in doing the work.

Several institutions within Austria—the Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) at the Austrian Academy of Sciences, the Centre for Information Modelling—Austrian Centre for Digital Humanities (ZIM-ACDH) at the University of Graz, and the Digitalisierung und elektronische Archivierung (DEA) group at the University of Innsbruck foremost among them—have taken on the task of supporting Austrian researchers by providing training for scholars new to digital methods, offering archiving of the digital sources, producing systems for transcription of digitized images, and providing solutions for publication of the finished result on the Web. Edition projects and the expertise that goes with them are settled at many other Austrian universities, and these are often an early port of call for the scholar who would begin a new digital edition.

The plethora of services around digital editions in Austria brings us to a first important point: although there exist a great variety of tools for various steps in the process of creating a digital edition, there will almost certainly not be a complete suite of tools that will take the scholar from the beginning of her project all the way through to publication, without the intervention of a programmer hired for the purpose or a specialist consultant, unless the scholar is herself conversant enough with the various relevant technologies to provide her own technical support. We can illustrate this point by considering the steps toward the production of such a “typical” edition, reviewing what tools are most helpful for each of these steps, and observing what is still missing.

### 3 Transcription as an act of data modelling

The first stage of a digital edition is to express the content of each source document in some sort of digital model. This expression usually takes the form of *embedded markup*, in which semantic information about the text is entered directly alongside the text itself. Although there are a number of options for embedded markup, such as LMNL (Piez 2014) or a system currently in development known as TagML (Haentjens Dekker et al. 2018), by far the most predominant toolkit for the expression of this model is the Text Encoding Initiative (TEI). The first step for the vast majority of trainee digital philologists is to become familiar with the TEI Guidelines, and how they are expressed in XML. We must stop short of calling the TEI itself a model; although it defines a vast number of scholarly concepts related to text and its features, many of these concepts are intentionally flexible in their definitions (e.g. text block definitions such as `<ab>` and `<div>`), and there are often multiple possible ways to

express the same textual feature (e.g. lines of text, which can either be enclosed in a `<line>` element or separated by an `<lb/>` [line break] milestone element). This is the reason that the authors of the TEI Guidelines advise all editors to produce a custom schema, specific to the needs of the project itself, at its outset.

Since the Text Encoding Initiative has based its work almost entirely on the framework of XML and its related technologies, the first consequence for the newly-digital philologist is that, from the very outset of the transcription work, she must learn a great deal about XML. This includes not only the basics of its grammar, but also the subtleties of how that grammar is employed via the TEI Guidelines to create a document that would be considered “valid”. Here too, the editor must understand the technical mechanism by which validity is ensured: this involves the creation and configuration of a custom XML schema using a tool such as Roma (Mittelbach, Rahtz, and Bernevig 2018), for which the editor will need to understand the technical contents of the TEI modules that she wishes to use. If the editor wishes to extend the TEI to deal with features of her text that are not adequately covered in the Guidelines, she will need to have an even better understanding of the principles of schema description in order to add or modify the required elements, attributes, or dependencies.

In many cases, the editor will do the transcription using an XML editor (the most commonly used editor is *oXygen*, though there are open-source alternatives) configured to incorporate her custom schema for validation checking. The very fact that XML editors are the primary tool of choice for creating TEI-XML transcriptions of source texts constitutes strong evidence of the impossibility of providing the comprehensive and user-friendly software tools that scholars so often wish for. Industry programmers, as well as colleagues from the field of computer science, are (in our experience) almost invariably stunned to discover that digital philologists write XML directly in an editor—although XML is a text format that is comprehensible by humans, its syntax is exacting enough that software developers almost never write it directly if they can avoid it. Rather, they expect that data in XML format is generated by some sort of intermediate software, and only edited by hand *in extremis*. This manual process is, however, almost inevitable when the very schema against which the transcription is checked varies from project to project. It is not uncommon for user-friendly alternatives to be developed within the bounds of individual projects—both the *Transcribe Bentham* crowdsourcing project (Causer and Wallace 2012) and the *New Testament Virtual Manuscript Room* (Institut für neutestamentliche Textforschung n.d.) implement WYSIWYG-style transcription interfaces, for example—but none of these interfaces makes a claim to widespread utility.

We say it is “almost inevitable” that the editor will write XML by hand—for the scholar who has facsimile images of her sources and wishes to transcribe them line by line with a view to publication of both facsimile and transcription, there are tools available that handle transcription of the text, together with line-by-line linking of

transcriptions to facsimile images. One of the best-known tools for this, particularly in Austria, is *Transkribus* (Kahle et al. 2017). *Transkribus* is offered as a desktop application, backed by a central data storage and processing service, for transcription of manuscripts and printed texts directly from images. It offers a plethora of useful tools; these include automated detection of regions of text in an image and lines within those regions (which allows for the association of segments of transcribed text with their corresponding places on the facsimile image), handwritten text recognition (if enough of a particular document has been manually transcribed), and output of the transcription data into several formats, including TEI-XML.

Given the claim above that it is more or less impossible to write a graphical user interface for production of TEI-XML encoded texts, we should look more closely into what *Transkribus* does offer. Their data model, tuned as it is for recognition of text blocks on page facsimiles and association of text with these blocks, saves information using a standard known as PAGE XML (Pletschacher and Antonacopoulos 2010) that was developed for automated document analysis and text recognition. Any conversion to TEI must therefore involve a transformation of the model of a text as conceived by PAGE into some TEI-compatible model. Indeed, in their introductory How-To guide, the developers note that “*Transkribus* is [...] more than a TEI editor, but also less (we will not support all peculiarities of TEI but just those which are necessary to create a good, standardized transcription)” (“How to Use Transkribus—in 10 Steps (or Less)” 2015).

What this means in practice is that the commonly recommended option of creating a custom XML schema for each individual project is not an option here. The *Transkribus* developers have had to make certain decisions of their own about how the PAGE XML model can be expressed in terms of TEI-XML concepts, and the user has little choice but to learn and understand the respective models if she wishes to use the TEI-XML output of *Transkribus*. For instance, the PAGE XML concept of a TextRegion (that is, the area on the facsimile that comprises a text block, or a line inside a block) demands that the content of the line be nested inside this element. This in turn disrupts the usual use of text-structural tags such as `<p>` (paragraph), `<q>` (quotation), or `<head>` (heading), which now cannot be used for text that spans multiple lines without violating the strict hierarchical principle of XML. Even if the user chooses to export a version of TEI-XML that employs `<lb/>` (line beginning) milestone tags instead of `<line>` tags containing the line content, she will find that no markup has been allowed to cross a line boundary.

The developers, perhaps sharing the ideals of Stefan Hagel in wanting to ease the complexity of the data model and its expression as far as possible, have pre-configured a few markup tags for the convenience of the user. In some cases these are direct analogues of TEI elements, but in other cases they are simplified pseudo-TEI elements that are converted to their more complex equivalents, sometimes with a

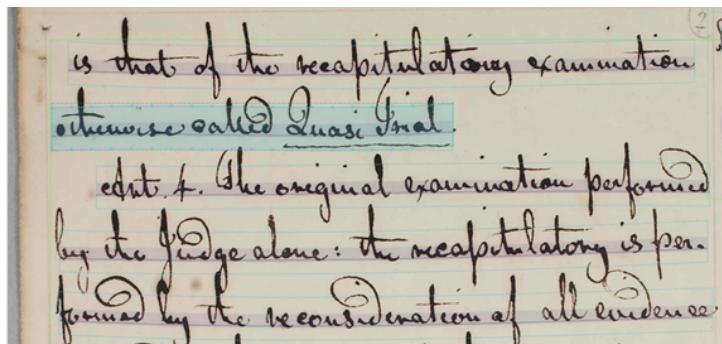


Figure 1: Adding XML-style annotations to Transkribus text.

loss of explicit information. For example, the user might find the `<abbrev>` element useful for marking up a line of text.

In Figure 1 the transcriber has provided two tags. Line 1–2 contains a technical term tagged with the `<term>` element, and line 1–3 contains an abbreviated word “Art.” (for “Article”) tagged as an abbreviation, with the entire word noted using the “expansion” property. The tagging interface would lead the XML-aware user to expect output such as

```
<lb n='2'>otherwise called <term rend='underline'>Quasi Trial</term>.  
<lb n='3'><abbrev expansion='Article'>Art.</abbrev> 4. The original examination  
performed
```

This, however, is not actually how abbreviations are expressed according to the TEI guidelines. In order to bridge the gap, Transkribus uses an internal XSLT stylesheet to convert the PAGE XML-based transcription data into valid TEI. The conversion results in this:

```
<lb facs='#fac1_r112' n='N002'>otherwise called <term rend='underline'>Quasi  
Trial</term>.  
<lb facs='#fac1_r113' n='N003'><choice>  
<expan>Article</expan>  
<abbr>Art.</abbr>  
</choice> 4. The original examination performed
```

Now that the abbreviation is represented as a choice, the implicit distinction between the abbreviation (which actually appears in the text) and the expansion (which does not appear as such in the text, but is an interpretation provided by the transcriber by means of a property on the element) has been lost. The distinction, and the

information about which of the variants in the `<choice>` element actually appeared in the document, can be arrived at only by understanding the process by which `<abbrev>` became `<choice>`.

The purpose of this seemingly arbitrary dive into the details of how TEI-XML is employed in one particular transcription tool must not be taken as a critique of the software, nor of the decisions the developers made. It is, rather, meant to illustrate a point that is already evident from the recommendations of the TEI Guidelines that a custom schema be produced for each edition project. The point is this: the process of converting text into data does not have a single natural outcome. As such, the technical details of that conversion within any particular software tool cannot be entirely hidden from the user, since these details cannot be inferred by the developers of other software tools that the user might wish to employ. It is the user of the respective tools who must translate the assumptions made by one tool into the assumptions made by the other tool, and this translation can only be done by a user who understands the model and expression of the data she is providing to the tools.

In the case of transcription of manuscript sources, editors who begin by reading the TEI Guidelines will usually adopt an approach that centers around the content of the text and its logical divisions, and treat the visual layout of the text as a secondary concern. On the other hand, editors who wish to have computational assistance with the transcription process—which after all involves the conversion of visual representations of text into records of its logical content—will almost certainly have to work with a form of data that begins with the visual layout, and fits the textual information into that frame. In order for the editor to proceed with further edition of the text, she will need to understand not only what kind of data has been produced, but also the implications of the choices that were made in its production.

## 4 Text-analysis tools and more complex editions

Eventually the editor will have one or more transcriptions of texts, probably in TEI-XML format. The next step in the process depends very much on the kind of edition being made. For some editions, once the text (or the corpus of texts) is transcribed with appropriate encoding for any notable features, and appropriate commentary or translation encoded alongside in the same document, all that remains is to publish it online in a suitably helpful format.

Many editions, however, are more complex. Some texts may require more structured information, for example to link names and places to relevant external sources of information. Other texts may benefit from algorithmic enrichment, such as morphological analysis, named-entity recognition, or various methods of stylometric analysis. In the case where the same text is preserved in multiple manuscripts, the

editor may require a full word-by-word collation of the variant texts and some form of stemmatic analysis.

The multiplicity of different kinds of possible editorial work, and the sheer variety of tools that exist in one form or another to assist with different sorts of work, is one of the great strengths of digital methods for scholarly edition. At the same time, this is by far the greatest bar to existence for the all-inclusive digital edition software package that the scholar may have envisioned. While some of these tools do advertise varying levels of support for TEI-XML encoded texts, others expect to operate only on plain text. The output of these different tools can come in a variety of different forms; the scholar must then decide whether the tool's output should be incorporated somehow into the existing TEI document, or whether the informational content of the edition will have to be stored in multiple complex forms. We can illustrate this with a few examples.

*Recogito* (Simon et al. 2017) is an online tool for annotating named entities (places, people and events) that appear in a text. Its central feature is the ability to link the place names to entries in online gazetteers such as *Pleiades* (Bagnall et al. 2016). The user may upload texts either as plain text or TEI-XML; in the latter case, there will be a set of assumptions, immutable and invisible to the user, about how to translate the structure defined in a TEI file into 'the text' that should be annotated. Here the scholar will have to exercise caution in ensuring that *Recogito* has, in fact, correctly derived the running text. Once the annotations have been added, the user has a variety of options for extracting them from *Recogito*'s system; these include CSV for tabular data (spreadsheet), RDF for Linked Open Data exchange, formats common in geography and cartography, formats for use as training data in machine learning applications, and, for textual scholars, TEI-XML. The TEI export has certain limitations, however: only place names are exported, and any annotations that would result in overlapping hierarchies (which would cause an XML parser to fail) are not included. The user who would like to enrich a TEI-encoded text with a set of potentially-overlapping semantic annotations prepared in Recogito about all three of people, places, and events, will first need to export these annotations into a separate file in RDF format. Integrating these annotations into the existing TEI document would then be a fairly complex programming task.

Other programs for text analysis, such as the web-based *Voyant* tools (Sinclair, Rockwell, and Voyant Tools Team 2012-) for exploration of corpora with distant-reading techniques, command line tools such as *Mallet* for topic modelling (McCallum 2002), or programming libraries such as the *Stylo R* package for stylometric analysis (Eder et al. 2016) expect to do their work with plain text supplied by the user. In each of these cases it is up to the user to extract 'the text' to be analyzed from the XML documents that contain them. *Voyant* provides a straightforward and flexible means of doing this directly in their interface, by allowing the user to specify an XPath

expression that will extract the text. This, of course, implies that the user is familiar not only with the structure of the XML document in question, but also the XPath syntax for referring to specified portions of the document. Stylo allows import of TEI-XML documents as well as HTML ones, but claims that these input options “have not been extensively tested so far” (Eder et al. 2017). Mallet expects plain text files to be prepared before the tool is run. For any sort of text analysis to have scholarly value, it is clear that the scholar must be able to control the text that is analyzed, ensuring, for example, that transcribers’ annotations are not inadvertently included as “original text”, or that scribal corrections are handled appropriately. Extraction of the intended text from a TEI-XML file thus involves a set of scholarly decisions that cannot easily be delegated to a non-scholar programmer; as such the scholar must be able to “speak the language” of the model she is using, be it XPath or R.

As a final example, we can take text collation. For editions that involve the reconstruction of a text from multiple manuscript witnesses, there are a number of tools available for collation and comparison of the source texts. Perhaps the two most well-known are *CollateX* (Dekker and Middell 2011), a program that is usually run either from the command line or from within a Python programming environment, and *JuXta* (Performant Software n.d.), a software package with a more user-friendly graphical interface for visualization of variation among texts.

To use *CollateX* it is very important to understand its core data model, which is entirely different from the strictly tree-hierarchical models developed primarily for single-document texts and used by the TEI. Rather, *CollateX* is based on the concept of the *variant graph*, a data structure intended specifically to represent agreement and variation among copies of the same text. The nodes that make up the variant graph—that is, the individual readings in the text—are known as *tokens*, and *CollateX* provides a great deal of flexibility both for the kind of data these tokens can include, and for the criteria that determine whether two tokens ought to be collated with each other. Both the graph and the token structure inform the data output options of *CollateX*. These include tabular formats expressed in CSV and JSON, the graph itself expressed in GraphML, which is an XML format that has no relation to TEI, and a custom form of XML that is modelled on, but does not entirely conform to, the TEI parallel-segmentation model of encoding text variants. Each of these output methods carries with it some trade-off; for example, GraphML is the only format that indicates where *CollateX* has detected transposed readings, but JSON is the only format that preserves extra data that was passed in along with the text of the reading. A textual scholar using *CollateX* must be prepared for these tradeoffs and choose the output format that best matches her needs for further analysis and publication, which itself implies that the scholar must understand what those needs are.

While the core feature of *JuXta* is its visualization options, it also provides a form of output data about the text variation that can be used in further digital processing.

Here the output is a list of variants given in HTML format. Just as with *CollateX*, only the textual content itself is given in the output list; any XML markup in the original TEI document is not preserved. Here too the scholar must evaluate for herself how to use this output, or alternatively, whether to confine the use of *JuXta* to the insights revealed by its visualizations.

The examples given here serve to demonstrate two points. First, the editor has an almost infinite variety of options concerning what to do with the texts that she has transcribed. Just as there is no single set of rules for how to engage in textual criticism, there is no single set of tools that all digital textual scholars will be expected to employ. Second, and following from the sheer variety of methods and needs, the use of any of these tools can only proceed from a solid technical understanding both of the data model that the tool employs, and the interface language needed to communicate with it.

## 5 From data model to publication

Having read through the section above on text analysis, some editors may at this point be tempted to breathe a sigh of relief. These are the editors whose text is a single document or a collection of similar documents that were transcribed in the same way and using the same model, who have included editorial comments directly in the respective XML document, and who now require only a place to store the TEI file(s) and a suitable means of conversion to a format understood by a modern Web browser.

Throughout Europe there are a growing number of services for storage and preservation of scholarly data, including transcribed text. The two primary options for digital textual scholars based in Austria are the ARCHE repository of the ACDH-ÖAW, and the GAMS repository of the ZIM-ACDH at the University of Graz. Other possibilities include intra-university IT services and commercial hosting, but these do not carry the long-term sustainability benefits that are provided by the dedicated research data repositories.

There are a great many options for conversion of TEI-XML files to a Web-readable format. Perhaps the two simplest, in that they do not require any additional software to run, are the XSLT files provided by the *TEI Boilerplate* project (Walsh, Simpson, and Moaddeli 2012–) and the CSS and Javascript library provided by the *CETEIcean* project (Cayless and Viglianti 2018). Other options, such as *TEIPublisher* (eXist Solutions 2015–), *EVT* (Rosselli Del Turco et al. 2014), and *dsebaseapp* (Andorfer 2016–) are fuller-featured software packages that require the installation and maintenance of software on a Web server. Like the simpler options, these software packages provide reasonable default presentations but, in order to produce an edition that is satisfactory to the

scholar, these defaults and the models that inform them will need to be understood and modified.

Although all of these options can be configured easily enough by a technologist who grasps the data model used for text encoding, a word of caution here to the optimistic “non-technical” scholar is in order. Although it is commonplace far beyond the world of digital philology to outsource the development of Web publications to professionals, the user interface to a digital edition communicates a great deal—perhaps more than the unwary scholar realizes—about the context, mood, and import of the text (Andrews and van Zundert 2018; Dillen 2018). The implication here is that, even if the scholar is fortunate enough to have the resources to procure a professional online publication, she will need to understand enough about the possibilities of the medium and about the correspondence between encoded text features and interface functionality to be able to retain control over the message that the interface will broadcast about the text. As such, the scholar who wishes to retain control over the message that is sent by her digital scholarly edition will need to understand the logical models and technical capabilities of the medium in which she publishes.

## 6 Conclusion

Having made the promised journey through the landscape of tools, we hope not to have lost our readers along the way. The digital medium is extraordinarily full of promise and possibility, and the very possibility that makes digital editions so exciting—the dizzying variety of options and capabilities that could not have been imagined in a print paradigm—also demands a substantial investment of time and understanding. The digital medium is dynamic and interactive in a way that has no parallels in print. It follows that the print paradigm, in which we editors are responsible only for the content which is delivered to a publishing house to take its form, cannot entirely hold. Nevertheless, we continue to hope and trust that the dynamism of the digital will continue to inspire textual scholars to be bold enough to engage with it.

## Bibliography

- Andorfer, Peter. 2016–. “Dsebaseapp.” Accessed September 30, 2019. <https://github.com/KONDE-AT/dsebaseapp>.
- Andrews, Tara L. n.d. “Guidelines for XML File Uploads to Stemmatweb.” Accessed September 30, 2019. <https://stematweb.net/stematweb/help/input#cte>.
- Andrews, Tara L., and Joris J. van Zundert. 2018. “What Are You Trying to Say? The Interface as an Integral Element of Argument.” In *Digital Scholarly Editions as Interfaces*, ed. by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 3–33. Norderstedt: BoD.
- Bagnall, Roger, Richard Talbert, Tom Elliott, Lindsay Holman, Jeffrey Becker, Sarah Bond, Sean Gillies et al. 2016–. *Pleiades: A Gazetteer of Past Places*. Accessed September 30, 2019. <https://pleiades.stoa.org>.
- Bordalejo, Barbara. 2018. “Digital versus Analogue Textual Scholarship or The Revolution Is Just in the Title.” *Digital Philology: A Journal of Medieval Cultures* 7: 7–28. doi:10.1353/dph.2018.0001.
- Causer, Tim, and Valerie Wallace. 2012. “Building A Volunteer Community: Results and Findings from Transcribe Bentham.” *Digital Humanities Quarterly* 6 (2). Accessed October 1, 2019. <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>.
- Cayless, Hugh, and Raffaele Viglianti. 2018. “CETEIcean: TEI in the Browser.” In *Proceedings of Balisage: The Markup Conference 2018*, Balisage Series on Markup Technologies, vol. 21, Washington, DC. doi:10.4242/BalisageVol21.Cayless01.
- Dekker, Ronald Haentjens, and Gregor Middell. 2011. “Computer-Supported Collation with CollateX: Managing Textual Variance in an Environment with Varying Requirements.” Paper presented at the meeting *Supporting Digital Humanities 2011*. Copenhagen.
- Dillen, Wout. 2018. “The Editor in the Interface: Guiding the User through Texts and Images.” In *Digital Scholarly Editions as Interfaces*, ed. by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 35–59. Norderstedt: BoD.
- Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. “Stylometry with R: A Package for Computational Text Analysis.” *R Journal* 8 (1): 107–21.
- Eder, Maciej, Mike Kestemont, Jan Rybicki, and Steffen Pielström. 2017. “‘Stylo’: R Package for Stylistic Analyses.” Accessed September 30, 2019. <https://github.com/computationalstylistics/stylo>.
- eXist Solutions. 2015–. “TEI Publisher.” Accessed September 30, 2019. <https://teipublisher.com>.
- Haentjens Dekker, Roland, Elli Bleeker, Bram Buitendijk, Astrid Kulsdom, and David J. Birnbaum. 2018. “TAGML: A Markup Language of Many Dimensions.” In *Proceedings of Balisage: The Markup Conference 2018*, Balisage Series on Markup Technologies, vol. 21, Washington, DC. doi:10.4242/BalisageVol21.HaentjensDekker01.
- Hagel, Stefan. 1997–. “Classical Text Editor.” Accessed September 30, 2019. <https://cte.oeaw.ac.at>.
- . 2007. “The Classical Text Editor. An Attempt to Provide for Both Printed and Digital Editions.” In *Digital Philology and Medieval Texts*, ed. by Arianna Ciula, and Francesco Stella, 77–84. Accessed Sept. 30, 2019. <http://www.infotext.unisi.it/upload/DIGIMED06/book/hagel.pdf>.

- “How to Use Transkribus—in 10 Steps (or Less).” 2015. Accessed September 30, 2019.  
<https://transkribus.eu/Transkribus/docs/How%20to%20use%20TRANSKRIBUS-0.1.5.pdf>.
- Institut für neutestamentliche Textforschung. n.d. “New Testament Virtual Manuscript Room – INTF.” Accessed October 1, 2019. <http://ntvmr.uni-muenster.de/home>.
- Kahle, Philip, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. 2017. “Transkribus – A Service Platform for Transcription, Recognition and Retrieval of Historical Documents.” In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 4, 19–24. doi:10.1109/ICDAR.2017.307.
- Klug, Helmut W., Selina Galka, and Elisabeth Steiner, eds. 2021. “Tools.” In *Weißbuch Digitale Editionen*. Graz: HRSM-Projekt Kompetenznetzwerk Digitale Editionen. <https://www.digitale-edition.at/archive/objects/context:sdef:Context/get?mode=tools>.
- McCallum, Andrew Kachites. 2002. “MALLET: A Machine Learning for Language Toolkit.” Accessed September 30, 2019. <http://mallet.cs.umass.edu>.
- Mittelbach, Arno, Sebastian Rahtz, and Ioan Bernevig. 2018. “Roma: Generating Customizations for the TEI (version 5.0.0)” Accessed September 30, 2019. <https://roma2.tei-c.org>.
- Performant Software. n.d. “Juxta: Collation Software for Scholars.” Accessed September 30, 2019. <https://web.archive.org/web/20220307201635/http://juxtasoftware.org>.
- Piez, Wendell. 2014. “TEI in LMNL: Implications for Modeling.” *Journal of the Text Encoding Initiative* 8 (December). doi:10.4000/jtei.1337.
- Pletschacher, Sefan, and Apostolos Antonacopoulos. 2010. “The PAGE (Page Analysis and Ground-Truth Elements) Format Framework.” In *2010 20th International Conference on Pattern Recognition*, 257–60. doi:10.1109/ICPR.2010.72.
- Rosselli Del Turco, Roberto, Giancarlo Buomprisco, Chiara Di Pietro, Julia Kenny, Raffaele Masotti, and Jacopo Pugliese. 2014. “Edition Visualization Technology: A Simple Tool to Visualize TEI-Based Digital Editions.” *Journal of the Text Encoding Initiative* 8 (December). doi:10.4000/jtei.1077.
- Sahle, Patrick. 2016. “What Is a Scholarly Digital Edition?” In *Digital Scholarly Editing: Theories and Practices*, ed. by Matthew James Driscoll, and Elena Pierazzo, 19–40. Open Book Publishers. doi:10.11647/OPB.0095.02.
- Simon, Rainer, Elton Barker, Leif Isaksen, Pau De Soto CaÑamares. 2017. “Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2.” *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives* 13: 111–32. doi:10.1080/15420353.2017.1307303.
- Sinclair, Stefan, Geoffrey Rockwell, and Voyant Tools Team. 2012–. “Voyant Tools.” Accessed October 1, 2019. <https://voyant-tools.org>.
- Vogeler, Georg. 2019. “Digitale Editionspraxis. Vom pluralistischen Textbegriff zur pluralistischen Softwarelösung.” In *Textgenese in der digitalen Edition*, ed. by Anke Bosse, and Walter Fanta, 117–36. Berlin, Boston: De Gruyter. doi:10.1515/9783110575996-008.
- Walsh, John, Grant Simpson, and Saeed Moaddeli. 2012–. “TEI Boilerplate.” Accessed October 1, 2019. <http://teiboilerplate.org>.
- Zundert, Joris van, and Peter Boot. 2011. “The Digital Edition 2.0 and the Digital Library: Services, Not Resources.” In *Digitale Edition Und Forschungsbibliothek*, ed. by Christiane Fritze, Franz Fischer, Patrick Sahle, and Malte Rehbein, 141–52. Wiesbaden: Harrassowitz.



# **Editionen und Graphentechnologie: Vorteile und Hürden digitaler Editionstechniken abseits von TEI-XML**

Peter Hinkelmanns

## **Abstract**

The article offers an overview of the use of graphs for the representation of texts. The focus is on variant graphs for genetic editions and the embedding of RDF data in edition projects based on TEI-XML. A possible technology stack for such an edition project with connection to the Semantic Web is also presented. Finally, a guideline offers a decision-making aid for the conception of an edition as a variant graph.

## **Zusammenfassung**

Der Beitrag bietet einen Überblick über den Einsatz von Graphen zur Repräsentation von Texten. Dabei stehen besonders Variantengraphen für genetische Edition und die Einbindung von RDF-Daten in Editionsprojekte auf TEI-XML-Basis im Fokus. Vorgestellt wird auch ein möglicher Technologie-Stack für ein solches Editionsprojekt mit Anbindung an das Semantic Web. Abschließend bietet ein Leitfaden eine Entscheidungshilfe für die Konzeption einer eigenen Edition als Variantengraph.

Für die Abbildung von Bezügen zwischen Texten aber auch zu anderen Datensätzen eignen sich Graphen in besonderer Weise. Ein ‚Graph‘ ist ein Datenmodell, das aus ‚Knoten‘ und diese verbindenden ‚Kanten‘ besteht. Ein Knoten ‚Buch‘ kann etwa über die gerichtete Kante ‚hat Autor‘ mit einer ‚Person‘ verbunden sein. Den Einsatz von Graphen in den digitalen Geisteswissenschaften hat Andreas Kuczera (Kuczera 2017) vorgestellt. Er hält fest, dass „Graphentechnologien für die Weiterentwicklung der digitalen Geisteswissenschaften sehr interessante Perspektiven bieten.“ (Kuczera 2017, 196) Auch Texteditionen, allen voran an textgenetischen Fragestellungen orientierte, können von der Realisierung mit Graphen profitieren.

Der vorliegende Beitrag bietet einen Überblick über den Einsatz von (Varianten-) Graphen für Editionsprojekte und stellt dies gängigeren Methoden wie der Realisierung mit TEI-XML gegenüber. Im Mittelpunkt stehen dabei folgende Fragen: Welche Vorteile kann eine graphbasierte Edition besonders in Hinblick auf genetische Editionen bieten und welche Risiken sind mit einem zu TEI-XML alternativen Verfahren

verbunden? Wie kann ein entsprechender Technologie-Stack aussehen? Welche Möglichkeiten bietet darüber hinaus eine Kombination von TEI-XML mit RDF-Daten? Als Zusammenfassung soll ein kleiner Leitfaden Unterstützung für die Konzeption eines eigenen Editionsprojektes mit Einbezug von Graphdaten bieten.

## 1 Modelle der Textrepräsentation

Grob lassen sich Modelle zur Repräsentation und Annotation von Text in die zwei Ansätze ‚hierarchischer Graph‘ und ‚gerichteter Graph‘ einteilen. *Markup*, also das Annotieren von Text, kann dabei interlinear (*inline*) oder separat mit Textbezug (*stand-off*) erfolgen. Das bei den meisten Editionen eingesetzte Modell der *Text Encoding Initiative* (TEI) ist ein *Markup*-Modell, bei dem ein hinterlegter Text *inline* mit Annotationen versehen wird. Das Codebeispiel unten und in Abbildung 1 zeigen einen kurzen mit linguistischer Annotation der Wortart versehenen Ausschnitt aus Lewis Carrolls *Alices Abenteuer im Wunderland*:

```
<p>
  <lb/>
  <w pos="NE">Alice</w>
  <w pos="VFIN"> fing</w>
  <w pos="APPR"> an</w>
  <w pos="PRF"> sich</w>
  <w pos="PTKZU"> zu</w>
  <w pos="VVINF">
    lang<pc>--</pc>
    <lb break="no"/> weilen
  </w>
  <pc> ; </pc>
  <w pos="PPER"> sie</w>
  <w pos="VFIN"> saß</w>
  <w pos="ADV"> schon</w>
  <w pos="ADV"> lange</w>
  <w pos="APPR"> bei</w>
  <w pos="PPOSAT"> ihrer</w>
  <w pos="NN"> Schwester</w>
  <w pos="APPRART"> am</w>
  <w pos="NN"> Ufer</w>
  [...]
</p>
```

Realisiert wird das Datenmodell der TEI mit XML, einer Markupsprache, bei der ein Knoten des beschriebenen Graphen jeweils genau einen Elternknoten hat und es einen einzigen Wurzelknoten für das gesamte Dokument geben muss. Wenn also etwa ein Wort, wie im obigen Beispiel ‚langweilen‘, über das Zeilenende hinausläuft und sowohl die einzelnen Zeilen als auch das Wort in einem Knoten repräsentiert werden sollen, führt dies zu einer Überlappung, die das Prinzip der geforderten strengen Hierarchie des XML-Baums verletzt. Mehrere Strategien sind in TEI eingeführt worden, um Probleme der Hierarchie zu lösen. So können etwa leere Elemente und *Standoff*-

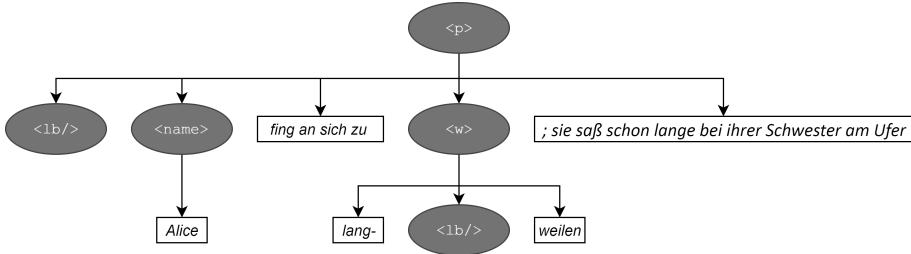


Abbildung 1: Darstellung des Code-Beispiels als hierarchischer Graph.

*Markup* eingesetzt werden, um Überlappungen zu ermöglichen. Im Beispiel sind die Zeilenanfänge mit dem leeren Element `<1b/>` markiert und nicht die gesamte Zeile von einem Element umschlossen, da die Dokumentstruktur in TEI als nachgeordnet eingestuft und durch leere Elemente realisiert wird. Somit kann die Worttrennung beim Beispielwort ‚langweilen‘ mit einem umschließenden Element `<w>` zeilenübergreifend markiert werden, ohne die Hierarchie formal zu verletzen.

Mit der Erweiterung von TEI um für genetische Editionen erforderliche Elemente hat sich eine Arbeitsgruppe befasst, die ihre Ergebnisse in einem Entwurf vorgestellt hat (Burnard et al. 2010). Teile dieses Entwurfes sind in die TEI-Guidelines übernommen worden (Text Encoding Initiative Consortium 2019; vgl. Wout Dillen 2016). Mit dem Modell der TEI lassen sich somit auch komplexe genetische Editionen realisieren. Ein kurzes Beispiel aus einer Greifswalder Handschrift (Abbildung 2) demonstriert eine Möglichkeit zur Transkription von intertextueller Varianz in Codebeispiel 2:

```

<1b/> [...] das man
<1b/> an dem ort der ſtadt ſo alzeit wußte iſt, noch
<1b/> ferner lieber wolt eine <del>Einode vnd</del> Wuſtenei
<1b/> am Cloſter
<subst>
  <del seq="1">ſehen</del>
  <add seq="2">wiffen</add>
</subst> oder daßelbige einen Pau
<1b break="no"/>ren liebere gonnen [...]
  
```

Ob die auf dem XML-Format gründenden strengen Hierarchien ein Problem für die Codierung von Texten darstellen, ist seit Veröffentlichung des TEI-Modells auf XML-Basis intensiv diskutiert worden.<sup>1</sup> Mittels der drei Elemente `<subst>`, `<add>`

<sup>1</sup> Ein umfassender Beitrag zur Problematik der Überlappungen liegt mit DeRose (2004) vor. Weitere Auseinandersetzungen mit dem Thema finden sich, chronologisch geordnet, etwa bei Huitfeldt 1994, 237; Buzzetti 2002, 76; Ide und Romary 2006, 227; Burghardt und Wolff 2009, 54; Schmidt und Colomb 2009, 498–99; Di Iorio, Peroni und Vitali 2011; Stührenberg 2012, §42; Schmidt 2012, 129–30; Sahle 2013, 103; Kuczera 2016; Efer 2016, 36–38; Haentjens Dekker und Birnbaum 2017; Bleeker 2017, 85; Turska und Spandini 2018; Bruder und Teufel 2018, 161.

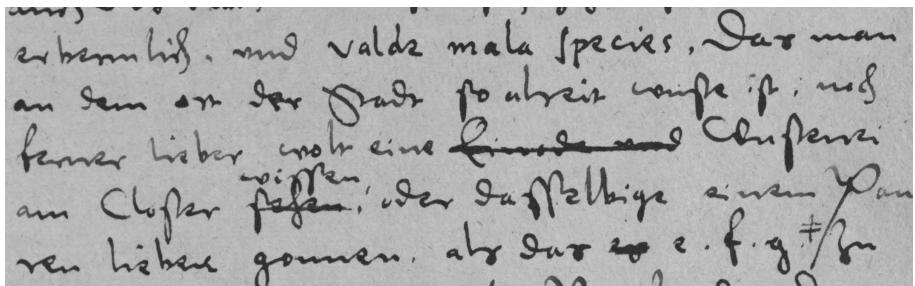


Abbildung 2: Textpassage (Universitätsarchiv Greifswald, Altes Rektorat Hbg. 134, 16r).

und <del> können Textvarianten innerhalb eines Textzeugen nach Regeln der TEI ausgezeichnet werden. Intertextuelle Varianz kann in TEI etwa über einen Apparat über das Element <app>, dokumentiert werden, um etwa separate Transkriptionen mehrerer Textzeugen aufeinander zu beziehen.<sup>2</sup> Mit GODDAG (Sperberg-McQueen und Huitfeldt 2004), GrAF (Ide und Suderman 2007) oder TAGML (Haentjens Dekker und Birnbaum 2017; Haentjens Dekker et al. 2018) sind Markupmodelle als Alternative zu XML vorgestellt worden, die überlappende Hierarchien erlauben. Dass TEI trotz der Basis XML überlappende Hierarchien bewältigen kann, hat James Cummings herausgestellt (Cummings 2018, 13–14). Gleichwohl bedarf es dazu einer Erweiterung des *Inline-Markups* durch *Standoff*-Annotationen.

## 2 Genetische Editionen als Variantograph

In der Editions- und Literaturwissenschaft erfreuen sich textgenetische Editionen anhaltender Popularität. Vertreter der jüngeren Forschungsgeschichte sind etwa die *Faust-Edition* (Bohnenkamp, Henke und Jannidis 2016), die *Edition der Manuskripte Samuel Becketts* (van Hulle und Nixon 2011–), die *Edition der Werke Arthur Schnitzlers* (Burch et al. 2016) oder auch die *Edition des Mann ohne Eigenschaften* Robert Musils (Bosse, Boelderl und Fanta 2016–), die eine vollständige Wiedergabe der Textgenese anstreben. ‚Textgenese‘ bezeichnet das Nachvollziehen der Entstehung von Texten. Dabei kann ein ‚Text‘ Bearbeitungen mehrerer Schreiberinnen und/oder Schreiber umfassen, sich über unterschiedliche Textträger erstrecken und sich in parallele Textfassungen verzweigen. Das Wesen einer genetischen Edition beschreibt Paolo D’Iorio treffend: „Das Ziel einer genetischen Edition ist es, die existierenden Textdokumente

<sup>2</sup> Vgl. z. B. das Datenmodell der Software Versioning Machine (Vetter et al. 2016) oder jenes, das Projekt Welscher Gast Digital (Šimek 2014, 4) benutzt.

zueinander in Beziehung zu setzen und so zu kommentieren, dass die Textgenese nachvollziehbar wird.“ (D’Iorio 2017, 196)

Alternativen zum TEI-Modell, auch für genetische Editionen, sind von der Forschung wiederholt diskutiert worden. Die Arbeitsgemeinschaft *Graphentechnologien in den Digitalen Geistes- und Sozialwissenschaften* hat das Thema auf Tagungen und in Workshops diskutiert (Kuczera et al. 2017–). Ausführlich beschäftigen sich auch die Dissertationsschriften Thomas Efers (Efer 2016) und Elli Bleekers (Bleeker 2017) mit der Abbildung von Texten als Graphen bzw. mit genetischen Texteditionen. Gemein ist den Ansätzen eine Abkehr vom *Markup*-Ansatz, bei dem ein Text gleich dem analogen Markieren mit dem Textmarker annotiert wird.

Eingeführt wurde der Variantengraph zur Textcodierung von Desmond Schmidt und Robert Colomb (Schmidt und Colomb 2009). Sie beschreiben den Variantengraph als azyklischen, gerichteten Graphen mit einem Start- und einem Endknoten (Schmidt und Colomb 2009, 501). Dies bedeutet, dass es vom Start- zum Endpunkt zahlreiche Wege geben kann, die aber keine Rückkehr an einen zuvor besuchten Punkt erlauben. Ihr Modell bietet die Möglichkeit, Textvarianten unterschiedlicher Textzeugen darzustellen. Der Text erscheint dabei auf den Kanten des Graphen, Knotenpunkte stellen Verzweigungen in der Textüberlieferung dar: Für jeden Textzeugen existiert genau ein Pfad durch den Graphen, der dessen Text ergibt, wie zum Beispiel für *E* in Abbildung 3.

Auch Hinzufügungen, Löschungen, Ersetzungen und Umstellungen lassen sich aus dem Graphen direkt ableiten. So ersetzt *scandito* nach Knoten 1 in Version C: *certo* aus Version B während DEFHI den Text an dieser Stelle ersatzlos streichen. Umstellungen zwischen Versionen können durch weitere Kanten gekennzeichnet werden, die als gestrichelte Kanten dargestellt werden.

*CollateX* (Haentjens Dekker et al. 2015), welches seit 2010 als Nachfolgeprojekt der Software *Collate* entwickelt wird, ermöglicht die automatische Kollationierung unterschiedlicher Varianten eines Textes. Datenmodell ist ein Variantengraph (vgl. Abbildung 4), dessen Knoten den Text abbilden und dessen Kanten den Verlauf des Textes zeigen. Anders als im Modell von Schmidt und Colomb 2009 werden die Textzeugen tokenisiert, die einzelnen Wörter bilden also die Knoten des Graphen.

Ein Modell, das auf dem von *CollateX* basiert (Andrews und Mace 2013, 506), wird für das Projekt *Stemmaweb* genutzt (*The Stemmaweb Project* 2012–). Die verschiedenen Types der Tokens der Textzeugen bilden die Knoten des Graphen in Abbildung 5. „Tokens“ sind die einzelnen Vorkommen der Wortformen im Text, „Types“ hingegen das Inventar dieser Tokens. Gerichtete Kanten zeigen den Verlauf des Textes in den unterschiedlichen Textzeugen an. Zwischen den einzelnen Type-Knoten kann durch weitere Kanten die Variante näher bestimmt werden. Möglich ist etwa die Annotation orthographischer, grammatischer oder lexikalischer Varianz (Andrews und Mace 2013, 508).

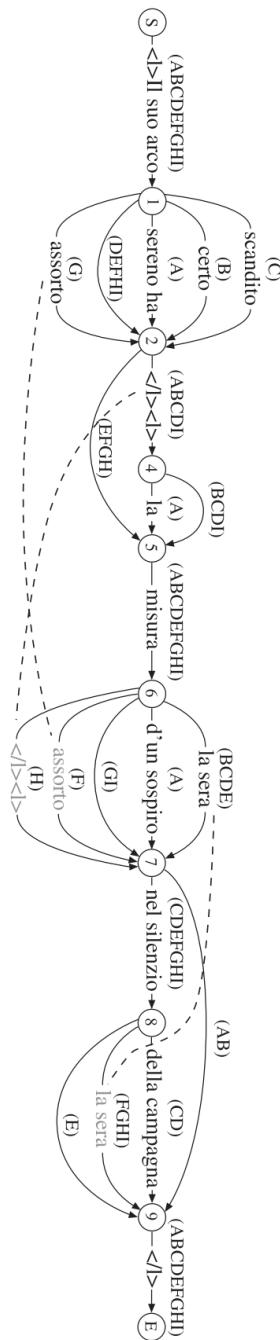


Abbildung 3: Ein Variantengraph (Abbildung aus Schmidt und Colomb 2009, 502).

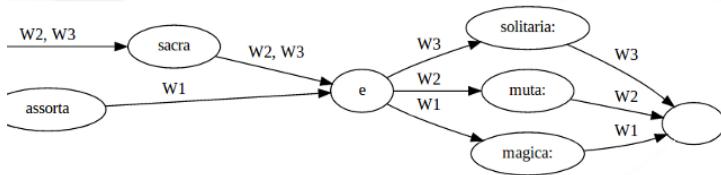


Abbildung 4: Variantengraph CollateX (Abbildung aus Haentjens Dekker und Middell 2010–).

Das graphbasierte Recherchesystem *Kadmos* ist von Thomas Efer vorgestellt worden (Efer 2016, 73–74). Das Modell unterscheidet zwischen Tokens und Types. Zwar stellt das Modell, das Abbildung 6 zeigt, keinen Variantengraph mehrerer Textzeugen dar, anders als die bereits vorgestellten Modelle besitzt es jedoch einen Knoten für jeden im Text vorkommende Type. Der Graph zeigt darüber hinaus, dass auch weitere Annotationen, wie die Zuordnung von Tokens zu Sätzen in einem Graph realisiert werden können. Konkurrierende Annotationsschichten sind dabei, anders als bei XML-*Inline-Markup*, voneinander unabhängig.

Auch der Frage der Visualisierung von Variantengraphen haben sich mehrere Untersuchungen gewidmet. Zu nennen ist hier der Ansatz des Projekts *Text Re-use Alignment Visualization* (TRAViz: Jänicke et al. 2015), weiters die Beiträge Jänicke und Wrisley 2018; Bleeker und Kelly 2018; Bleeker, Buitendijk und Haentjens Dekker 2019. Aufgrund des Fokus dieses Beitrags kann eine tiefergehende Darstellung des Themas ‚Visualisierung‘ leider nicht erfolgen, es bleibt jedoch festzuhalten, dass mit zunehmender Variation des Textes auch die Anforderungen an eine Visualisierung steigen. Und zwar unabhängig davon, ob TEI-XML oder ein gerichteter Graph als Basis der Edition gewählt worden ist.

### 3 Technischer Vergleich: XML und RDF

Die Textdarstellungen als hierarchischer Graph und als gerichteter Graph setzen auf unterschiedliche Technologie-Stacks. Das Format der *Text Encoding Initiative* baut in der gegenwärtigen Umsetzung auf XML auf. Im Vorwort der *P5 Guidelines* heißt es jedoch:

However, the TEI encoding scheme itself does not depend on this language; it was originally formulated in terms of SGML (the ISO Standard Generalized Markup Language), a predecessor of XML, and may in future years be re-expressed in other ways as the field of markup develops and matures. (Text Encoding Initiative Consortium 2019, iv. About These Guidelines)

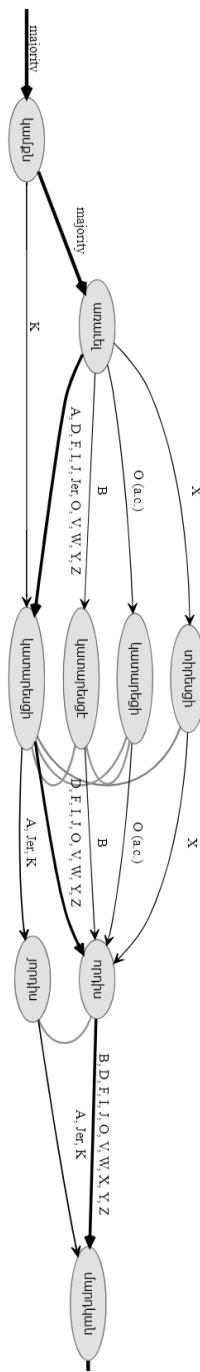


Abbildung 5: Screenshot des Text Relationship Mappers von Stemmaweb mit einem Ausschnitt aus dem ersten Abschnitt der Chronik Matthias, von Edessa (Abbildung aus *The Stemmaweb Project 2012*-).

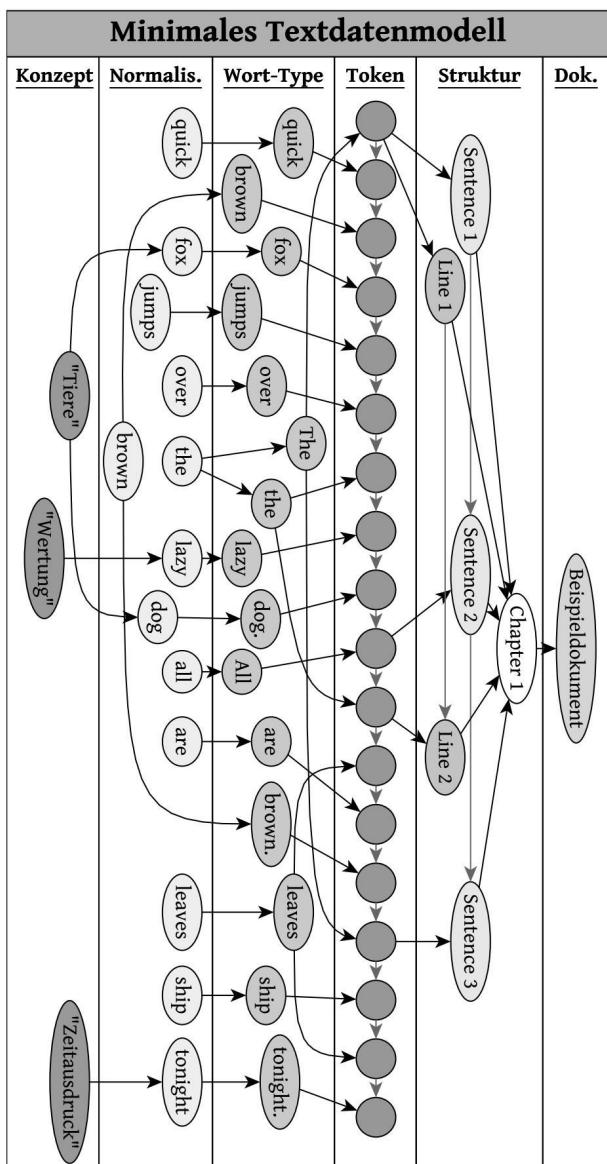


Abbildung 6: Schematische Darstellung der Instanzdатенsätze und Verknüpfungen eines kurzen Beispieldokuments bei minimalistischem Textdatenmodell (Abbildung aus Efer 2016, 76).

TEI ist damit nicht an eine Serialisierung gebunden, sodass eine an die Wahl des Formats XML gerichtete Kritik zukünftig gegenstandslos werden könnte. XML bietet für wesentliche Funktionalitäten Lösungen, die es attraktiv für den Einsatz in einem Editionsprojekt machen. XML wird in Form von Textdateien gespeichert, die für die Langzeitarchivierung und -lesbarkeit besser geeignet sind als etwa Binärdateien. Mit der Sprache *XML Schema* (Gao, Sperberg-McQueen und Thompson 2012) kann die Struktur von XML-Daten beschrieben werden und XML-Daten können auf Validität gegenüber einem Schema geprüft werden. Ein Schema definiert etwa, dass ein TEI-Dokument einen Kopfbereich haben muss, in dem ein Dokumenttitel angegeben ist. Zur Abfrage von Inhalten, aber auch, um weitreichendere Operationen zu ermöglichen, wurde *XQuery* (Robie, Dyck und Spiegel 2017b) eingeführt. Mittlerweile unterstützt *XQuery* nicht nur Abfragen auf XML-Daten, sondern auch JSON. Zur Transformation eines gesamten Dokumentes gedacht ist *XSLT* (Saxonica 2007). *XSLT* ermöglicht etwa die Umwandlung eines TEI-XML-Dokuments in ein HTML-Dokument. Die für *XML Schema*, *XQuery* und *XSLT* notwendige Navigation im Baum erfolgt über *XPATH* (Robie, Dyck und Spiegel 2017a).

Die Erstellung einer auf TEI-XML basierenden Edition unterstützen zahlreiche Werkzeuge wie XML-Editoren und XML-Datenbanken. Unter den Editoren hervorzuheben sind *Atom*<sup>3</sup> und *Oxygen*.<sup>4</sup> XML-Datenbanken sind etwa *eXist-db*<sup>5</sup> oder *BaseX*<sup>6</sup> wobei sich die Zugriffsgeschwindigkeiten der Datenbanken durchaus unterscheiden.<sup>7</sup>

Nachdem eine Standardisierung für Editionen auf Graphbasis noch aussteht, könnten an dieser Stelle auch beliebige andere Graphdatenmodelle genannt werden. Als Beispiel eines Technologie-Stacks für eine auf einem gerichteten Graph basierende Edition soll hier RDF vorgestellt werden. Das *Semantic Web* ist die Umsetzung einer Idee von Tim Berners-Lee (Berners-Lee 2006; zusammenfassend auch Stein 2014), um semantisch ausgezeichnete Daten maschinenlesbar über das Netz verfügbar zu machen. Im Fokus steht also die Vernetzung aller in das System eingespeister Daten, ein auch für Editionsprojekte nützlicher Vorteil. Wesentliche Technologien sind das *Resource Description Framework* (Gandon und Schreiber 2014), die Web Ontology Language (OWL: W3C OWL Working Group 2012) und die *SPARQL Protocol And RDF Query Language* (W3C SPARQL Working Group 2013). RDF bildet die Basis, um Triple bestehend aus Subjekt – Objekt – Prädikat zu beschreiben, und kann in unterschiedlichen Textformaten wie *XML-RDF* oder *N3* mit dessen Subset *Turtle* serialisiert werden. Für eine Langzeitarchivierung gelten damit dieselben Voraussetzungen wie für XML-Daten. OWL beschreibt Objektklassen und deren Beziehungen unter-

<sup>3</sup> <https://atom.io>.

<sup>4</sup> <https://www.oxygenxml.com>.

<sup>5</sup> <http://exist-db.org>.

<sup>6</sup> <http://baseX.org>.

<sup>7</sup> Für einen Überblick über die Performance und weitere Eigenschaften vgl. Chau et al. 2019.

	XML	RDF / Semantic Web
<b>VORANNAME</b>	Closed-World	Open-World
<b>ARCHIVIERUNG</b>	Textdatei	Textdatei
<b>MODELLIERUNG</b>	XML Schema	OWL
<b>VALIDIERUNG</b>	XML Schema	SHACL
<b>ABFRAGE</b>	XQuery	SPARQL
<b>TRANSFORMATION</b>	XSLT	-(SPARQL Construct)
<b>SPEZIFISCHES VOKABULAR FÜR</b>	TEI	-(domänen spezifische Ontologien)
<b>EDITIONEN</b>		

Tabelle 1: Technologie-Stacks im Vergleich.

einander und SPARQL ermöglicht schließlich das Abfragen der erzeugten Graphen. Anders als XML geht RDF von einem *Open-World*-Ansatz aus. Wesentliches Prinzip ist, dass nicht vorausgesetzt werden kann, dass im *Semantic Web* ein Teilnehmer das gesamte Netz kennt. Daher kann eine Aussage mit einer OWL-Ontologie nicht auf deren Wahrheitsgehalt geprüft werden, wie dies durch den *Closed-World*-Ansatz von XML mit einem Schema möglich ist. Für die Datensätze eines Editionsprojektes können mittels der *Shapes Constraint Language* (SHACL: Knublauch und Kontokostas 2017) allerdings Regeln definiert werden, die von einem *Closed-World*-Ansatz im Projektrahmen ausgehen und daher die Validierung der Daten ermöglichen.

Implementationen von RDF-*Triple-Stores*, die zum Technologie-Stack des *Semantic Webs* gehören, sind etwa *Apache Jena*<sup>8</sup> oder *GraphDB*.<sup>9</sup> Die Entscheidung für eine Graphdatenbank unterliegt zahlreichen Kriterien, etwa den Kosten, der Performance oder der verfügbaren Bibliotheken für unterschiedliche Programmiersprachen. Eine Übersicht über die Eigenschaften verschiedener *Triple-Stores* bieten Banane und Belangour (2019).

Eine mit TEI vergleichbare Ontologie zur Beschreibung von Texten und zahlreichen weiteren Daten, wie Personen oder Orten, existiert im Kontext des *Semantic Webs* zwar nicht, jedoch können einzelne, domänen spezifische Bereiche, wie Personendaten, nach normierten Vokabularen bzw. Ontologien beschrieben werden. Ein Beispiel ist die Ontologie der *Gemeinsamen Normdatei* (GND), die unter anderem ein Vokabular zur Beschreibung von Personen, Orten, Werken und Körperschaften bereitstellt (Haffner 2012–). Abschließend sollen die skizzierten Möglichkeiten der beiden Technologie-Stacks tabellarisch gegenübergestellt werden (Tabelle 1).

Sowohl XML als auch RDF bieten eine Vielzahl von Werkzeugen, die wichtige Funktionen der Datenerhebung, des Datenabfragens und des Datenarchivierens er-

<sup>8</sup> <https://jena.apache.org>.

<sup>9</sup> <https://graphdb.ontotext.com>.

möglichen. Im Bereich der Normierung ist TEI-XML für Editionen und zahlreiche damit in Verbindung stehende Daten etabliert, während für das *Semantic Web* eher kleinteilige, domänenspezifischere Ontologien von Relevanz sind. Eine Ontologie zur Abbildung von Texten als Graph existiert bislang nicht bzw. nur in Ansätzen.<sup>10</sup> Obwohl TEI-XML auf anderen Technologien basiert, sind bereits Vorschläge gemacht worden, TEI auch als Ontologie verfügbar zu machen (Ciotti 2018; Ciotti und Tomasi 2016).

## 4 Graphen und (TEI-)Editionen

Neben einem Einsatz von gerichteten Graphen als Datenmodell für die Textwiedergabe können auch auf TEI-XML basierende Editionen mittels Graphen erweitert werden. Die TEI Guidelines enthalten für die direkte Einbettung von RDF-Daten etwa das Element <xenoData>. Für die Vernetzung archivalischer bzw. geisteswissenschaftlicher Ressourcen bieten die Vokabulare und Ontologien des *Semantic Webs* vielfältige Möglichkeiten, wie etwa Christopher Pollin und Lina M. Zangerl (2020) zeigen. Ontologien zur Auszeichnung geisteswissenschaftlicher Ressourcen sind unter anderem das *Datenmodel der Europeana* (Isaac 2013), die bereits zuvor genannte *Ontologie der Gemeinsamen Normdatei* oder das *CIDOC Conceptual Reference Model* (Le Boeuf et al. 2017) – letzteres ist als *Top-Level-Ontologie* ein Bezugsmodell für den Entwurf eigener Ontologien und Schnittstelle zwischen unterschiedlichen Datenmodellen.

Die Verbindung von TEI-Editionen mit Graphdaten wird auch in der *Mittelhochdeutschen Begriffsdatenbank* (MHDBDB: Universität Salzburg 1992–) umgesetzt, an deren Entwicklung und Konzeption der Autor beteiligt ist. Aufgrund der vielfältigen Annotationsschichten, die auf den Texten dieses onomasiologischen Projekts zum mittelhochdeutschen Wortschatz liegen, konnte nicht ausgeschlossen werden, dass eine einfache Inline-Annotation zu Verletzungen der Hierarchie von XML führt. Weiters stellt das Begriffssystem einen Thesaurus dar, der idealerweise mit dem RDF-Vokabular *Simple Knowledge Organization System* (SKOS: Miles und Bechhofer 2009) beschrieben wird. Aus diesen Gründen werden in der MHDBDB Texte mit TEI-XML codiert, aber auf Tokenebene mit RDF-Graphdaten vernetzt.

Die Vernetzung von TEI-basierten Briefeditionen ist im Projekt *correspSearch* umgesetzt worden (Dumont 2016). Zwar entspricht das Verfügbarmachen über ein *Application Programming Interface* (API) nicht den *Semantic-Web*-Technologien, es zeigt jedoch sehr gut, welches Potential in der Verbindung unterschiedlicher Editionsprojekte steckt. Die Transformation unter anderem von TEI-Daten in ein RDF-Modell ist ein Ziel der Ontologie der *Digital Edition Publishing Cooperative for Historical*

---

<sup>10</sup> An dieser Stelle sei auf die Überlegungen des Autors zu einer *Text Graph Ontology* hingewiesen (Hinkelmanns 2019).

*Accounts* (DEPCHA: Pollin und Vogeler 2019; Vogeler 2019). Die Ontologie ermöglicht es, Rechnungsbucheinträge semantisch auszuzeichnen und somit aufeinander zu beziehen. Wie in *correspSearch* können so die Daten unterschiedlicher Projekte zusammengeführt werden.

## 5 Leitfaden und Ausblick

Bislang sind auf einem Variantengraph basierende Editionen die Ausnahme. Die laufenden Projekte zeigen jedoch, dass Graphen besonders für genetische Editionen und für die Vernetzung von Editionen unterschiedlicher Projekte einen interessanten Ansatz darstellen. Da eine Standardisierung bislang nur in Grundzügen existiert, ist es auch für graphbasierte Editionen ratsam, einen Export nach TEI-XML zur Archivierung der eigenen Daten einzusetzen. In der Vorbereitungsphase zu einer variantengraph-basierten Edition sollte eine Reihe von Fragen gestellt werden:

1. **Wie soll der Text segmentiert werden?** Genügt die Similarität von Zeichenketten aus unterschiedlichen Textzeugen als Segmentierungskriterium (Schmidt und Colomb 2009)? Soll eine Tokenisierung an Weißraum durchgeführt werden (Andrews und Mace 2013)? Sollen auch Phänomene unterhalb der Wortgrenze betrachtet werden?
2. **Welche Art von Varianz soll analysiert werden?** Sollen etwa Varianzen innerhalb eines Textträgers oder auch zwischen verschiedenen Fassungen eines Textes analysiert werden?
3. **Welche Datenobjekte neben der eigentlichen Textedition soll meine Edition beinhalten?** In zahlreichen Projekten gehören Metadaten zu Entitäten wie Personen, Orten oder Objekten zu den Anforderungen. Eine graphbasierte Edition sollte hier auf Ontologien wie *Europeana* (Isaac 2013) basieren und erforderliche eigene Ontologien auf das Referenzmodell CIDOC-CRM (Le Boeuf et al. 2017) beziehen.
4. **Mit welchen Normdatenrepositorien und anderen Editionsprojekten möchte ich meine Edition vernetzen?** Durch das *Semantic Web* und Projekte wie *correspSearch* (Dumont 2016) wird die Möglichkeit eröffnet, die eigene Edition mit externen Daten, etwa aus der *Gemeinsamen Normdatei* (Deutsche Nationalbibliothek) anzureichern.
5. **Welche technologische Basis soll meine digitale Edition haben?** Neben den erwähnten verfügbaren Infrastrukturen von Editionsprojekten können auch Eigenentwicklungen eingesetzt werden. Von Vorteil ist die Erstellung der Graphedition mit *Semantic-Web*-Technologien.

**6. Wie soll meine Edition visualisiert werden?** Die Visualisierung des Textes ist elementarer Bestandteil einer textgenetischen Edition, deren Komplexität einen eigenen Beitrag rechtfertigen würde, zusammenfassend dazu etwa Bleeker und Kelly (2018).

Zum gegenwärtigen Zeitpunkt gibt es noch keine einheitliche Vorgehensweise für die Umsetzung einer Edition als Variantengraph. Das in den letzten Jahren gestiegene Interesse an Graphen und der Vernetzung über das *Semantic Web* könnte gleichwohl auch neue Entwicklungen im Bereich der Editionswissenschaft nach sich ziehen. Aber auch eine Verbindung von TEI-XML mit RDF-Graphdaten kann eine lohnenswerte Vorgehensweise für das eigene Editionsprojekt sein. Auf der einen Seite profitiert das Projekt so von der erprobten Textkodierung mit TEI-XML, auf der anderen Seite werden die eigenen Daten über das *Semantic Web* erschlossen.

Einige wenige Projekte bieten bereits die Möglichkeit, eine eigene graphbasierte genetische Edition umzusetzen. Frei nutzbar ist die Infrastruktur des Projektes *Stemmaweb* (*The Stemmaweb Project* 2012–). Eingespeiste Transkriptionen in unterschiedlichen Formaten können automatisch kollationiert und deren Stemmata sowie der daraus entstandene Variantengraph manuell nachbearbeitet werden. Ebenfalls zur freien Nutzung freigegeben ist *CollateX*, das es ermöglicht, mehrere Textversionen zu vergleichen und einen Variantengraphen zu generieren (Haentjens Dekker und Middell 2010–).

Es kann festgehalten werden, dass es – zum Glück! – nicht die eine Lösung für die Umsetzung einer Edition als Graph oder unter Einbeziehung von Graphdaten gibt. Dieser Beitrag und Leitfaden soll demnach nicht abschließend sein, sondern die vielfältigen Möglichkeiten mit und jenseits TEI-XML für Editionsprojekte aufzeigen und ermutigen, auch die eigenen Projekte von diesen Möglichkeiten profitieren zu lassen; Gründe dafür gibt es viele, exemplarisch genannt können etwa die im Vergleich zu TEI-XML bessere Abbildung von Textstufen in Variantengraphen oder auch die Anbindung an das *Semantic Web* sein.

## Literatur

- Andrews, T. L. und C. Mace. 2013. „Beyond the tree of texts: Building an empirical model of scribal variation through graph analysis of texts and stemmata.“ *International Journal of Human-Computer Studies* 28 (4): 504–21. doi:10.1093/llc/fqt032.
- Banane, Mouad und Abdessamad Belangour. 2019. „A Comparative Study of RDF Triple Stores.“ *SSRN Journal*. doi:10.2139/ssrn.3349399.
- Berners-Lee, Tim. 2006. „Linked Data.“ Zugriff: 4. September 2017. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Bleeker, Elli. 2017. „Mapping invention in writing: Digital infrastructure and the role of the genetic editor“ Dissertation, Universität Antwerpen. hdl:10067/1556760151162165141.

- Bleeker, Elli, Bram Buitendijk und Ronald Haentjens Dekker. 2019. „From graveyard to graph.“ *International Journal of Digit Humanities* 1: 141–163. doi:10.1007/s42803-019-00012-w.
- Bleeker, Elli und Aodhán Kelly. 2018. „Interfacing Literary Genesis.“ In *Digital Scholarly Editions as Interfaces*, hg. v. Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber und Gerlinde Schneider, 193–218. Norderstedt: Books on Demand.
- Bohnenkamp, Anne, Silke Henke und Fotis Jannidis. 2016. *Johann Wolfgang Goethe: Faust: Historisch-kritische Edition*. 2. Beta-Version. Zugriff: 26. April 2017. <http://beta.faustedition.net/>.
- Bosse, Anke, Artur Boelderl und Walter Fanta. 2016–. *Musil online*. Zugriff: 30. September 2019. <http://musilonline.at/>.
- Bruder, Daniel und Simone Teufel. 2018. „Data models for Digital Editions: Complex XML versus Graph structures.“ In *Kritik der digitalen Vernunft. Abstracts zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum, 26.02.–02.03.2018 an der Universität zu Köln, veranstaltet vom Cologne Center for eHumanities (CCeH)*, hg. v. Georg Vogeler, 158–62. Köln: Universität zu Köln.
- Burch, Thomas, Stefan Büdenbender, Kristina Fink, Vivien Friedrich, Patrick Heck, Wolfgang Lukas, Kathrin Nühlen et al. 2016. „Text[ge]schichten: Herausforderungen textgenetischen Edierens bei Arthur Schnitzler.“ In *Textgenese und digitales Edieren: Wolfgang Koeppens „Jugend“ im Kontext der Editionsphilologie*, hg. v. Katharina Krüger, 87–105. Berlin, Boston: de Gruyter.
- Burghardt, Manuel und Christian Wolff. 2009. „Stand off-Annotation für Textdokumente: Vom Konzept zur Implementierung (zur Standardisierung?).“ In *Von der Form zur Bedeutung: Texte automatisch verarbeiten*, hg. v. Christian Chiarcos, Richard Eckart de Castilho und Manfred Stede, 53–59. Tübingen: Narr.
- Burnard, Lou, Fotis Jannidis, Elena Pierazzo und Malte Rehbein. 2010. „An Encoding Model for Genetic Editions.“ Zugriff: 25. März 2019. <http://www.tei-c.org/Activities/Council/Working/tcw19.html>.
- Buzzetti, Dino. 2002. „Digital Representation and the Text Model.“ *New Literary History* 33 (1): 61–88. <https://www.jstor.org/stable/20057710>.
- Chau, Kien Tsong, QianLing He, XuQing Hu und Rui Wu. 2019. „Comparison on Performance of Text-Based and Model-Based Architecture in Open Source Native XML Database.“ In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, 340–44: IEEE Xplore. doi:10.1109/SIPROCESS.2019.8868709.
- Ciotti, Fabio. 2018. „A Formal Ontology for the Text Encoding Initiative.“ *Umanistica Digitale* 2 (3). doi:10.6092/issn.2532-8816/8174.
- Ciotti, Fabio und Francesca Tomasi. 2016. „Formal Ontologies, Linked Data, and TEI Semantics.“ *Journal of the Text Encoding Initiative* 9. doi:10.4000/jtei.1480.
- Cummings, James. 2018. „A world of difference: Myths and misconceptions about the TEI.“ *Digital Scholarship Humanities* 34 (Supplement 1): i58–i79. doi:10.1093/lcc/fqy071.
- D'Iorio, Paolo. 2017. „Die Schreib- und Gedankengänge des Wanderers. Eine digitale genetische Nietzsche-Edition.“ *Editio* 31 (1): 191–204. doi:10.1515/editio-2017-0011.
- DeRose, Steven J. 2004. „Markup Overlap: A Review and a Horse.“ In *Extreme Markup Languages 2004 Proceedings*. Zugriff: 25. März 2019. <http://xml.coverpages.org/DeRoseEML2004.pdf>.

- Deutschen Nationalbibliothek. „Gemeinsame Normdatei: (GND).“ Zugriff: 8. Juni 2022. [https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html).
- Di Iorio, Angelo, Silvio Peroni und Fabio Vitali. 2011. „A Semantic Web approach to everyday overlapping markup.“ *Journal of the American Society for Information Science and Technology* 62 (9): 1696–1716. doi:10.1002/asi.21591.
- Dumont, Stefan. 2016. „correspSearch: Connecting Scholarly Editions of Letters.“ *Journal of the Text Encoding Initiative* 10. doi:10.4000/jtei.1742.
- Efer, Thomas. 2016. „Graphdatenbanken für die textorientierten e-Humanities.“ Dissertation, Universität Leipzig. urn:<https://nbn-resolving.org/nbn:de-bsz:15-qucosa-219122>.
- Gandon, Fabien und Guus Schreiber. 2014. „RDF 1.1 XML Syntax.“ Zugriff: 23. August 2019. <http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225>.
- Gao, Shudi, C. M. Sperberg-McQueen und Henry S. Thompson. 2012. „W3C XML Schema Definition Language (XSD).“ Zugriff: 18. Februar 2020. <http://www.w3.org/TR/xmlschema11-1>.
- Haentjens Dekker, Ronald und David J. Birnbaum. 2017. „It's more than just overlap: Text As Graph.“ In *Proceedings of Balisage: The Markup Conference 2017. Balisage Series on Markup Technologies* 19. doi:10.4242/BalisageVol19.Dekker01.
- Haentjens Dekker, Roland, Elli Bleeker, Bram Buitendijk, Astrid Kulsdom und David J. Birnbaum. 2018. „TAGML: A markup language of many dimensions.“ In *Proceedings of Balisage: The Markup Conference 2018. Balisage Series on Markup Technologies* 21. doi:10.4242/BalisageVol21.HaentjensDekker01.
- Haentjens Dekker, Ronald und Gregor Middell. 2010–. „CollateX: Software for Collating Textual Sources.“ Documentation. Zugriff: 23. September 2019. <https://collatex.net/doc/>.
- Haentjens Dekker, Ronald, Dirk van Hulle, Gregor Middell, Vincent Neyt und Joris van Zundert. 2015. „Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project.“ *Digital Scholarship Humanities* 30 (3): 452–70. doi:10.1093/ll/cfqu007.
- Haffner, Alexander. 2012–2019. „GND Ontology.“ Zugriff: 30. September 2018. <http://d-nb.info/standards/elementset/gnd>.
- Hinkelmanns, Peter. 2019. „Text Graph Ontology.“ Zugriff: 5. Januar 2020. <https://github.com/Wolkenstein/tgo-ontology>.
- Huitfeldt, Claus. 1994. „Multi-dimensional texts in a one-dimensional medium.“ *Computers and the Humanities* 28 (4–5): 235–41. doi:10.1007/BF01830270.
- Ide, Nancy und Laurent Romary. 2006. „Representing Linguistic Corpora and Their Annotations.“ In *Proceedings of the Fifth Language Resources and Evaluation Conference (LREC'2006)*, hg. v. Nicoletta Calzolari, et al., 225–28. Genua: European Language Resources Association. Zugriff: 18. Februar 2020. [http://www.lrec-conf.org/proceedings/lrec2006/pdf/562\\_pdf.pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/562_pdf.pdf).
- Ide, Nancy und Keith Suderman. 2007. „GrAF: A Graph-based Format for Linguistic Annotations.“ In *Proceedings of the Linguistic Annotation Workshop*, hg. v. Branimir Boguraev et al., 1–8. Prague: Association for Computational Linguistics. Zugriff: 18. Februar 2020. <https://aclanthology.org/W07-1501.pdf>.
- Isaac, Antoine. 2013. „Europeana Data Model Primer.“ Zugriff: 18. Februar 2020. [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf).

- Jänicke, Stefan, Annette Geßner, Greta Franzini, Melissa Terras, Simon Mahony und Gerik Scheuermann. 2015. „TRAViz: A Visualization for Variant Graphs.“ *Digital Scholarship in the Humanities* 2013 (2): 183–199. doi:10.1093/lhc/fqv049.
- Jänicke, Stefan und David J. Wrissley. 2018. „Interactive Visual Alignment of Medieval Text Versions.“ In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, IEEE Xplore. doi:10.1109/VAST.2017.8585505.
- Knublauch, Holger und Dimitris Kontokostas. 2017. „Shapes Constraint Language (SHACL).“ Zugriff: 18. Februar 2020. <https://www.w3.org/TR/shacl/>.
- Kuczera, Andreas. 2016. „Digital Editions beyond XML: Graph-based Digital Editions.“ In *Proceedings of the 3rd HistoInformatics Workshop, Krakow, Poland, 11 July 2016*, hg. v. Marten Düring, Adam Jatowt, Johannes Preiser-Kappeller und Antal van Den Bosch, 37–46. [http://ceur-ws.org/Vol-1632/paper\\_5.pdf](http://ceur-ws.org/Vol-1632/paper_5.pdf).
- , Aline Deicke, Julian Jarosch. 2017–. „Graphentechnologien: Graphentechnologien in den digitalen Geistes- und Sozialwissenschaften.“ Zugriff: 30. September 2019. <https://graphentechnologien.hypotheses.org>.
- . 2017. „Graphentechnologien in den Digitalen Geisteswissenschaften.“ *ABI Technik* 37, Nr. 3: 179–96. doi:10.1515/abitech-2017-0042.
- Le Boeuf, Patrick, Martin Doerr, Christian Emil Ore und Stephen Stead. 2017. „Definition of the CIDOC Conceptual Reference Model: Version 6.2.2.“ Zugriff: 1. Februar 2022. [https://www.cidoc-crm.org/sites/default/files/2017-09-30%23CIDOC%20CRM\\_v6.2.2\\_esIP.pdf](https://www.cidoc-crm.org/sites/default/files/2017-09-30%23CIDOC%20CRM_v6.2.2_esIP.pdf).
- Miles, Alistair und Sean Bechhofer. 2009. „SKOS Simple Knowledge Organization System: Reference.“ Zugriff: 18. Februar 2020. <http://www.w3.org/TR/skos-reference>.
- Pollin, Christopher und Georg Vogeler. 2019. „Datenmodell für historische Rechnungsunterlagen: Version 1.1.“ Zugriff: 30. September 2021. <https://gams.uni-graz.at/context:depcha>.
- Pollin, Christopher und Lina M. Zangerl. 2020. „Der Nachlass als Netzwerk: Zur Entwicklung einer Nachlass-Ontologie am Beispiel des Projekts ‚Stefan Zweig digital‘.“ In *Digital Humanities Austria 2018: Empowering researchers*, hg. v. Marlene Ernst, Peter Hinkelmanns und Lina M. Zangerl, 123–27. doi:10.1553/dha-proceedings2018s123.
- Robie, Jonathan, Michael Dyck und Josh Spiegel. 2017a. „XML Path Language (XPath) 3.1.“ Zugriff: 18. Februar 2020. <https://www.w3.org/TR/xpath-31/>.
- . 2017b. „XQuery 3.1: An XML Query Language.“ Zugriff: 18. Februar 2020. <https://www.w3.org/TR/xquery-31/>.
- Sahle, Patrick. 2013. *Digitale Editionsformen 3: Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels: Textbegriffe und Recodierung*. Norderstedt: BoD.
- Saxonica, Michael K. 2007. „XSL Transformations (XSLT) Version 2.0.“ Zugriff: 18. Februar 2020. <http://www.w3.org/TR/xslt20/>.
- Schmidt, Desmond. 2012. „The Role of Markup in the Digital Humanities.“ *Historical Social Research* 37 (3): 125–46. <http://www.jstor.org/stable/41636601>.
- Schmidt, Desmond und Robert Colomb. 2009. „A data structure for representing multi-version texts online.“ *International Journal of Human-Computer Studies* 67 (6): 497–514. doi:10.1016/j.ijhcs.2009.02.001.

- Šimek, Jakub. 2014. Welscher Gast Digital. TEI-Handbuch. Version 0.6. Zugriff: 5. Januar 2021. [http://digi.ub.uni-heidelberg.de/wgd/pdf/TEI-Handbuch\\_0-6.pdf](http://digi.ub.uni-heidelberg.de/wgd/pdf/TEI-Handbuch_0-6.pdf).
- Sperberg-McQueen, C. M. und Claus Huitfeldt. 2004. „GODDAG: A Data Structure for Overlapping Hierarchies.“ In *Digital Documents: Systems and Principles*, hg. v. Peter King und Ethan V. Munson, 139–60. Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-39916-2\_12.
- Stein, Christian. 2014. „Linked Open Data: Wie das Web zur Semantik kam.“ *Bibliothek. Forschung und Praxis* 38 (3): 447–55.
- Stührenberg, Maik. 2012. „The TEI and Current Standards for Structuring Linguistic Data.“ *Journal of the Text Encoding Initiative* 3. Zugriff: 17. Juni 2016. <http://jtei.revues.org/523>.
- Text Encoding Initiative Consortium. 2019. „P5: Guidelines for Electronic Text Encoding and Interchange: Non-hierarchical Structures.“ Zugriff: 5. Februar 2019. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html>.
- The StemmaWeb Project: Tools and techniques for empirical stemmatology*. 2012–. Zugriff: 30. September 2019. <https://stemmaweb.net>.
- Turska, Magdalena und E. Spandini. 2018. „XML/TEI Stand-off Markup: One step beyond.“ *Digital Philology: A Journal of Medieval Cultures* 8 (2): 225–39. doi:10.1353/dph.2019.0025.
- Universität Salzburg. 1992–. „Mittelhochdeutsche Begriffsdatenbank.“ Zugriff: 18. Februar 2020. <http://www.mhdbdb.sbg.ac.at/>.
- van Hulle, Dirk und Mark Nixon. 2011–. „Samuel Beckett: Digital Manuscript Project.“ Zugriff: 30. September 2019. <https://www.beckettarchive.org>.
- Vogeler, Georg. 2019. „The ‚assertive edition‘.“ *International Journal of Digital Humanities* 1: 309–22. doi:10.1007/s42803-019-00025-5.
- Vetter, Lara, Jarom McDonald, Sean Daugherty, Tanya Clement, Susan Schreibman, Roman Bleier und Joshua D. Savage. 2016-01-21. Dokumentation: Versioning Machine 5.0. Zugriff: 05. Januar 2021. <http://v-machine.org/documentation>.
- W3C OWL Working Group. 2012. „OWL 2 Web Ontology Language: Document Overview (Second Edition).“ W3C Recommendation 11 December 2012. Zugriff: 8. Juni 2018. <https://www.w3.org/TR/owl2-overview>.
- W3C SPARQL Working Group. 2013. „SPARQL 1.1 Overview: W3C Recommendation.“ Zugriff: 05. Januar 2021. <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>.
- Wout Dillen. 2016. „Sequentiality in Genetic Digital Scholarly Editions: Models for Encoding the Dynamics of the Writing Process.“ In *Digital Humanities 2016: Conference Abstracts*, hg. v. Maciej Eder und Jan Rybicki, 174–75. doi:10.17613/M6GB9B.

# **Digitale Editionen aus der Perspektive von Expert:innen und User:innen – Rezensionen der Zeitschrift RIDE im Meta-Review**

Claudia Resch

## **Abstract**

This article deals with the acceptance of digital editions among scholars. In order to identify recurring points of criticism of digital editions, seven volumes of the online journal *RIDE* (a review journal for digital editions and resources) published between 2014 and 2019 are analysed as a text corpus and subjected to a meta-review. By identifying specific lexemes and phrases that introduce or support critical statements, it can be shown in which aspects digital editions generally (still) do not sufficiently meet the assessment standards and expectations of their reviewers. The results of the analysis may provide recommendations for future digital editions.

## **Zusammenfassung**

Vorliegender Beitrag beschäftigt sich mit der Akzeptanz, die digitale Editionen in der Fachwelt erfahren. Um wiederkehrende Kritikpunkte an bislang publizierten digitalen Editionsprodukten zu ermitteln, werden sieben im Zeitraum von 2014 bis 2019 erschienene Bände der Online-Zeitschrift *RIDE* (A review journal for digital editions and resources) als Textkorpus analysiert und einem Meta-Review unterzogen. Durch die Identifikation spezifischer Lexeme und Phrasen, die kritische Äußerungen einleiten oder stützen, kann in der Zusammenschau gezeigt werden, in welchen Aspekten digitale Editionen den Beurteilungsmaßstäben und Erwartungshaltungen ihrer Rezensent:innen generell (noch) zu wenig entsprechen. Die Ergebnisse der Analyse sind dazu geeignet, Empfehlungen für künftige digitale Editionen abzuleiten.

*I do not doubt that the editors had a clear idea of what they were doing, but they have not documented it in the edition.*

---

Richard Hadden

Vorliegendem Beitrag geht die Beobachtung voraus, dass künftige User:innen bei der Planung und Erstellung von digitalen Editionen nur selten einbezogen und nach ihren grundsätzlichen Erkenntnisinteressen oder konkreten Forschungspraktiken gefragt werden. So arbeiten Editor:innen meist jahrelang und mit großer Sorgfalt an der digitalen Erschließung ihres Quellenmaterials, ohne abschätzen zu können, welche Akzeptanz ihre Edition erfahren wird. Ihre editorischen Anstrengungen werden indessen von einer hohen Erwartungshaltung seitens der Nutzer:innen begleitet, die sich – gerade durch den Einsatz digitaler Technologien – viel (mehr) von der entstehenden Edition versprechen.

Dabei steht außer Zweifel, dass sowohl Editor:innen als auch User:innen unter den Bedingungen des digitalen Wandels mit neuen Anforderungen konfrontiert sind. Gerade im Fall von digitalen Editionen hätten erstere – vielmehr als bei klassischen Printeditionen – die Möglichkeit, bereits im Entstehungsprozess und darüber hinaus auf Bedürfnisse von Nutzer:innenn einzugehen, sodass idealerweise sogar eine Interaktion stattfinden kann, wie sie von Patrick Sahle in seinem Werk zu digitalen Editionsformen beschrieben wird: „Es ist ein wechselseitiger Prozess: Veränderungen im Angebot und in der Nutzung bedingen sich gegenseitig und müssen gegenseitig aufgegriffen werden“ (Sahle 2013, Bd. 2, 268). Editor:innen könnten damit die Akzeptanz und Weiterverwendung der von ihnen erstellten digitalen Ressource sichern und die Edition als „Grundlagenarbeit für die weitere Forschung“ (vgl. Vogeler 2017, 12) ihrem eigentlichen Zweck zuführen. Im Gegensatz dazu können vergebene Chancen zur Interaktion zwischen Editor:innen einerseits und User:innen andererseits schlimmstenfalls die (für alle unerfreuliche) Ablehnung eines neuen digitalen Angebots bedeuten.

## 1 RIDE-Rezensionen im Meta-Review

Aus der Sicht der Wissenschaftlerin, die eigene digitale Editionen verantwortet und andere für ihre Erkenntnisinteressen nutzt, soll in diesem Beitrag ausgelotet werden, in welchen Bereichen die Ersteller:innen von digitalen Editionsprodukten noch stärker auf ihre Anwender:innen eingehen sollten, damit diese gerne und mit zufriedenstellendem Ergebnis davon Gebrauch machen. Um die wiederkehrenden Kritikpunkte an bislang entstandenen digitalen Editionen zusammenfassend zu ermitteln, sollen

im Rahmen dieser Untersuchung die bislang vorliegenden Rezensionen der Online-Zeitschrift *RIDE* (Akronym für: *Review Journal of the IDE*) einem Meta-Review unterzogen werden. Das vom Institut für Dokumentologie und Editorik (IDE) begründete Rezensionsorgan bespricht seit 2014 in regelmäßigen Abständen digitale Editionen und Textsammlungen und möchte damit – gemäß der Eigendefinition auf der Website – „ExpertInnen ein Forum zur kritischen Auseinandersetzung“ mit Editionen bieten und dazu beitragen, „die gängige Praxis zu verbessern und die zukünftige Entwicklung voranzutreiben“ (*RIDE* 2019). Der Anspruch, aus den einzelnen erschienenen Rezensionen also auch Implikationen für die Zukunft abzuleiten, wird in diesem Zitat besonders deutlich und kann durch den im vorliegenden Beitrag präsentierten methodischen Zugang und seinen Ergebnissen im günstigsten Fall präzisiert werden.

Die untersuchten *RIDE*-Rezensionen basieren auf einem ausführlichen Kriterienkatalog,<sup>1</sup> der auf den Seiten des IDE als Leitfaden online zur Verfügung steht. Er dient in erster Linie als „Bewertungsgrundlage für Gutachter“<sup>2</sup> (Henny 2018), formuliert aber auch Minimalanforderungen für zeitgemäße Editionen (vgl. Schnöpf 2013, 75). Mit diesem Bewertungssystem setzt das Institut auf „ein Wechselspiel verschiedener stärker und weniger stark formalisierter Evaluationsmethoden“ (Henny 2018): Zusätzlich zu einem formalisierten *Factsheet*, das von *RIDE* selbst für statistische Auswertungen und Datenvisualisierungen<sup>3</sup> herangezogen wird, reichen eingeladene Expert:innen ausführliche Rezensionen von jeweils mindestens 2000 Wörtern ein. Im Vergleich sind die in *RIDE* veröffentlichten Besprechungen also deutlich umfangreicher als jene anderer Fachzeitschriften (vgl. Henny 2018).

---

<sup>1</sup> Der Kriterienkatalog wurde als Richtlinie zur Besprechung und Evaluation digitaler wissenschaftlicher Editionen von Patrick Sahle und unter Mitarbeit von Georg Vogeler und den Mitgliedern des IDE erarbeitet. Die Version 1.1 von Juni 2014 steht unter <http://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/> (letzter Zugriff am 20. April 2020) in mehreren Sprachen zur Verfügung.

<sup>2</sup> Als GutachterInnen gefragt sind insbesondere Forschende mit „Doppelqualifikation, weil sie einerseits fachspezifische Inhalte beurteilen und andererseits den methodischen Rahmen und seine Umsetzung aus der Warte der ‚Digital Humanities‘ betrachten müssen“ (aus dem Vorwort). Eventuell wäre sogar von einer ‚Dreifachfunktion‘ zu sprechen, wenn man davon ausgeht, dass Rezensent:innen in der Begutachtung digitaler Editionen ja auch die Perspektive von User:innen einnehmen und vertreten – wenngleich „mit einer klaren individuellen Position“ (Henny 2018).

<sup>3</sup> Das IDE verwertet die Gesamtheit der erschienenen Rezensionen bislang dahingehend, dass einzelne Aspekte der sogenannten Factsheets in Diagrammen dargestellt werden – vgl. hierzu Philipp Steinkrüger: „Charts: Scholarly Editions.“ *RIDE*. <https://ride.i-d-e.de/data/charts/> (letzter Zugriff am 20. April 2020). „Die Auswertungen deuten an, welche Bereiche der digitalen Editorik – gemessen an der Zahl der entstandenen Editionen – bisher den Schwerpunkt ausmachen und wo der Raum für Entwicklungen – auch von Empfehlungen für Ersteller digitaler Editionen, Gutachter und Rezensenten – noch größer ist“ (Henny 2018).



Abbildung 1: Übersicht der Rezensionen im *Voyant Prospect Viewer*. Je höher der Balken, desto umfangreicher ist der Text der jeweiligen Rezension, wobei der kürzeste etwa 2.100 und der längste Text fast 11.000 Token zählt.

Die „prinzipiell freie Textform“ wird dabei zwar durch den oben genannten und Gutachter:innen empfohlenen<sup>4</sup> Kriterienkatalog thematisch geleitet, doch ist den RIDE-Herausgeber:innen daran gelegen, „die Diskussion der digitalen Editionen nicht zu sehr anzuleiten und einzuengen“ (Henny 2018), was bedeutet, dass Kritik frei verbalisiert werden kann. Diese offen gestalteten Textteile sind für vorliegende Untersuchung besonders interessant: Im Folgenden werden daher insbesondere jene Zitate der Rezendent:innen ausgewertet und klassifiziert, die sich mit nennenswerten Mängeln und davon abgeleiteten Verbesserungsvorschlägen digitaler Editionen auseinandersetzen.<sup>5</sup> Hierfür sollen genau jene Passagen identifiziert werden, in welchen die Rezendent:innen ihr Expertenurteil abgeben und dabei (auch stellvertretend für andere User:innen) formulieren, an welchen neuralgischen Punkten sie anderes erwarten, irregleitet werden oder uninformativ und mit dem Quellenmaterial auf sich gestellt bleiben. Neben diesen Kritik enthaltenden Textstellen wird in der Auswertung besonders auf jene geachtet, die Überlegungen dazu enthalten, welche verbessern den Schritte aus Sicht der Expert:innen zu setzen wären und wie man dadurch die Akzeptanz von spezifischen digitalen Editionen in Zukunft begünstigen könnte.

## 2 Methodik: 35 Rezensionen als Untersuchungskorpus

Zum Zeitpunkt der Untersuchung waren im Zeitraum zwischen 2014 und 2019 bereits sieben Bände mit insgesamt 35 Rezensionen in deutscher oder englischer Sprache

<sup>4</sup> Vgl. die Hinweise zum Verfassen und Einreichen von Rezensionen unter: <https://ride.i-d-e.de/reviewers/submission-guidelines/> und <https://ride.i-d-e.de/reviewers/writing-guidelines/> (Zugriff am 20. August 2022).

<sup>5</sup> Von der Untersuchung aller bislang rezensierten Editionen soll jedoch klarerweise kein Anspruch auf allgemeine Repräsentativität abgeleitet werden.

zu „Wissenschaftlichen Editionen“<sup>6</sup> unter ride.i-d-e.de/ veröffentlicht, die auf GitHub unter <https://github.com/i-d-e/ride> zum Download zur Verfügung stehen. Diese gesammelten Rezensionen ergeben ein Untersuchungskorpus im Umfang von etwa 155.000 Token, das sich noch weiter reduzieren ließe, wenn man davon ausgeht, dass für ‚Verbesserungsvorschläge‘ beziehungsweise *Suggestions for improvement* laut Kriterienkatalog insbesondere Punkt 5.7 im Fazit vorgesehen ist. Allerdings zeigt ein exemplarisches *Close Reading* mehrerer Rezensionen, dass Rezensent:innen ihre Kritikpunkte nicht nur – wie im Katalog empfohlen – abschließend unter Punkt 5.7 anbringen, sondern gegebenenfalls auch an früherer Stelle und unter Aufhebung der von RIDE vorgeschlagenen Anordnung nach Themenpunkten auf diese zu sprechen kommen. Vor allem deshalb schien es also sinnvoll, das Gesamtkorpus im Volltext auszuwerten und darin jene Lexeme und Phrasen zu identifizieren, die signifikant häufig kritische Äußerungen einleiten oder stützen und daher für diese Untersuchung von besonderer Relevanz sind. Damit soll letztlich auch die Methodik dieses Meta-Reviews dargelegt und dessen Ergebnis nachvollziehbar werden.

Um in allen Rezensionen zeitsparend und gezielt nach solchen Äußerungen suchen zu können und verlässlich an relevante Textstellen zu gelangen, wurden die einzelnen Texte als *plain text files*<sup>7</sup> in das Korpusanalyseprogramm *AntConc*<sup>8</sup> eingespeist. Auf Basis der sorgfältigen Lektüre exemplarischer und der kurorischen Lektüre weiterer Rezensionen wurden zuerst mögliche Kandidaten<sup>9</sup> für die Suche sowohl in deutscher und als auch in englischer Sprache identifiziert. Unter diesen im Korpus abgefragten Lexemen und Lexemverbindungen waren beispielsweise sowohl auf Fehler oder Mängel verweisende Nomen wie *problem*, *limitation*, *inconvenience* oder *shortcoming* als auch deren deutsche Entsprechungen. Außerdem wurde nach der Phrase *lack of* gesucht, welche zu folgenden Ergebnissen führte.

Abbildung 2 zeigt die Konkordanzen der gesuchten Phrase *lack of* im *Key Word In Context*-Format (KWIC). Nach dem ersten rechten Nachbarn über *KWIC-Sort* alphabetisch geordnet, kann diese Phrase nun in ihrer direkten kontextuellen Umgebung betrachtet werden.

<sup>6</sup> Die Bände zu „Textsammlungen“ werden nicht in die Untersuchung einbezogen, weil hierfür (wie auch zur Begutachtung von *Tools and Environments*) ein anderer Kriterienkatalog vorgesehen ist als für „Wissenschaftliche Editionen“.

<sup>7</sup> Möglich wären freilich auch Einzelabfragen über den *search ride*-Button auf der Website von RIDE, allerdings bieten die Rohdaten entsprechend bessere Auswertungsoptionen.

<sup>8</sup> Abrufbar unter <http://www.laurenceanthony.net/software/antconc/> (letzter Zugriff am 20. April 2020).

<sup>9</sup> So lautete die Liste an englischen Keywords (in alphabetischer Reihenfolge): *confusing*, *could*, *critique*, *cumbersome*, *desirable*, *difficult*, *flaw*, *has to*, *hopefully*, *improve*, *inconvenience*, *is forced to*, *issue*, *lack of*, *limitation*, *not able to*, *not ideal*, *odd*, *problem*, *sadly*, *shortcoming*, *should*, *unclear*, *uncomfortable*, *unfortunately*, *wish* und *would* – ebenfalls abgefragt wurden deren deutsche Übersetzungen. Belegstellen, die auf diese Weise gefunden wurden, sich für die Fragestellung aber als nicht relevant erwiesen, mussten ausgeschlossen werden.

of this outdatedness is its lack of a straightforward, consistent, and comprehensive navigation especially problematic is the lack of a synoptic line-up of manuscript and transcription. That is certainly true. The lack of a table of contents, however, as you would need to counterbalance the lack of an independent study, although it is fair to note that flaw, however, is the lack of any possibility to re-use the online material 918, nor performed. This lack of attention from both theatre production companies and publishers, factual errors, the lack of clear bibliographic citation for the various editions, the lack of this edition is the lack of complementary texts, such as a document explaining the denied, even if there is a lack of creation of digital tools that are useful for some navigational issues, lack of documentary overview and questionable points of reference. At the same time, the lack of documentation and explicitly stated editorial methodology regarding the edited text. This lack of documentation, no doubt, is one of weakest points on the 'Links' page. The lack of documentation on the technical labor and the single team. Although there is a lack of documentation regarding the methodology followed in the general. The aforementioned lack of downloadable TEI files of the transcriptions makes it

Abbildung 2: Ausschnitt der Key Word In Context-Abfrage zu *lack of* in AntConc.

Wenn Rezendent:innen ihre individuellen, mitunter negativen Erfahrungen mit der jeweiligen Edition beschreiben, verwenden sie oftmals auch Adjektive wie *confusing* oder *unclear* und benennen einzelne Schritte im Arbeitsprozess, welche sie als schwer oder problematisch empfunden haben (vgl. die Phrasen in Abbildung 3: *difficult to access / to compare / to find / to guess / to identify / to process usw.*).

Abgefragt wurden des Weiteren Adverbien wie *unfortunately* oder *leider* sowie weitere Ausdrücke des Bedauerns, die verlässlich zu jenen Textstellen führen, welche aus Sicht der Rezendent:innen Unerfreuliches enthalten. Hierzu zählt auch das etwas positivere *hopefully* beziehungsweise *hoffentlich*, dem zumeist Phrasen zu möglichen Verbesserungsmaßnahmen folgen. Mit der Suche nach dem Wortstamm *expect* respektive *erwart* gelangt man darüber hinaus zu jenen für die Untersuchung aussagekräftigen Textstellen, an denen Rezendent:innen über ihre Annahmen schreiben und darlegen, wo ihren Erwartungen nicht entsprochen wurde – in Form von Wendungen wie *However, one would expect that ..., Rather unexpectedly, ...* oder *XY seems not to fulfill the expectations*.

Um neben den Erwartungen auch Wünsche der Rezendent:innen in der Textmenge fassbar zu machen und weitere konkrete, verbessende Maßnahmen zu identifizieren, lohnt sich außerdem die systematische Abfrage nach dem Modalverb *would*<sup>10</sup> beziehungsweise seinem deutschen Äquivalent. Konkrete Vorschläge oder alternative Vorgehensweisen verbergen sich auch in der dezenten Phrase *could also* und in jenen

---

<sup>10</sup> So führen Phrasen wie [...] *would have been nice*, *It would be good to [...]* oder *It would be ideal to [...]* in vielen Fällen zu Empfehlungen und Ratschlägen.

difficult to access the whole of the dataset, diminishing the  
difficult to access in by any other means, including in-  
difficult-to-access collections around the world. Using an  
difficult to add the translations in situ: the setting describes  
difficult to add the translations in situ: the setting  
difficult to assess the quality of the information since it  
difficult to capture the tentativeness of possible textual links  
difficult to compare versions of the text in their entirety.  
difficult to decipher, since, due to the age and materiality  
difficult to determine (at least, without Godwin giving an  
difficult to find). One may be able to generate the  
difficult to find and it corresponds to the critical text  
difficult to find). One may be able to generate the  
difficult to go from an encoded text to an edition  
difficult to 'guess' which reading is the original and correct  
difficult to identify the website as either an online archive or  
difficult to interpret. This factor can be attributed to the  
difficult to know where gaps have been filled, how those  
difficult to know where gaps have been filled, how those  
difficult to locate. Identifiers, Dates, Locations, etc.

difficult to process some of the information presented to  
difficult to process because units like sentences or words  
difficult to read. One of only three word glosses, the  
difficult to reconstruct linear text. The TEI encoding  
difficult to reuse work from other scholars. The fact that  
difficult to run old SGML-based software on modern systems  
difficult to transition back and forth between persons involved.

Abbildung 3: Ausschnitt der Key Word In Context-Abfrage zu *difficult to* in *AntConc*.

Phrasen, die den Wortstamm *improve* beziehungsweise *verbesser* enthalten und die jeweils rezensierte Edition mit Hinweisen und Empfehlungen ausstatten. Als wenig sinnvoll erweist sich dagegen die Suche nach *user experience*, weil Rezendent:innen ihre Erfahrungen zwar formulieren, meist jedoch nicht unter expliziter Nennung dieses Stichworts.

Zu all diesen korpusbasierten Abfragen ist festzuhalten, dass freilich jede Belegstelle in ihrem Kontext betrachtet werden muss, um bewertet werden zu können. Hierfür lässt sich in *AntConc* unter den Einstellungen entweder das Fenster der Kontextanzeige erweitern oder man gelangt mit einem Klick auf die Belegstelle wieder in die jeweilige Textdatei und damit in den längeren Textzusammenhang. Mit den genannten und weiteren Stichworten lässt sich aller Voraussicht nach nicht jede Art von Kritik an

der Textoberfläche fassen, zumal es sich um narrative Texte von professionellen Schreiber:innen handelt, die ihre Kritik in vielfacher Art und Weise zu formulieren gewohnt sind. Dennoch geht die Autorin davon aus, dass durch die Kombination von sorgfältiger Lektüre und punktueller Suche ausreichend Textstellen identifiziert werden können, um zu aussagekräftigen Auswertungen<sup>11</sup> zu kommen, wie sie im folgenden Kapitel nun zusammengefasst werden.

Dem ist nur noch vorauszuschicken, dass es in dieser Untersuchung nicht darum geht, einzelne Produkte einer nochmaligen Prüfung zu unterziehen, sondern in der Zusammenschau zu beurteilen, in welchen Bereichen die besprochenen Editionen generell (noch) nicht den Erwartungen der Rezendent:innen entsprochen haben.<sup>12</sup>

### **3 Ergebnisse aus den untersuchten Rezensionen**

Um konkreter feststellen zu können, welche Bereiche insbesondere von der Kritik der Rezendent:innen betroffen sind, wurde die Fülle von Belegstellen einem oder mehreren der folgenden sechs – am RIDE-Kriterienkatalog orientierten – Bereiche<sup>13</sup> zugeordnet:

1. Präliminarien und Kontext
2. Inhalte und Auswahl
3. Methodik
4. Dokumentation und Paratexte
5. Umsetzung und Präsentation
6. Nachnutzbarkeit

Zu Punkt 1 gibt es in den untersuchten Editionen – vor allem verglichen mit den noch folgenden Punkten – verhältnismäßig wenige kritische Äußerungen. In Einzelfällen würden Rezendent:innen gerne mehr über den Entstehungsprozess erfahren („one desires to be able to read more about the thinking and process behind this work“<sup>14</sup>) sowie über die grundlegenden Ziele („more extensive overview of the principles and goals“<sup>15</sup>

<sup>11</sup> Für die Unterstützung beim Auswerten der Belegstellen danke ich Nina C. Rastinger.

<sup>12</sup> Um den Eindruck von Einzelbewertungen zu vermeiden, waren in einer ersten Fassung des Beitrags Belege aus dem Korpus den individuellen Editionen nicht näher zugeordnet; aus Gründen der Transparenz sind sie jetzt aber wie Zitate behandelt und daher in den Fußnoten ausgewiesen.

<sup>13</sup> So betrifft zum Beispiel der Vorschlag eines *layer-models* sowohl die Methodik als auch die Umsetzung und Präsentation einer digitalen Edition.

<sup>14</sup> Andrew Dunning über Petrus Plaoul. *Commentarius in libros Sententiarum: Editiones electronicas*.

<sup>15</sup> Frederike Neuber über *The Shelley-Godwin Archive*.

„more information on the development of the project“<sup>16</sup>), allerdings lässt sich dies nicht verallgemeinern. Interessant ist dennoch, dass mehrere Rezensionen einen Mangel an theoretischer Reflexion beobachten, was die Textgenese und -gestalt angeht.

Ähnlich zurückhaltend wie Präliminarien und Kontext werden Inhalte und Auswahl digitaler Editionen kommentiert: Nur gelegentlich wird ein weiterer Textzeuge zur Aufnahme empfohlen oder auf ähnliche, ebenso bedeutsame Quellen verwiesen, die man als Kontextmaterial in die jeweilige Edition auch hätte einbinden können.

Mehr Kritik findet sich hingegen im Bereich der Methodik: Laut untersuchter Rezensionen lassen hier mehrere Editionen die Darlegung eines Konzeptes (*editorial statement*) vermissen. So heißt es etwa seitens der Verfasser:innen: „Für den editorischen Bereich fehlen jedenfalls die editorischen Richtlinien“<sup>17</sup> beziehungsweise „it [the edition] lacks clear documentation concerning textual editorial methodology“<sup>18</sup> – womit insbesondere kritisiert wird, wenn editorische Entscheidungen entweder nicht vorhanden oder für Außenstehende nicht ausreichend transparent sind. Dieses Informationsdefizit, das den Ersteller:innen von Editionen oft gar nicht bewusst ist, formuliert ein Reviewer so deutlich, dass sein Zitat vorliegendem Beitrag in gekürzter Form als Motto vorangeht. Es lautet in der Langversion: „I do not doubt that the transcribers and editors had a clear idea of what they were doing, but they have not documented it in the edition and so the user (and the reviewer) can only retrospectively deduce what the idea might have been.“<sup>19</sup> Auf ähnliche methodische Schwächen verweisen Rezendent:innen auch, wenn sie den Eindruck haben, dass keine weitreichende inhaltliche Erschließung erfolgt ist oder weiterführende Kommentierungen ausstehen. In konkreten Fällen wären etwa Quellenverweise und Kommentare sowie Indizes (z. B. von Personen, Sachen und Werken) willkommen gewesen. Diese sehr individuellen Hinweise zur Erschließung beziehen sich freilich nur auf einzelne Projekte und lassen sich nur begrenzt auf andere übertragen.

Was hingegen editionsübergreifend eingefordert wird, ist die Möglichkeit, Quellen in verschiedenen Formaten und Ansichten verfügbar zu halten. Hier sprechen mehrere Reviews von weiteren Textschichten beziehungsweise einem *layer-model*, welches mehrere parallele Sichtweisen auf Texte bietet. Darunter fällt letztlich auch die Beigabe von qualitativ hochwertigen Faksimiles, deren Fehlen gegebenenfalls beanstandet wird („*inavailability of digital facsimiles*“<sup>20</sup>).

Ein ebenfalls stark kritisierte und von Editor:innen offenbar vernachlässigte Bereich sind Dokumentation und Paratexte. Dabei haben Rezendent:innen vor allem jene User:innen vor Augen, welche mit digitalen Editionen noch weniger vertraut sind

<sup>16</sup> Helmut W. Klug über *Welscher Gast digital*.

<sup>17</sup> Friederike Wein über *Der Zürcher Sommer 1968: Die digitale Edition*.

<sup>18</sup> Richard Hadden über *William Godwin's Diary*.

<sup>19</sup> Ebenda.

<sup>20</sup> Magdalena Turska über *The Correspondence of James McNeill Whistler*.

und umfassender Informationen bedürfen, um sich zurecht zu finden („*to guide a user not familiar with the project*“<sup>21</sup>). Die in den Rezensionen geforderte Dokumentation bezieht sich primär auf zwei Aspekte: erstens auf editorische Entscheidungen bei Transkription und Annotation und zweitens auf technische Implementierungen. Sind hierzu keine (leicht auffindbaren) Richtlinien vorhanden, kritisieren Gutachten „stillschweigende Eingriffe in den Text“<sup>22</sup>, fragen nach nachvollziehbaren Konzepten zum Mark-up und der „Offenlegung der Auszeichnungsregeln“<sup>23</sup> und stellen dann Folgendes fest: „*the project lacks easily accessible documentation of the textual tradition chosen, as well as the transparency of the editorial decisions made.*“<sup>24</sup> Genauso wird auf technische Funktionalitäten hingewiesen, die nicht beschrieben sind beziehungsweise das Fehlen eines detaillierten Abschnittes zur technischen Umsetzung und Architektur der Website moniert („*more information about the technical architecture of the website*“<sup>25</sup>).

Mehrere dieser Rezensionen, die eine ausführliche Ergänzung der Dokumentation fordern, stellen den Editor:innen gleichzeitig eine maßgebliche Verbesserung ihrer Produkte in Aussicht, wie aus den folgenden beiden Zitaten deutlich hervorgeht: „*The addition of documentation and description of the text encoding could contribute significantly to the improvement of the edition*“<sup>26</sup> schreibt etwa eine Rezensentin, während es an anderer Stelle heißt: „*The expansion of the documentation and a transcription of the encoding model of the text would notably increase the value of this edition.*“<sup>27</sup> – worin sich die Tatsache spiegelt, dass digitale Editionen noch lange nicht selbsterklärend sind.

Ein weiterer, in den untersuchten Rezensionen besonders frequent kommentierter Bereich betrifft die Umsetzung und Präsentation von digitalen Editionen. Obwohl natürlich jede Edition anders gestaltet ist, lässt sich hier doch beobachten, wie wichtig unter anderem die Struktur eines Interfaces ist, um User:innen den Zugang zu den Daten zu erleichtern und ihnen „*navigational issues*“<sup>28</sup> zu ersparen. Diesbezüglich bieten die Expert:innen konkrete Verbesserungsvorschläge, indem sie den Gebrauch von Sitemaps, schlüssige Navigationswege („*more consistent routes of navigation*“<sup>29</sup>) und klare Anleitungen („*clear instructions on how to use the site*“<sup>30</sup>) empfehlen. Sie

<sup>21</sup> Ulrike Henny-Krahmer über *P. S. Post Scriptum*.

<sup>22</sup> Ulrike Henny-Krahmer über *Leal Conselheiro*.

<sup>23</sup> Friederike Wein über *Der Zürcher Sommer 1968: Die digitale Edition*.

<sup>24</sup> Merisa A. Martinez über *The Fleischmann Diaries*.

<sup>25</sup> Ulrike Henny-Krahmer über *P. S. Post Scriptum*.

<sup>26</sup> Susanna Allés Torrent über *La entretenida by Miguel de Cervantes: A Digital, Annotated Edition and an English Translation*.

<sup>27</sup> Antonio Rojas Castro über *La dama boba: edición crítica y archivo digital*.

<sup>28</sup> Tessa Gengnagel über *Faust Edition*.

<sup>29</sup> Aodhán Kelly über *Digital Thoreau*.

<sup>30</sup> Kendal Crawford und Michelle Levy über *The William Blake Archive*.

geben allerdings auch ihre ersten Eindrücke von Interfaces wieder („verwirrend“<sup>31</sup>) und beklagen, wenn relevante Informationen nicht auffindbar sind („buried in the table of contents“<sup>32</sup>). Zudem bedauern sie, wenn weniger vorteilhafte Entscheidungen getroffen worden sind („some counter-intuitive interface choices“<sup>33</sup>), und propagieren ein – freilich unwidersprochenes und erstrebenswertes – Ideal der intuitiven Benutzerführung. Ein anderes mehrfach besprochenes Thema ist die Suchfunktion, da diese entweder nicht vorhanden ist oder Insuffizienzen und Einschränkungen aufweist, welche sich aber etwa durch den Einsatz von Filteroptionen beheben lassen würden. Erneut wird in diesem Zusammenhang die fehlende Erläuterung von Suchmaschinen erwähnt.

Nicht zuletzt hängen die Möglichkeiten zur Umsetzung einer digitalen Edition mit der bereits genannten Modellierung und Erschließung zusammen – als Ideal gelten „verschiedene Präsentationsmodi“<sup>34</sup> wie es in einer der Rezensionen heißt. Dabei ist es für User:innen von Vorteil, wenn ihnen mehrere Textversionen parallel angezeigt werden (*side-by-side text* oder *synoptic view*), um sie miteinander vergleichen zu können, worauf in den Rezensionen mehrfach hingewiesen wird.

Obwohl der Kriterienkatalog an sich keine Bewertung des graphischen Designs von Oberflächen vorsieht, vermerken Rezensent:innen immer wieder, wenn ihnen etwas missfällt („*simplistic interface*“; „*slightly dated layout*“<sup>35</sup>), und geben Hinweise, wo eine zeitgemäße Aktualisierung sinnvoll erscheint („*It would be ideal if the website were re-organized and re-designed*“<sup>36</sup>). Wenig Anklang in der Begutachtung finden insbesondere jene digitalen Editionen, die nicht regelmäßig auf ihre Funktionalitäten getestet werden („*broken links*“<sup>37</sup>) oder sich zu sehr an bestehenden Druckerzeugnissen orientieren („*digital proxies of existing print works*“<sup>38</sup>), wodurch sie das Potential des digitalen Paradigmas nicht annähernd ausschöpfen.

Der sechste und letzte Punkt betrifft die Nachnutzbarkeit von digitalen Editionen. Die Rezensent:innen sprechen sich hier ganz generell für die Bereitstellung der Daten in mehreren Formaten aus, wobei sich zahlreiche Hinweise konkret auf XML-Files beziehen, welche User:innen viele Möglichkeiten der Nachnutzung eröffnen. So wird etwa nach der Bereitstellung von Grunddaten in XML oder der Bereitstellung des Schemas und der XML-Quelltexte verlangt (vgl. „*access to the XML files*“<sup>39</sup>; „*down-*

<sup>31</sup> Maja Hartwig über *Mozart Briefe und Dokumente – Online-Edition*.

<sup>32</sup> Misha Broughton über *A London Provisioner's Chronicle, 1550–1563*.

<sup>33</sup> Ebenda.

<sup>34</sup> Ulrike Henny-Krahmer über *Leal Conselheiro*.

<sup>35</sup> Magdalena Turska über *The Correspondence of James McNeill Whistler*.

<sup>36</sup> Elise Hanrahan über *Mark Twain's Letters, 1853–1880*.

<sup>37</sup> Mellissa Hinton über *Crossing Brooklyn Ferry: An Online Critical Edition*.

<sup>38</sup> Misha Broughton über *A London Provisioner's Chronicle, 1550–1563*.

<sup>39</sup> Philipp Steinkrüger über *Nietzschesource*.

*loadable XML/TEI files*“<sup>40</sup>, „allow greater access to their core markup“<sup>41</sup> u. v. m.). Diese nachdrücklichen Appelle bestätigen nicht nur, dass großes Interesse daran bestünde, die Daten der untersuchten Editionen nachzunutzen, sondern machen auch darauf aufmerksam, dass ein derart offener Umgang mit dem edierten Material in der Praxis noch lange nicht Usus ist: So zeigen die eingangs erwähnten Diagramme und Charts von Philipp Steinkrüger, die RIDE selbst auf Basis der *Factsheets* veröffentlicht, dass der überwiegende Teil der Editionsprojekte bislang eben noch keine Daten – sei es vollständig oder in Auszügen – zur Verfügung stellt.<sup>42</sup> Das ist wohl mit ein Grund, weshalb Rezensent:innen auch stellvertretend für die User:innen darauf drängen, einen möglichst uneingeschränkten Zugriff auf die Daten zu bekommen, und auf diesbezügliche Versäumnisse hinweisen, wie es etwa an folgenden Stellen der Fall ist: „*The edition offers no access to its original document-descriptive data encoding, no API for interfacing the text with other corpora or critical engines, and no explanation for or explication of these lacks.*“<sup>43</sup> Wiederholte taucht zudem die Frage auf, ob es möglich wäre, „Daten herunterzuladen oder über eine API zu beziehen. Daran mangelt es zurzeit.“<sup>44</sup> So suchen Rezensent:innen häufig umsonst nach Optionen, Daten vom Interface herunterladen zu können, jedoch: „*To do so would require software and training to sift through the raw XML files [...] or some kind of technical interface or API, which currently does not exist.*“<sup>45</sup>

Der potentielle Mehrwert für die Forschung wird in diesen Empfehlungen deutlich formuliert: „Die Zurverfügungstellung der Transkriptionen in XML oder einem anderen für die Nachnutzung der Daten geeigneten Format wäre wünschenswert und wertvoll“,<sup>46</sup> heißt es etwa in einem der Gutachten. Als Voraussetzung hierfür werden die Editorinnen bzw. Editoren der untersuchten Projekte mehrfach daran erinnert, für ihre Daten offene Lizenzen vorzusehen („*An open licence would have been desirable for data reuse.*“<sup>47</sup>) beziehungsweise die Bedingungen zur Weiternutzung unmissverständlich darzulegen („*vague licensing terms*“<sup>48</sup>).

---

<sup>40</sup> Francisco Javier Álvarez Carballo über *The digital edition of the Becerro Galicano de San Millán de la Cogolla*.

<sup>41</sup> Misha Broughton über *A London Provisioner’s Chronicle, 1550–1563*.

<sup>42</sup> Vgl. insbesondere die beiden Diagramme zu den Fragen: *Is the raw data accessible, either for the individual parts of the edition or as a whole?* und *Can the raw data of the edition be downloaded (as a whole)?* Philipp Steinkrüger: Charts: Scholarly Editions. RIDE <https://ride.i-d-e.de/data/charts/>. Zugriff 20. April 2020).

<sup>43</sup> Misha Broughton über *A London Provisioner’s Chronicle, 1550–1563, by Henry Machyn*.

<sup>44</sup> Anna Busch über *EMA – Erich Mendelsohn Archiv. Der Briefwechsel von Erich und Luise Mendelsohn 1910–1953*.

<sup>45</sup> Joshua Kruchten über *The Casebooks Project*.

<sup>46</sup> Martina Scholger über *Paul Klee – Bildnerische Form- und Gestaltungslehre*.

<sup>47</sup> Ulrike Henny-Krahmer über *P. S. Post Scriptum*.

<sup>48</sup> Christof Schöch über *Théâtre classique*.

## 4 Fazit und Empfehlungen

Aus eigener langjähriger Erfahrung weiß die Autorin, wie schwierig es ist, digitale Editionen aufzubauen. Es wäre daher kurzsichtig, die punktuellen Defizite einzelner Editionsprojekte exemplarisch ins Treffen zu führen, denn schließlich sind diese unter ganz unterschiedlichen Voraussetzungen entstanden und nach Maßgabe ihrer individuellen Möglichkeiten, was den personellen, finanziellen und infrastrukturellen Aufwand betrifft. Anhand der zusammengeführten Ergebnisse sollte aber zumindest deutlicher geworden sein, wo Editionsprojekte in der Praxis derzeit stehen. Mit der – zugegebenermaßen: wenig populären – *ex negativo*-Erhebung lässt sich jedenfalls besser abschätzen, welche Komponenten digitaler Editionen bislang aus Sicht der Rezendent:innen (noch) unzureichend behandelt oder ganz vernachlässigt wurden. Die enthaltenen Verbesserungsvorschläge sind daher nicht an einzelne Projekte adressiert, sondern geben sachdienliche und richtungsweisende Hinweise darauf, welche Aspekte digitaler Editionen generell unverzichtbar sind, wo weiterhin nutzer- und nutzerinnenzentrierte Arbeit zu leisten wäre und wie sich die digitale Editionslandschaft aus der Perspektive von Expert:innen und User:innen in Zukunft weiterentwickeln könnte.

Die Ergebnisse zeigen etwa sehr deutlich, wie wichtig es ist, dass Editor:innen jene Überlegungen und Methoden, welche zu ihren Produkten führen, explizit beschreiben und nicht User:innen rückblickend darüber mutmaßen lassen, welche Idee hinter ihrer Arbeit gestanden haben könnte. Diese Art der verortenden „Selbstdefinitorik“ (vgl. Schnöpf 2013, 72) kann dabei nur unter Bezugnahme auf künftige User:innen erfolgen: „*who are ,they‘, the people one hopes will receive a project, a post, a game, an archive, an interactive experience?*“ fragt Kramer (2016) und spricht zugleich auch den Wunsch nach Akzeptanz an, den Editor:innen bei der Erstellung ihrer Produkte hegen: „*How do makers wish to treat their audiences?*“

In der Pluralisierung der zitierten *audiences* wird unter anderem erkennbar, dass die Zielgruppen im digitalen Medium potenziell um ein Vielfaches diverser sind, als es bei Printpublikationen jemals der Fall gewesen ist. Den Nutzer:innenn werden bestimmte Kompetenzen zugeschrieben (vgl. etwa *basic users*, *advanced users* und *meta-users* bei Dillen (2018, 40) oder ein ähnliches Konzept bei Rasmussen (2016, 127)) und damit einhergehend unterschiedliche, auch fachdisziplinär geprägte Informationsbedürfnisse, die berücksichtigt werden wollen. Editor:innen haben es demzufolge zunehmend schwer(er), eine digitale Edition zu konzipieren, deren Akzeptanz in allen Zielgruppen wahrscheinlich ist. Dabei werden der Anspruch, „etwas für alle“ (Resch 2017) zu bieten, und Harveys (2001, 11) ehemalige Vorgabe („*a proper edition should be adequate for them all, and for many others besides*“) inzwischen kontrovers diskutiert, wie etwa bei Dillen (2018, 56), der davor warnt, allen Wünschen gerecht werden zu

wollen: „*The temptation here is to shower the user with possibilities, to try to cater to her every need, and to make all these different options available to her at all times.*“

Um die gewünschte Akzeptanz von digitalen Editionen dennoch zu fördern, scheint es sinnvoll, künftige Userinnen- und User(gruppen) nach ihren Erwartungshaltungen und Nutzungsintentionen, die der eigentlichen Nutzung vorgelagert sind, zu fragen. Der von User:innen wahrgenommene Nutzen einer Edition ist schließlich einer der ausschlaggebenden Faktoren für deren tatsächliche Verwendung: „*If digital resources fit well with what they want to do with them, users will adopt them*“, so Warwick (2012, 18).

Bislang werden künftige User:innen – wie eingangs festgestellt – bei der Planung und Erstellung von digitalen Editionen aber nur selten einbezogen.<sup>49</sup> Dabei wäre die Anwendung von – im Grunde naheliegenden, aber kaum getesteten (Shaw u. Milewicz 2017) – partizipativen Methoden in mehreren Phasen eines Editionsprojektes denkbar und möglich. Je früher eine erste Interaktion zwischen Editor:innen und User:innen stattfindet, desto eher könnten theoretisch beide Seiten profitieren. Allerdings zeigt sich in der Praxis, dass Projekte, sofern sie sich überhaupt auf eine Interaktion einlassen, damit eher zuwarten: „*Too often user testing, both in academic projects and industry, is left until late in the project; users are only asked for their opinion when the resource is built and a prototype is being tested*“ (Warwick 2012, 4).<sup>50</sup> Dann freilich kann es für (zum Teil zeit- und personalintensive) Anpassungen bereits zu spät sein.

Von den Ergebnissen solcher anwenderzentrierten Überlegungen ist letztlich abhängig, wie User:innen an das Quellenmaterial herangeführt werden können, wie ‚präsent‘ Editor:innen dabei sein müssen (vgl. Dillens „*different degrees of editorial presence and regulation*“ 2018, 56) und wie sehr sie ihre Argumente und ihre Sichtweise auf das Material vermitteln sollen – etwa durch ein hierfür wesentliches Interface (vgl. Andrews u. van Zundert 2018, 4). Weil dieses von User:innen als erstes wahrgenommen wird, empfiehlt sich seine gleichermaßen attraktive wie intuitive Gestaltung.

<sup>49</sup> Eine Ausnahme ist das go!digital 2.0-Projekt *Das Wien[n]erische Diarium: Digitaler Datenschatz für die geisteswissenschaftlichen Disziplinen*, das von der Autorin am Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) an der Österreichischen Akademie der Wissenschaften geleitet wird (Resch 2018 und 2019).

Es empfiehlt sich jedenfalls, tatsächlich Außenstehende für Feedback zu gewinnen und nicht den ohnehin im Projekt Involvierten Rollen zuzuweisen, wie Hewing/Mandl/Womser-Hacker (2016, 5) es versucht haben: „*In these workshops, the two historians working in the project represent the user group.*“ Ganz ähnlich sieht das Claire Warwick (2012, 9), wenn sie von Doppelfunktionen abrät: „*the model of designer-as-user is not advisable*“.

<sup>50</sup> Claire Warwick (2012, 1) verweist in diesem Zusammenhang auf – mittlerweile revidierte! – Vorannahmen, die eine Einbeziehung künftiger User:innen wenig aussichtsreich erscheinen ließen: „*there was little point asking them [users] what they needed, because they would not know, or their opinion about how a resource functioned, because they would not care. It was also assumed that technical experts were the people who knew what digital resources should look like, what they should do and how they should work.*“

Diese will sorgfältig überlegt sein, weshalb Fred Gibbs' und Trevor Owens' (2012) diesbezüglicher Rat lautet: „*Allocate more resources to user interface development. The user interface for many digital projects often seems developed as an afterthought, thrown together after completing the core functionality.*“

Konsens besteht darüber, dass User:innen vermutungsweise wenig Zeit aufwenden (wollen), um sich in Editionen einzuarbeiten: „*Most users, especially humanities academic users*“, schrieb Warwick (2012, 9) noch vor einigen Jahren, „*do not want to have to be trained to use digital resources, regarding it as a waste of time.*“ Das mag sich inzwischen geändert haben, doch lautet auch heute die entscheidende Frage, wie Nutzer:innen in Zukunft an digitale Editionen herangeführt werden sollen, denn „[k]omplexe digitale Ressourcen stellen auch an die Benutzer höhere Anforderungen“ (Sahle 2013, Bd. 2, 262). Um den methodischen Wert einer Edition zu erklären, empfehlen sich Fallstudien („*clear—but not overly simplistic—sample use cases*“, vgl. Gibbs u. Owens 2012), die User:innen anhand von konkreten Anwendungsszenarien Hinweise geben, wie eine Edition zu nutzen ist. Idealerweise stammen diese Beispiele aus unterschiedlichen Disziplinen. Die Kosten hierfür wären zwar nicht gering, umso größer aber der Nutzen: „*The cost outlay to create such content is not negligible, but the benefit would be disproportionately high*“ (Gibbs u. Owens 2012).

Ergänzend zu einer Optimierung des Interfaces besteht – wie in den untersuchten Rezensionen häufig nachgefragt – die Möglichkeit, weiterverwendbare Daten zur Nachnutzung zur Verfügung zu stellen. Hierbei haben sowohl diese Auswertung als auch die von *RIDE* erstellten Diagramme deutlich gezeigt, wie sehr die begutachteten Editionen in diesem Punkt noch vom Ideal abweichen. Diese Diskrepanz ist bemerkenswert, zumal in den letzten Jahren<sup>51</sup> dahingehend bereits Veränderungen stattgefunden haben, wie sie Roman Bleier und Helmut Klug (2018, VII) beschreiben:

Over the last decade the already mentioned APIs [application programming interface] have gained importance for DSEs [Digital Scholarly Edition] as editors increasingly see their editions not only as static texts published online, but as data that can be linked to other data to answer interesting research questions.

Wohl ist es so, dass nicht nur Editor:innen ihre Daten anschlussfähig machen wollen, sondern auch User:innen großes Interesse an der (über die ursprünglich intendierte Verwendung hinausgehende) Nachnutzung bekunden, was entweder durch den vollständigen Download der zugrundeliegenden Daten oder durch das Anbieten von

---

<sup>51</sup> In diesem fünf Jahre umfassenden Meta-Review wurden die jeweiligen Erscheinungsdaten der Editionen und Rezensionen nicht in die Auswertung einbezogen. Bei der Evaluation künftiger Rezensionen könnte sich dieser zeitliche Aspekt vor allem in Bezug auf mögliche Entwicklungen als durchaus interessant erweisen.

Schnittstellen unterstützt werden kann. In der Praxis schlägt sich dies allerdings noch kaum nieder, zumindest nicht in den ausgewerteten Editionen. Wie aus der anfangs zitierten Beobachtung Patrick Sahles (2013, Bd. 2, 268) hervorgeht, wäre demnach zu hoffen, dass „Veränderungen im Angebot und in der Nutzung“ auch „gegenseitig aufgegriffen werden.“

Um dieses Angebot der Interaktion zu ermöglichen und künftige Nutzer:innen über den (vorläufigen) Abschluss digitaler Editionen zu informieren, empfiehlt es sich, neue Ressourcen auf DH-Plattformen und DSE-Verzeichnissen zu bewerben, um deren Sichtbarkeit für verschiedene Zielgruppen zu erhöhen. Gleichzeitig kann die Qualität eines neuen Produktes durch eine Rezension bestätigt oder idealerweise noch gesteigert werden, sofern sie ein umfassendes und fundiertes Feedback von Expert:innen bietet. Die Frage, ob Rezensionsorgane wie *RIDE* dazu geeignet sind, die allgemeine Glaubwürdigkeit und Akzeptanz von digitalen Editionen erhöhen, lässt Elena Pierazzo, die kürzlich einen Beitrag mit dem Titel „*What future for digital scholarly editions?*“ veröffentlicht hat (Pierazzo 2019), derzeit noch offen:

To address this question, *RIDE*, a journal completely devoted to the reviewing of scholarly editions was launched in 2014; however, in spite the fact that, at the time of writing, they have produced forty-five reviews of very high quality, examining in detail both the scholarly and the digital component of the editions, it is not yet clear whether or not these reviews have been able to provide any further academic credibility to the editions they discuss.

Mit Sicherheit aber trägt *RIDE* als Rezensionsorgan seit seiner Gründung dazu bei, die Qualität einzelner digitaler Editionsprodukte festzustellen, und setzt mit den – im Kriterienkatalog formulierten – Bewertungsgrundlagen erstrebenswerte Maßstäbe. Je mehr einzelne Editionen den Anspruch haben, den Orientierung gebenden Anforderungen zu genügen, umso mehr Akzeptanz und Kredibilität wird das dynamische Forschungsfeld des zeitgemäßen digitalen Edierens künftig wohl auch generell erfahren.

In Bezug auf die digitale Editionslandschaft in Österreich zeigt sich bei der Durchsicht der bislang in *RIDE* begutachteten Erzeugnisse, dass Forschende von österreichischen Digital Humanities-Instituten zwar immer wieder als Rezentinnen bzw. Rezensenten oder Jury-Mitglieder eingeladen und tätig werden, bis jetzt aber nur eine einzige digitale Edition aus Österreich einer eingehenderen Qualitätsprüfung durch *RIDE* unterzogen worden ist. Dies könnte sich jedoch erfreulicherweise in absehbarer Zeit ändern, wenn man die Liste<sup>52</sup> von Editionen betrachtet, die derzeit

---

<sup>52</sup> Das Projekt *Hugo von Montfort* der Universität Graz ist bereits begutachtet, der *Briefwechsel Sauer-Seuffert* der Österreichischen Nationalbibliothek befindet sich derzeit in Begutachtung und mehrere weitere Projekte sind hierfür vorgeschlagen, etwa die *Moralischen Wochenschriften* der Universität

entweder bereits in Begutachtung oder dafür vorgeschlagen sind. Für sie und die bereits rezensierten digitalen Editionen stellt die Autorin eine neuerliche korpusbasierte Auswertung in Aussicht, die sich – im Gegensatz und vor allem als Ausgleich zu der vorliegenden *ex negativo*-Untersuchung – vorwiegend mit positivem Feedback und dem bereits Erreichten befassen könnte, denn auch hierüber ließe sich ausführlich berichten.

## Literatur

- Andrews, Tara und Joris van Zundert. 2018. „What Are You Trying to Say? The Interface as an Integral Element of Argument.“ In *Digital Scholarly Editions as Interfaces*, hg. v. Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber und Gerlinde Schneider, 3–33. Norderstedt: Books on Demand.
- Anthony, Laurence. 2019. *AntConc – A freeware corpus analysis toolkit for concordancing and text analysis. (Version 3.5.8)*. Tokyo: Waseda University. Zugriff: 20. April 2020. <https://www.laurenceanthony.net/software>.
- Dillen, Wout. 2018. „The Editor in the Interface: Guiding the User through Texts and Images.“ In *Digital Scholarly Editions as Interfaces*, hg. v. Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber und Gerlinde Schneider, 35–59. Norderstedt: Books on Demand.
- Gibbs, Fred und Owens, Trevor. 2012. „Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs.“ *Digital Humanities Quarterly* 6 (2). Zugriff: 20. April 2020. <http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html>.
- Henny, Ulrike. 2018. „Reviewing von digitalen Editionen im Kontext der Evaluation digitaler Forschungsergebnisse.“ In *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, hg. v. Roland S. Kamzelak und Timo Steyer. doi:10.17175/sb002\_006.
- Hewing, Ben, Mandl, Thomas und Womser-Hacker, Christa. 2016. „Methods for User-Centered Design and Evaluation of Text Analysis Tools in a Digital History Project.“ *ASIS&T Annual Meeting Proceedings* 53 (1): 1–10.
- Kramer, Michael J. 2016. „The Digital Humanities Reader.“ Zugriff: 20. April 2020. <http://www.michaeljkramer.net/the-digital-humanities-reader/>.
- Pierazzo, Elena. 2019. „What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter.“ *International Journal of Digital Humanities* 1: 209–220. Zugriff: 20. April 2020. doi:10.1007/s42803-019-00019-3.
- Rasmussen, Krista Stinne Greve. 2016. „Reading or Using a Digital Edition? Reader Roles in Scholarly Editions.“ In *Digital Scholarly Editing*, hg. v. Matthew James Driscoll und Elena Pierazzo. Zugriff: 20. April 2020. [https://www.openbookpublishers.com/htmlreader/978-1-78374-238-7/ch7.xhtml#\\_idTextAnchor019](https://www.openbookpublishers.com/htmlreader/978-1-78374-238-7/ch7.xhtml#_idTextAnchor019).
- Resch, Claudia. 2017. „»Etwas für alle« – Ausgewählte Texte von und mit Abraham a Sancta Clara digital.“ *ZfdG* 2. Zugriff: 20. April 2020. [http://www.zfdg.de/2016\\_005](http://www.zfdg.de/2016_005).

---

Graz bei den digitalen Editionen sowie das *ABaC:us – Austrian Baroque Corpus* der Österreichischen Akademie der Wissenschaften und das *Repertorium digitaler Quellen zur österreichischen und deutschen Rechtsgeschichte in der Frühen Neuzeit* bei den digitalen Textsammlungen.

- . 2018. „Zeitung Lust und Nutz‘ im digitalen Zeitalter. Partizipative Ansätze zur Erschließung historischer Ausgaben der Wiener Zeitung.“ *medien & zeit. Kommunikation in Vergangenheit und Gegenwart* 2: 20–31.
  - . 2019. „Das Wien[n]erische Diarium und seine digitale Erschließung oder ‚Was Zeitungsleser vor Geräte haben müssen?‘“ In *Wiener Geschichtsblätter* 2: 115–129.
- RIDE – A review journal for digital editions and resources*. 2020. Zugriff: 20. April 2020.  
<https://ride.i-d-e.de/>.
- Sahle, Patrick. 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 1–3: Befunde, Theorie und Methodik*. Norderstedt: Books on Demand.
- Schnöpf, Markus. 2013. „Evaluationskriterien für digitale Editionen und die reale Welt.“ *HiN – Humboldt im Netz. Internationale Zeitschrift für Humboldt-Studien* XIV (27): 69–76. Zugriff: 20. August 2022. [https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/6887/file/hin27\\_online-ss69-76.pdf](https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/6887/file/hin27_online-ss69-76.pdf).
- Shaw, William und Liz Milewicz. 2017. „Access for Whom? Rethinking Audience and Editorial Decisions through Project Vox.“ Conference Paper, DH 2017. Zugriff: 20. April 2020. [https://pdfs.semanticscholar.org/d7c5/a44647341219374dad0cabf21e0687275f69.pdf?\\_ga=2.24659002.1807356829.1564488265-2052193220.1564488265](https://pdfs.semanticscholar.org/d7c5/a44647341219374dad0cabf21e0687275f69.pdf?_ga=2.24659002.1807356829.1564488265-2052193220.1564488265).
- Sinclair, Stéfan und Geoffrey Rockwell. 2016. *Voyant Tools*. Zugriff: 20. April 2020.  
<http://voyant-tools.org/>.
- Steinkrüger, Philipp. 2021–. „Charts. Scholarly Editions.“ In *RIDE – A review journal for digital editions and resources*. Zugriff: 20. April 2020. <https://ride.i-d-e.de/data/charts/>.
- Vogeler, Georg. n.d. „Digitale Edition. Grundlagen und Workflow.“ Zugriff: 20. April 2020. [https://www.i-d-e.de/wp-content/uploads/2017/11/01\\_Digitale\\_Editionen\\_Vogeler.pdf](https://www.i-d-e.de/wp-content/uploads/2017/11/01_Digitale_Editionen_Vogeler.pdf).
- Warwick, Claire. 2012. „Studying users in digital humanities.“ In *Digital Humanities in Practice*, hg. v. Claire Warwick, Melissa Terras und Julianne Nyhan, 1–22.

# **Experience with a Workflow using MS Word and a DOCX to TEI Converter**

Markus Ender, Joseph Wang

## **Abstract**

In this contribution we summarize our experience in edition projects with the conversion from DOCX files to TEI-XML files. A tool called DOCX2TEI, developed in our team, is capable of transforming DOCX files automatically to TEI files without the need of manual post-processing of conversion results by editors. However, for this to function properly pertinent programming knowledge is required to make the necessary adaptations on DOCX2TEI. We compare DOCX2TEI with other tools capable of producing TEI files out of DOCX files. At the end of the paper we will present the lesson learned from deploying and using this tool.

## **Zusammenfassung**

In diesem Beitrag schildern wir unsere Erfahrungen in Editionsprojekten mit der Konvertierung von DOCX-Dateien zu TEI-XML. DOCX2TEI ist ein Tool, das im Rahmen dieser Projekte entwickelt worden ist; es verwandelt DOCX-Dateien automatisch in TEI-Dateien, ohne dass die Resultate nachbearbeitet werden müssen. Damit dieser Prozess richtig funktioniert, muss eine Adaptierung des Tools vorgenommen werden, die einschlägige Programmier-Kenntnisse verlangt. Wir stellen DOCX2TEI anderen Tools gegenüber, die ebenfalls dazu verwendet werden können, um TEI-Dateien aus DOCX-Dateien zu erzeugen. Am Schluss des Beitrags geben wir unsere Lehre aus unseren Erfahrungen mit diesem Workflow wieder.

Since the introduction of TEI P5 in 2007 (TEI-Consortium 2019, ch. i.) creating TEI-XML data has become a common task in humanities research. Today most editing projects make use of TEI-XML for the modelling and annotation of primary sources. However, for those scholars not familiar with the praxis of coding, creating TEI-XML can be a frightening task. There are many different reasons for individual scholars to be intimidated by the TEI guidelines. One of them most certainly is that the learning curve for TEI is steep. Since XML is the underlying technology of TEI, a common TEI course usually starts with a basic XML introduction where participants soon get

the feeling that concepts like *element*, *attribute*, *node* and *comment* are confusing and maybe even a little bit frightening. The course will then focus on explaining the different parts of the TEI guidelines. In the end participants will realize that—even though it is not necessary to memorize all 576 elements of the TEI guidelines (TEI-Consortium 2019, Appendix C.1)—coding in XML is still a very different task from using a common word processor. Another difficulty that arises when people start to use TEI is the paradox of choices. While we do welcome the flexibility of the TEI guidelines (the ability of using different approaches to solve one problem is a key feature of TEI), being able to choose is sometimes an excessive demand, especially for beginners. Instead of a single possibility to carry out one task, e.g., how double underlined text passages should be encoded, there are always different solutions to a single problem. These and many more reasons keep humanities scholars from using XML editors.

In this paper we want to share our experience with using an application to convert MS Word DOCX files to TEI (DOCX2TEI).<sup>1</sup> There are benefits and challenges using such a converter and we hope that by publishing our experience readers can benefit from our know-how.

## 1 History of using a DOCX to TEI converter in the research institute Brenner-Archives

In the Research Institute Brenner-Archives (Forschungsinstitut Brenner-Archiv, FIBA)<sup>2</sup> of the University of Innsbruck, TEI had a slow start. While members of FIBA have always embraced digital technology for their editing activities, TEI was first introduced through the cooperation of FIBA and the Wittgenstein Archives in Bergen (WAB) on the project *Culture and Value Revisited* (FWF P19022, led by Allan Janik). Vemund Olstad, then a member of WAB, gave the first formal introduction to XML and the TEI in 2006. However, aside from this cooperation no other projects have adopted TEI.

During the project, project staff also acquired the technical skills for XML processing. These skills were essential when it was decided that the results of the projects *Critical Online-Edition of the Works of Otto Weininger* (FWF P17975), *Critical On-Line-Edition of Otto Weininger's Correspondence* (FWF P20301) and *A Documentary Volume on Otto Weininger and his Works* (FWF P22168, all lead by Allan Janik) should be published as online-accessible digital editions in 2008.

---

<sup>1</sup> As soon as Ficker's correspondence is published online, the entire source code of DOCX2TEI will be publicly accessible. We will also license it under a free license, so other projects may re-use DOCX2TEI or any parts of it for their own purpose.

<sup>2</sup> <https://www.uibk.ac.at/brenner-archiv/>.

At that time, Intelex, the publisher who wanted to publish the edited volumes of the *Weininger* projects, could only process data in FolioView format or in TEI-XML. The documents of the *Weininger* projects, however, were available only as MS *Word* files. Furthermore, the editors have employed some functions of MS *Word* in order to “mimic” the behavior of web pages, such as adding comments and hyperlinks. The commenting function was used to add additional information to different lemmata and the hyperlinking function was used to link names to biographies and lists of works in arts and literature. Already at that time, MS *Word* (Version MS *Word* 2007) was able to export a *Word* document to XML. The XML used there is a predecessor of the OOXML specification (Kelly 2003). The best technical approach seemed therefore to write an XSLT stylesheet to convert the XML produced by MS *Word* to TEI-XML.

Unfortunately, the conversion from *Word* XML to TEI succeeded only in parts. We had to manually correct several passages of the conversion. One major issue was that the authors have used a special font-face for Greek letters, e.g. the letter ‘a’ in this font-face encodes for ‘α’; and the font-face also used zero space diacritics to form composite accented characters. Therefore, we had to manually check and emend the conversion errors. After the emendations and two years of work, the results of the *Weininger* projects were published using Intelex’s publishing platform in 2010 (Hirsch 2011).

After the project, the FIBA technical staff soon realized the potential inherent in the conversion of a text processor format to TEI-XML. Therefore, they started to generalize the conversion workflow so that it could also be used for other scenarios. At the same time, MS *Word* moved their standard format from DOC to DOCX (ECMA-International 2016). In doing so they have also opened up the possibility of transforming *Word* files automatically to TEI. The DOCX format is basically a set of XML files (together with other media files) that are compressed into a ZIP file. Thus, when the project *Ludwig von Ficker as a Cultural Transmitter* (FWF P24283, led by Eberhard Sauermann) and later its follow-up project (*Ludwig von Ficker: Critical Online Edition of the Correspondence and Monograph*, FWF P29070, under the lead of Ulrike Tanzer)<sup>3</sup> were granted, we developed routines to convert DOCX files to TEI-XML and then send the TEI data to an XML database. There were two main reasons for using these conversion routines. The first reason was that about half of the correspondences were already available as transcriptions in *Word* files.<sup>4</sup> Instead of manually copying and pasting the vast amount of data into XML files, it was, of course, easier to convert the data automatically. The second reason was that there simply are too many letters to be manually encoded. Judging from the amount of correspondence there is not enough

<sup>3</sup> One of the main aims of this project was to make Ficker’s correspondence (about 17,300 letters) available to scholars and the wider public.

<sup>4</sup> Aside from four published volumes (Zangerle 1986–1996), Ficker’s correspondences were constantly being transcribed at FIBA.

time nor personnel resources to encode every aspect of a letter. Therefore, we had to strictly confine our transcriptions and commentaries, and so *MS Word* (even with its limited features) worked fine for the project.

It is worth noting that the nature of the conversion has changed since we moved from the *Weininger* projects to the *Ficker* projects. In the *Weininger* projects, the editing process had already been finished – essentially, the edition texts were not changed anymore. The conversion from DOCX to TEI-XML was therefore a one-time process and afterwards all corrections to the edition – when there had been any – were manually recoded. Furthermore, we were only interested in converting a few *Word* features like character and paragraph formatting and hyperlinking. In contrast to this, the aim of the DOCX to TEI-XML conversion for the *Ficker* projects have substantially changed. The transcriptions of the letters appear in different qualities and stages throughout the edition. Some letters had already been edited and enriched with comments; some were only transcribed and collated; some were just transcribed without the necessary step of collation. The biggest bulk of the letters, however, was not even transcribed. Thus, our expectation was that the majority of files would be changed constantly. Furthermore, at the beginning of the project the direction and goals of the edition (besides providing a reading text) were not yet clearly defined. With these considerations in mind, the conversion process had to be more flexible and it needed to be easily adjustable in case the editing guidelines were revised.

## 2 The current atatus of the *Ficker* projects and the reuse of the routines for other projects

The main aim of the routine we call DOCX2TEI is to extract data from files in DOCX format written in a specific way.<sup>5</sup> Each file contains at least one letter, but potentially as many letters as possible between Ficker and one of his correspondence partners. Each letter starts with a normalized heading containing the metadata of the letter. Paragraphs concerning provenance and archival notes follow the heading. Then the file contains the letter proper. The letter itself consists of a transcription and annotations on the transcription. The annotations can be either references to places, people, and subjects; or they contain editorial remarks to a lemma. The conversion should also retain paragraphs and character formatting. Each letter should result in one TEI-XML file. Since we will have more than 17,300 TEI-XML files at the end of

---

<sup>5</sup> There are other features (or requirements) that are not part of the conversion, although they are part of TEI data creation. For example, the linking of references to an authority file is done through a separated mapping tool that maps the unique keys to, e.g., *Gemeinsame Normdatei* (GND, Deutsche Nationalbibliothek 2019). Adding GND URIs to TEI data is done only when full TEI-XML data is required, e.g. for users who want to download the data for themselves, but not during the conversion of DOCX to TEI.

the projects, we have agreed that any correction of the data should happen in the DOCX files. The conversion should reach a quality that manual changes to the TEI document after conversion are not necessary anymore.

The basic idea for the conversion is quite simple: We are using XSL-Transformation for the conversion process. A small number of XSLT stylesheets is used in this process and each of those transforms a single task of the conversion routine. The conversion process starts at the main entry point of the DOCX file. Then, a series of XSL transformations is initiated, each using the results of the previous XSL transformation as input. This routine enables us to debug and improve the transformation step by step.<sup>6</sup> The result of the last transformation will then be sent through an enterprise integration solution to a backend where the data is stored in a database and further data processing that is not considered part of the conversion (like adding certain metadata and authority controlled data) takes place. There is also a frontend that uses the Web-API provided by the backend application. The frontend, however, does not enable scholars to change the TEI data. The frontend is meant for the purpose of monitoring the conversion and doing some data processing, e.g., searching or aggregating data. The frontend provides users with a table of content of all TEI files, a HTML and a PDF representation of every TEI-XML file, a full-text search, and a registry of annotated TEI data.<sup>7</sup>

A few examples should illustrate how a set of stylesheets is organized.

1. One stylesheet is specialized in separating letters from the whole set of correspondence. This stylesheet is capable of creating different TEI files from a single input document/file. This is actually a special case where we have not only one, but many TEI-XML files as a result. Each resulting TEI-XML file is passed on to the following stylesheets individually.
2. A second stylesheet is used to create the TEI header element by parsing the heading and all metadata contained in the first lines. Since the heading (title) of each letter is standardized, we can write a parser to extract information from

---

<sup>6</sup> It should be noted here that changing the order of stylesheets will change the outcome. It could happen that after changing, e.g., stylesheet no. 2, stylesheet no. 4 will not work anymore. Because of this, stylesheets that change the general data structure should actually be applied later in the stylesheet series.

<sup>7</sup> Since we are only concerned with the conversion process in this paper, we will just briefly mention the post-processing and the hardware configuration: the enterprise integration solution is based on *Apache Camel* (<https://camel.apache.org/>). We have created a backend application based upon *Apache Sling* (<https://sling.apache.org/>) which also serves as the data storage for the XML data. *Camel* is able to periodically check directories for changed files and process them. The results are then sent to *Sling*. Additionally, we created our own frontend using *AngularJS* (<https://angularjs.org/>). At the time of writing, we are migrating the backend from *Apache Sling* to a standalone java application using *eXist* (<http://exist-db.org/exist/apps/homepage/index.html>) as data storage. The frontend will be based upon the *Angular Framework* (<https://angular.io/>).

the headings. Some effort was put into the parsing of dates, to make sure that uncertain dating is encoded correctly.

3. Several stylesheets are responsible for converting comments to references of people and places. To encode a certain reference, we need to encode both the referring words and the referred entity. In the TEI guidelines, these references are to be written as <rs>, <persName>, <placeName> or <date> using the lemma as the text of the XML elements, and the referenced entity is provided as value of an attribute (either @ref or @key) of the XML element. If we use foot- or endnotes to annotate references in DOCX files we lose the starting point of the annotation. MS *Word* allows users to add comments to a document where the comments consist of a starting XML-element, an ending XML-element and one element for the note itself.<sup>8</sup> We have created a stylesheet for the conversion of these elements to a single XML element. Other stylesheets are responsible for converting these XML elements further to the correct TEI elements. We introduced a special syntax for encoding references, e.g., a comment starting with “P:” denotes a reference to a person and a comment starting with “O:” denotes a reference to a place. What follows “P:” or “O:” is the normalized name of the person respectively the place (see Figures 1 and 2). The normalized names of people and places also act as the @key value of the referencing element. If editors need to add additional information on the lemma, they get the possibility of adding this information in a new paragraph within the comment. The information in the paragraph will be converted automatically to a <note>-element.

Special care is needed when converting the delimiting milestones (and the nodes between them) into an element. At first, we have simply replaced the milestones with a starting and an ending tag. However, in some cases (e.g. when the commented lemmata are overlapping) we have produced overlapping elements and the result is not a well-formed XML anymore (Wang 2016).

As one can see, by organizing the stylesheets in this way, we can add, switch or omit stylesheets in order to adapt the conversion process for other projects. For example, if in a project each DOCX should also be converted to a single <TEI> we can omit the separation stylesheet. Similarly, we can adapt the conversion of comments by adding new rules of conversion. If, e.g., in a project people need to annotate literary works in a certain way, we can easily introduce a new syntax for this task (e.g., “Bibl: the name of author: the title of the work”).

Since the beginning, we have also improved several other features of our approach. For instance, the way we extracted metadata from each letter using a parser soon

---

<sup>8</sup> Note that the starting and the ending points of a MS comment are marked up using XML elements (and not a start- and end-tag).

Sehr geehrter Herr Professor Ficker!

Ich bin seit einiger Zeit damit befasst, ein [!] kurze Biographie meines entfernten Vetters Ludwig Wittgenstein zu schreiben. Erst am letzten Tage meines Urlaubes, von dem ich eben zurückgekehrt bin, erfuhr ich in Wien, dass Sie für den BRENNER einen Aufsatz über Wittgenstein in Vorbereitung haben. Es läge mir sehr daran, diesen Aufsatz möglichst bald zu sehen, da ich niederschreiben vielleicht eine Skizze von E sehr viele Jahre seines verschliefen versucht habe. Ich beabsichtige, für etwaige Entwürfe Ihnen Aufsat

### Kommentare

-  C62464 26. Mai 2020  
P: Wittgenstein, Ludwig
  - Antworten  Auflösen
-  C62464 26. Mai 2020  
O: Wien
  - Antworten  Auflösen

Mit dem Ausdruck der vorzüglichsten Hochachtung,

Figure 1: Word document with annotations.

```
<p>Sehr geehrter Herr Professor Ficker!<lb/>Ich bin seit einiger Zeit damit befasst, ein [!] kurze Biographie meines entfernten Vetters Ludwig Wittgenstein zu schreiben. Erst am letzten Tage meines Urlaubes, von dem ich eben zurückgekehrt bin, erfuhr ich in Wien, dass Sie für den BRENNER einen Aufsatz über Wittgenstein in Vorbereitung haben. Es läge mir sehr daran, diesen Aufsatz möglichst bald zu sehen, da ich meine Arbeit, wenigstens im ersten Entwurf, in den nächsten Wochen niederschreiben möchte. Ich wäre Ihnen sehr dankbar, wenn Sie mir zu diesem Zweck <seg rend="strikethrough">be</seg> vielleicht einen Bürstenabzug oder einen Durchschlag Ihres Aufsatzes zukommen lassen könnten. Auch für alle anderen Mitteilungen über Wittgenstein, die für eine biographische Skizze von Bedeutung sein könnten, wäre ich sehr dankbar. Ich habe Wittgenstein zwar durch sehr viele Jahre aber nie wirklich gut gekannt, und bin im Allgemeinen auf Nachrichten von seinen verschiedenen Freunden aus den diversen Perioden seines Lebens angewiesen.<lb/>Ich beabsichtigte, den ersten Entwurf der Skizze zunächst <seg rend="strikethrough">zukommen</seg> hektographieren zu lassen und für etwaige Berichtigungen und Ergänzungen unter W.'s Freunden zu zirkulieren. Ich werde mir dann erlauben, Ihnen ein Exemplar zuzusenden. Ich wäre sehr dankbar wenn Sie mir Ihren Aufsatz und etwaige sonstige Mitteilungen noch [xxx] zugänglich machen könnten.<lb/>Mit dem Ausdruck der vorzüglichsten Hochachtung,<lb/>Ihr sehr ergebener<lb/>FAvHayek<lb/>(Prof. Dr. F.A. von Hayek)</p>
```

Figure 2: View of the TEI data in Oxygen XML Editor.

seemed to be quite laborious. Instead of parsing the header (a text string) we should have created tabular data at the beginning of each letter containing metadata like “date”, “sender”, “recipients”, and “archival note”. The editors could have then entered the data directly into the table. By doing so, the stylesheet for creating `<teiHeader>` would be much easier to write and maintain. Unfortunately, at the time we realized

this the transcription process in the Ficker project was already far too advanced to change the transcription workflow. But other projects have already benefited from this insight!

The correctness of the DOCX to TEI conversion needs to be checked by scholars after the conversion process. On the one hand, the conversion routine may contain bugs and produce incorrect artifacts, and on the other hand, the DOCX file itself may contain errors (e.g. typos and transcription mistakes) as well. An XML schema validation provides some help here, but the editors need more control over the conversion process. To be able to detect all errors, scholars must have the possibility to see the result of the conversion as soon as possible. In the Ficker projects, the application responsible for the conversion is also able to render TEI-XML data as HTML pages.

As one can see, transforming DOCX to TEI using specific designed stylesheets is quite powerful. There are other constraints in DOCX that can be utilized for TEI conversions: font-faces, font-styles, text-colors, the correction-tracking feature, etc. If one would want to utilize these features in DOCX2TEI, developers only need to embed other stylesheets that make use of these features.

Using DOCX, scholars will not have to interact with an XML editor. Since the conversion is mainly determined by the stylesheets, people will not have to worry about choosing the correct TEI annotations, if the stylesheet does not allow different conversion for the same feature.

### 3 Comparison to *OxGarage* and similar tools

There are also other solutions that are capable of converting DOCX files to TEI. The most renowned tool for this task is surely *OxGarage*,<sup>9</sup> which is maintained by the TEI Consortium itself. *OxGarage* is not only able to convert DOCX to TEI, but reads many different formats and can convert texts, spreadsheets, and even presentations. *OxGarage* also offers a REST-API (TEI Consortium 2020)<sup>10</sup> through which other applications can execute conversions without the user interface of *OxGarage*. This is especially convenient if one wants to use the conversion routines of *OxGarage*<sup>11</sup> programmatically.

Internally, *OxGarage* uses the XSLT stylesheets. It does offer additional functionality of pre- and post-processing files, but essentially we can regard it as an interface for the execution of stylesheets that are provided by the TEI-Consortium. From the perspective of someone who only needs the conversion of DOCX files to TEI, *OxGarage* and DOCX2TEI – and potentially also other tools that convert DOCX files

<sup>9</sup> <https://oxgarage2.tei-c.org>.

<sup>10</sup> Cf. the README file of the *OxGarage* Github Repository under <https://github.com/TEIC/oxgarage>.

<sup>11</sup> We acknowledge that *OxGarage* is a much greater conversion tool than simply a DOCX to TEI converter, however, in this article, we will solely concentrate on the DOCX to TEI conversion function of *OxGarage*.

to TEI through XSL transformations – are quite similar. In both cases the content of DOCX files is put into an XML file using TEI vocabulary and the conversion does not aim for a faithful reproduction of a text-page, but for data conversion of text models. However, the main difference lies in the desired workflow and with that the different behavior of the editors using these tools. In our opinion, *OxGarage* is the appropriate tool for scholars who want to get a text with its structure as TEI data quickly. The enrichment of the data (e.g., creating the metadata for the TEI header or the semantic annotations) is done later by using other XML-editors. With *OxGarage*, people can take a “short-cut” in the creation of TEI-data.

DOCX2TEI, on the other hand, aims at the substitution of an XML editor. The conversion creates as clean TEI data as possible without any need of manually editing its results. Scholars and editors use *MS Word* as a tool both for the transcription and the annotation. We can even boldly claim that DOCX2TEI makes *MS Word* a project specific TEI editor. This short comparison shows that *OxGarage* and DOCX2TEI require different efforts from their users. For *OxGarage*, editors need to know the XML rules, XML editors, and the TEI guidelines. For the technicians, on the other hand, the programming efforts are reduced.

We can see that while DOCX2TEI and *OxGarage* are both capable of converting DOCX files to TEI, their user scenarios are quite different. DOCX2TEI only works well if it is customized to meet the needs of a specific project. This means that if another project wants to re-use DOCX2TEI, the chance is very high that the conversion routines must be adapted as goals and focuses of different editions will vary. *OxGarage*, on the other hand, produces uniformed conversion results for a wide range of use cases.<sup>12</sup> If one wants to adapt the *OxGarage* conversion for their own project, the customization must be done *after* the conversion.

What about other tools that are capable of converting DOCX files to TEI data? Since conversion data from DOCX- and ODT-files (i.e. the standard file format for *Open-* and *LibreOffice* text processors) is considered as a standard workflow in academic publishing (McGlone 2014), we can expect that there is a wide range of tools available for this functionality. A *Google*-search reveals that several projects are working on this issue as well.<sup>13</sup> The handling of the conversion may differ (some require users to execute a command in the console, other require a preinstalled software library), but the conversion itself is usually done through XSLT. Thanks to the fact that most of these projects are open-sourced, scholars can adapt them to meet their own need.

---

<sup>12</sup> This is not entirely true. The online version of *OxGarage* has been updated several times and the stylesheets used for data conversion are also improved. However, if we stick to the same version of *OxGarage*, the conversion result will not vary.

<sup>13</sup> For instance *Pandoc* (MacFarlane 2006–2022), or *TEI Publisher* (2020). The *Oxygen XML Editor* can convert DOCX to TEI-XML (SyncRO Soft 2019), too.

Some tools also provide customization features to ease the conversion.<sup>14</sup> Thus, one can use these tools as a quick way to produce preliminary TEI-XML, or one can even use them to make *MS Word* a TEI editor. Furthermore, it should be emphasized here that several digital editing projects<sup>15</sup> have adapted a workflow that takes DOCX files to produce TEI files and consequently *MS Word* can be a viable alternative to XML editors.

## 4 Insights from working with DOCX2TEI

We have been using DOCX2TEI since 2010, have adapted DOCX2TEI for different projects, and we have learned from our practical experience. In the following section we want to share some insights we gained in the process of developing and improving our workflow. We believe that readers can profit from our report as other project teams using tools similar to DOCX2TEI may have similar challenges we have encountered in the past.

1. The main rule is: editorial guidelines need to be clear and exact! Consequently, the editing experience (whether using an XML editor or other tools like *MS Word*) is more efficient and customizing stylesheets for DOCX2TEI becomes easier too. Just like in printed editions, editorial guidelines will reflect features from the primary source that are to be edited as well as the goal of the edition. If editors have clear guidelines, the editing process itself will become easier, and developers, too, can come up with solutions for transformation and presentation more easily. Of course, every now and then situations come up that are not covered by the editorial guidelines. While a printed book can usually find ‘ad hoc solutions’, e.g., a deviation from the editorial guidelines, applying these kinds of deviations from the editorial guidelines may also involve the need of changing software code. They are therefore more difficult to deal with when one uses (and is committed to) DOCX2TEI. There might also be situations for which an appropriate solution is not found, in those cases we either forfeit the desired ad hoc solution or we have to manually encode the respective solution.
2. Good and regular communication between scholars and technicians is essential. This is especially obvious when one wants to implement customizations. This

---

<sup>14</sup> The *Cirilo* client, for example, can use customization to improve the result of its DOCX to TEI conversion, see: <http://gams.uni-graz.at/o:gams.doku#custom>.

<sup>15</sup> One can find several work reports of digital editing projects that have the DOCX to TEI conversion as subject, e.g. Lehečka (2019), or the following blog post by Dot Porter: <http://www.dotporterdigital.org/workflow-ms-word-to-tei/>. The software TEI publisher can also ingest DOCX-files and convert them to TEI.

requires scholars stating their intentions clearly (cf. rule 1) to enable the technicians to adapt routines and codes accordingly. This is, however, not the only case where good communication helps to resolve misunderstandings. This can be best achieved simply by talking with each other. Especially at the beginning of a project, meetings of project members should be scheduled more often, e.g. once in two weeks. Later, when people have more routine, these meetings will become less frequent. However, keeping the possibility of individual meetings (e.g. between scholars and technicians) is essential. We have made good experience using a ticket system for documenting problems, concerns, and their solutions.<sup>16</sup> The ticket system (or other online collaborative tools used for documentation purposes) should be openly available for all project members. The methodology proposed by the so-called Agile Software Development (Beck 2001, Dingsoyr et al 2012) can also be helpful, as it emphasizes the importance of embedding users (clients) into the design and development processes.

3. Any problem reported by the users should be taken seriously, but not personally. There can be unusual bug fixes and workarounds. We have experienced that scholars have tried to compensate for programming errors by simply not using specific features. While the programming error was very easy to correct, scholars did not realize that it was something that can be “fixed” (cf. Rule 2). This is truly a remarkable symptom of an endeavor that involves multiple parties. This kind of behavior certainly can be minimized by respecting each member of the editing team and caring about her or his problems.
4. Even when one uses DOCX2TEI, scholars still must acquire some knowledge of TEI. At least they must know the difference between structured metadata and XML data that is semi-structured. They need to understand the difference between data and data representation. We do not encourage using tools like DOCX2TEI without giving scholars a proper introduction to TEI. The reason for this suggestion is that digital editing requires a different, more large-scale understanding of ‘edition’ than editing a printed volume. In a printed volume, scholars work in terms of printed pages. In a digital edition employing TEI-XML (or any other technical means with which the editors are not familiar), however, scholars must think in terms of data-models that represent the primary source. Furthermore, these models themselves are represented in other data-models to produce HTML views or printed pages. Learning TEI is the best way to grasp these basic concepts.

---

<sup>16</sup> In our project we use *GitLab* provided by our university for both version control of the source code and for the documentation of the issues within projects.

5. Reusing DOCX2TEI for other projects is doable, even in different settings. In the last six years we have also ported DOCX2TEI to other projects. One of them is the creation of a literature map that is a part of the project *The Tyrol / The South Tyrol – A literary Topography* (FWF P26039, led by Johann Holzner). The aim of the literature map is to create a database that contains references to geographical places in literature. Such a reference could be, e.g., a description of a village in South Tyrol, or a poem written about a mountain peak.

For this project, we have prepared DOCX files so that scholars only have to fill in the data in MS Word and upload it to the database. After the conversion (and if the result of the conversion is valid), the TEI data is further processed together with data coming from a separate database on geographical locations, containing data on coordinates coming from *GeoNames*<sup>17</sup> to produce an interactive map that is now accessible in *LiteraturTirol*.<sup>18</sup>

## 5 Conclusion

This chapter introduced a DOCX to TEI-XML conversion that is successfully used in several projects at the University of Innsbruck. The discussed workflow is particularly well-suited for projects with a large amount of material to be transcribed or material already available as DOCX files. In such cases, a DOCX to TEI-XML workflow can save quite some time, however, the conversion tool plays a central role in the process and the quality of the result depends not only on the edited text in the DOCX file(s), but on the conversion tool and the intermediary stylesheets too. We are quite positive that for certain projects our DOCX2TEI-approach can be more efficient than using other conversion tools, due to its modularity and adaptability to a project's needs. Compared to *OxGarage*, DOCX2TEI shows certain benefits, mainly due to the fact that data enrichment is realized prior to the conversion. However, users must keep in mind that MS Word is not a full-blown XML editor, thus in all projects that employ a DOCX to TEI conversion workflow the data modeling is subject to limitations. An edition project that relies solely on DOCX as an editing tool cannot use the full flexibility that TEI provides, and furthermore MS Word does not include the options to check wellformedness and validity in accordance with the TEI.

---

<sup>17</sup> <https://www.geonames.org>.

<sup>18</sup> <https://literaturtirol.at/landkarte>.

## Bibliography

- Beck, Kent, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning et al. 2001. “Manifesto for Agile Software Development.” <https://agilemanifesto.org>.
- Deutsche Nationalbibliothek, eds. 2019. “Gemeinsame Normdatei (GND).” Last modified June 25, 2019. Accessed September 30, 2019. [https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html).
- Dingsøyr, Torgeir, Sridhar Nerur, VenuGopal Balijepally, and Nils Brede Moe. 2012. “A decade of agile methodologies: Towards explaining agile software development.” *The Journal of Systems and Software* 85: 1213–21.
- ECMA International, eds. 2016. “Standard ECMA-376. Office Open XML File Formats.” Accessed September 30, 2019. <http://www.ecma-international.org/publications/standards/Ecma-376.htm>.
- Hirsch, Waltraud, ed. 2010. “Otto Weininger: Kritische Online-Ausgabe der Werke und Briefe.” Charlottesville: Intelex. Accessed September 30, 2019. [http://library.nlx.com/xtf/view?docId=weininger\\_de/weininger\\_de.00.xml;chunk.id=div.weininger.pmpreface.1;toc.depth=2;toc.id=div.weininger.pmpreface.1;hit.rank=0;brand=default](http://library.nlx.com/xtf/view?docId=weininger_de/weininger_de.00.xml;chunk.id=div.weininger.pmpreface.1;toc.depth=2;toc.id=div.weininger.pmpreface.1;hit.rank=0;brand=default).
- Kelly, Peter. 2003. “Microsoft Office System and XML: Bringing XML to the Desktop.” Accessed September 30, 2019. [https://docs.microsoft.com/en-us/previous-versions/office/developer/office-2003/aa159914\(v=office.11\)?redirectedfrom=MSDN](https://docs.microsoft.com/en-us/previous-versions/office/developer/office-2003/aa159914(v=office.11)?redirectedfrom=MSDN).
- Lehečka, Boris. 2019. “Using Microsoft Word for preparing XML TEI-compliant digital editions.” Accessed September 30, 2019. <https://zenodo.org/record/3451430#.X7zHUVAxmUk>.
- MacFarlane, John. 2006–2022. “Pandoc, a Universal Document Converter.” Accessed April 1, 2020. <https://pandoc.org/index.html>.
- McGlone, Jonathan, 2014. “Preserving and Publishing Digital Content Using XML Workflows.” In *Library Publishing Toolkit*, ed. by Allison P. Brown, IDS Project Press, 97–108.
- Porter, Dot. 2018. “Workflow: MS Word to TEI. March 21, 2018.” Accessed February 1, 2021. <http://www.dotporterdigital.org/workflow-ms-word-to-tei>.
- SyncRO Soft. 2019. “TEI P5 Document Type (Framework).” Accessed April 1, 2020. <https://www.oxygenxml.com/doc/versions/22.0/ug-editor/topics/author-teip5-doc-type.html>.
- TEI Consortium, eds. 2019. “TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.6.0.” Accessed September 30, 2019. <http://www.tei-c.org/Guidelines/P5>.
- TEI Consortium. 2020. “OxGarage.” Accessed April 1, 2020. <https://oxgarage.tei-c.org>. Source code available under <https://github.com/TEIC/oxgarage>.
- TEI Consortium. 2020. “TEI Publisher, the Instant Publishing Toolbox.” Accessed April 1, 2020. <https://teipublisher.com/exist/apps/tei-publisher/index.html>.
- W3-Consortium, eds. 2003. “Mathematical Markup Language (MathML) Version 2.0 (Second Edition) W3C Recommendation 21 October 2003.” Accessed September 30, 2019. <https://www.w3.org/TR/MathML2>.
- W3-Consortium, eds. 2007. “XSL Transformations (XSLT) Version 2.0. W3C Recommendation 23 January 2007.” Accessed September 30, 2019. <https://www.w3.org/TR/2007/REC-xslt20-20070123>.

- W3-Consortium, eds. 2008. “Extensible Markup Language (Xml) 1.0 (Fifth Edition). W3c Recommendation 26 November 2008.” Accessed September 30, 2019. <https://www.w3.org/TR/2008/REC-xml-20081126>.
- W3-Consortium, eds. 2009. “Namespaces in XML 1.0 (Third Edition). W3C Recommendation 8 December 2009.” Accessed September 30, 2019. <https://www.w3.org/TR/REC-xml-names>.
- W3-Consortium, eds. 2011. “Scalable Vector Graphics (SVG) 1.1 (Second Edition). W3C Recommendation 16 August 2011.” Accessed September 30, 2019. <https://www.w3.org/TR/SVG11>.
- Wang, Joseph. 2016. “From DOCX via TEI to Literature Map.” Poster. Access on September 30th 2019. <http://tei2016.acdh.oeaw.ac.at/sites/default/files/Joseph%20Wang%20From%20DOCX%20via%20TEI%20to%20Literature%20Map.pdf>.
- Zangerle, Ignaz, Walter Methlagl, Franz Seyr, Anton Unterkircher, and Martin Alber, eds. 1986–1996. *Ludwig Ficker. Briefwechsel*, four volumes. Salzburg: Otto Müller and Innsbruck: Haymon.

# A Linked Data Vocabulary for Intertextuality in Literary Studies, with some Considerations Regarding Digital Editions

Bernhard Oberreither

## Abstract

Editing a text often includes pointing out its manifold connections to other texts—be it cultural contexts, sources, influences, or texts evoked by quotations and allusions—, all of which can be subsumed under a broad notion of ‘intertextuality’. In light of the opportunities provided by the Semantic Web, this paper presents an OWL 2 ontology for the modelling of these intertextual relations—with respect to the fact that the identification of said relations is in itself an act of interpretation that must be reflected upon as such.

## Zusammenfassung

Einen Text zu edieren beinhaltet nicht zuletzt, seine vielfältigen Verbindungen zu anderen Texten aufzuzeigen, seien das kulturelle Kontexte, Einflüsse, zitatweise oder in Anspielungen aufgerufene Texte etc.; alles das fällt unter den (weiter gefassten) Begriff der Intertextualität. Angesichts der Möglichkeiten des Semantic Web wird hier eine OWL 2-Ontologie für die Modellierung dieser intertextuellen Verknüpfungen vorgestellt, die dabei die Tatsache berücksichtigt, dass jede Feststellung von Intertextualität zugleich einen interpretativen Akt darstellt, der als solcher reflektiert werden muss.

## 1 Edition, commentary, and interpretation

Every edition is an interpretation. Every conclusion drawn by the editor based on his or her material can be seen as an interpretive act. This is perhaps most obvious regarding the commentary. As a consequence, over the past decades, very different paths have been proposed for the scholarly edition: some argue that the ideal editor has to exercise interpretive abstinence wherever she can (e.g., Martens 1993), restraining herself especially when it comes to adding a commentary to the edited text. Others suggest that commentary is not to be avoided, even more so: it is not to be restrained to selective (for example, biographical, historical) data on individual text passages.

Instead, the commentary should also discuss semantic and poetic dimensions of the text as a whole, which cannot be done without proposing hypotheses (e.g., Windfuhr 1991, 173f.)—thus engaging in the business of interpretation. Some have drawn radical conclusions, arguing for a section titled ‘interpretation’ as an equal part of a scholarly edition, pointing out that what is usually called ‘commentary’ can only to a very small degree be clearly separated from an interpretation (Stüben 1993).

While there seems to be no clear distinction between these two, in order to determine where commentary/interpretation is appropriate one could refer to a traditional differentiation: the differentiation between ‘primary obscurity’, where obscurity is a function of, for instance, poetics and interpretation should be left to the reader, and ‘secondary obscurity’, where meaning has become obscure due to historical and/or cultural distance and the reader might be misled without the help of the editor (Fuhrmann 1985; Martens 1993, 40f.). In practice, this opens a grey area—and by doing so, transfers the problem of blurred boundaries from the question of commentary and interpretation to yet another unclear distinction.

Commentary cannot be kept free from interpretation; this is also—perhaps especially—true for a specific sub-domain of commentary: accounting for the intertextual relations of the text in question, traditionally under the notion of ‘sources’. This domain has been perceived as “the one closest to the actual task of the editor, publishing a text and its variants” and thus has been used to argue in favour of the justification of the commentary section (Groddeck 1993, 2; transl. B.O.<sup>1</sup>). Still, even the apparently simple act of pointing out a source can interfere with underlying structures of meaning and, consequently, effect the text’s reception (cf. Groddeck 1993, Kanzog 1999).

Linking a text to its sources has always had certain theoretical implications: the term ‘source’ can have a broader or a narrower definition (a narrower distinguishing it from meanings like ‘influence’, ‘reminiscence’, ‘allusion’; Woesler 1997), but the theory behind it usually includes, broadly speaking, an author reading a text, which then in some way affects her writing of a text. The term ‘source’ implies some kind of chronology, a positivistic cause-and-effect-chain, and maybe an insight into the author’s intentions. However, concepts of text-text-relations such as these were complicated by the emergence of intertextuality theories. The relevance of other, more distinct types of intertextual relations was pointed out, and the editorial preference of explicit quotes to implicit ones was questioned (Zittel 1997).<sup>2</sup>

However, the more radical theories of intertextuality, especially the poststructuralist (where the term—not the notion or investigation—of intertextuality was coined), were widely considered impractical and bearing the danger of arbitrariness; among

---

<sup>1</sup> “[D]ie Aufgabe, Zitate im Text nachzuweisen, liegt der eigentlichen Aufgabe des Editors, den Text mit seinen Varianten herauszugeben, am nächsten.” (Groddeck 1993, 2).

<sup>2</sup> Groddeck aims to ‘infect’ scholarly editing with the ‘virus of intertextuality theory’ (Groddeck 1993, 2).

editors, more restrictive classification systems for intertextual relations were deemed applicable, to make explicit the exact nature of any text-text-link (Pfister/Broich and the clearly structured catalogue of the different aspects of an intertextual relationship in Pfister 1985 seemingly met the editors' taste; see Groddeck 1993, Kanzog 1999, Zittel 1997).

It is probably safe to say that it was not only the danger of arbitrariness that limited the amount (and type) of intertextual relations accounted for in an edition but also issues closely related to the print medium: readability and limited space. Thus, in the eighties and early nineties, hypertext-theory and its realisation in HTML were welcomed as the solution for editorial problems such as these. Among other things (like the direct linking of the text and commentary/apparatus, interactive user interfaces and the possibility of multiple perspectives on a text), the representation of intertextual relations and the integration of unlimited amounts of source material were expected to be facilitated and enabled by the digital medium (Sahle 2013, 41–52). Gunter Martens's concept of the 'intertextual commentary' from 1993 already pointed towards what was later discussed under the label 'edition as archive', where an edition formerly referencing external archive material now includes these materials (Martens 1993, Sahle 2007). Martens used the term 'intertextuality' to elegantly evade the problem of interpretation. In his concept of an 'intertextual commentary', intertexts of any kind and the edited text would be published side by side, without specification of the respective relation: "The processing of these materials is not to be forced upon the reader. The drawing of conclusions and integrating them in a reading of the poetic text is left to the recipient's own interpretive operation" (Martens 1993, 49; transl. B.O.<sup>3</sup>).

However, what looks like a new technology—the hypertext—and a methodological concept—the 'intertextual commentary'—coming together to solve the problem leaves one crucial point untouched: Martens's proposal helps avoid explicit interpretation, but not only leaves the interpretation to the reader, but also has its own theoretical implications. Unless the 'intertextual commentary' consists of an unfiltered and unstructured amount of texts (which would render it useless for its intended purpose, but would make for a useful tool of its own; Baßler/Karczewski 2009), it has to undergo processes of selection and combination, processes that contain an interpretive aspect and certainly draw from theoretical prerequisites.

Wherever one stands in this discussion about which intertextual relations to account for and which materials to include—it is fair to say that as soon as an intertextual relation is pointed out, some kind of theoretical framework plays a role and should be made explicit to uphold scholarly transparency.

---

<sup>3</sup> "Die Verarbeitung solcher Materialien ist dabei dem Leser nicht vorgegeben. Schlüsse daraus zu ziehen und in die Deutung des poetischen Textes einzubringen, ist der eigenen interpretierenden Arbeit des Rezipienten überlassen." (Martens 1993, 49).

## 2 Building and transcending the ‘data silo’

Let us assume a basic digital edition concerned with a text’s intertexts (for instance, the author’s sources): it would usually include, besides the edited text, a document containing said sources/intertexts (maybe as a list of bibliographical references). These two components can have the form of two TEI-XML documents. Specifying theoretical and methodological implications in the most general way comes down to specifying the nature of the link between a passage of the edited text in the one document and the data representing its intertext in the other. For this purpose, TEI provides several options (TEI-Consortium 2020)—an obvious one would be to add the attributes `@type` and `@subtype` to a `<quote>` element. The explanatory power of these possibilities is, however, quite limited: attribute values can be used in a self-explanatory way, leading to a limitation to common terms without regard for the theoretical background. Otherwise, attribute values can point to a more sophisticated classification of intertextual relations that then needs to be underpinned in a statement in natural language somewhere in the document’s metadata or a website’s editorial note.

This is one example of why ‘traditional’ digital editions tend to fall under the verdict of being ‘data silos’: of not representing the published statements (about the edited text, its sources, etc.) in an interoperable or—ideally—a machine-readable way. In this respect, such an edition is no different from a printed one (cf. for example, van Zundert 2016, 101). The intertextual relations accounted for in the edition are neither classified as such nor differentiated in any way that, to a non-human recipient, would not appear as white noise. Furthermore, the notion of the data silo also includes what I would like to call a ‘theory silo’ since without classification an assessment of theoretical implications is impossible.

There lies a huge potential for interconnectivity in an edition’s list of intertexts. This is the place where an edition opens itself to literary or cultural discourse. In many cases, the referenced texts might already exist in a digitized form somewhere on the Internet. Linked Data is an obvious choice to represent these intertextual relations.<sup>4</sup> If an edition exposes its intertexts through Linked Data, short- or medium-term benefits would be interoperability, data-retrieval and -visualization. A visionary long-term perspective would include a decentralized database of intertextual relations throughout numerous online editions and archives; a Linked-Data-Cloud of literature and literary studies.

---

<sup>4</sup> Early on, a conceptual analogy between the RDF-triple and intertextual relation was noticed—both consist of two elements linked by a specified relation (Tennis 2004). Fabio Ciotti has pointed out that enrichment of data based on Linked Data vocabularies “has similarities with the humanities tradition of annotation and comment” (Ciotti 2014).

Projects and concepts in the digital humanities that tackle the data silo problem build on Linked Data technologies. The *Dante Sources Project*, for example, provides RDF-data on Dante's primary sources, cited authors and more via a SPARQL-endpoint (Bartalesi and Meghini 2017). Interoperability, however, is just the most obvious benefit: as Fabio Ciotti has pointed out, terms and concepts in humanities and cultural studies are often vague and used ambiguously, relying on the reader's prior knowledge and intellectual ability to infer intended meaning. In the course of semantic modelling and formalising, this kind of implicit knowledge is made explicit (Ciotti 2015); the vocabulary to do so is, as part of the technical prerequisites of Linked Data, accessible and provides context for any term or concept inside a conceptual hierarchy. In addition, ontologies in the Semantic Web are meant to represent only a segment of the world as perceived from one specific epistemological point of view (and explicitly so).<sup>5</sup> Every 'thing' in the Semantic Web, every URI, can be subject to numerous classifications based on various ontologies, giving way to a coexistence of epistemologies.

There are a couple of preliminary studies concerned with the utilization of Linked Data in literary studies—mostly of a merely recommending nature. In 2010, Adrian Pohl in a blog-entry emphasized the potential of bibliographic metadata in Linked Data formats, suggesting 'participative catalogues,' where implicit relations between texts could, by way of crowdsourcing, be added to the data (Pohl 2010). In a 2011 article, Sarah Bartlett and Bill Hughes illustrated a possible Linked Data approach to intertextuality based, like Pohl's proposal, on Gérard Genette's terminology (Genette 1997). Charlotte Brontë's *Jane Eyre* is used as a test case in a dry-run-modelling of the text's intertextual (and other) relations (Bartlett and Hughes 2011). In a paper from 2014, Pauline Rafferty evaluated available options for the representation of intertextuality in Linked Data. This possibility is provided by web platforms like 'LibraryThing,' albeit, according to Rafferty, with some conceptual shortcomings. She has suggested a typology that would expand the 17 types of intertextuality provided there (Rafferty 2014).

There are also projects concerned with intertextuality that already implement Linked Data technology. One of them is the *Dante Sources Project*, which even provides a vocabulary for this purpose (and was awarded a DH Award in 2015 in the category 'Best DH Tool or Suite of Tools'). It allows for searching a number of Dante's works as well as their sources from many different angles, starting from Dante's work, cited literature and authors, thematic areas and so on. The ontology created for these intertextual relations is a tailor-made solution strictly for the project's own means. It provides three distinct types of intertextual relation or citation: `citesAsExplicitCitation`, `citesAsStrictCitation`, and `citesAsGeneralCitation`. The first is a citation made

---

<sup>5</sup> Elsewhere, I called it the "immanent constructivism" of the Semantic Web; proposals for a better wording are still welcome (Oberreither 2019).

explicit by the author himself, the second is a citation indicated by a scholar, the third is a reference to a concept instead of a specific source (Bartalesi and Meghini 2017). While this ontology fits the project's research question, its domain regarding intertextuality is limited to these three types of intertextual relations, reducing its reusability to projects with matching research interests. This limitation is also an issue in approaches suggested by Rafferty or Bartlett/Hughes.

### 3 Opportunities and desiderata

Thinking about intertextuality in digital literary studies (and here especially editing), we have already come across a couple of things to consider when going forward in constructing a Linked Data vocabulary for this purpose. One aspect worth pointing out is that the concepts and projects referred to above are based on a certain notion of intertextuality that limits its scope to a fixed set of relation types, or to relations inside the realm of literature. If interpreted in a broader way, however, the notion of intertextuality covers not only a possibly infinite number of types of said relations but also the relations between different types of texts. These could concern literary and non-literary texts (for example, scholarly) as well as relations across this divide, when, for instance, a scholarly text analyses a literary text. Intertextuality can also be determined between a text and a text corpus (when questions of classification regarding literary epochs, movements, schools, discourses, etc. arise) as well as between a text and abstract concepts like rhetorical devices, motifs, etc. Thus, modelling intertextuality means modelling a wide variety of scholarly activities. This aspect adds to the relevance of the domain as well as to the scope of an ontology covering it. Before proceeding to the architecture of a vocabulary that would take all this into account, let us sum up the desiderata we have come across so far, and elaborate:

**Interoperability:** Linked Data is a promising strategy to address the data silo problem. Ontologies, as the cornerstone of interoperability, provide the machine-readable vocabulary that allows for merging different data sets. Interoperability, in this respect, relies on standardization, and it is exactly this aspect that has caught the critical attention of humanists. Standardization opposes individuality of approach; standardization based on the epistemology of one field of research may not fit the needs of another. While the main standard in digital text editing—TEI-XML—is not a standard, since in practice it is open to adaptation and one and any problem can be approached in many different ways, machine-readability calls for stronger restrictions. There are several vocabularies in fields neighbouring literary studies that can serve as points of contact or reference. Alignment to existing standards is a key aspect in Linked Data, but these standards have to be evaluated with respect to their methodological/epistemological and practical implications.

**Domain specificity:** In respect to the applicability to the domain of literary studies (and probably other research areas inside the humanities as well), any existing Linked-Data-vocabulary—if an alignment makes sense—will have to be extended for the humanities scholars’ use. Since links between, to and from literary texts concern more than just the wording of a text, classes and relations tailored for the domain of literary studies will have to address the seemingly simple, yet profound question: what is it that we are actually talking about when we talk about (literary) texts?

**Universality:** Any vocabulary for the purpose of representing intertextuality in literary studies has to include as wide a variety of methodologies as possible, in order to not involuntarily exclude relevant ones. This means finding a modelling practice that allows for unlimited types of intertextuality; this also means looking for a common denominator of relevant methodological frameworks, as well as sorting out aspects that, if implemented in the modelling, would limit its reach.

These three sections already hint at the design of an ontology covering the domain in question; at the same time, addressing them sometimes means weighing one against the other two (maybe every vocabulary can be characterized by a specific balance of these aspects). Accordingly, the following subchapters give an insight into the considerations leading to the architecture of the proposed ontology, and possibly some points of thought for other, related ventures. Afterwards, an important final aspect will be addressed: the question of practicability, with special attention to the applicability for scholarly editions.

### 3.1 Interoperability

Looking for a vocabulary to model statements about intertextuality, one soon stumbles upon CiTO, the *Citation Typing Ontology*. This ontology is part of the SPAR ontology suite, a collection of ontologies for the purpose of enriching publications with Semantic Web-metadata (Peroni/Shotton 2018). CiTO provides a list of no less than 43 types of citation, ranging from agrees with to uses method in, including critiques, discusses, parodies, plagiarizes and so forth. The ontology, as these examples show, is designed to represent citations inside academic discourse. Since citation practices are very much standardized in academic practice, the extensive list provided likely covers its domain to a large extent. However, it is still a closed list, and one that is, by design, not meant to cover the myriad ways of intertextual reference found in literature, nor the types of intertextuality conceptualised in literary theory. CiTO’s citation types are modelled as object properties, linking a subject and an object, such as a citing and a cited, a plagiarizing and the plagiarized text (see Figure 1 [a]). As can be seen from these examples, relations modelled with CiTO are unidirectional. This makes sense if applied to a scholarly corpus, where one wants to model chains of citations running through a text corpus as chronological chains of cause and effect. In

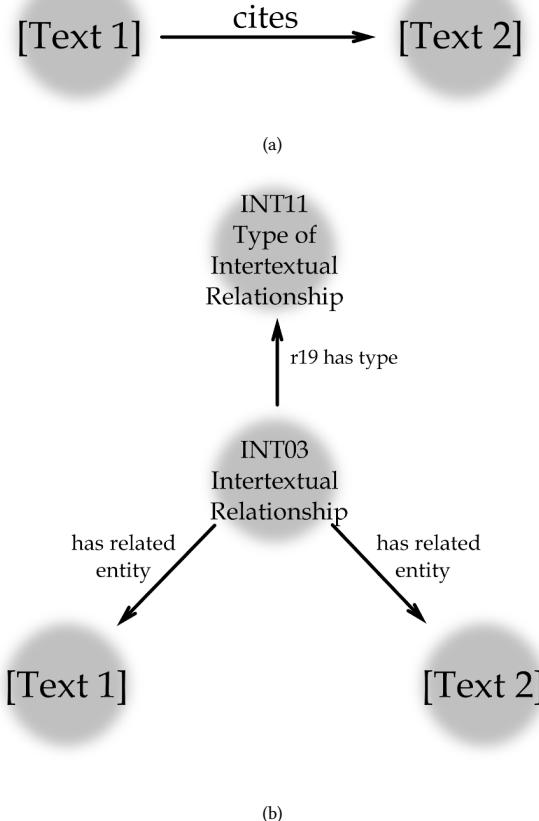


Figure 1: Intertextual relations modelled as object properties or as instances with an attached type.

literary studies, however, bi- and a-directional relations are also quite common. As a most prominent example for bi-directional relations, one can point to Julia Kristeva's notion of a textual universe where 'meaning' in texts is the product of mutual "absorption and transformation" (Kristeva [1967]). An a-directional intertextual relation might be found in relations that entirely depend on a scholar's reading (Bronfen's 'cross mapping' could be interpreted as an example for a methodology of this kind; Bronfen 2002).

As a consequence of the above, the proposed vocabulary first of all provides a class, not object properties, for the intertextual relation (thus reifying the rela-

tion;<sup>6</sup> see Figure 1 [b]). To specify the relation, the user creates an instance of Type of Intertextual Relationship, linking it to the relation. It should be mentioned at this point that CiTOs object properties—the list of 43 types of citation—can be re-introduced here as instances of this class. Secondly, the object properties linking the relation to the texts constituting the (at least) two anchor points of the relation do not specify its direction; this is done only in the respective sub-properties.

An important aspect is the question of what an intertextual relation relates to, meaning: where to take the classes and properties to model the instances in question (such as texts). Several vocabularies for this purpose, built for the representation of bibliographic metadata in a library context, already exist. The most elaborate one is FRBRoo (Bekiaris et al. 2016), which offers classes for texts on every ontological level, from the Work as the most abstract idea and the Expression as a work in a structured, yet immaterial form, down to its material Manifestation.<sup>7</sup>

FRBRoo is aligned to CIDOC-CRM (ICOM/CRM Special Interest Group 2020): its core classes like Work or Expression, but also the dynamic ones like Work Conception or Expression Creation are subclasses of respective classes from the CRM. However, the ontology proposed is linked to this model not only indirectly via FRBRoo but also directly. This is due to the broad domain of CIDOC-CRM: it covers the domain of cultural heritage from a museologist's, archivist's or archaeologist's perspective and offers, at a high hierarchical level, a class for Conceptual Objects (as opposed to Physical Things; s. Le Boeuf et al. 2019). The relevance of this for our case is beyond question, whenever we talk about abstract concepts like textual features or intertextual relations.

### 3.2 Domain specificity

A characteristic of literature is the eventual indecisiveness of meaning; this includes the ‘meaning’ of a quotation. Thus, a characteristic of literary and cultural studies is the more or less general agreement that statements in these fields of research are ‘just’ interpretations, intrinsically tied to their epistemology, their methodology and

---

<sup>6</sup> As part of its alignment to the Open Annotation Standard, this was also implemented in CiTO (Shotton 2013).

<sup>7</sup> One of its most useful classes in terms of bibliographic information certainly is the Publication Expression—the text in an immaterial form, yet so specific as to refer to a distinct edition. This makes for a text with year and place of publication as cited in any bibliography. If I am not mistaken, this train of thought leads to another important notion of text, which however is maybe not supported by FRBRoo: a text that is part of a larger text, yet has all the bibliographical information linking it not only to the larger text but also to time and place of publication. This class is needed when modelling a research paper published in a journal, for instance. A Publication Expression can, according to domain/range of the relevant property (P165), incorporate another Publication Expression, but the examples given for this class only apply to complete publications, not their parts (like a text in an anthology or a periodical).

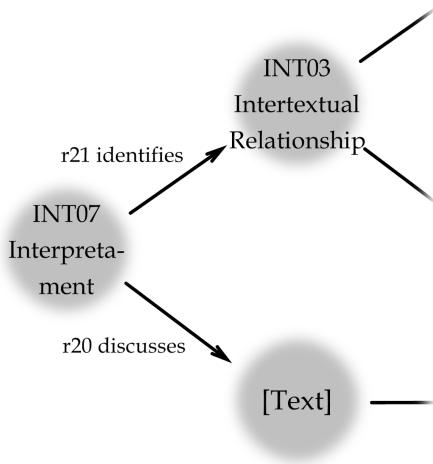


Figure 2: The Interpretament as a means to model the scholarly act.

theoretical background. The humanities have a strong tendency towards reflection on the question of perspective. Thus, a certain feature should not be directly tied to an investigated object. Instead, an instance is needed that accounts for the fact that every attribution is an interpretation. For this instance, the proposed vocabulary provides the class Interpretament. Instances of this class serve as the entities to which the identification of features and intertextual relations as well as the more general notion of ‘discussing’ (a text, a text passage etc.) are tied (see Figure 2). They introduce the meta-level necessary for the discussion of methods and theoretical fundamentals and may be tied to a research output (like a paper, or an RDF document) that they stem from.<sup>8</sup> By making Interpretaments explicit, sources and other research findings such as supporting evidence can be linked to them.

The next step is to ask what the Interpretament identifies when it identifies a feature of a text. The feature? Hardly ever. Usually, what is being identified is the specific form in which a textual phenomenon resembles other phenomena in other

<sup>8</sup> CIDOCinf, an extension of CIDOC-CRM, models the mental process (the ‘inference making’) that leads to academic output (Stead et al. 2018); it offers the means to capture argumentation within the descriptive and empirical sciences historically, and thus scientific progress. However, the Interpretament excludes the temporal facet represented in the event of inference making; it is rather a representation of its textual traces (where historical markers are only in the respective text’s metadata).

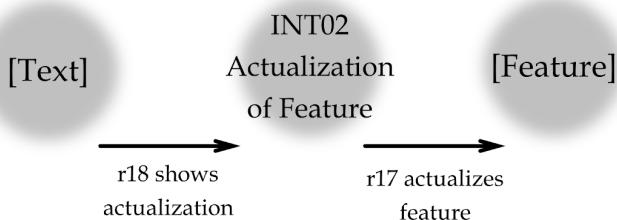


Figure 3: The Actualization of Feature as an intermediary between text and feature.

texts, an abstraction of which is called a textual feature. In the case of the phenomenon in a specific text, we do not talk about features but about the *actualization* of a feature in a text. Thus, a text can never ‘have’ a feature, it can show its actualization, which constitutes an intermediary between the text and the feature (see Figure 3). This also makes sense for modelling reasons: this way, one can model the common view that the same feature—like a ‘Faustian character’—can appear in different texts, and different forms (or actualizations). The class *Actualization of Feature* is the link between the text and the feature in question. Also—as is its main purpose here—the class is linked to the intertextual relation, in the many cases where intertextuality does not mean word-by-word-quotation, but instead refers to reoccurring themes, motifs, rhetoric, characters, etc. For example, this allows for the statement of the fact that a fictional text derives one of its characters (maybe a Hanswurst-, Faust- or Sherlock-character) from another text or from some literary tradition.<sup>9</sup>

<sup>9</sup> CIDOC-CRM provides the event E13 *Attribute Assignment*, which in a way merges the classes *Interpretament* and *Actualization of Feature*. However, following this model would be inadequate as we cannot reduce interpretaments to events in time (but rather latencies in text), but also, since the *Attribute Assignment* is defined as “directly relating the respective pair of items or concepts” (ICOM/CRM Special Interest Group. 2020, 10)—with a built-in shortcut, so to speak, while we insist on detours. “Per default”, the CRM states, all instances of properties described in a knowledge base “are the opinion of the team maintaining the knowledge base. This fact must not individually be registered for all instances of properties provided by the maintaining team, because it would result in an endless recursion of whose opinion was the description of an opinion.” (ibid., 12)—However, we want to individually register these facts, and this recursion is kind of what we are after. Despite the great opportunities provided by CIDOC-CRM, and the prominence of the model, one has to give thought to a concern expressed by Jörg Wetzlaufer: “Ontologies (just like controlled vocabularies) can’t completely defuse the problem of subjectivity of world view ... can we really grasp the humanities in a concept which was originally meant for a museum context?” (Wetzlaufer 2018)

Consequently, a class is required that provides textual features, as well as subclasses for specification. The idea of building a conceptual hierarchy that covers every possible textual feature relevant in literary studies is one that could easily cause vertigo ('things to talk about when talking about a text' is more an abyss than a class). However, first efforts have been made: a superclass called *Receptional Entities*, itself a subclass of *Conceptual Object* with the subclasses *Semantic*, *Rhetoric* and *Formal Feature*. Subclasses to these three are only a provisional part of the ontology, and currently, externalising this into, for example, SKOS-classifications is being contemplated.

The last question in this regard is this: where are all these features actualized? This sometimes takes place in the text (like genre-specific features, or characters relevant throughout the text), most of the time in parts of the text—text passages. Consequently, a class for the text passage was introduced to the model (see Figure 4). On the other end of the spectrum, one might want to relate a text (or passage) not to another text, but to an abstract entity that is considerably larger than a text—an oeuvre perhaps, a discourse, an epoch. For these entities, which can all be seen as (or reduced to) text corpora, the class *Architextual Entity* was introduced (utilizing the term 'architext' as coined in Genette's fundamental study on intertextuality—Genette 1997), again as a subclass of *Conceptual Object*.

### 3.3 Universality

As mentioned above, the attribution of a feature, the identification of a relation can be modelled as a conceptual object. This is the case in the ontology proposed, but not in the important standard it is aligned to, CIDOC-CRM. At the heart of both CIDOC and FRBRoo, there are events. Both can be seen as divided into a static and a dynamic half, where the events from the dynamic half serve as the 'semantic glue' for all the other things (such as people, places, archive materials). The declaration/classification of an intertextual relation or any kind of text feature in this vocabulary does not rely on events (as a consequence of which one could call it only half an alignment). The thought behind this is simple: the ontology is meant to represent statements typically made in texts like commentaries, forewords, research papers, etc. These texts contain declarations and classifications, not as acts, which would be events in time, but as textual phenomena that probably, but not unconditionally, are the conserved results of events (intellectual acts), and ultimately are relevant only in their textual form. The conceptual bonus of this modelling lies in its integrative potential: no matter what one believes about how an *interpretament* got *into* a text—via an authorial intention, a certain reading, a contextual structure—we can all agree that somewhere there is a text.

This is one example of a path a vocabulary designed to include as many theoretical frameworks as possible has to choose: reduction, or, to be more precise: reduction of

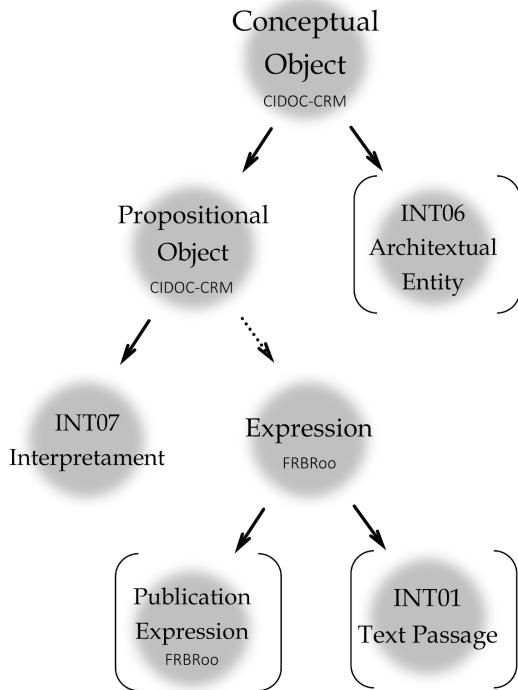


Figure 4: Notions of “Text” in hierarchical context.

elements that by default limit methodologies. As another example, one could point at the class *Type of Intertextual Relationship*. Instances of this class can have various methodological implications, and many of them could have been modelled more explicitly. However, this would have undermined the universal applicability intended by the ontology. Look at the difference between, say, Harold Bloom’s theory of ‘anxiety of influence’ and the more recent theorem of the ‘cultural archive’ by Moritz Baßler. Both of these count as intertextuality theories (or theories including a subtheory concerned with intertextuality) but reside on quite distant points of the spectrum: whereas one of them captures a writer’s resistance against his literary father-figures (Bloom 1973), the other one renders intertextual relations as, to put it in a nutshell, any kind of parallel between two texts that are part of the same

synchronous text corpus (Baßler 2005). Going for the exact representation of either of these two makes the representation of the other one difficult, if not impossible. While Bloom's theory, to name a significant difference, implies a notion of time and events, Baßler's does not. The proposed vocabulary has to take this wide variety of theories into account, by reducing all theoretical particulars under the notion of a simple 'Type of Intertextual Relationship,' which means: reducing them to instances of a class of this name (or, more likely: one of its subclasses or sub-subclasses).

### 3.4 Practicability: detail, granularity, and perspectivity

At this point, I'd like to return to the problem outlined at the beginning of this paper: the impossibility of distinguishing between commentary and interpretation. This is a fact that needs to be taken into account in any modelling of intertextual relations, all the more so for scholarly editions. The goal to produce a valid version of a text (whatever 'valid' means in the respective project context) goes hand in hand with the need to explicitly identify each act of interpretation as such—such as the declaration of an intertextual relation.

However, the following is also true: the fact that identifying intertexts is an interpretive act does by no means imply that an edition is obliged to do all the interpretation and model it in-depth. Providing knowledge on the edited text's intertexts is a service for a community of readers, who in turn bring their own interpretations to the text (see again Martens 1993, 49). Also, a practical matter must not be neglected: editions often deal with large quantities of texts, thus even larger quantities of intertextual relations (one may think of literary collages of quotations like T. S. Eliot's *The Waste Land* or Karl Kraus' *Third Walpurgis Night*), the in-depth analysis of which would exceed the scope (and, as one could argue, the purpose) of an edition project.

Every edition project concerned with intertextuality has to deal with this issue and find a solution from a wide spectrum of possibilities, ranging from the rudimentary to the in-depth modelling of intertextual relations. Thus, a vocabulary built for this purpose needs to provide flexibility.

As mentioned before, the modelling of intertextuality can take many forms, depending on what kinds of texts and what dimension of text (whole texts, text passages, text features, 'architexts') one wants to link and what additional information one wants to provide. This wide scope of different modelling options can be differentiated by three aspects: detail, granularity and perspectivity.

Perspectivity comes into play when a scholar's research, the edition or commentary itself, is included in the modelling. In that case, modelling the interpretive acts as *Interpretaments* (which then, in another step, could be linked to the scholarly discourse) would be useful (see Figure 5).

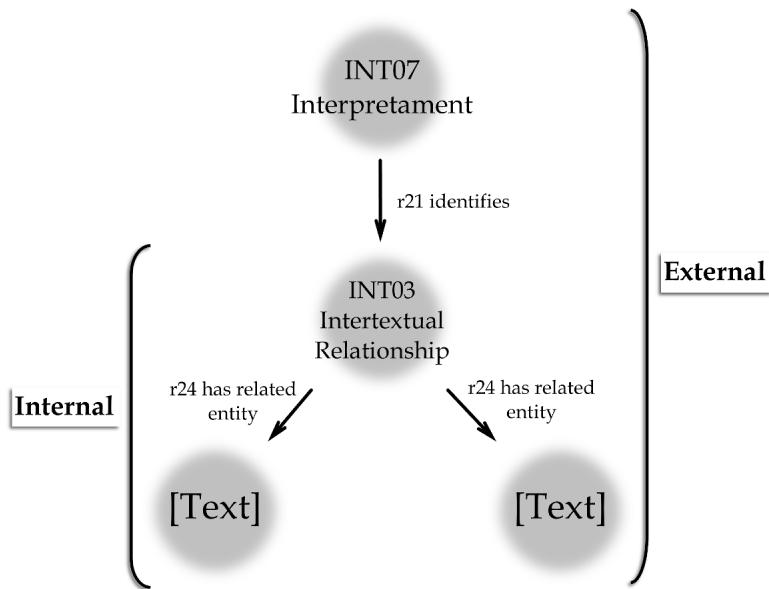


Figure 5: Internal and external perspectives on a scholarly act.

In terms of granularity, a single intertextual relation can be determined between two texts as a whole (low granularity) or numerous relations between their most minute fragments or features (high granularity). Finally, one can ask how much detail the projected model of, say, a digital edition's intertextual aspect should provide. The detail could range from the simple statement that an edited text refers to some source, to a modelling that includes a classification of the relation, to one that includes a source for a definition of said classification. Furthermore, it could be more specific in terms of what the intertextual relation actually relates to: the edited text as a whole, a text passage, or a certain feature shown by a passage (see Figure 6).

## 4 Conclusion

At more than one point in reading this paper, one may have been taken by surprise by yet another dissection of yet another aspect of the seemingly straightforward act of determining and classifying intertextual relations. (This experience then would mirror quite well the process of creating the vocabulary.) As the last remarks probably

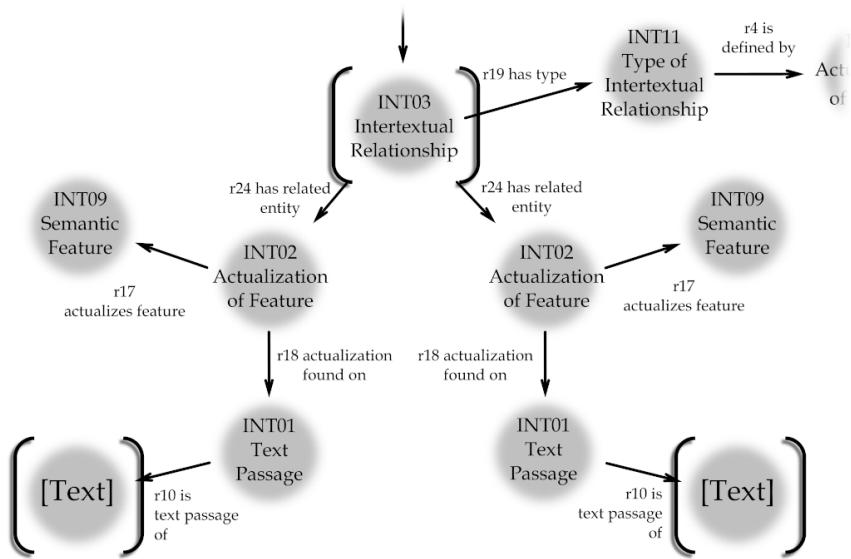


Figure 6: A modelling in higher detail; a lower detail modelling would make do with the three instances in brackets.

made clear, the vocabulary intends to provide the means for two things: to represent the respective findings and scholarly activities relatively in-depth—considering the sometimes hidden complexity of, for example, a literary scholar’s day-to-day business—but also, if necessary, to model intertextuality in a minimalistic way. The latter then could serve as a starting point for a gradual enrichment.

The proposed ontology *INTRO – an Intertextual Relationships Ontology for literary studies* (Oberreither 2018–) is in an advanced beta-state right now (early 2020). A current version is available online, as well as a test case, modelling a paragraph of one of Elisabeth Bronfen’s ‘cross mappings’ and a documentation of central classes.

## Bibliography

- Bartalesi, Valentina and Carlo Meghini. 2017. "Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study." *Semantic Web Journal* 8 (3): 385-94. Accessed February 15, 2021. <https://content.iospress.com/articles/semantic-web/sw198>.
- Bartlett, Sarah, and Bill Hughes. 2011. "Intertextuality and the Semantic Web: Jane Eyre as a test case for modelling literary relationships with Linked Data." *Serials: The Journal for the Serials Community* 24 (2): 160-5. doi:10.1629/24160.
- Baßler, Moritz. 2005. *Die kultурpoetische Funktion und das Archiv. Eine literaturwissenschaftliche Text-Kontext-Theorie*. Tübingen: Francke.
- , and Rainer Karczewski. 2009. "Computergestützte Literaturwissenschaft als Kulturwissenschaft. Eine Wunschliste." *Jahrbuch für Computerphilologie* 9: 27–34. Accessed February 15, 2021. <http://computerphilologie.digital-humanities.de/jg07/BassKarcz.html>.
- Bekiaris, Chryssoula, Martin Doerr, Patrick Le Boeuf, and Pat Riva, eds. 2016. Definition of FRBRoo. A Conceptual Model for Bibliographic Information in Object-Oriented Formalism. Version 2.4. Accessed February 15, 2021. [http://www.cidoc-crm.org/frbroo/sites/default/files/FRBRoo\\_V2.4.pdf](http://www.cidoc-crm.org/frbroo/sites/default/files/FRBRoo_V2.4.pdf).
- Bloom, Harold. 1973. *The Anxiety of Influence. A Theory of Poetry*. New York: Oxford University Press.
- Bronfen, Elisabeth. 2002. "Cross-Mapping. Kulturwissenschaft als Kartographie von erzählender und visueller Sprache." *Kulturwissenschaften. Forschung – Praxis – Positionen*, ed. by Lutz Musner, 110–34. Wien: WUV.
- Ciotti, Fabio. 2014. "Digital literary and cultural studies: the state of the art and perspectives." *Between* IV (8). Accessed February 15, 2021. <http://ojs.unica.it/index.php/between/article/view/1392/1183>.
- . 2015. "Digital methods for Literary Criticism." [Presentation at the "Digital Humanities Autumn School 2015", Trier]. Accessed February 15, 2021. <http://didattica.uniroma2.it/files/scarica/insegnamento/161783-Informatica-Umanistica-Lm-Per-Il-Llea/37175-Slide>.
- Connell, Sarah, and Julia Flanders. 2017. "Intertextual Networks: Theorizing and Encoding Textual Connections in Early Women's Writing." *Women Writers Project. The Blog*. Accessed February 15, 2021. <http://web.archive.org/web/20210128174049/https://wwp.northeastern.edu/blog/dh2017>.
- Fuhrmann, Manfred. 1985. "Kommentierte Klassiker?" In *Warum Klassiker? Ein Almanach zur Eröffnung der Bibliothek deutscher Klassiker*, ed. by Gottfried Honnefelder, 37–57. Frankfurt am Main: Deutscher Klassiker-Verlag.
- Genette, Gérard. Palimpsests. 1997 [1982]. *Palimpsests. Literature in the Second Degree*, trans. by Channa Newman and Claude Doubinsky. Lincoln: University of Nebraska Press.
- Groddeck, Wolfram. 1993. "'Und das Wort hab' ich vergessen.' Intertextualität als Herausforderung und Grenzbestimmung philologischen Kommentierens, dargestellt an einem Gedicht von Heinrich Heine." In *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition, 4.–7. März 1992, autor- und problembezogene Referate*, ed. by Gunter Martens, 1–10. Tübingen: de Gruyter.

- ICOM/CRM Special Interest Group. 2020. "Volume A: Definition of the CIDOC Conceptual Reference Model." Version 7.0.1. Accessed February 15, 2021. [http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM\\_v.7.0.1\\_%2018-10-2020.pdf](http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v.7.0.1_%2018-10-2020.pdf).
- Kanzog, Klaus. 1999. "Baustein, Kontext, Intertextualität. Zur Beziehung von Eigen- und Fremdtext als typologisches und editorisches Problem." In *Produktion und Kontext. Beiträge der internationalen Fachtagung der Arbeitsgemeinschaft für Germanistische Edition im Constantijn Huygens Instituut, Den Haag, 4. bis 7. März 1998*, ed. by Hendricus Theodorus Maria van Vliet, 329–35. Tübingen: Niemeyer.
- Kristeva, Julia. 2003 [1967]. "Bachtin, das Wort, der Dialog und der Roman." In *Texte zur Literaturtheorie der Gegenwart*, ed. and commented by Dorothee Kimmich, Rolf Günter Renner, and Bernd Stiegler, 334–48. Stuttgart: Reclam.
- Le Boeuf, Patrick, Martin Doerr, Christian Emil Ore, and Stephen Stead, eds. 2019. *Definition of the CIDOC Conceptual Reference Model. Volume A. Version 6.2.6*. Accessed February 15, 2021. <http://www.cidoc-crm.org/Version/version-6.2.6>.
- Martens, Gunter. 1993. "Kommentar – Hilfestellung oder Bevormundung des Lesers?" *editio 7*: 36–50.
- Oberreither, Bernhard. 2018– INTRO – an Intertextual Relationships Ontology for literary studies. Accessed February 15, 2021. <https://github.com/BOberreither/INTRO>.
- . 2019. "Zwei Überlegungen zur Konzeption einer Linked-Data-Ontologie für die Literaturwissenschaften." In *digital humanities austria 2018. empowering researchers*, ed. by Marlene Ernst et al., 134–9. 2020. doi:10.1553/dha-proceedings2018s134.
- Peroni, Silvio, and David Shotton. 2018. "The SPAR Ontologies." *Submitted to the 17th International Semantic Web Conference. Proceedings, Part II*. Accessed February 15, 2021. <https://w3id.org/spar/article/spar-iswc2018>.
- Pfister, Manfred. 1985. "Konzepte der Intertextualität." In *Intertextualität. Formen, Funktionen, anglistische Fallstudien*, ed. by Ulrich Broich and Manfred Pfister, 1–30. Tübingen: Niemeyer.
- Pohl, Adrian. 2010. "Partizipativer Katalog, Intertextualität und Linked Data." *Übertext: Blog*. Last modified February 2, 2010. Accessed February 15, 2021. <http://www.uebertext.org/2010/02/>.
- Rafferty, Pauline. 2016. "Genette, Intertextuality, and Knowledge Organization". Accessed February 15, 2021. [https://www.researchgate.net/publication/290310957\\_Genette\\_intertextuality\\_and\\_knowledge\\_organization](https://www.researchgate.net/publication/290310957_Genette_intertextuality_and_knowledge_organization).
- Sahle, Patrick. 2007. "Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung." In *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, ed. by Michael Stolz, Lucas Marco Gisi, and Jan Loop. Accessed February 15, 2021. [https://www.germanistik.ch/scripts/download.php?id=Digitales\\_Archiv\\_und\\_digitale\\_Edition](https://www.germanistik.ch/scripts/download.php?id=Digitales_Archiv_und_digitale_Edition).
- . 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik*. Norderstedt: Books on Demand.
- Shotton, David. 2013. "Extending CiTO to enable use of the Open Annotation Data Model to describe citations." *Semantic Publishing*. Accessed May 5, 2021. <https://semanticpublishing.wordpress.com/2013/07/03/extending-cito-for-open-annotations>.

- Stead, Stephen, Martin Doerr, Christian-Emil Ore et al. 2018. "CRMInf: the Argumentation Model. An Extension of CIDOC-CRM to support argumentation. Version 0.9 (Draft)." Accessed February 15, 2021. <https://cidoc-crm.org/crminf/ModelVersion/version-0.9>.
- Stüben, Jens. 1933. "Interpretation statt Kommentar. Ferdinand von Saar und Marie von Ebner-Eschenbach: Kritische Texte und Deutungen." *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition, 4.–7. März 1992, autor- und problembezogene Referate*, ed. by Gunter Martens, 99–107. Berlin: de Gruyter.
- TEI Consortium. 2020. "TEI P5: Guidelines for Electronic Text Encoding and Interchange." Accessed February 15, 2021. <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.
- Tennis, Joseph F. 2004. "URIs and Intertextuality: Incumbent Philosophical Commitments in the Development of the Semantic Web." In *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference*, ed. by Ia C. McIlwaine, 103–8. Würzburg: Ergon.
- Van Zundert, Yoris. 2016. "Barely beyond the book." In *Digital Scholarly Editing. Theories and Practices*, ed. by Matthew James Driscoll, and Elena Pierazzo, 83–106. Cambridge, UK: Open Book Publishers.
- Wettlaufer, Jörg. 2018. "Der nächste Schritt? Semantic Web und digitale Editionen." In *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, ed. by Roland S. Kamzelak and Timo Steyer. doi:10.17175/sb002\_007.
- Windfuhr, Manfred. 1991. "Zum Verständnis von Kommentar und Genese." *editio* 5: 173–7.
- Woesler, Winfried. 1997. "Der Autor und seine Quellen aus der Sicht des neugermanistischen Editors." In: *Quelle – Text – Edition. Ergebnisse der österreichisch-deutschen Fachtagung der Arbeitsgemeinschaft für germanistische Edition in Graz vom 28.2.–3.3.1996*, ed. by Anton Schwob, Erwin Streitfeld, and Karin Kranich-Hofbauer, 3–19. Berlin: de Gruyter.
- Zittel, Claus. 1997. "Von den Dichtern. Quellenforschung versus Intertextualitätskonzepte, dargestellt anhand eines Kapitels aus Friedrich Nietzsches *Also sprach Zarathustra*." In *Quelle – Text – Edition. Ergebnisse der österreichisch-deutschen Fachtagung der Arbeitsgemeinschaft für germanistische Edition in Graz vom 28.2.–3.3.1996*, ed. by Anton Schwob, Erwin Streitfeld, and Karin Kranich-Hofbauer, 315–31. Berlin: de Gruyter.



# **“Don’t Worry, We’re also Doing a Book!” – A Hybrid Edition of the Correspondence of Bernhard and Hieronymus Pez OSB**

Daniel Schopper, Thomas Wallnig, Victor Wang<sup>1</sup>

## **Abstract**

Although digital methods have become an undeniable requisite for scholarly editing, only few projects attempt to standardize data and to envision reusable modules. The learned correspondence of the Austrian Benedictine historians Bernhard and Hieronymus Pez has been object of scholarly efforts for more than fifteen years, transitioning from a traditional print edition to a digital paradigm. The article outlines in detail the features of a hybrid solution for the Pez correspondence edition: a pilot workflow that allows for print and digital publication in the series *Quelleneditionen des Instituts für Österreichische Geschichtsforschung*, while envisioning long-time data storage in the ARCHE repository of the Austrian Academy of Sciences.

## **Zusammenfassung**

Obwohl digitale Methoden heute unerlässlich für wissenschaftliches Edieren sind, bemühen sich doch wenige Projekte um Datenstandardisierung und Wiederverwendbarkeit von Modulen. Die gelehrte Korrespondenz der österreichischen Benediktinerhistoriker Bernhard und Hieronymus Pez ist seit mehr als fünfzehn Jahren Gegenstand editorischer Bemühungen, die eine Transformation von einer klassischen Printedition zu einem digitalen Paradigma mit sich gebracht haben. Der Artikel beschreibt im Detail die Beschaffenheit einer hybriden Lösung für die Pez-Korrespondenz-Edition: einen exemplarischen Arbeitsablauf, der eine gedruckte und digitale Edition in der Editionsreihe *Quelleneditionen des Instituts für Österreichische Geschichtsforschung* ermöglichen soll, während zugleich die Langzeitspeicherung der Daten im Repository ARCHE der Österreichischen Akademie der Wissenschaften gewährleistet wird.

---

<sup>1</sup> This text was collectively conceived and written by the authors, albeit with distributed roles regarding the individual chapters, and the creation process was supported by Herwig Weigl, Patrick Fiska, and Ursula Huber. It reflects the state of discussion in 2019/20.

## 1 Introduction: on hybrid scholarly editing

In many cases [...] print is still the medium of choice for the publication of the fruits of scholarly endeavour, and digital editions have not yet been accepted among the scholarly community; they well may be used behind the curtains, but when it comes to citations and referencing, it is the printed edition that takes centre stage, thereby depriving the producer of the digital edition of legitimate acknowledgement. (Driscoll and Pierazzo 2016, 15).

Expressed in 2016 in one of the groundbreaking publications on critical scholarly editing, this assertion has lost none of its validity to this day. There may in fact even be an inverse trend, at least concerning the historical disciplines: while digital humanities topics and methods have gradually started entering the scholarly mainstream, the multitude and potentially confusingly broad range of related approaches—from topic modelling to historical GIS and network analysis—have shifted some attention away from digital editing, or at least have not contributed to convincing those hitherto reluctant to engage with it.

Hybrid scholarly editions—that is, editions published digitally *and* in print—hold a particularly difficult position here. To those arguing the cause of the printed book, any effort invested in the digital aspect of an edition may seem a waste of resources, while the opposite may be the case for those advocating purely digital editing. When we argue for the benefits of a hybrid edition in the following, we do so in the full awareness that these benefits depend greatly on the nature of the published source, the target audience and, as we will show, the specific configuration of institutions involved. By putting hybrid editions on the agenda and devoting a separate working group (AG10) to them, the consortium of the Austrian Hochschulraumstrukturmittel project *Kompetenznetzwerk Digitale Edition* (KONDE) has clearly acknowledged the relevance of this field. Moreover, as a chapter of a comprehensive book on critical scholarly editing, this text represents one aspect of a broader context in which digital scholarly editing itself is thoroughly discussed.

We can therefore assume that the reader is familiar with the various types of text encoding as well as with infrastructural matters, with the types of scholarly questions to be answered with the help of digital editions, and with the various issues connected to the creation and maintenance of user interfaces. What we would like to demonstrate on the following pages is how a hybrid edition can be the best answer to specific needs in an appropriate case. Pointing out these needs and sketching out the possible answers is a goal in itself, since other scholars may find different (and perhaps better) solutions to the same problems. We will therefore attempt to frame these problems, explain our choices, and describe the process—but at the same time we would hope not to be understood as unduly advertising a product or preaching a

gospel. This is particularly important when it comes to the prominent participation of a publishing company: we are not seeking publicity, rather trying to work out a solution that assigns each actor a specific role.

One may rightly argue, however, that not all hybrid editions require a publishing company as a key partner in the design of the technical editing process. Some of the respective KONDE partners have found different solutions: the Werner Kofler hybrid edition (Straub and Dürr 2019) is built around an existing copyrighted print edition for the broader public, which it supplements by way of an extensive commentary. Similar approaches in terms of anthologies have been applied to the works of Peter Handke (Kastberger 2015) and Ernst Jandl (Hannesschläger 2015) to respect the copyright protection of initial prints valid for 70 years after an author’s death. Conversely, in the case of Robert Musil (Bosse, Boelderl, and Fanta 2016), the entire oeuvre along with its commentary can be part of the hybrid design.

Compared to the Pez letter edition (Wallnig, and Stockinger 2010; Wallnig, et al. 2015), which is the main topic of this chapter, the aforementioned projects represent a different way of thinking about hybrid editions. They *supplement* a printed edition mostly designed as a ready-to-use showcase publication for the interested public; they target distinct audiences (scholarly or other) and feature material of varied legal status. Neither is the case with the Pez edition, which, for reasons we will explain presently, deliberately aims to create a digital product that is equivalent to its printed counterpart. Both approaches are consistent with the relevant deliberations in Patrick Sahle’s magisterial treatise on the matter, which, after introducing practices of data outsourcing (*Auslagerung*) under the heading of hybrid editing, also acknowledges (though somewhat reluctantly) the legitimacy and validity of calls for printed books (Sahle 2013 2, 61–69).

What we will present on the following pages is a description of precisely the scenario drafted by Sahle: the print publication as a derivative of a digital editorial environment, created primarily as a result of social and academic habits and practices rather than of an intrinsic technical exigency. This context must be taken seriously, however, for what Sahle and others appear to overlook at times is the close connection between the lack of sympathy for the “real-life” analogue aspect among digital editors and the rightly lamented marginalization of digital editions within the scholarly community.

## 2 The Pez ecosystem

Although digital methods have become a key feature of current humanities projects and an undeniable requisite for editions, only few projects seriously attempt anything but individual solutions when it comes to standardizing data and envisioning reusable

modules. There is no doubt that various attempts and viable practices exist in this regard, but the issue has yet to be addressed from the broader perspective of all the instances constituting the entirety of the workflow of a digital scholarly edition: researcher, hosting institution and series editor, publisher and repository. As will be pointed out on the following pages, it was precisely the relations between these partners, between their perspectives and requirements, that ultimately led to the conclusion that the optimal solution for the edition of the Pez correspondence is that of a hybrid one.

To be sure, there is another fundamental instance in the workflow. For several years now, the Austrian Science Fund (FWF) has been working hard to keep up with the fast-evolving landscapes of digital development that often diverge significantly between specific disciplines. One of the results of this work is the “innovative publication formats” funding scheme; one of several, however, when it comes to planning a hybrid edition: projects also have to take the FWF policies regarding open access and research data management as well as the institution’s general research funding schemes into account.

The tangible output of the envisaged project (approved in 2021 as FWF PUD 23) will thus be twofold: its primary goal is the publication of a hybrid scholarly edition of the *full* correspondence of Bernhard and Hieronymus Pez, while the secondary goal is the development of a TEI-based model generic enough to cater to the needs of other editions to be published in the QIÖG (*Quelleneditionen des Instituts für Österreichische Geschichtsforschung*) series. As indicated in the previous paragraphs, the following text will use a pragmatic definition of *hybrid edition* as the publication of the same scholarly content—edited text, commentary and auxiliary material contextualizing and “unlocking” the edited source—in a print and a digital publication.

Although it is generally evident that the modes of access and thus the nature of the two publication formats differ fundamentally,<sup>2</sup> the chosen hybrid approach implies that—on a general level—the versions in both media contain the same scholarly content without favouring one format over the other. While this is a major contrast to a model of hybrid editions where one medium is complementary to the other (e.g. by providing raw transcriptions or facsimile images digitally while the book remains the sole medium for the edited, canonized text),<sup>3</sup> we argue for a clear conceptual and functional distinction between the two publication forms which, in our understanding, qualifies

---

<sup>2</sup> We find this distinction in many different gradations: Sahle (2013, 66) differentiates between “lesen” and “benutzen” (i.e. “reading [a fixated, edited text] vs. using [a digital tool to make both sources and editorial decisions accessible]”); Pierazzo (2014, 151) takes up the difference between “reading” and “seeking information” described by Ciula and Lopez (2009), who see the usage of the digital and analogue publication as a continuum (“a comprehensive process”).

<sup>3</sup> Since there is no clear-cut categorization of such intermedial relations in hybrid editions, we take the spectrum of “use cases” collected in Pierazzo 2014, 150–3. Sahle similarly lists examples, but focuses on the variety of publication media rather than on their degree of overlap or complementarity (2013).

our undertaking as a hybrid one.<sup>4</sup> The digital format fosters an “explorative access” to the edition by offering native digital features like a full text search, hypertextualization of its constituents, or orthogonal views on the letters by relating them through common index terms—means of interaction that are inconceivable in printed form. However, these digital enhancements do not represent a disruption but rather a continuum with what the audience of the QIÖG series expects in terms of editorial practice.

Our editorial choice is thus not merely predetermined by the existence of already printed volumes of the Pez correspondence edition or the existence of a print series. It more broadly reflects the needs of a scholarly community accustomed to working with multiple editorial manifestations of the same text.<sup>5</sup> The need to combine a book with an online resource results from everyday practice and can be described as follows:

- a) as the need for a synoptic view of resources that can hardly be achieved digitally—one can lay out twenty open books in a room, but not reasonably maintain an overview of twenty open browser windows;
- b) the need to use a physical book copy as a container for comments, annotations and corrections collected over time—annotating and re-annotating the same *digital* document has simply not proven to be practical for the majority of scholars;
- c) the need to collocate a book in a broader physical context representing a specific topic matter—no tagging system can substitute the intuitive topicality of a book shelf; and
- d) the wish to safeguard sustainability—while this should not be treated as a religious matter, it is equally evident that there are sufficient examples for digital non-sustainability to justify such concerns.

In conclusion, Sahle (cf. footnote 2) is certainly right to point to a juxtaposition of “reading” (continuous texts conceived as such by their authors, in this case of historical sources) and “using” (bits and pieces of screen-size information); in precisely this

<sup>4</sup> “In die Gruppe der Hybrideditionen fallen solche digitalen Editionsformen, die in einer solchen Weise in verschiedenen Medien präsentiert werden, dass es zu einer inhaltlichen und konzeptionellen Rollenverteilung kommt. [...] Entscheidend ist, dass bestimmte Medien die Inhalte und die konzeptionelle Haltung zur Edition beeinflussen.” (Sahle 2013, 68).

<sup>5</sup> This is also the case with the Pez letters themselves: digital images (scans) of large parts of the original correspondence kept at the abbey of Melk are published in an adjacent but separate project (Rabl 2013) aiming at a digital representation of the Pez brothers’ papers. In this way, it becomes possible to interlink both datasets on an integration layer that draws on common models like RiC (Llanes-Padrón and Pastor-Sánchez 2017).

sense, reading and interpreting texts in their *entirety* has been and still is considered by many as part of the craft and identity of historians. The preferred medium for this is the book.

If one takes all these considerations and practices seriously, the hybrid approach can become a key factor in the current phase of scholarly transition towards a digital paradigm. It can help to reassure the community by way of a “regulatory authority” (Deegan and Sutherland 2009, 72) that counterbalances an often perceived “instability [of digital editions].” Furthermore, it can help to recruit experienced editors reluctant to leave their field of expertise and get them to invest their time in a medium that has yet to prove its sustainability, as well as young editors whose careers will depend on the acceptance of their work by the broader scholarly community.

Besides this user group, however, which we imagine will use the book and the online edition in parallel, we envision that the data itself will be of interest in many other as yet unconceived digital contexts—up to the point where it merges into what Gabler (2017, §9) calls “relational webs of discourse, energized through the dynamics of the digital medium into genuine knowledge sites.” To this end, the data, which is also used to produce the digital edition, is being enriched (e.g. by using common identifier systems for named entities) in ways that allow for its easy consumption. Given the epistolary nature of the topic, this can mean services like *correspSearch* but also its potential deeper integration into the LOD paradigm (Ciotti and Tomasi 2016), thereby making it a distinct digital artifact serving a different audience. This separation of user groups with regard to both interest and digital literacy is reflected in the institutional setup of the project: The data consisting of the correspondence of Bernhard and Hieronymus Pez will be archived in ARCHE,<sup>6</sup> the CTS-certified repository for humanities research data run by ACDH-CH (Austrian Centre for Digital Humanities and Cultural Heritage). The user interface will be developed by the Böhlau publishing house and will take into account the needs of further publications in the QIÖG series.

We thus envisage creating a distributed architecture on a sustainable technical foundation that enables interaction with a rich, standards-based dataset in an intuitive manner on the one hand and producing an accompanying print edition that maintains the high standards of the existing series on the other. These goals proactively address the fact that complementary ways of working with editions have emerged as a consequence of the digital turn: besides close reading and the prevalent usage of print publications, the digital humanities have recently brought about programmatic ways of accessing an edition’s content. To safeguard the sustainability of the results, we will follow a modular, service-oriented approach with well-defined interfaces between the systems maintained by differentiated institutions, which will allow us to

---

<sup>6</sup> <https://arche.acdh.oeaw.ac.at/browser>.

replace individual modules in the future without having to reimplement the system as a whole.

### 3 The transforming edition of the Pez correspondence

The following diagram (Figure 1) shows the draft outline of such an environment serving as the blueprint for the development in the project.

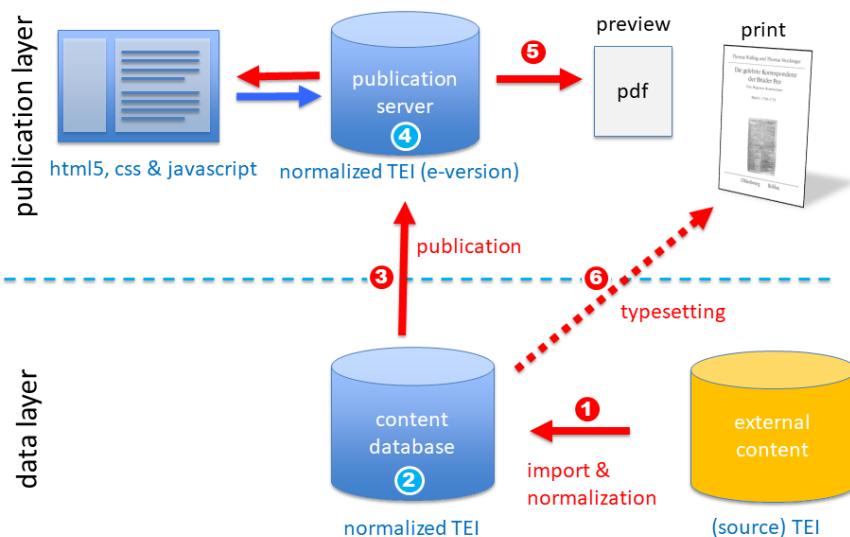


Figure 1: Conceptual diagram.

#### 3.1 Content: the Pez correspondence and the QIÖG series

The Pez correspondence consists of roughly 1150 letters preserved primarily at Melk abbey as well as in various locations scattered all across Europe. It has been the object of editorial efforts since 2004, which to date have led to the *print* publication of roughly half of the letters within the QIÖG series.<sup>7</sup> The ongoing FWF project P28016

<sup>7</sup> Open access: doi:10.7767/9783205794134.

(2016–2020),<sup>8</sup> however, is engaged with preparing the *remaining* letters for *digital publication* while reworking the extant publication into digital formats in parallel.

This process is the background against which the endeavour described in this paper should be viewed, for it reflects the need to create a model fit for digital publishing that simultaneously allows us to preserve previous “non-digital” editorial choices as well as to replicate the print layout adopted for the first two volumes of the Pez correspondence so as to maintain the overall print appearance of the series.

QIÖG is a series edited by the Institute of Austrian Historical Research and published by Böhlau / Vandenhoeck & Ruprecht (since 2021: Brill), with the first volume issued in 2008. The series mostly features stand-alone sources from pre-modern Central Europe whose specific nature is hardly appropriate for the larger series of source editions like the MGH. Each QIÖG volume includes a thorough introduction to the source itself and its genre and individual characteristics, as well as to the specific editorial rules that need to be adopted for it. Though oriented towards best-practice models, these solutions often break new ground and draw on an informed dialogue between the series editors and the individual scholars working with the respective source. This sometimes experimental approach also includes deliberations on digital edition and publication models, and the QIÖG series has therefore added two requirements to the present use case portfolio:

1. it insists, for institutional reasons, on a model allowing for print and online publication; and
2. it wants such a model to be reusable, ideally for all following QIÖG volumes to come.

In order to better understand this second requirement, some of the normalization described below in Chapter 3.2 has also been outlined for other QIÖG publications (namely Volumes 3 and 8).

The Pez correspondence itself is a source relevant to the early modern intellectual history of Central Europe, as it aptly displays the specific Catholic variant of religious antiquarianism and historical criticism. The letters are relevant in regard to the historical—medieval—source material they talk about, the international eighteenth-century networks of scholars traceable through their correspondence, the specific Latinity of Catholic late humanist erudition, and the specific framings of emergent discourses of “Austrian” and “German” history.

These specific features led to certain editorial decisions at the outset of the Pez project that determined the “data model” long before any discussion on digital matters; these decisions are described in detail at the beginning of the two printed volumes.

---

<sup>8</sup> [www.pezworkshop.org](http://www.pezworkshop.org).

For the purposes of this paper, it should be remembered that

- a) all preserved and inferred letters are documented by way of basic metadata;
- b) each preserved letter is included with its full text, a German summary and an extensive commentary discussing its contents in detail;
- c) some bio-bibliographical information is outsourced to the main index referring to the letter number, not the page; and
- d) there are additional appendices and elements extracted from the letters, among which is a list of mentioned third-party letters and of material objects sent along with individual letters.

Footnotes are used for the critical apparatus, while the commentary is placed after the respective text passage it refers to. All three parts—letter text, summary and commentary—are divided into associated content segments so as to facilitate orientation within longer letters. This division serves as the edition’s document-internal reference system and provides a hook between the three parts relating to each letter.

In the context of Pez volumes 1 and 2, this system was implemented in MS *Word* files by way of standard formats for the individual elements (metadata, summary, text, commentary, literature etc.). These files form the basis for the print layout and PDF files as well as for the digital transformation described in the following chapters.

### **3.2 Technical execution: implemented and envisaged workflows**

#### **Data creation and editorial process**

Given the fact that two published volumes already exist, the adoption of a hybrid approach implies establishing two parallel workflows for integration into a single data format:

1. semi-automatic retro-encoding of the existing volumes into TEI-XML, and
2. establishment of an XML-first editing workflow for the letters to be edited in the ongoing project.

Both workflows need to implement a data model expressive enough to be useful as a dataset in its own right within the DH community. The focus in this regard lies on modelling a TEI customization with rich descriptive metadata and sound semantic and structural markup that makes implicit editorial conventions of the printed volumes explicit by means of TEI-encoding. On the other hand, the model needs to be flexible

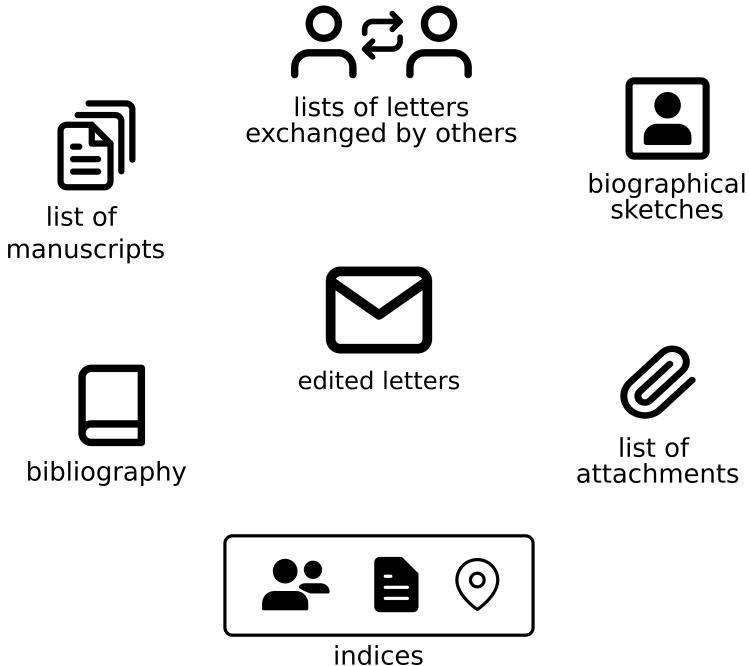


Figure 2: Data architecture of the Pez edition.

enough to support conventions established in the previous published volumes without requiring unreasonable effort. Lastly, the model has to encompass the full resource architecture, which includes material necessary to contextualize and “unlock” the letters as historical sources (see Figure 2).

Naturally, the edited letters are at the centre of the architecture, with each letter represented as a separate TEI document. The document encoding is modelled on the structure of the printed edition and therefore contains a `<front>` matter with the summary, the edited letter in the `<body>`, and a `<back>` matter including the commentary (i.e. a sequence of `<note>` elements attached to segments in the letter’s text via a `@target` attribute).

The abovementioned *content segments* are tagged as corresponding `<seg>` elements both in the summary and the letter text, thus making them easily addressable TEI structures. It is important to mention that the editors already refrained from using page numbers as references in the existing print volumes (both in the index and the commentary), instead opting for the letter numbers along with a segment number

```

<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
  </teiHeader>
  <text>
    <front>
      <div type="regest">
        <p>
          <seg type="context" n="1"> ... </seg>
        </p>
      </div>
      <div type="editorialNote">
        <!-- ... -->
      </div>
    </front>
    <body>
      <div type="editionText">
        <seg type="context" n="1"> ...
          <seg type="Lemma" xml:id="l_d4618e6102">Adiutores igitur in
            partem operis vocati</seg>
        </seg>
      </div>
    </body>
    <back>
      <div type="commentary">
        <note target="#l_d4618e6102">
          <ref type="context">1</ref>
          <label>Adiutores <gap/> vocati</label>
          <p>... vgl. <ref type="letter">32 <ref type="context">3</ref></ref>. </p>
        </note>
      </div>
    </back>
  </text>
</TEI>

```

Figure 3: Pez TEI model.

where appropriate. This decision allows existing references to be effortlessly migrated into digital links and ensures the ability to cite across systems and media for both the digital- and analogue-born content.

In our model (Figure 3), most parts of the so-called *editorische Notiz* (roughly “editorial notes”) are integrated in the `<teiHeader>`: Literature (including prior editions of the letter in question) is listed as `<relatedItem>` elements of various types, while contextual information is embedded in `<correspDesc>`.

In some cases, several writing stages of a letter are preserved. These variants are recorded by means of double-end-point-attached `<app>` elements placed in a separate division in the `<back>`. The majority of the existing apparatus entries found in the footnotes of the print edition can be mapped to standard TEI markup (mostly adopting the mechanisms provided by the *transcr* module) without loss of information, however in some cases such formalization is unfeasible or unreasonable because flat text in a page-oriented medium requires different modes of expression. The ability to reference null-positions may serve as a simple example: while it is easy from a data-oriented

proxima Deo favente exitura. His<sup>u</sup>

<sup>u</sup> Davor durchgestrichen Tum C.

```
<!-- inside <body> -->
<anchor xml:id="app021"/>His<anchor xml:id="app021e"/> curis ...
<!-- inside <back><div type="apparatus"> -->
<app from="#app021" to="#app021e">
  <rdg wit="#C"><del>Tum</del> His</rdg>
  <witDetail wit="#C">Davor durchgestrichen <mentioned>Tum</mentioned></witDetail>
</app>
```

Figure 4: Critical apparatus (Example 1).

```
<!-- inside <body> -->
... <anchor type="app-start" xml:id="app019"/>quam suscepi, librorum
sancti Irenaei episcopi Lugdunensis et martyris Adversus
haereses<anchor xml:id="app019e" type="app-end"/>, ...
<!-- inside <back><div type="apparatus"> -->
<app from="#app019" to="#app019e">
  <lem wit="#B">quam suscepi <gap/> Adversus haereses</lem>
  <rdg wit="#C">a me concinnata operum sancti Irenaei episcopi
Lugdunensis ac martyris</rdg>
</app>
```

Figure 5: Critical apparatus (Example 2).

view to document that a specific word exists in a witness other than the base text, translating this piece of information into a comprehensible footnote generally requires the addition of a location indicator and thus potential custom wordings (see Figure 4).

Similarly, the lemmata within a printed apparatus (i.e. the words of the base text repeated at the beginning of an apparatus entry to assist the reader in locating the variation) can be easily constructed programmatically from the TEI markup in simple cases, but in cases of ambiguity, editorial intervention is necessary to select a reference that is also meaningful in a printed form. Here too, the model aims to fully consider both the data-centric and the print-centric approach, which comes at the slight cost of the duplication of information—or more precisely, of repeated expression of the same information (see Figure 5).

Besides the letters, the auxiliary material has likewise been converted into TEI-XML from the print edition’s source documents (.docx files) by means of custom XSLT scripts. Further curation was necessary to enrich or correct this intermediate step:

- a) The (hierarchically structured) indexes from Volumes 1 and 2 were merged into a single index and their entries categorized as referring to persons, organizations /places, manuscripts or works in a simple web interface. Wherever possible, reference resource IDs were manually added to the index entries.

- b) The bibliographies of both volumes were extracted and imported into a shared Zotero library for further refinement.
- c) The letters were semantically marked up with pointers to the merged index and the bibliography, and first steps towards encoding the critical apparatus were taken.

Since the model has evolved into a largely stable form and has been implemented as a TEI-ODD customization, it is ready to be employed as the foundation of the XML-first workflow that is being set up to encode the remaining letters in the ongoing FWF project. At its end, the completed dataset will be archived and published under a Creative Commons license in the ACDH-CH’s repository ARCHE. The digital objects in ARCHE are referenceable via *handle* PIDs and made available through so-called “dissemination services”, i.e. specialized web services that render the archived data. Building upon the data provided by such a service, the Böhlau publishing house plans to develop both an intuitive user interface and a workflow to create the print versions of the remaining volumes using the same conventions and typographic features as the existing parts.

Realization of this step requires a substantial investment in technical and social infrastructure, however—an investment that a commercial provider can only make if a critical number of similar publications can be produced with minimal additional overhead, thus leveraging scaling effects. The key thus lies in achieving the level of data homogeneity required for the software components of the publishing ecosystem to be reusable at little cost. Since the data model of any digital edition aims to reproduce its source documents as faithfully as possible (not necessarily in a visual dimension, but at least semantically and/or structurally), the data for volumes in a series like QIÖG will inevitably vary to a great extent. Based on our discussions regarding the Pez correspondence edition, we have drafted a strategy that should help to deal with this intrinsic diversity within the economic boundaries of a publishing house without forcing editors and DH scholars to dispense with the semantic richness required by their particular objects of interest.

### **Data normalization / mapping of input TEI to publication TEI**

In order to make data processing as modular and reusable as possible, we propose a multi-level normalization process that distinguishes between three levels of data preparation:

1. The starting point is **source data** resulting from an edition project. This source or import data must be in TEI markup and represents the digital primary source.

In case of the Pez edition, we assume the primary data to be delivered by the digital long-time preservation repository ARCHE.

2. The first and crucial step of **normalization** prepares the TEI data for the actual publication process. The normalized data remains compatible with TEI, generating document structures or serializing document content in the form required by the publication concept. This corresponds to step (1) in the conceptual sketch in Figure 1.
3. The third step completes the initially media-neutral processed data with regard to an **electronic version**. This corresponds to step (3) in the conceptual sketch in Figure 1.

In the following, the two normalization steps will be presented in more detail (see also Table 1).

Ad 1: Examination of the TEI data is a requirement necessary for all XML processes. In this project, however, it refers to more than mere technical parsing according to the TEI scheme. It is also about examining all detailed structures that are important for publication as well as about detecting semantic inconsistencies. Examples of such checks are reference-relevant structures (links), important metadata, or document hierarchies. All errors must be removed before the next step.

Ad 2: Normalization and enrichment of the data are at the heart of the data preparation process. The goal of normalization is to reduce the variants in TEI tagging. Defined semantic structures (such as names of places or persons) are normalized so that they can later be used uniformly in the publication layer. The concept of data enrichment follows the observation that due to the rigid semanticization, TEI often does not correspond to the desired presentation and sequence in the publication layer. An example of this is metadata that is found in the header in TEI but is usually needed in the text at a defined point (see Figure 6). This necessitates additional effort to prepare the data for presentation. The theory here is that the two media basically have similar requirements and this step can therefore be completed for both of them together. As a result, normalization and enrichment means that significant problems of the publication layer have already been resolved, thereby simplifying the actual print and online publication process.

Ad 3: Accordingly, the last step of data preparation is easily explained. The normalized data can be fed directly into both media forms since all essential intrinsic issues of document delimitation and serialization of the text elements have been addressed. What remains is to implement the technical reference-target-structure and transform it into the desired publication format. For print media, this means an XML typesetting process based on established automatic typesetting techniques (XSL-FO, La-TeX etc.).

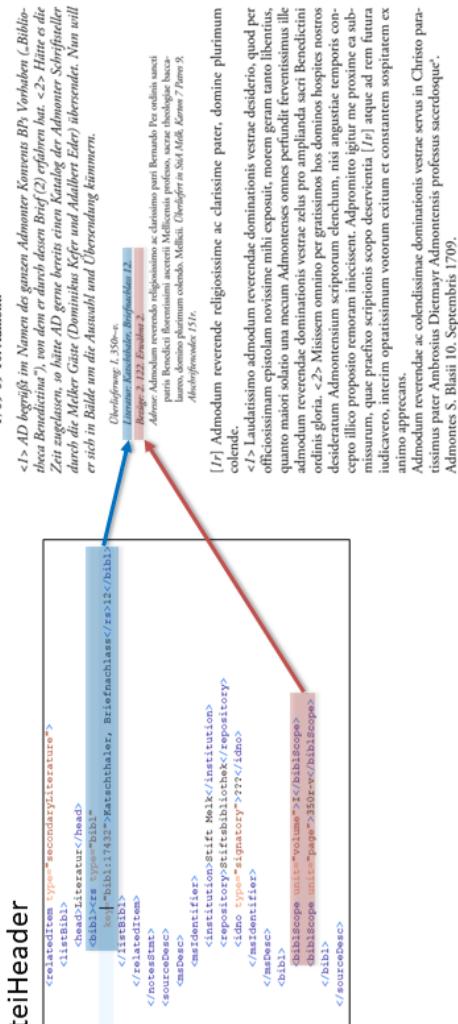


Figure 6: Transformation of metadata into text.

	<b>Source TEI</b>	<b>Normalized TEI = step (1)</b>	<b>Normalized TEI (e-version) = step (3)</b>
<b>A. General</b>	<ul style="list-style-type: none"> <li>.Representation of primary source</li> <li>.Pure semantic markup</li> </ul>	<ul style="list-style-type: none"> <li>.Flattening of the semantic markup</li> <li>.Addition of the edition-specific texts</li> </ul>	Technical implementation for e-media
<b>B. Media-specific</b>	Text-critical apparatus	Media-specific implementation	Technical implementation for e-media
<b>C. Semantic normalization</b>	<p>Example</p> <p>Deletion of text</p> <p>Semantic ambiguity in TEI markup across projects</p>	<p>Generation of a (foot)note with deletion remark</p> <p>Normalized TEI semantics</p>	<p>Implementation of the (foot)note as an ID-IDREF link with its content</p>
<b>D. Serialization</b>	<p>Example</p> <p>Different TEI tagging variants for persons</p> <p>Semantic tagging scatters information across different places in the markup</p>	<p>Unified TEI tagging of persons</p> <p>Serialization of all text parts for the purpose of the critical edition</p>	
<b>E. Reference and linking mechanism</b>	<p>Example</p> <p>Metadata with state of research, literature and text references is separated from the editorial text</p>	<p>Metadata is inserted as generated text in the desired text flow</p> <p>Selection and marking of the semantics for which reference and target structures are created</p>	<p>Technical implementation of reference and target structures</p>
	<p>Example</p> <p>Context-based position and counting of lemmata</p>	<p>Concrete counter of lemmata for representation issues</p>	<p>Lemmata with target ID corresponding to concrete counter</p>

Table 1: Overview of normalization tasks.

For online publishing, the normalized TEI can easily be transformed into HTML5 plus CSS, or other electronic formats.

### 3.3 Publication layer

We focus on the publication layer for the digital medium as the concept of a printed text edition is well-known and established. The publication layer represents the interface between content and user. It therefore has to meet all of the needs of the user, whom we assume to belong to an academic community.

Even for a resilient user group like humanities scholars, a digital publication should be not simply functional but also user-friendly. In the following, we present mockups/sketches of a possible user interface that also include practical experience from the publisher’s perspective.

#### Text representation and readability

A central component of text-based content is the readability of the text. In addition to well-thought-out typographic implementation, it is also important to consider the specific needs of the user.

Fundamentally, the corpus of the edition as well as the printed medium are presented in a similar fashion. Since the on-screen readability of extensive text documents such as the Pez letters suffers quickly, especially because the early modern syntax of various languages is often characterized by long sentences, not all parts of the text are displayed in the initial view. Instead, the document is displayed in a condensed representation providing the user with a quick overview (see Figure 7). Long text passages not needed for this overview are hidden, but the main text structures with their headings are displayed. The hidden text sections can then be displayed with a simple click if desired.

For a more in-depth comparison between e.g. summary and text, a “synoptic representation” and thus parallel access to the different contents is envisioned. Contextual information presented in footnotes in printed editions should ideally be visible in the respective context. Depending on the semantics of such footnotes, they can either be placed directly behind the respective text and displayed, for example, as a popup or in a sidebar next to the edited text. In the case of the Pez edition, footnotes are mostly textual variants that could sensibly be implemented as pop-ups following an appropriately marked passage of text.

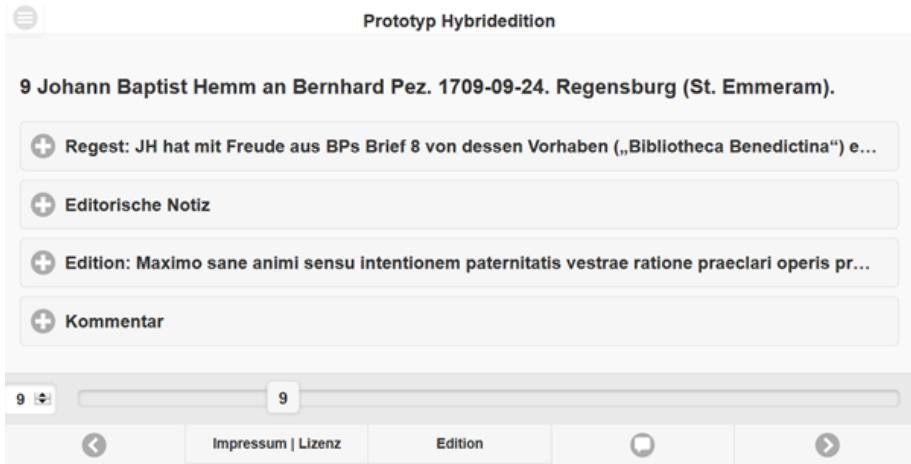


Figure 7: Condensed text representation of a document.

## Navigation

In order to guide the user by means of specific information featured in the document, navigation for the usual semantic structures such as persons, chronology, places, works etc. should be provided. These navigation tools fulfil the function of a classic index in the printed edition. A special navigation system makes the individual documents of the edition accessible and shows where the user is located. Hierarchical navigation trees or breadcrumb navigation can be used as display options. In the case of the Pez edition, a simple list of letters is sufficient (see Figure 8).

## Other user features and information

Finally, a number of basic features that are nowadays very common in content-based user interfaces should be mentioned—for the very reason that they are practically taken for granted:

- Navigation between documents: it must be possible to navigate between documents directly, once in the form of the usual “back” and “forward” buttons, but also by reference to the preceding or subsequent document of the correspondence.
- Print functionality: although a print function seemingly represents a media rupture, experience has shown that it has its legitimacy in digital use as well. It is crucial for this printing function to be simple and not replace the printed edition.

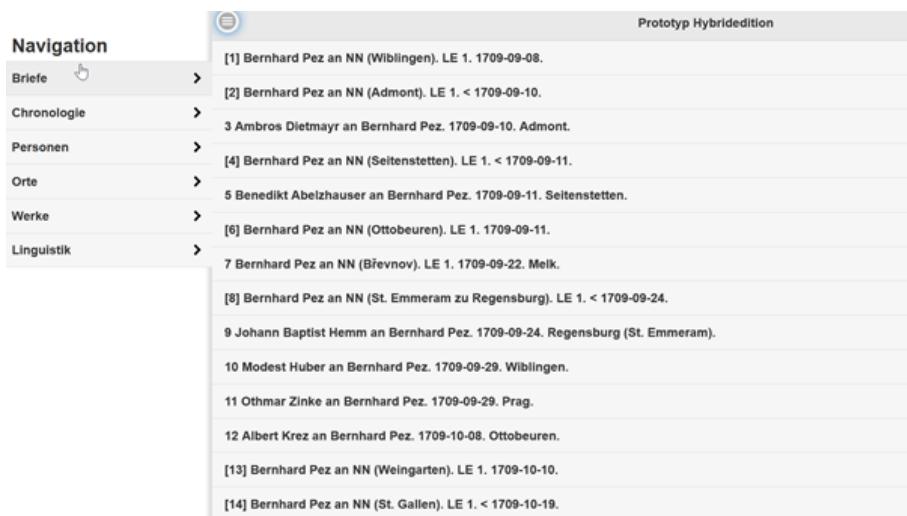


Figure 8: Navigation mockup.

- Simplifying citation: a citation button provides the user with a complete citation proposal for the current document that can be used immediately via the clipboard.
- Full text search: a simple and intelligent full text search should be able to find concatenated search terms as well as recognizing and finding terms with diverging spellings.
- User-specific views: while the order and selection of the letters/documents in the printed edition is fixed and unchangeable, users of the digital edition may use their own document selection or document view. For example, one could restrict the display of letters/documents using specific search criteria and thus create and keep an individual selection of documents. It would also be conceivable to generate an individual view by displaying or hiding certain information that corresponds to the researcher’s current question.

Finally, certain information making the context of the document accessible should be displayed. This includes:

- keywords or register information;
- the necessary legal and copyright information of the publication, which includes the imprint and a clear indication of the assigned rights of use;

- ideally, the document should also feature a technically unique and unchangeable identifier. Given its prevalence in the publishing industry and the corresponding familiarity to readers, the use of a DOI seems an obvious choice here, though it is simultaneously debatable for reasons of cost and effort.<sup>9</sup>

### **Export instead of integrated analysis**

The discussion so far has shown that complex search functions are not truly necessary since academic users generally have very specific questions. Instead of such functions, we therefore favor a variety of export functions:

- Export of defined standard lists displaying common search structures (people, places, works, etc.) and their contexts. These indexes should be easily accessible even for non-specialist users and contain links to the respective documents containing occurrences of the index entry.
- In order to allow custom queries as well as for reasons of data transparency, the underlying TEI data should be exportable.<sup>10</sup> Other export options could be provided as well, for example a plain-text version to be used with linguistic tools like *Voyant* (Sinclair et al. 2016) or *AntConc* (Anthony 2014).<sup>11</sup>

## **4 Conclusion and strategic outlook: on institutional and technical sustainability**

The complexity of the outlined design is a response to some of the concerns frequently associated with digital-first publications: From the point of view of the editors, relying on the expertise and reputation of a well-established publishing house can be an important factor. Even though *digital* (and *hybrid*) has hitherto been limited to page-oriented digital media like e-books or screen-optimized PDFs in many publishing houses, the benefit of optimizing workflows and having a single contact for producing both a book and a web-based edition can outweigh the costs of technical expertise yet to be developed.

<sup>9</sup> In case of the Pez edition discussed here, the source documents are imported into the publication system from the ARCHE repository, which uses the handle.net system to assign PIDs for the archived resources. Although we can imagine a scenario where PIDs for the frontend view of the documents in the publication system are created during ingest and persisted in the <teiHeader> of the archived document, there may be cases where the publisher's system is the sole source of data, so that it must be capable of citable identifiers itself.

<sup>10</sup> Although the fully encoded version of the Pez edition's data can be retrieved directly from ARCHE, we cannot assume this for any future publication served by the system.

<sup>11</sup> This would be beneficial since workflows like the ones described in Fröstl 2019 are outside of the scope of non-expert users without technical support.

Separating the roles in such a process may also have the advantageous side effect of a higher level of sustainability not only in regard to the source data (which is provided in a well-established standard format and maintained by a certified repository hosted by an academic institution with public support), but also in regard to the much more volatile user-facing front end. Dissevering these concerns requires modularization and a clear-cut definition of procedural and technical interfaces, which can become key factors in the future, should parts of the infrastructure grow obsolete—but only if both the data and the implementation of the publishing platform are published under a license permitting reuse and adaptation by third parties. Bridging the differences between commercial interests, public funding policies, and scholars in both DH and other disciplines will be a matter of discussion for the years to come along with the development of licensing schemes that respect and foster every actor’s effort in a complex and dynamic field subject to continuous transformation.

## Bibliography

- Anthony, Laurence. 2014. *AntConc: Version 3.4.1w* [computer software]. Tokyo: Waseda University. Accessed June 15, 2018. <http://www.laurenceanthony.net>.
- Bosse, Anke, Artur Boelderl, and Walter Fanta, eds. 2016. *Musil Online*. Klagenfurt: Robert Musil Intitut / Kärntner Landesarchiv. Accessed March 7, 2021. <http://musilonline.at>.
- Ciotti, Fabio, and Francesca Tomasi. 2016–2017. “Formal Ontologies, Linked Data, and TEI Semantics.” *Journal of the Text Encoding Initiative* 9, online since 24 September 2016. doi:10.4000/jtei.1480.
- Driscoll, Matthew James, and Elena Pierazzo, eds. 2016. *Digital Scholarly Editing. Theories and Practices*. Cambridge, UK: Open Book Publishers.
- Fröstl, Michael. 2019. “Digital-linguistische Annotation neulateinischer Texte. Theorie und Praxis für Editions- und Geschichtswissenschaft.” MA Thesis, University of Vienna (Institute of Austrian Historical Research). doi:10.25365/thesis.58660.
- Gabler, Walter. 2018. “Theorizing the Digital Scholarly Edition.” In: *Text Genetics in Literary Modernism and Other Essays* [Online]. Cambridge: Open Book Publishers, 2018. Accessed March 7, 2021. <http://books.openedition.org/obp/5453>. [originally published in 2010 *Literature Compass* 7 (2)].
- Hannesschläger, Vanessa, ed. 2015. *Ernst Jandl Online*. Wien: Österreichische Nationalbibliothek / Ludwig Boltzmann Institut für Geschichte und Theorie der Biographie. Accessed March 7, 2021. <https://jndl.onb.ac.at>.
- Kastberger, Klaus (with the collaboration of Katharina Pektor and Christoph Kepplinger-Prinz), eds. 2015. *Handkeonline*. Accessed March 7, 2021. <https://handkeonline.onb.ac.at>.
- Llanes-Padrón, Dunia, and Juan-Antonio Pastor-Sánchez. 2017. “Records in Contexts: the Road of Archives to Semantic Interoperability.” *Program* 51 (4): 387–405. doi:10.1108/PROG-03-2017-0021.

- Pierazzo, Elena. 2014. *Digital Scholarly Editing: Theories, Models and Methods*. Grenoble. Accessed June 21, 2020. <http://hal.univ-grenoble-alpes.fr/hal-01182162>.
- Rabl, Irene. 2013. "Der digitalisierte Nachlass der Brüder Bernhard und Hieronymus Pez. Ein Projektbericht." *MIÖG* 121: 437–44.
- Sahle, Patrick. 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik*. Norderstedt: Books on Demand. Accessed March 7, 2021. <https://kups.ub.uni-koeln.de/5352>.
- Stockinger, Thomas, et al. 2015. *Die gelehrt Korrespondenz der Brüder Pez. Text, Regesten, Kommentare, 2: 1716–1718*. Quelleneditionen des Instituts für Österreichische Geschichtsforschung 2/2. Wien: Böhlau. Accessed March 7, 2021. <https://www.oapen.org/search?identifier=576951>; <http://www.oapen.org/search?identifier=576952>.
- Straub, Wolfgang, and Claudia Dürr, eds. 2019. *Werner Kofler. Kommentar zur Werkausgabe*. Accessed March 3, 2021. <http://wernerkoehler.at>.
- TEI Consortium. 2019. "TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.5.0." Accessed April 23, 2019. <https://tei-c.org/Vault/P5/3.5.0/doc/tei-p5-doc/en/html>.
- Van den Heuvel, Charles, et al. 2019. "Transcribing and Editing text." In *Reassembling the Republic of Letters in the Digital Age. Standards, Systems, Scholarship*, ed. by Howard Hotson, and Thomas Wallnig, 237–64. Göttingen: GUP. doi:10.17875/gup2019-1146.
- Wallnig, Thomas, and Thomas Stockinger, eds. 2010. *Die gelehrt Korrespondenz der Brüder Pez. Text, Regesten, Kommentare, 1: 1709–1715*. Quelleneditionen des Instituts für Österreichische Geschichtsforschung 2/1. Wien: Böhlau. Accessed March 7, 2021. <http://www.oapen.org/search?identifier=445402>.

# Digital Scholarly Editions of Alchemical Texts as Tools for Interpretation

Sarah Lang

## Abstract

Digital Scholarly Editions commonly provide annotations and visualization tools for Named Entities, dates or basic quantitative text analysis using methods based on bags of words and word counting. This paper argues that such methods are not enough to provide meaningful insight into texts from the history of science and alchemy especially, using two case studies from the alchemical texts by iatrochymist Michael Maier (1568–1622). A custom-made method for the interpretation of alchemical terms is presented which relies on the polyvalent semantic annotation of relevant yet ambiguous alchemical Decknamen and the analysis of their semantic contexts using a digital knowledge organization system.

## Zusammenfassung

Digitale Editionen bieten üblicherweise Annotationen und Tools zur Visualisierung von Named Entities, Daten oder simplen Auswertungen quantitativer Textanalyse an. In diesem Beitrag wird argumentiert, dass diese weit verbreiteten Tools für eine sinnvolle Analyse im Bereich der Wissenschafts- und Alchemiegeschichte nicht ausreichen. Anhand von zwei Fallstudien aus dem Werk des Iatrochymikers Michael Maier (1568–1622) wird eine Methode zur polysemantischen Annotation uneindeutiger alchemistischer Decknamen vorgestellt und gezeigt, wie damit deren semantische Kontexte mithilfe eines Wissensorganisationssystems analysiert werden können.

The iatrochymist<sup>1</sup> Michael Maier (1568–1622)<sup>2</sup> has attracted considerable attention over the last few years. Yet Maier’s works are mentioned very often, but studied very little, some scholars have complained.<sup>3</sup> This has caused research on Maier to be highly dependent on secondary literature which is, in part, quite outdated. Criticized by Erik

<sup>1</sup> While the term alchemist is still used for medieval times, alchemical practitioners of the early modern period are nowadays referred to as chymists. A iatrochymist is a chymist with a medical background who creates alchemical medicines as opposed to engaging in chrysopoeia, i.e. gold-making. On the subject of alchemy see Principe 2013.

<sup>2</sup> On Maier see Tilton 2003; Leibenguth 2002; Tilton 2005.

<sup>3</sup> Wels especially criticizes this even regarding Maier’s most famous and most well researched work, the *Atalanta Fugiens* (1617) (Wels 2010, 149).

Leibenguth in 2002,<sup>4</sup> Volkhard Wels finds the situation had hardly improved by 2010 despite the renewed interest in Maier and his work (Wels 2010, 149). Although the popular emblem book *Atalanta Fugiens* (1617/8)<sup>5</sup> has recently been the subject of a Digital Scholarly Edition, the rest of Maier's corpus is not well known and is even less studied.<sup>6</sup> Much research on Maier's vita has been done in recent years<sup>7</sup> but apart from *Atalanta Fugiens* only few of his texts have been studied in detail.<sup>8</sup> Text mining Maier's corpus could help in solving this problem.

The language of alchemy is known for its obscurity and riddles. These are, however, not chaotic or illogical as it has been claimed in the past, prominently so by Umberto Eco (2016). Rather they represent stylistic devices typical of alchemical communication, the most important of which are the so-called *Decknamen*.<sup>9</sup> Furthermore, the techniques of *paraphesis* and 'dispersion of knowledge' are widely used in alchemical texts and will be addressed in this article.<sup>10</sup> All those verbal deceits ultimately serve

<sup>4</sup> Leibenguth stresses: "Charakteristisch für die Maierforschung sind mit wenigen [...] Ausnahmen eine Abhängigkeit und teils sinnentstellende Rezeption von Sekundärquellen sowie die weitgehende Unkenntnis der lateinischen Primärtexte." Cf. Leibenguth 2002, 8–9.

<sup>5</sup> On *Atalanta Fugiens*, see for example: Maier 1617a; Jong 1969; Wels 2010; Nummedal and Bilak 2020.

<sup>6</sup> The digital representation of the corpus of his printed works yields 3500 PDF pages.

<sup>7</sup> On Maier's vita cf.: Leibenguth 2002; Tilton 2003; Purš and Hausenblasová 2016; Tilton 2020.

<sup>8</sup> Detailed studies exist regarding Maier's doctoral theses (*Theses de Epilepsia*, 1596), *Examen* (1617), *Atalanta Fugiens* (1617) und *Cantilenae Intellectuales* (1622). Maier 1596; Maier 1617a; Maier 1617b; Maier 1622; Beck 1991; Stiehle 1991; Leibenguth 2002; Nummedal and Bilak 2020.

<sup>9</sup> In the historiography of alchemy, the term *Decknamen* has come to refer to cover names used by the alchemists. In the case of Michael Maier, these often come from classical mythology. Case study 2 will serve to better illustrate their full extent than a theoretical definition. However, it is important to note that the term as it has become established in the research discussion is meant as a neutral one in English, without the German connotation of covering something up, potentially with a bad intent. Principe elaborates: "*Decknamen* function as a kind of code, [...] serving a dual purpose: they maintain secrecy, but they also allow for discreet communication among those having the knowledge or intelligence to decipher the system. They simultaneously conceal and reveal. Consequently, *Decknamen* have to be *logical*, not arbitrary, so that they can be deciphered. If *Decknamen* could not be deciphered by readers, then total secrecy would be the result; and if the intent were to conceal information entirely, it would be far simpler for alchemists to have written nothing at all." (2013, 18).

<sup>10</sup> Newman explains the terms in his concrete example: "Not only does our author employ dispersion de la science and the use of *Decknamen*, he uses two complementary techniques that I shall call syncope and paraphesis. These terms, although altered from their usual Greek sense, will serve to characterize certain techniques of concealment within an alchemical context. By syncope I mean the elliptical description of an alchemical process, substance, or even apparatus, with the intent to conceal. We observed this in the highly abbreviated recipe for the philosophical mercury that omitted both antimony and silver, and the technique is implicit in all of Philalethes' recipes that fail to mention the role of silver or copper as a 'mediator' for making antimony metal amalgamate with mercury. By paraphesis, on the other hand, I mean the heaping-up of synonyms for a given process, substance, or apparatus, again with the intention of bewildering the reader. Such paraphasis is present in the profusion of names used by Philalethes for antimony in its several forms, as we remarked above." (1996, 180). Syncope, that is intentionally leaving out relevant bits of information, thus far eludes digital analysis and consequently is not addressed in this article. 'Dispersion of knowledge' means spreading out relevant knowledge over multiple books or

to make sure that ‘the worthy’—that is the alchemical expert with sufficient chemical background knowledge—can follow what’s being said while ‘the unworthy’ are excluded from the communication. This serves both to foster community amongst the alchemical authors and to hide valuable information which can be monetized on the market but also for self-fashioning and marketing purposes in the so-called ‘economy of secrets’. In the ‘New Historiography of Alchemy’, William Newman and Lawrence Principe established a methodology for decoding *Decknamen* by reading them chemically and validating the interpretation by chemical experiment.<sup>11</sup>

With ongoing digitization efforts, considerable numbers of alchemical texts have become available online.<sup>12</sup> Of Maier’s works, the *Arcana* (1614) (Maier 1614; Maier 2009) were digitized by the Early English Books Online (EEBO) project in TEI-XML; Herzog-August-Bibliothek Wolfenbüttel digitized Maier’s *Symbola* (1617) (Maier 1617d; Maier 1617e) and *Examen* (1617) (Maier 1617b; Maier 1617c) in a project concerning alchemical sources of HAB with a transcription being available in TEI-XML (Feuerstein-Herz 2017). A first digital scholarly edition of Maier’s *Atalanta Fugiens* has been provided by the recently completed *Furnace and Fugue* project (Nummedal and Bilak 2020).<sup>13</sup> Alchemy research however, has not yet considered the potential of digital scholarly editions for providing analytical tools to address the specificities of alchemical language.<sup>14</sup>

---

documents which ultimately need to be linked back together again so one book can explain the other. This is easily addressed using semantically annotated digital editions. With dispersion of knowledge, it is imperative that key terms can be found over the whole corpus and that it can be detected when they correlate with similar other terms or else, when they seem completely out of place which is a common marker of dispersion of knowledge.

<sup>11</sup> For the “New Historiography of Alchemy,” see for example: Principe and Newman 2001; Tilton 2003, 9–18; Martínón-Torres 2011. Its agenda is broad yet the methodology for decoding *Decknamen* is most relevant for the present discussion. Essentially, the success of this method is measured in terms of coherence. The same is true for any kind of cryptological methods: In the process of encrypting, a string is deprived of its coherence. The process of decryption renders the garbled string coherent once again. Of the many possible reorderings of a garbled string, the ‘correct one’ is the one which is most coherent or meaningful. Especially with anagrams, which are common in alchemy too, many coherent ‘solutions’ are possible. Similarly, if a coherent result can be achieved in a chemical recreation of an experiment, it is deemed likely and probable that a Deckname has been correctly identified, thus conducting a best explanation ‘proof’ using chemistry as an ‘encryption key’.

<sup>12</sup> As of 15.08.2019, a VD-17 search currently yields 1484 hits of which 937 are digital facsimiles. Staatsbibliothek zu Berlin — Preußischer Kulturbesitz, Bayerische Staatsbibliothek München, and Herzog August Bibliothek Wolfenbüttel 2017.

<sup>13</sup> As of the time of writing, the digital edition was not accessible to the public yet but the author had access.

<sup>14</sup> For example: Martínón-Torres 2011, 233. Jockers criticizes this “disciplinary habit of thinking small”: “The traditionally minded scholar recognizes value in digital texts because they are individually searchable, but this same scholar, as a result of a traditional training, often fails to recognize the potentials for analysis that an electronic processing of text enables.” This habit, he continues, was also perpetuated in the types of digital tools commonly available for literary scholars in the past (Jockers 2013, 17).

An evaluation of effective strategies for their digital editions is thus in order. It is common for Digital Scholarly Editions to provide annotations and visualization tools for Named Entities or dates which are meant to help users explore the texts semantically. It has also become fairly established practice to offer tools for quantitative text analysis using methods based on bags of words and word counting. In this paper, two case studies relating to alchemical literature are presented: Firstly, the adequacy of word count based interpretation in the context of alchemical literature is discussed on the example of a keyword search for the Named Entity Paracelsus in the corpus of Michael Maier's printed works. Secondly, a custom-made method for the interpretation of alchemical terms is presented which relies on the polyvalent semantic annotation of relevant yet ambiguous alchemical terms and the analysis of their semantic contexts, that is the groupings these words appear in, using a digital thesaurus.

## 1 Case study 1: looking for Paracelsus using bags of words

A good example of what Leibenguth criticized as patchy research regarding Maier (Leibenguth 2002, 8–9) is the recurring assertion that Maier was a Paracelsian. This belief was taken for a fact until it came under scrutiny from the 1990s onwards.<sup>15</sup> As one of four possible answers to this question given by previous research, Leibenguth deduces Maier was a Paracelsian for his devotion to medieval alchemy which Paracelsus integrates in his theories (Leibenguth 2002, 71–73). However, Maier mentions Paracelsus very little in general and not all comments can be read as apologies. No general overtly positive narrative on Paracelsus is present. Thus, Wels argues, Maier was not a Paracelsian since Maier does not agree with contemporaries who were clearly Paracelsians on relevant theoretical aspects such as the doctrine of *tria principia* (Wels 2010, 158, 160–63, 188). Maier also praises Thomas Erastus, an opponent of Paracelsianism.<sup>16</sup> Thirdly, as Paracelsus was a *persona non grata*, it could be that Maier does not mention Paracelsus despite being influenced by him.<sup>17</sup> On the other hand, Maier does mention Paracelsus in his works, albeit infrequently, so it might be that Maier just does not care much about Paracelsus. His comments are

<sup>15</sup> First examples are: Beck 1991; Stiehle 1991.

<sup>16</sup> See also: "Zwar feierte Maier gelegentlich in Paracelsus eine Koryphäe von Weltrang und anerkannte bestimmte 'medicamenta [...] Chymica vel Paracelsica'. Von einem Paracelsisten (so z.B. Weyer, 1992, 279) kann freilich schwerlich die Rede sein (siehe Leibenguth, 1992, 68, 71f., 107; Wels, 2010): Allein schon seine Kritik an der Paracelsistischen Drei-Prinzipien- und Imaginationslehre, vor allem aber der äußerst geringe Anteil von Paracelsismen an seinem Gesamtwerk und sein Lob auf seinen 'praeceptor' [...] Th. Erastus [...] verraten Maiers feste Verwurzelung in der galenistischen Schulmedizin und im späthumanistischen Aristotelismus." (Kühlmann and Telle 2013, 956–57).

<sup>17</sup> Thanks to PD Dr. Ute Frietsch of Herzog-August-Bibliothek Wolfenbüttel for pointing out this possibility to me. See also: Frietsch 2013, 338.

ambiguous; some can be interpreted as positive; others are quite clearly pejorative. This might simply mean Paracelsus is not an especially important influence on Maier, yet he still mentions him due to Paracelsus's being a central figure of the German alchemical tradition.

The following questions will now be approached using readily available digital methods based on word counts like they are offered in many existing Digital Scholarly Editions:

1. Can we judge Maier's attitude towards Paracelsus using methods based on word frequencies?
2. How far do we get with these methods in general? Are they sufficient to grasp the specificities of alchemical texts? Can they, for instance, be used to help understand the meaning of the highly complex concept 'Mercurius'?

Maier's corpus was machine transcribed using the *Transkribus* Optical Character Recognition. *Voyant Tools* can be used to quickly evaluate the potential of string-based word counting. Queries of typical Paracelsian words<sup>18</sup> yield hardly any results: many do not appear at all, others just once or twice in the whole corpus. Querying name variants for Paracelsus, however, is more effective.<sup>19</sup> Figures 1 and 2 compare other authorities mentioned in Maier, such as Geber, Sendivogius, Albertus Magnus, Galen, Hippocrates, or Lullus. The results were visualized using the 'bubble lines' tool. This makes it quite clear that Maier mentions Paracelsus significantly less than all of the other authorities, especially the ones he seems to appreciate. Of course, instances where Maier talks about someone without mentioning their name cannot be tracked this way. Also, a certain error rate has to be expected with the transcriptions not having been corrected.<sup>20</sup> Even taking standard concerns on precision of these numbers into account, the quantitative analysis adds new insights to Leibenguth's observation, that Paracelsus is rarely cited.<sup>21</sup> The quantitative analysis not only shows that Paracelsus is cited very rarely by Maier, as Leibenguth had already pointed out. It also shows that Maier's mentions of Paracelsus are considerably less frequent than citations of other authorities Maier holds in high regard. This quantitative analysis

<sup>18</sup> In Paracelsian theory, there are quite a few otherwise uncommon terms like 'Archeus' or 'astral' powers, etc. If Maier was a Paracelsian, he would have used these *termini technici*.

<sup>19</sup> He is called not only by his chosen pseudonym 'Paracelsus', but also Theophrastus Bombastus von Hohenheim. Thus the query included Paracelsus, Theophrastus and Bombastus.

<sup>20</sup> Even though the corpus was neither double-key corrected or cleaned up, with a rate of 1–2 errors per page, it is of sufficient quality to obtain decent results when analyzing it using *Voyant Tools*.

<sup>21</sup> Unlike Franco Moretti whose concept of *Distant Reading* implicitly includes the idea that serial reading alone would suffice to interpret a corpus, Jockers's methodology of macroanalysis stresses that the "zooming in and out" between the overview over a big corpus and tiny details produces "findings, not interpretations" (30). Moretti 2013; Moretti 2005; Jockers 2013, 8–10, 19, 23, 29–30.

has provided arguments “to solidify intuitions we may already have had or [...] help identify what you might want to read” (Sinclair and Rockwell 2016, 282).

But to what extent can such an out-of-the-box keyword analysis using Voyant Tools be put to use to interpret alchemical *Decknamen*? Does a word count or a visualization suffice to answer research questions relevant to the historiography of alchemy? As Franco Moretti states: “Quantitative research [...] provides *data*, not interpretation” (Moretti 2005, 9). *Voyant Tools* helps by creating a concordance of all the sentences about Paracelsus. However, the method does not go beyond this compilation of data.<sup>22</sup> The collation of relevant paragraphs yields no immediate information on their semantic relevance.<sup>23</sup> This can be demonstrated in tracking the concept of ‘Mercurius’: the word ‘Mercurius’ occurs very frequently but covers an enormous range of possible meanings in alchemy. The string querying approach counts all mentions of ‘Mercurius’ as the same when, in fact, they have different word-meaning-relations. Among the additional shortcomings of the software are, for example, the fact that the font size is so small given many text samples and the fact that the colours are not consistent in multiple analyses. Even in two virtually similar queries, *Voyant Tools* often uses completely different colours, many times even switching the colours of concepts to one previously indicating the last one. This is extremely confusing and potentially misleading. The colourings also do not behave in a very predictable way so that it would be possible to easily achieve some form of ‘visual continuity’. This makes it very hard to interpret the visualizations in a comparative way. All that the following figures really show is that in a variety of personal names of scientific or alchemical authorities mentioned by Maier, Paracelsus does not stand out at all, quite the opposite. Therefore, a method is needed which allows for word sense disambiguation to take the specificities of alchemical language and its *Decknamen* into account.

## 2 Second case study: a digital thesaurus of alchemical concepts on the example of Saturn and the dragon’s blood

A simple example can show that Michael Maier himself is very aware of the ambiguity of alchemical concepts. Let’s consider a statement about ‘Saturnus’:

---

<sup>22</sup> “Computational analysis may be seen as an alternative methodology for the discovery and the gathering of facts. Whether derived by machine or through hours in the archive, the data through which our literary arguments are built will always be a movement from facts to interpretation of facts. The computer is a tool that assists in the identification and compilation of evidence. We must, in turn, interpret and explain that derivative data.” (Jockers 2013, 30)

<sup>23</sup> Sinclair and Rockwell warn: “Don’t expect much from the tools. Most tools at our disposal have weak or nonexistent semantic capabilities; they count, compare, track, and represent words; but they do not produce meaning—we do.” (Sinclair and Rockwell 2016, 288)

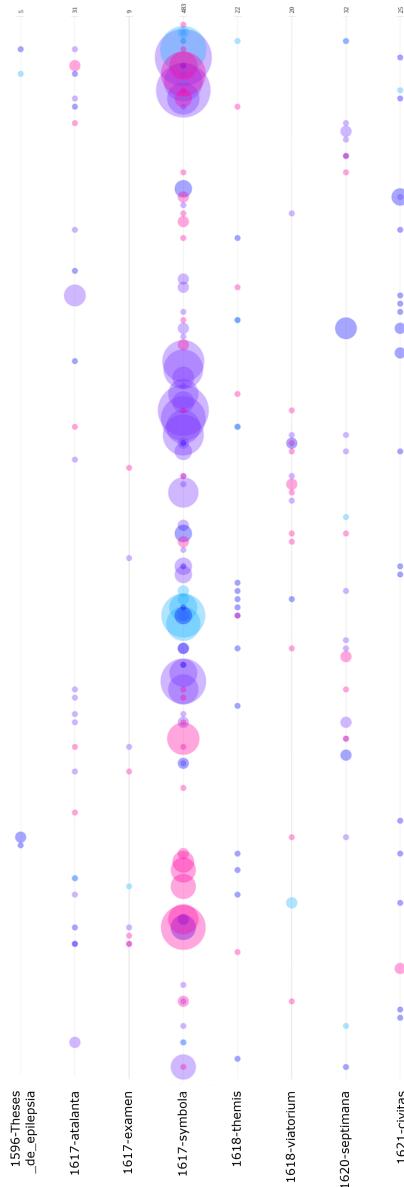


Figure 1: Overview of authorities mentioned in the Maieriana, reduced to works in which Paracelsus is mentioned at all.  
Paracelsus is visualized in blue, Galen in pink.

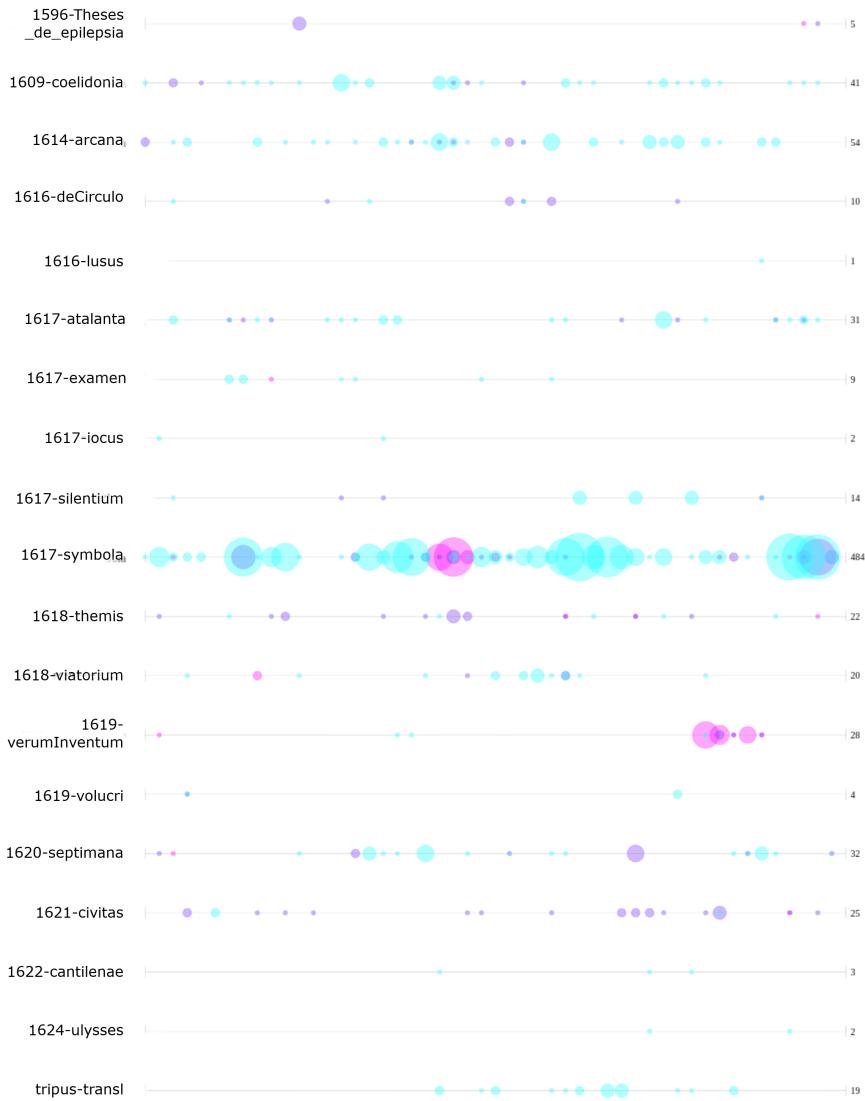


Figure 2: Overview of authorities mentioned in the Maieriana. Paracelsus is visualized in pink, Galenus and Hippocrates in purple, and all other authorities mentioned in blue.

```

215 :tinctura a :chemicalSubstance .
216 :tinctura :hasColour :red .
217 :tinctura :transformsTo :aurum .
218
219 :myth rdfs:subClassof :alchemicalSymbol ;
220 :hasContext :contextMythological .
221 :mythologicalFigure rdfs:subClassof :myth ; a :person .
222 :mythologicalObject rdfs:subClassof :myth ; a :object .
223 :mythologicalPlace rdfs:subClassof :myth ; a :place .
224
225 :red :hasChemicalProperty :tinctio .
226 :tinctio :hasChemicalProperty :tinctio .
227
228 :perfectio rdfs:subClassof :chemicalProperty .
229 :aurum :hasTheoreticalProperty :perfectio .
230 :aurum :hasColour :citrinitas .
231
232 :aurum a :chemicalSubstance .
233 :aurum rdfs:subClassof :aurum .
234
235 :fixum rdfs:subPropertyOf :chemicalProperty .
236 :agens rdfs:subPropertyOf :chemicalProperty .
237
238 < 7 :elephas a :animal .
239 :draco a :animal .
240 :medicina :partOf :iatrochymia .
241 :medicinaPhilosophica :partOf :iatrochymia .
242 :medicinaPhilosophica :hasContext :contextAlchemicalTheory .
243 :sanguis a :alchemicalSymbol ; a :chemicalSubstance ;
244 rdfs:label "cruor" .
245 :sanguis :hasColour :red .
246
247 :terreum :hasElement :earth .
248
249 :tason a :mythologicalFigure .
250 :Apollo a :mythologicalFigure ; :hasLinkTo :ladon .
251 :Ladon a :mythologicalFigure ; a :draco .
252 :MythologicalMercury :hasAttribute :caduceus .
253 :caduceus a :mythologicalObject ;
254 :consistsOf :draco .
255
256 :Cadmos a :mythologicalFigure ;
257 :hasRelatedStory :draco .
258 :vellusAureum a :mythologicalObject ;
259 :madeOf :aurum ;
260 :hasRelatedStory :Iason .
261
262 :AtalantaAndHippomenesRace a :myth ;
263 :partOf :AtalantaFugiensIntro ;
264 :guardedBy :draco .
265 :AtalantaAndHippomenesRace .
266
267 :hasActor :Atalanta , :Hippomenes ;
268 :hasObject :goldenApples .
269 :gardenOfTheHeleperides a :mythologicalPlace ;
270 :hasRelatedStory :AtalantaAndHippomenesRace .
271 :goldenApples a :mythologicalObject ;
272 :goldenApples .
273 :guardedBy :draco .
274 :hasPlace :gardenOfTheHeleperides ;
275 :hasRelatedStory :AtalantaAndHippomenesRace .
276
277 :Mercurius a :alchemicalSymbol .
278 :Sulphur a :alchemicalSymbol .
279 :Mercurius rdfs:subClassOf :Mercurius ;
280 :PhilosophicalMercury rdfs:subClassOf :Mercurius ;
281 :theoreticalSubstance a :theoreticalSubstance .
282 :PhilosophicalSulphur rdfs:subClassOf :Sulphur ;
283 :CommonMercury rdfs:subClassOf :theoreticalSubstance .
284 :CommonMercury rdfs:subClassOf :Mercurius ;
285 :Hydraryrum a :CommonMercury .
286 :CommonSulphur rdfs:subClassOf :Sulphur ;
287 :CommonSulphur a :chemicalSubstance .
288
289 :MythologicalMercury rdfs:subClassOf :Mercurius ;
290
291 :HermesTrismegistos a :theoreticalSubstance .
292 :HermesTrismegistos rdfs:subClassOf :Mercurius ;
293 :alchemicalAuthority a :alchemicalAuthority .
294
295 :Saturnus a :alchemicalSymbol .
296 :MythologicalSaturn rdfs:subClassOf :Saturnus ;
297 :mythologicalFigure a :mythologicalFigure ;
298 :PlanetsSaturnus rdfs:subClassOf :Saturnus ;

```

Figure 3: Snippet view of the thesaurus of alchemical terms.

Saturnum hic intellegimus non planetam illum mundi supremum, nec Deorum Ethnicorum quasi abavum & primum saturem (de quo in Hieroglyphicis nostris suo loco sat egimus) sed metallicam substantiam, quae plumbi vel ejus minerae, nomine appellatur, prout communiter recipitur. (Maier 1618, 44)

By Saturnus, I mean here neither the outmost planet of the universe nor the pagan God, so-to-say the forefather and first founder (about whom I talked about enough in his place in my Hieroglyphics), but the metallic substance which is called by the name of lead and its minerals, like it is commonly understood. (translated by the author)

In this example, Maier himself explains which of the different possible meanings he intends. Such explanations, however, are exceedingly rare in most alchemical texts and *Decknamen* can be much more difficult than this one.

As we have seen, for infrequent terms with a reduced amount of ambiguity, like the personal name of Paracelsus, a concordance already gives us a good impression on how the word is used in a text. For the more complex concept ‘Mercurius’ however, the mere finding of a string is not complex enough to provide any meaningful context (at least not in a way which does not require copious *close reading* in order to make interpretation possible). A specific method which goes beyond existing out-of-the-box tools for quantitative text analysis is needed.

Michel Butor stated in his essay on alchemical language that alchemical terms are essentially arbitrary when used without context, but their grouping is not (Butor 1990, 22). After all, how could chymical experts have communicated amongst each other if their language had been completely arbitrary? If it is the groupings which determine the meaning of alchemical *Decknamen*, those need to be made explicit: Fortunately, groupings of terms can easily be analysed in an annotated corpus, and relevant relations can be made explicit in a digital knowledge organization system. In the following, a method for the enrichment of digital editions using the annotation of alchemical terms and a digital thesaurus is presented on the example of a peculiar allegory in Michael Maier’s *Viatorium* (1618).

Alchemical language functions via analogies. Analogical thinking in chains of associations can be modelled digitally. By determining the context in which a term is uttered, we can gather evidence for the register of word usage, i.e., whether a chymical author is currently talking about practical chymistry or whether they use *Decknamen* to describe a more philosophical idea. This can be achieved by linking annotated strings in the text to concepts in a digital thesaurus (see Figure 3): The digital resource in question is neither a lexicon nor glossary. It was termed ‘thesaurus’ because this fits best with the definitions found in information science and it is the term used

to describe SKOS resources. Still, the term ‘thesaurus’ is ambiguous concerning the different research communities involved. Here, by ‘thesaurus’ we mean a digital knowledge organization resource which can be used as a hermeneutic and analytical tool. It does not aim to offer a controlled set of keywords to tag library catalogs, nor does it want to provide universal definitions for lemmata as in a glossary or lexicon. It does not want to be an introductory source to alchemical knowledge either. Its purpose is not to collect and condense all current knowledge on a topic in short entries. It is not a handbook either which serves as an entry point to research or for reference. It is a resource for the analysis of a text corpus. It is a hermeneutic, analytical tool to examine text and to bring a stronger philological text-centered focus to historical research which has been focusing on contexts of works and their authors rather than the work as text and its contents. How it is put to use will be elaborated in this second case study.<sup>24</sup>

The general method is as follows: A concordance is created for the occurrences of all labels from the thesaurus as they appear in the lemmatized Michael Maier corpus. This allows the creation of a specialized ‘Keyword-In-Context’ (KWIC) list restricted to the words in the text linked to the thesaurus. This is not a common KWIC, which provides linguistic context but rather an analysis of surrounding alchemical terms, providing a ‘context of *Decknamen*’. Using their respective links to the thesaurus, these surrounding alchemical terms can provide context for the word occurrence in question. These links are annotations. The annotated *Decknamen* in a text could also be used to quickly trace the subject of a work or to determine how present a certain topic is in the larger opus of an alchemical author, etc. Annotation of alchemical terms allows us making more targeted analyses within alchemical digital editions.<sup>25</sup>

In his *Sanguinis Draconis Descriptio*, in the second chapter of his 1618 *Viatorium*, “On the mountain of Saturn,” Maier recounts an allegory, an etiological explanation of the alchemical concept of ‘dragon’s blood’, meant to underpin the chemical interpretations laid out before. In this allegory, he uses the stylistic device of parathesis, that is the heaping up of *Decknamen* in order to confuse a reader but, possibly, also to further illustrate his point. In example code 1, a grouping of alchemical terms around the *Deckname* ‘Saturno’ in the passage in question has been encoded in XML.<sup>26</sup> As the terms are annotated over the whole corpus, a large number of such specialized KWIC

<sup>24</sup> The method is also explained in Lang 2022, 2021, 2020.

<sup>25</sup> For example, citations in the primary source can be excluded. Maier often makes use of extensive citations. Those most certainly do not offer insight into his use of *Decknamen*, so we might want to exclude *Decknamen* coming from inside citations.

<sup>26</sup> For better readability, both the XML and RDF examples are reduced to incomplete or pseudo-code versions. For example, in the XML, references from the strings annotated as <term> to their related thesaurus concepts are omitted as not to clutter the example so much that readability of the text is impaired. The examples, thus, do not show finished data but are meant rather as tools to help explain the functioning of the method.

results can be created for the purpose of semantic analysis. The example code below shows an annotated text snippet around the *Deckname* ‘Saturno’ from Maier 1618, 55-56, in XML pseudocode:

```
<example n="1" ref="Maier.Viatorium.55">
    ... qui post <term>DRACONIS</term> (licet potior pars
    <term>elephantis</term> sit) dicitur, ad <term>medicinam</term>
    aliasque necessitates hominum non inutilis.
    <term>Draconem</term> hunc, cuius <term>sanguis</term> ad
    <term>medicinam philosophicam</term> assumitur, oportet
    quem probè agnoscere, ne <pb n="56"/> <term>terreum</term> quid,
    seu ex <term>terrae</term> fundo depromptum, aut aliud haud
    conveniens in ejus locum substituat.
    Est autem ex eo genere, quod <deckname>Saturno</deckname>,
    <term>Apollini</term>, <term>Aesculapio</term> et
    <term>Mercurio</term> asscribitur, quod <term>Cadmi</term>
    socios interfecit, et ab eo interfectum, quod <term>vellus
    aureum</term> observavit, quod <term>pomis aureis</term>
    <term>Hesperidum hortis</term> invigilavit, ...
</example>
```

Its blood is then called [dragon’s / snake’s] blood (even though it is rather part of the elephant), quite useful to medicine and other necessities of humankind. This snake whose blood is taken as philosophical medicine needs to be properly understood, that is, not as something earthly taken from the ground, so that it is not wrongly substituted.

It is of the kind, which is attributed to Saturn, to Apollo, to Aesculapius and Mercurius. It is that which killed Cadmos’ comrades and was killed by him, which guarded the Golden Fleece and the apples from the Garden of the Hesperides [...]. (Maier 1618, 55-56, translated by the author.)

We can now obtain probability values indicating the most likely current semantic context by linking the words annotated as `<term>` with their related thesaurus concepts and calculate the probabilities using the following formula:

$$\text{current\_context} = \max(\text{sum\_of\_occurrences\_per\_context})$$

Each term can have multiple contexts. For example, *Saturnus* can have the semantic context ‘mythological’, ‘chemical’, ‘planet’, or ‘metal’. The count of possible contexts will thus be greater than the total number of terms considered. In the example, the search term for the KWIC ‘Saturno’ is the *Deckname* in question, surrounded by 15 alchemical terms in close proximity.

To make the process clear, a complete count will be given for this example (see the table in the Appendix): The concept `:draco` is defined as having `:contextAllegorical` and appears twice in the example of code (see the table in the Appendix). Plus one instance of `:elephas` which has the same semantics in

the thesaurus, this yields a total of 3 hits for :contextAllegorical for the passage considered.

Two occurrences of medicine ('medicina' and 'medicina philosophica') yield 2x :contextIatrochymia; 'sanguis' of type :chemicalSubstance adds one :contextChemical. 'Terreum' and 'terrae', belonging to the concept :earth yield 2x :contextElements. 'Saturno' as mentioned above yields one each of :contextMythological, :contextChemical, :contextMetals and :contextPlanets. The mythological figures, which follow, give a total of 8 :contextMythological with :Mercurius yielding one of :contextChemical, :contextPlanets, :contextAlchemicalTheory, :contextAlchemicalTradition, :contextHistorical as well as :contextMetals because the concept of :Mercurius is extremely rich in possible meanings in alchemy. For a total of 16 terms considered (including 'Saturno') the statistics are as follows:

:contextMythological	8
:contextAllegorical	3
:contextChemical	3
:contextElements	2
:contextIatrochymia	2
:contextPlanets	2
:contextMetals	2
:contextAlchemicalTheory	1
:contextAlchemicalTradition	1
:contextHistorical	1

If we wanted a percentage, the number of occurrences of :contextMythological divided by the number of total terms considered is:  $(8/16) = 50\%$ .

It has to be noted, however, that, as Willard McCarty points out, digital models are "temporary states in a process of coming to know" in which computers are not "knowledge jukeboxes" but "representation machines" and hermeneutic tools (McCarty 2004, 255). The probabilities they offer are not 'interpretations' but 'data'. In the end, they only serve to guide the human interpreter in making informed decisions. In this example, statistically the context is mythological in total, but it is also interesting to note that following the sequence of words after 'Saturno' all other terms can be mythological in context and all share a semantic relationship with the concept :draco. Preceding 'Saturno', we can see that the context is mostly allegorical,

with the same :draco as a dominant intersection which appears later as a common relation after ‘Saturno’. This is a good example of a semantic context switch used as an indicator in the alchemical stylistic and cryptographical device termed *parathesis*. *Parathesis* is defined as the ‘heaping up of Decknamen’ in which terms related to the concept or substance to which the author really wants to allude are cluttered around the text. It is meant to confuse readers who are not familiar with alchemical cryptographic devices but also possibly, in Maier’s case, as a form of ‘dispersion of knowledge’ and demonstration of erudition. In this device, one book explains another. The context switch indicates that Maier wants the informed reader to collect hints on the substances the symbol :draco stands for in the present allegory in his other works. That is especially in his *Arcana Arcanissima* (1614) where he explains the Greco-Roman myths as alchemical allegories and recounts in detail what he thinks they mean chemically.

A use case of the method illustrated above for readers who are not extremely familiar with classical mythology is to figure out what the common relation of the current concordance is, i.e., :draco. For a reader who can see this immediately, a semantically enriched digital edition improves the hermeneutical process: For each of the single annotated entities in the short snippet (there can be a much longer *paratheses*), the corpus of Maier’s printed works yields at least 300 hits each. However, only those where the terms appear in combination with a :draco are actually relevant to the analysis. A simple text query for “draco” would not help because there are many mythological beasts which qualify as :draco which are rather referred to by name and not by the string ‘draco’. For example, Apollo slays a :draco named ‘Python’ and the dragon guard of the golden apples in the garden of the Hesperides is called ‘Ladon’. If an instance of a :draco is encoded in RDF, excluding passages not relevant for the interpretation of :Saturn becomes much more efficient. Otherwise, a scholar would have to go through all the possible intertexts to determine which ones are relevant for the analogy at hand. Whether the present *parathesis* of mythological allusions in the allegory of the draco’s blood adds to Maier’s chemical argument has yet to be decided by chemists. However, finding possible intertexts which might contain additional information following the principle of ‘dispersion of knowledge’ becomes much simpler using a semantically enriched digital edition.

The next example appears immediately before the last one in Maier’s *Viatorium*. But immediately before the scene discussed previously, Maier had offered an allegory of a draco and an elephant in Africa to illustrate how he thinks some individuals misinterpret the alchemical symbol of “dragon’s blood”:

```
<example n="2" ref="Maier.Viatorium.55">
Cui, ut nihil est ab omni parte beatum, continuò se invidum
et hostem internecinum <term>draco</term> praegrandis offert.
Hic ut tam vastae beluae, cui viribus longè sit impar, damnum ingerat,
non audens aperto <term>Marte</term> cum ea configere,
```

```

rem ex insidiis aggreditur hoc modo. Observat quâ viâ
<term>elephas</term> aquatu redditurus sit, eandem, ubi arctior est,
praeoccupat, seseque inferiori loco abscondit, donec <term>elephas</term>
<term>aquis</term> praetumidus accedat, cuius pedibus nexus et
gyros suo corpore sinuoso injiciens <term>sanguinem</term> libat
et ebbit. Tandem ubi viribus defecerit <term>elephas</term>,
<term>draconem</term> circum sese in nodum reflexum, jam
<term>cruore</term> saturatum, deque victoria contra hostem
obtenta securum, suo pondere in terram corrueens, ipsa morte
vindictam sumens conterit et extinguit, omni regurgitante
et offusso ejus <term>sanguine</term>, ...
</example>
```

Because nothing is entirely good, a huge draco often crosses his path, hostile and murderous. In order to cause damage to such an immense beast to whose strength it [the snake] is no match at all, not daring to attack it openly (aperto Marte), it attacks him in an ambush. Watching the path by which the elephant returns from drinking, the draco sets itself up where it is straightest, hidden down below until the elephant, swollen with water, approaches around whose feet and sinuous body it then binds itself, spilling [the elephant's] blood and drinking it up. At last, when the elephant's strength has waned, taking revenge for its own death, it kills the draco who, blood-saturated lays curled up around him by falling down to the floor with its whole weight; the draco, assured of its own victory, regurgitating and sprayed in [the elephant's] blood. (Maier 1618, 55, translated by the author.)

With the dominant terms ‘elephas’ and ‘draco’, the context here is clearly allegorical rather than mythological as in the following text (a detailed account will not be given for this example due to space constraints). Due to the frequent occurrence of ‘sanguis’, :contextChemical is the runner-up. However, it is notable that the concept of draco is already present here which reappears indirectly later in the mythological analogies of the *parathesis*. If we go back even further in the chapter on “The mountain of Saturnus” before the *Descriptio Sanguinis Draconis*, we find a completely different context:

```

<example n="3" ref="Maier.Viatorium.52">
Quod si verò à <term>sulfure</term> <term>auri</term>
<term>fixo</term> et <term>agente</term>, hoc est, <term>calido</term>
et <term>sicco</term>, <term>tingente</term> <term>Mercurius</term>
<term>plumbi</term> <term>coaguletur</term>, nullum est dubium ex eo,
ut et aliâs <term>aurum</term> <term>perfectissimum</term> fieri posse,
de quo hic non est disserendi propositum.
</example>
```

But if it is true that Mercurius Plumbi is coagulated by the fixed and active Sulfur Auri, that is by the hot and dry which tints, there is no doubt that

other [substances] can also be perfected into gold. An argument of which we won't speak here. (Maier 1618, 52, translated by the author.)

The concept :aurum is very present both via direct mentions as well as in the relations of other terms, such as 'perfectissimum' or 'tingens'. Many different contexts appear of which none is dominant. However, they all have contexts which are subclasses of theoretical alchemy (:contextAlchemicalTheory). Combined with the dominance of gold, it is safe to interpret that the current subject is a theoretical discussion relating to gold-making). A close reading shows that it is very likely the *sulfur auri* mentioned in this sentence which is the (theoretical) substance Maier later refers to in the *draco* allegory.

These three examples illustrate how the presence of a common theme can be disguised in alchemical texts by switching the contexts of the symbols used to refer to the substance in question. However, the new symbols still share enough associations with the old ones that machine reasoning can help make this invisible Ariadne's thread resurface again.

### 3 Conclusion

A Digital Scholarly Edition of an alchemical text enriched as such provides a much-needed hermeneutical tool for the interpretation of alchemical literature. The method described allows the presence of multiple possible contexts at the same time. Therefore, it is adequate to the specificities of alchemical communication styles which consciously employed polyvalent semantics as a stylistic device. Alchemical terms cannot be disambiguated in a strict sense because their ambiguity is intentional and an integral part of the semantic message. Digital annotation offers an environment for making such ambiguities explicit and machine-processable without forcing a binary disambiguation. Clearly, this method customized to the specific needs of alchemical texts provides a much deeper and more adequate interpretation than would be possible with mere word counting without making semantic interrelations explicit. Text enrichment with semantic relations provides a hermeneutical tool which is of great help when dealing with the complexities of alchemical language. Even though alchemical literature is a relatively narrow historical genre, possible applications for the method proposed are manifold even outside the realm of alchemy, such as the analysis of poetic allusions.

## Bibliography

- Beck, Wolfgang. 1991. "Michael Maiers Examen Fucorum Pseudo-Chymicorum – Eine Schrift wider die falschen Alchemisten". München: PhD diss., TU München.
- Butor, Michel. 1990. *Die Alchemie und ihre Sprache. Essays zur Kunst und Literatur*. Frankfurt: Fischer.
- Eco, Umberto. 2016. "Il discorso alchemico e il segreto differito." In Umberto Eco. *I limiti dell'interpretazione*. Milano. 97–116.
- Feuerstein-Herz, Petra. 2017. "Alchemie Portal der Herzog August Bibliothek Wolfenbüttel. Entwickelt im Rahmen des DFG-Projektes Projekt 'Erschließung Alchemischer Quellen in der Herzog August Bibliothek' (2014–2017)." Herzog August Bibliothek Wolfenbüttel. Accessed February 28, 2022. <http://alchemie.hab.de>.
- Frietsch, Ute. 2013. *Häresie und Wissenschaft. Eine Genealogie der paracelsischen Alchemie*. München: Wilhelm Fink.
- Jockers, Matthew L. 2013. *Macroanalysis. Digital Methods and Literary History*. Chicago: University of Illinois Press.
- Jong, Helena de. 1969. *Michael Maiers's Atalanta Fugiens. Sources of an Alchemical Book of Emblems*. Lake Worth: Nicolas-Hays.
- Jütte, Daniel. 2011. *Das Zeitalter des Geheimnisses. Juden, Christen und die Ökonomie des Geheimen (1400–1800)*. Göttingen: Vandenhoeck & Ruprecht.
- Kühlmann, Wilhelm, and Joachim Telle, eds. 2013. "Michael Maier." In *Der Frühparacelsismus*. III / 1, ed. by Wilhelm Kühlmann and Joachim Telle. Berlin: de Gruyter.
- Lang, Sarah. 2022. "A Machine Reasoning Algorithm for the Digital Analysis of Alchemical Language and its Decknamen." *Ambix* 69 (1): 65–83. doi:10.1080/00026980.2022.2038428.
- Lang, Sarah. 2021. *Alchemische Decknamen digital entschlüsseln. Polysemantische Annotation und Machine Reasoning am Beispiel des Korpus des Iatrochymikers Michael Maier (1568–1622)*. Graz: PhD diss., Universität Graz.
- Lang, Sarah. 2020. "Digitale Annotation alchemischer Decknamen. 'Die Allegoriae werden uns nit mehr verborgen seyn.'" In *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, ed. by Julia Nantke und Frederik Schlupkothen, 201–19. Boston: De Gruyter.
- Leibenguth, Erik. 2002. *Hermetische Philosophie Des Frühbarock. Die "Cantilenae Intellectuales" Michael Maiers. Edition mit Übersetzung, Kommentar und Bio-Bibliographie*. Tübingen: Niemeyer.
- Maier, Michael. 1596. *Theses de Epilepsia*. Basel: Typis Conradi Waldkirchii.  
—. 1614. *Arcana Arcanissima*. London: Creede.  
—. 1617a. *Atalanta fugiens*. Oppenheim: de Bry.  
—. 1617b. *Examen Fucorum Pseudo-chymicorum Detectorum et in Gratiam Veritatis Amantium Succincte Refutatorum*. Frankfurt: de Bry.  
—. 1617c. *Examen Fucorum Pseudo-Chymicorum: Daten der Herzog August Bibliothek Wolfenbüttel*. Frankfurt: de Bry. Accessed February 28, 2022. <http://diglib.hab.de/drucke/46-med-4s/start.htm>.  
—. 1617d. *Symbola Aureae Mensae Duodecim Nationum*. Frankfurt: Jennis.

- . 1617e. *Symbola Aureae Mensae: Daten der Herzog August Bibliothek Wolfenbüttel*. Accessed February 28, 2022. <http://digilib.hab.de/drucke/46-med-1s/start.htm>.
  - . 1618. *Viatorium*. Oppenheim: de Bry.
  - . 1622. *Cantilenae intellectuales*. Rostock: Hallervord.
  - . 2009. “Arcana Arcanissima (Digitale Edition der 1613-Ausgabe): Ann Arbor, Mi; Oxford (UK): Text Creation Partnership, 2008–09 (EEBO-TCP Phase 1).” Accessed February 28, 2022. <http://name.umdl.umich.edu/A06751.0001.001>.
- Martinón-Torres, Marcos. 2011. “Some Recent Developments in the Historiography of Alchemy.” *Ambix* 58 (3): 215–37.
- McCarty, Willard. 2004. “Modeling: A Study in Words and Meanings.” In *A Companion To Digital Humanities*, ed. by Susan Schreibmann, Ray Siemens, and John Unsworth, 254–72. Oxford: Wiley Blackwell.
- Moretti, Franco. 2005. *Graphs, Maps, Trees. Abstract Models for a Literary History*. London: Verso.
- Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- Newman, William R. 1996. “Decknamen or Pseudochemical Language?” *Eirenaeus Philalethes and Carl Jung*.” *Revue d'histoire Des Sciences* 49: 159–88.
- Nummedal, Tara, and Donna Bilak, eds. 2020. *Furnace and Fugue: A Digital Edition of Michael Maier's Atalanta Fugiens (1618) with Scholarly Commentary*. Accessed February 28, 2022. <https://furnaceandfugue.org>.
- Principe, Lawrence M. 1992. “Robert Boyle's Alchemical Secrecy: Codes, Ciphers and Concealments.” *Ambix* 39 (2): 63–75.
- , and William R. Newman. 2001. “Some Problems with the Historiography of Alchemy.” In *Secrets of Nature: Astrology and Alchemy in Early Modern Europe*, ed. by William R. Newman and Anthony Grafton, 385–432. Cambridge/Massachusetts: MIT Press.
- . 2013. *The Secrets of Alchemy*. Chicago: University of Chicago Press.
- Ivo Purš, and Jaroslava Hausenblasová. 2016. “Michael Maier and His Prague Activities.” In *Alchemy and Rudolf II*, ed. by Ivo Purš, and Vladimir Karpenko, 335–66. Prag: Artefactum.
- Sinclair, Steffan, and Geoffrey Rockwell. 2016. “Text Analysis and Visualization: Making Meaning Count.” In *A New Companion To Digital Humanities*, ed. by Susan Schreibmann, Ray Siemens, and John Unsworth, 274–90. Oxford: Wiley Blackwell.
- Sinclair, Steffan, and Geoffrey Rockwell. “Voyant Tools.” Accessed February 28, 2022. <https://voyant-tools.org>.
- Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Bayerische Staatsbibliothek München, and Herzog August Bibliothek Wolfenbüttel. 2017. “VD17 — Das Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts.” Accessed February 28, 2022. <http://www.vd17.de>.
- Stiehle, Hans Roger. 1991. “Michael Maierus Holsatus (1569–1622) Alchemist und Arzt. Ein Beitrag zur Naturphilosophischen Medizin in seinen Schriften und zu seinem wissenschaftlichen Qualifikationsprofil”. PhD diss., München: TU München.
- Tilton, Hereward. 2003. *The Quest for the Phoenix. Spiritual Alchemy and Rosicrucianism in the Work of Count Michael Maier (1569–1622)*. Berlin, New York: De Gruyter.

- . 2005. “Maier, Count Michael.” *Dictionary of Gnosis and Western Esotericism*, Vol. I, ed. by Wouter J. Hanegraaff et al., 747–49.
- . 2020. “Michael Maier: An Itinerant Alchemist in Late Renaissance Germany.” In *Furnace; Fugue. A Digital Edition of Michael Maier’s Atalanta fugiens (1618) with Scholarly Commentary*, ed. by Tara Nummedal and Donna Bilak. Accessed February 28, 2022. <https://furnaceandfugue.org>. “Transkribus.” Accessed February 28, 2022. <https://transkribus.eu/Transkribus>.
- Vermeir, Koen, and Dánial Margócsy, eds. 2012. “Special Issue: States of Secrecy.” *The British Journal for the History of Science* 45 (2): 153–64.
- Wels, Volkhard. 2010. “Poetischer Hermetismus. Michael Maiers *Atalanta Fugiens* (1617/18).” In *Konzepte des Hermetismus in der Literatur der Frühen Neuzeit*, ed. by Peter-André Alt and Volkhard Wels, 149–94. Göttingen: V&R Unipress.

## 4 Appendix

	:contextMythological	:contextAllegorical	:contextChemical	:contextElements
<b>DRACONIS</b>		X		
<b>elephantis</b>		X		
<b>medicinam</b>				
<b>Draconem</b>		X		
<b>sanguis</b>			X	
<b>medicinam philosophicam</b>				
<b>terrum</b>				X
<b>terrae</b>				X
<b>Saturno</b>	X		X	
<b>Apollini</b>	X			
<b>Aesculapio</b>	X			
<b>Mercurio</b>	X		X	
<b>Cadmi</b>	X			
<b>vellus aureum</b>	X			
<b>pomis aureis</b>	X			
<b>Hesperidum hortis</b>	X			
<b>Sum</b>	8	3	3	2

	:contextIatrochymia	:contextPlanets	:contextMetals
<b>DRACONIS</b>			
<b>elephantis</b>			
<b>medicinam</b>	X		
<b>Draconem</b>			
<b>sanguis</b>			
<b>medicinam philosophicam</b>	X		
<b>terrum</b>			
<b>terrae</b>			
<b>Saturno</b>		X	X
<b>Apollini</b>			
<b>Aesculapio</b>			
<b>Mercurio</b>		X	X
<b>Cadmi</b>			
<b>vellus aureum</b>			
<b>pomis aureis</b>			
<b>Hesperidum hortis</b>			
<b>Sum</b>	2	2	2

	:contextAlchemicalTheory	:contextAlchemicalTradition	:contextHistorical
<b>DRACONIS</b>			
<b>elephantis</b>			
<b>medicinam</b>			
<b>Draconem</b>			
<b>sanguis</b>			
<b>medicinam philosophicam</b>			
<b>terreum</b>			
<b>terrae</b>			
<b>Saturno</b>			
<b>Apollini</b>			
<b>Aesculapio</b>			
<b>Mercurio</b>	X	X	X
<b>Cadmi</b>			
<b>vellus aureum</b>			
<b>pomis aureis</b>			
<b>Hesperidum hortis</b>			
<b>Sum</b>	1	1	1



# Versionierung von digitalen Editionen in der Praxis

Martina Bürgermeister

## Abstract

The aim of this paper is to find out whether versioning of digital editions is relevant for projects, which forms of versioning are used in practice and for whom and why versioning is used. In order to answer these questions, a quantitative inventory of digital editions that are versioned was made and their versioning strategies were examined in detail. The study of digital editions shows that versioning is not a marginal issue in the scholarly practice of editors. Every fourth edition project from the evaluated data set already implements forms of versioning.

## Zusammenfassung

In diesem Beitrag geht es darum herauszufinden, ob Versionierung von digitalen Editionen in Projekten eine Rolle spielt, welche Formen davon in der Praxis zum Einsatz kommen und für wen und warum versioniert wird. Um diese Fragen beantworten zu können, wurde eine quantitative Bestandsaufnahme von digitalen Editionen, die versioniert werden, gemacht und deren Versionierungsstrategien im Detail untersucht. Es wird gezeigt, dass Versionierung in der wissenschaftlichen Praxis von Editorinnen und Editoren kein Randthema darstellt. Jedes vierte Editionsprojekt aus dem evaluierten Datensatz setzt Formen der Versionierung bereits um.

Die Veränderlichkeit digitaler Medien und somit auch digitaler Editionen, ist ein Grund dafür, warum diese ihren Rezipienten, den Forscherinnen und Forschern, weniger vertrauenswürdig erscheinen als Druckeditionen. Digitalen Editionen fehlt im Gegensatz zu Druckeditionen oft ein finaler Zustand, ein spezifischer Moment der Veröffentlichung. Sie sind *fluid publications* (Sahle 2016, 29), deren autoritativer Charakter als *gesicherter Text* in Frage gestellt wird:

Digital editions are mutable objects: they change because the technology around them changes and therefore they are forced to adapt to it; they change because they can be changed; and they change because they are inherently mutable, interactive objects. (Pierazzo 2014, 198)

Aber gerade diese Wandel- und Veränderbarkeit von digitalen Editionen kann auch als Stärke gesehen werden, da der aktuelle Stand des editorischen Wissens eines

Projekts bei inkrementellen Publikationen in Echtzeit zugänglich ist. Vertrauenswürdige Ressourcen werden sie jedoch erst dann, wenn mit der Veränderlichkeit so umgegangen wird, dass Modifikationen nachvollziehbar sind und geänderte Inhalte referenzierbar bleiben. Eine Methode mit Änderungen in einem kontinuierlichen Entwicklungsprozess umzugehen ist Versionierung:<sup>1</sup>

By embedding a versioning system into a digital scholarly edition, for instance, the editorial work will become even more accountable and it would be possible for the user to follow the evolution of scholarship. (Pierazzo 2014, 200)

Versionierung kann Bearbeitungsschritte dokumentieren und wissenschaftliche Entwicklungen nachvollziehbar machen, weil die Methode es ermöglicht, Momentaufnahmen (Versionen) in einem Prozess festzuhalten. Auslöser für neue Versionen sind Veränderungen. An digitalen Editionen kann sich viel ändern, z.B. an den Inhalten, der Datenstruktur oder der Präsentation. Aber muss die Edition deshalb gleich versioniert werden? Wie relevant ist das Thema *Versionierung* überhaupt für Editionsprojekte? Ziel der vorliegenden Untersuchung war es, herauszufinden, ob Versionierung von digitalen Editionen in Editionsprojekten eine Rolle spielt, welche Formen davon in der Praxis zum Einsatz kommen und für wen und warum versioniert wird. Um diese Fragen beantworten zu können, wurde in einem ersten Schritt eine quantitative Bestandsaufnahme von digitalen Editionen, die versioniert werden, gemacht und in einem weiteren deren Versionierungsstrategien im Detail analysiert. Resultat der Analyse sind vier unterschiedliche Herangehensweisen, mit Versionen umzugehen, die im zweiten Teil des Beitrages präsentiert werden.

## 1 Bestandsaufnahme

Für die quantitative Bestandsaufnahme wurden alle Editionen, die im *Catalogue of Digital Editions*<sup>2</sup> (Franzini et al. 2016-2020) eingetragen sind, systematisch untersucht.<sup>3</sup> Der Datensatz wurde explorativ verwendet und stellt kein vollständiges Abbild der digitalen Editionslandschaft dar. Der Katalog wird von Greta Franzini, Peter Andorfer und Ksenia Zsaytseva unter CC-BY-SA-Lizenz herausgegeben. Sie betreiben eine sehr offene Aufnahmepolitik, weil potentielle Editionsprojekte jederzeit vorgeschlagen werden können und nach inhaltlicher Prüfung in den Katalog mitaufgenommen werden. Die seit 2016 verfügbare Webapplikation bietet zudem die Möglichkeit die

---

<sup>1</sup> Zum Mehrwert der Versionierung für die wissenschaftliche Praxis siehe Bürgermeister 2019.

<sup>2</sup> Stand März 2020.

<sup>3</sup> Eine weitere quantitative Analyse von digitalen Editionen, die vor allem die Zitierbarkeit von Editionen ins Zentrum rückt, findet man bei Bleier 2021.

Daten in verschiedenen maschinenlesbaren Formaten herunterzuladen, was die Analyse enorm erleichtert, während der weit umfangreichere Katalog von Patrick Sahle (2008–2019) mit mehr als 600 Einträgen, diese Funktion bisher nicht anbietet.<sup>4</sup> Jeder Katalogeintrag verzeichnet Metadaten wie Titel, Verfasser, Herausgeber, Webadressen und Lizenzierungsfragen zu der digitalen Edition. Der Gesamtkatalog umfasst im März 2020 309 Einträge, wovon nach Ausschluss von Editionen mit Zugangsbeschränkung und jenen, die nicht mehr auffindbar waren, 257 untersucht werden konnten (siehe Tabelle 1).<sup>5</sup>

	Anzahl der Projekte
<b>Gelistete Projekte</b>	309
<b>Als proprietär gelistet</b>	33
<b>Seitenladefehler</b>	19
<b>Untersuchte Projekte gesamt</b>	257

Tabelle 1: Einträge im *Catalogue of Digital Editions*.

Jede Edition wurde daraufhin analysiert, ob sie versioniert wird. Unter Versionierung wird nicht nur Versionskontrolle im Sinne des Software-Engineerings verstanden, sondern viel allgemeiner eine Methode, die einen Entwicklungsstand abbildet.<sup>6</sup> Das heißt konkret, dass Projekte, die gar nicht von Versionen in den editorischen Begleittexten sprechen, aber Strategien anwenden, methodisch mit Veränderungen umzugehen, dadurch mitgezählt werden. Die Entscheidung, ob eine digitale Edition versioniert, beruht auf meiner Arbeitsdefinition zu Versionierung:<sup>7</sup>

Versionierung ist das Kennzeichnen von Zuständen digitaler Ressourcen in Entwicklung.<sup>8</sup> Die Veränderungen im fortlaufenden Prozess werden dokumentiert und mit einem neuen Kennzeichen versehen. Das Kennzeichen

<sup>4</sup> Stand März 2020.

<sup>5</sup> Die Datenhaltung und der Funktionsumfang einer Edition können nur untersucht werden, wenn sie auch öffentlich zugänglich sind, was bedeutet, dass Editionen mit eingeschränktem Zugang nicht näher analysiert wurden.

<sup>6</sup> Grundsätzlich ist Versionskontrolle oder -management eine spezielle Form der Versionierung, die über Versionskontrollsystme, wie sie seit den 1970er Jahren in der Softwareentwicklung erfolgreich eingesetzt werden, ausgeübt wird. Diese Systeme sorgen dafür, dass unterschiedliche Bearbeitungszustände im Prozess der Softwareentwicklung in Form von Versionen gespeichert und jederzeit wiederherstellbar sind und mit anderen Versionen verglichen werden können (Collins-Sussman et al. 2004; Santacroce et al. 2016).

<sup>7</sup> In der Einleitung des Sammelbandes von Bleier und Winslow 2019 wird die Schwierigkeit einer einheitlichen Definition von Versionierung in den DH diskutiert (Bleier und Winslow 2019, V-XI).

<sup>8</sup> Diese Definition lehnt sich an jene des Onlinewörterbuchs *Duden* an: „Vergabe von [mehrteiligen]

eines Entwicklungsstandes kann eine Datumsangabe oder inkrementelle Versionsnummern sein. Diese unterschiedlichen Versionen beziehen sich auf eine gemeinsame Entität, in unserem Fall geben Versionen den Bearbeitungsstand der digitalen Edition oder konkreter Teile davon an. Versionierung ist erst dann eine Methode, digitale Ressourcen vertrauenswürdiger zu machen, wenn sie die Veränderungen einer Entität über die Zeit dokumentiert; aus diesem Grund ist hier weder die bloße Angabe des Datums der letzten Änderung eine Form der Versionierung, noch der Einsatz einer Versionsnummer ohne Bezug zu Vorgängerversionen.

Dieser Definition folgend, werden 65 der 257 Editionsprojekte als Ganzes oder in Teilen versioniert. Unter diesen Editionen konnten unterschiedliche Versionierungsstrategien herausgearbeitet werden, die sich in Versionierungstypen einteilen lassen. Diese unterscheiden sich nicht nur in der technischen Umsetzung, sondern auch in der Intention warum und für wen versioniert wird. In der folgenden Vorstellung dieser Typen führe ich einzelne Editionen als Beispiele an, um diverse Aspekte des jeweiligen Versionierungstyps besser veranschaulichen zu können.

## 2 Versionierungstypen

Die analysierten Editionen zeigen unterschiedliche Strategien, wie mit Änderungen umgegangen werden kann. Sie können prinzipiell auf vier Typen abgebildet werden, die aber auch kombiniert in den Editionen angewandt werden (siehe Anhang).

Typ eins wird hier *Änderungsbeschreibung* (vgl. Kapitel 2.1) genannt: In den meisten Fällen wird diese Form der Versionierung auf einer eigenen Webseite der Edition, die vom Hauptmenü aus angesteuert werden kann, umgesetzt. Bei Versionierungstyp zwei werden Veränderungen an der Edition in den Ausgangsdaten beschrieben. Dieser Typ wird hier als *Revision Description* (vgl. Kapitel 2.2) bezeichnet. Der dritte Versionierungstyp wurde über Verweise in der technischen Dokumentation bzw. in den jeweiligen Projektbeschreibungen der Editionen identifiziert, mehr dazu unter *Versionskontrolle* (vgl. Kapitel 2.3). Ein vierter Typ wurde definiert, der hier *Versionierung des Gesamtsystems* (vgl. Kapitel 2.4) genannt wird. Diese Form der Versionierung erschließt sich aus den jeweiligen Projektbeschreibungen oder über Hyperlinks mit eindeutigen Bezeichnungen, die zu Vorgängerversionen der Edition führen. In den folgenden Unterkapiteln wird anhand von konkreten Editionsbeispielen beschrieben, wie diese unterschiedlichen Versionierungsstrategien funktionieren, für wen diese Mechanismen gedacht sind und warum sie angewandt werden.

---

Versionsnummern zur Kennzeichnung und Unterscheidung der Entwicklungsstadien einer Software oder einer Datei“, verzichtet aber auf die Einschränkung der Kennzeichnung durch Versionsnummern.

## 2.1 Änderungsbeschreibung

Die Analyse hat ergeben, dass mehr als die Hälfte der Editionsprojekte (34), die Versionierung anwenden, chronologisch Änderungen der Edition beschreibt und mit Datum oder Versionsnummer bezeichnet. Diese Angaben machen Entwicklungsstadien der Edition unterscheidbar und bilden eine nachvollziehbare Versionskette. Einige Projekte kombinieren diese Form des Versionierens mit einer weiteren Art der Versio-nierung, andere setzen ausschließlich auf diese Strategie.<sup>9</sup> Änderungsbeschreibungen befinden sich entweder gut sichtbar als extra Menüpunkt in der Edition, oder werden direkt auf der Startseite integriert, eventuell sind sie aber auch im Vorwort zur Edition zu finden. Die Editionen verwenden sehr unterschiedliche Bezeichnungen für Änderungsbeschreibungen, die in der Regel aber das gleiche bezwecken. So führen Menüpunkte wie *Site Updates*, *Latest News*, *What's new Archive*, *Recent changes*, *Revision history* oder *Versionsarchiv* zu einem Verzeichnis von Änderungen der Edition. Das Verzeichnis besteht aus Versionsnummern oder auch Zeitstempel (Datumsangabe) gefolgt von einer freien Beschreibung dessen, was sich verändert hat. Dies kann die Information sein, dass der Edition neue Texte hinzugefügt wurden, dass das CSS erneuert wurde oder dass Bilder in höherer Auflösung verfügbar sind, die wiederum zu einer Überarbeitung von spezifischen Texten führte.

In der Edition *Auchinleck Manuscript* beispielsweise werden Änderungen auf Zeichenebene genau verzeichnet. David Burnley und Alison Wiggins edieren in diesem Projekt eine Londoner Handschrift aus den 1330er Jahren. Sie befindet sich in der schottischen Nationalbibliothek und wurde 1740 von Lord Auchinleck wiederentdeckt, der zum Namensgeber der Handschrift wurde. Das Projekt startete 2000 und wurde im Juli 2003 mit Version 1 publiziert. Der aktuelle Stand der Edition entspricht Version 1.1 von März 2004. Im Hauptmenü auf der linken Seite befindet sich der Link zur Änderungsbeschreibung, genannt *Archive of site updates*. Es besteht aus drei Listen: Eine mit Versionsnummer und Datumsangaben, eine weitere mit *Corrections to texts and to textual notes* und einer extra Rubrik *Record of other changes*, die etwa Änderungen zu Bibliographie und Seitenmenü listet. Die Herausgeberin Alison Wiggins erklärt, warum und wie die Änderungen verzeichnet werden:

All changes made to the content of this site are recorded here in the archive of site updates. Each time a batch of updates is added, the site is designated a new version number. This system ensure that users can keep track of any changes made and can reference the site materials accurately (Wiggins 2004).

Mit diesen einleitenden Worten in der Änderungsbeschreibung richten sich die Herausgeber vor allem an die Benutzerinnen und Benutzer, denen es damit ermöglicht

---

<sup>9</sup> Die unterschiedlichen Kombinationen sind in der Tabelle im Anhang ersichtlich.

werden soll, alle Änderungen an der Edition nachzuverfolgen. Die Versionierungsstrategie ist so angelegt, dass nicht jede einzelne Änderung zu einem Versionsupdate führt, sondern, eine neue Version wird erst dann veröffentlicht, wenn ein ganzes Bündel an Änderungen gemacht wurde, die in der Änderungsbeschreibung erfasst werden.

Ein Verzeichnis mit Änderungen gibt es auch im Editionsprojekt *The Online Froissart*, einer Edition der umfangreichen und in Europa und Nordamerika aufbewahrten handschriftlichen Chroniken Jean Froissarts zur Zeit des Hundertjährigen Krieges. Aufgrund der großen Menge an Material und dessen verstreuter Provenienz, haben die Herausgeber Peter Ainsworth und Godfried Croenen entschieden, nicht mit der Publikation der Edition zu warten bis alle Quellen ediert sind, sondern die Edition regelmäßig zu aktualisieren, die Änderungen zu verzeichnen und unter neuem Versionsnamen zu veröffentlichen. Der aktuelle Stand entspricht der Version 1.5 von Dezember 2013.

Versionierung als Änderungsbeschreibung ist nicht nur technisch einfach umzusetzen, sondern hat für die Projektverantwortlichen und die Benutzerinnen und Benutzer Vorteile. Wie Änderungsbeschreibungen ausgeführt werden, variiert aber von Edition zu Edition. In Abhängigkeit dazu ändert sich auch die Zielgruppe, für die diese Aufzeichnungen gemacht werden. Jene Projekte, die ihre Änderungen an der Edition als Neuigkeiten deklarieren und über Änderungen im Allgemeinen berichten, richten sich stärker an ein breites Publikum, während jene, die exakt festhalten, an welcher Stelle der Edition bzw. des edierten Textes sich etwas geändert hat, klar textkritisch interessierte Wissenschaftlerinnen und Wissenschaftler adressieren.

## 2.2 Revision Description

Änderungsbeschreibungen kommen auch auf der Ebene der Daten in den einzelnen Textdateien der Edition im Backend zum Einsatz. Wie Änderungsbeschreibungen (Kapitel 2.1), werden auch diese Aufzeichnungen als Form der Versionierung angesehen, weil sie die Entwicklung eines Dokuments mit Datum kennzeichnen und die beschriebenen Änderungen als chronologische Ereigniskette nachvollziehbar machen. Die *Text Encoding Initiative (TEI)* z.B. bietet dafür einen standardisierten Weg der Beschreibung der Änderungen im Dokument an. Die *Revision Description* dient laut Guidelines der Aufzeichnung von Änderungen an der Datei und beschreibt auf diese Art eine Dateihistorie (TEI Consortium, „<revisionDesc>“):

The final sub-element of the TEI header, the `revisionDesc` element, provides a detailed change log in which each change made to a text may be recorded. Its use is optional but highly recommended. It provides essential information for the administration of large numbers of files which are being updated,

corrected, or otherwise modified as well as extremely useful documentation for files being passed from researcher to researcher or system to system. (TEI Consortium „2.6 The Revision Description“)

Das TEI Konsortium empfiehlt Änderungen an der Datei sowohl für die Transparenz gegenüber Nutzerinnen und Nutzern als auch für die maschinelle Weiternutzung auf diese Art zu verzeichnen. Insgesamt konnten im evaluierten Datensatz 29 Editionen (53%) gefunden werden, die Änderungen in einer *Revision Description* mit Datum, der Angabe zur Autorin bzw. zum Autor und einer Beschreibung der Änderungen, festzuhalten. Das entspricht knapp 38 Prozent aller im Katalog gelisteten *Open Source XML/TEI* Projekte (77), die den Bearbeitungsstand in ihren Forschungsdaten auf diese Art festzuhalten.

Die Beschreibungen in der *Revision Description* fallen sehr unterschiedlich aus, von basal bis sehr ausführlich.<sup>10</sup> Jedes zweite Editionsprojekt, das von der *Revision Description* Gebrauch macht, legt in den editorischen Richtlinien fest, wie die Veränderungen festzuhalten sind. So zum Beispiel im *The Newton Project*. Die 1998 von Rob Iliffe und Scott Mandelbrote ins Leben gerufene Edition aller gedruckten und unveröffentlichten Schriften Isaac Newtons (1642/43–1727) weist in ihren Transkriptions- und Annotationsrichtlinien darauf hin, dass in der *Revision Description* die Bearbeitungsphasen eines Dokuments festgehalten werden:

Keeping an exact account of the dates on which the various stages of processing a document are started and completed is very important because each upgrade of the tagging policy means that previous mark-up also has to be upgraded to bring it into line with the new dispensation. This process can be largely automated, but only if we have a record of what policy was being followed in the first place. If this is not clear, the whole thing has to be proofread again from scratch. It also provides an easy way of verifying which is the latest version of a given document if any confusion arises in the back-up system (Young 2019).

Die Notwendigkeit in der *Revision Description* Entwicklungsschritte zu verzeichnen, wird auch in der Edition *Casebooks: The casebooks of Simon Forman and Richard Napier, 1596–1634* von Lauren Kassell et al. (2019) kommuniziert. Die Edition veröffentlicht die medizinischen Berichte der Astrologen Simon Forman und Richard Napier zu Beginn des 17. Jahrhunderts. Das Pilotprojekt startete 2008, im Mai 2019 wurde die digitale Edition offiziell publiziert.<sup>11</sup> Die *Revision Description* wird hier auch zum Nachweis

<sup>10</sup> Viele Bearbeitungseinträge findet man zum Beispiel in den Daten zur *Edition Briefe und Texte aus dem Intellektuellen Berlin um 1800*. Siehe Baillot.

<sup>11</sup> In der Navigation unter „About us“ > „Releases and events“ befinden sich die Änderungsbeschreibungen zur Edition. Darin der aktuellste Eintrag: Mai 2019 „Official launch of the Casebooks digital edition“.

der Qualitätskontrolle eingesetzt: „Proofreading a file counts as a <change> even if the proofreader has not in fact made any changes to it. The file itself may not have changed but its status has.“ (Young et al. 2019)

Der verbreitete Einsatz der TEI-Versionierung macht deutlich, wie relevant die Nachvollziehbarkeit von Änderungen im editorischen Prozess ist. Die Funktionalität, in den Dateien Änderungen zu kennzeichnen, kommt vor allem Editorinnen und Editoren und dem kollaborativen Projektfortschritt zugute. Eine direkte Verknüpfung der Information aus der *Revision Description* zum Webinterface der Edition konnte bei den untersuchten Editionen nur im Newton Projekt festgestellt werden.<sup>12</sup>

### 2.3 Versionskontrolle

Ein Viertel der Editionsprojekte, die versionieren, geben in ihren Projektdokumentationen an, dass sie ihre Forschungsdaten mithilfe von Versionskontrollsystmen über Plattformen wie *Github* oder *GitLab*<sup>13</sup> verwalten. Ein wichtiger Teil dieser Services liegt darin, dass sie Daten als Plaintext in Repositoryn speichern, automatisch versio-nieren und verwalten, sowie zum Herunterladen anbieten. Diese Serviceplattformen sind entstanden, um Softwareprojekte zu hosten, die über eine Versionskontroll-software (wie z.B. *Git*) verwaltet werden. Versionskontrollsoftware hat einen sehr praktischen Nutzen vor allem für die Softwareentwicklung. Jede Änderungsspeiche-rung wird zur neuen Version mit eindeutiger Kennung, einem Zeitstempel und der Angabe zu Autorin bzw. Autor. Es ist jederzeit möglich, festzustellen, wer wann was geändert hat. Jede Datei hat eine *Version history*, über die es möglich ist, Versionen zu vergleichen. Dieser Vergleich erfolgt zeilenweise und jede Änderung an der Datei, sei es das Hinzufügen von Leerzeichen oder das Ausbessern eines Schreibfehlers, wird von der Software erkannt und automatisch als neue Version gespeichert. Es ist möglich, jeden gespeicherten Textzustand wiederherzustellen. Ein großer zusätzlicher Vorteil von Versionskontrollsoftware wie *Git* besteht darin, dass sie kollaborativ ge-nutzt werden kann, das heißt mehrere User können gleichzeitig am selben Dokument arbeiten (Chacon und Straub 2014). Aus diesem Grund sind webbasierte Versionsver-waltungssysteme auch für die kollaborative Erstellung einer Edition interessant. Eine Technologie, die grundsätzlich zur Erleichterung der Softwareentwicklung eingesetzt wird, wird hier zur Verwaltung der editorischen Prozesse verwendet. In den Editionsprojekten sind diese Systeme kollaborative Arbeitsumgebung und Datenspeicher zugleich. So etwa im Fall der Edition *Les cours d'Antoine Desgodets*.<sup>14</sup> Die Edition setzt

<sup>12</sup> Bei der Einzelansicht der Textdokumente führt der Link „Revision history“ zum entsprechenden TEI/XML-Dokument.

<sup>13</sup> Es gibt zahlreiche Anbieter, aber diese werden hier genannt, weil sie auch in den untersuchten Editionen in Verwendung sind: Github: <https://github.com/>, GitLab: <https://about.gitlab.com/>.

<sup>14</sup> Repozitorium: <https://github.com/desgodets/desgodetsSources>.

sich mit den Vorlesungen des Architekten Antoine Desgodets auseinander, die er im ersten Drittel des 18. Jahrhunderts an der *Académie royale d'architecture* gehalten hat. Die Transkriptionen der Manuskripte werden über das Reppositorium bereitgestellt. Die Herausgeber schreiben im Einleitungstext zur Edition:

Les fichiers XML-TEI des cours sont téléchargeables depuis leurs pages de consultation. L'ensemble des sources du projet sont versionnées et librement téléchargeables depuis notre entrepôt GitHub. (Carvais 2014)

In der *Readme* Datei des Reppositoriums (emchateau 2015) wird die Ordner- und Dateistruktur zum Projekt beschrieben. Was aber weder in der Edition klar wird, noch aus dem *Readme*-Text hervorgeht, ist, wie diese beiden Komponenten in Zusammenhang stehen. Man weiß nicht, ob der Stand der Edition, jenem im Repository entspricht.

Zu diesem Aspekt wird im Editionsprojekt *Inscriptions of Israel/Palaestine* Stellung genommen. Die Edition wurde 2002 veröffentlicht, und es wird beständig daran weitergearbeitet. Mittlerweile werden darin über 10.000 Inschriften aus Israel und Palästina um 500 v. Chr. bis 614 n. Chr. zugänglich gemacht. Seit 2016 werden die Daten in einem Reppositorium über eine Versionskontrollplattform<sup>15</sup> bearbeitet. In der Projektdokumentation zur Edition beschreibt der Herausgeber Michael Satlow den Workflow, wie die Daten aus dem Reppositorium in die Edition aufgenommen werden:

There are two main repositories of our data. One lives on the production machine, a server maintained by the Brown University Library. This is the “working” repository accessed through our website. The second repository is on Github and is publicly accessible for download or use. The two repositories are substantially the same, although new inscriptions and corrections are first added to the Github repository, and then periodically synced to the repository on the production machine (Satlow, „Documentation“).

Demnach wird das Reppositorium als Arbeitsumgebung genutzt, um die Daten auf dem aktuellen Stand zu halten und in Intervallen ins Produktivsystem, also die Edition überführt.

Im Falle des *Shelley Godwin Archive* wird sowohl an den Daten als auch der Software in einem gemeinsamen Reppositorium<sup>16</sup> kollaborativ gearbeitet. Ziel des Archivs ist es, die physisch weit verstreuten Manuskripte von Percy Bysshe Shelley, Mary Wollstonecraft Shelley, William Godwin und Mary Wollstonecraft digital zu vereinen. Die Sammlung wächst beständig dank vieler Editorinnen und Editoren, die über die Nutzung der Versionskontrollplattform, Transkriptionen der Edition hinzufügen.

---

<sup>15</sup> Reppositorium: <https://github.com/Brown-University-Library/iip-texts>.

<sup>16</sup> Reppositorium: <https://github.com/umd-mith/sga>.

Schon seit 2013 dient das Repotorium der kollaborativen Projektentwicklung für das Archiv. Es heißt dazu in der Dokumentation auf *Github*:

The Shelley-Godwin Archive (SGA) uses the Git version control system to manage all files associated with transcriptions created for the project. Since SGA is using Git, we also use Github, a hosted service for Git projects which incorporates many useful project management features, to manage the transcription and encoding workflow (Viglianti 2014)

Leider wird auch hier der Zusammenhang von Repository und Edition nicht erwähnt. Ein Editionsprojekt über eine Plattform mit Versionskontrollsystem zu entwickeln, ist gerade für kollaborative Projekte vorteilhaft, da diese auch Projektmanagementfunktionalitäten, wie die Vergabe von *Issues* und das Definieren von *Milestones* anbieten. Jedoch werden bisher in den jeweiligen editorischen Richtlinien und Projektbeschreibungen die Vorteile, wie kontinuierliches Überarbeiten, kollaboratives Edieren, Nachverfolgen von editorischen Tätigkeiten, die durch den Einsatz von Versionskontrollsystmen möglich sind, gar nicht thematisiert. Die bisher gesetzten Hyperlinks, die von der Edition zum Repotorium führen, zeigen auf keine bestimmte Version einer spezifischen Datei, sondern immer nur auf die Startseite des verwalteten Gesamtprojektes. Die Hürde, als Benutzerin oder Benutzer Änderungen an der Edition in den Dokumenten des Repotoriums nachzu vollziehen, ist groß. Versionierung auf diese Art kommt nur Projektmitarbeiterinnen und -mitarbeitern zugute. Einen Sonderfall dieses Versionierungstyps stellt das Wikibook dar, weil allen Nutzerinnen und Nutzern der Editionsprozess transparent über das Interface zugänglich ist.

### **Sonderfall Wikibook**

Wie alle Seiten von *Wikipedia* werden auch die Seiten eines *Wikibooks*, da sie auf der gleichen Software *MediaWiki* (seit Version 1.5) beruhen, über ein Versionskontrollsystem verwaltet und sie verfügen über eine Revisionsgeschichte (unter dem Reiter: „View history“). Die Revisionsgeschichte in Form einer Tabelle enthält alle Bearbeitungen einer Seite im Wiki. Jede Änderung einer Seite erzeugt eine Änderungszeile, die Informationen über jene Person, die die Bearbeitung vorgenommen hat, den Zeitpunkt, zu dem die Bearbeitung vorgenommen wurde, und einen Verweis auf den neuen Wikitext in der Texttabelle beinhaltet. Elemente der Revisionstabelle bleiben dauerhaft erhalten, es sei denn, die Seite wird gelöscht. Die Edition *Devonshire Manuscript* wird von der *Devonshire Manuscript Group* als *Wikibook* herausgegeben. Das Manuskript ist eine Verssammlung aus dem 16. Jahrhundert, die von drei Frauen zusammengestellt wurde. In der Mehrzahl sind die Verse vom Dichter Sir Thomas Wyatt und die Handschrift ist deshalb eine wichtige Quelle zu seiner Dichtung aber

auch zum Selbstverständnis von Frauen in der literarischen Produktion dieser Zeit (Wikibooks contributors 2018).

Ziel der Edition ist es, die edierten Quellen möglichst breit zu diskutieren und die Rolle des Editionsteams von der alleinigen Autorität über den Text hin zu einer editorischen Textmoderation zu verschieben: „The social edition is a work that brings communities together to engage in conversation around a text formed and reformed through an ongoing, iterative, public editorial process.“ (Wikibooks contributors 2014, „A Note on this Edition“). Einer der Hauptverantwortlichen dieses Projektes ist Ray Siemens, seine Motivation die Edition mithilfe von sozialen Medien zu stützen, ist:

[S]uch tools facilitate a model of textual interaction and intervention that encourage us to see the scholarly text as a process rather than a product, and the initial, primary editor as a facilitator, rather than progenitor, of textual knowledge creation. (2012, 453)

Es geht demnach um den editorischen Prozess als iterative und kollaborative Tätigkeit. Insofern ist es für das Vorankommen des Projektes essentiell, alle Iterationen in Form von Versionen kommentiert verfügbar zu halten. In einem *Wikibook* wird Versionskontrolle ins Frontend gebracht. Alle Benutzerinnen haben die Möglichkeit den editorischen Prozess nachzuvollziehen und sich zu beteiligen.

## 2.4 Versionierung des Gesamtsystems

Dieser Versionierungstyp unterscheidet sich von den bisher besprochenen Typen dadurch, dass Vorgängerversionen der Edition als Gesamtsystem von Daten und Software aufrufbar sind. Wird auf eine veraltete Version Bezug genommen, dann ist gewährleistet, dass dieser Zustand noch immer zur Verfügung steht. Im Gegensatz zu der vorhin erwähnten Versionierung über Versionskontrollsysteme, die vor allem Editorinnen und Editoren in der kollaborativen Zusammenarbeit unterstützt, adressiert diese Form der Versionierung eindeutig die Benutzerinnen und Benutzer einer Edition. In Intervallen werden die Editionen unter neuem Versionsnamen aktualisiert, dieser Vorgang ist vergleichbar mit einem major release eines Softwareproduktes. In den Projekten wird entschieden, wann die Edition aktualisiert und die bisherige Version zur Archivversion gemacht wird.

Unter den analysierten Editionen konnten acht gefunden werden, die diesem Typ zugeordnet werden können (12%). Besonders eindrucksvoll ist die Versionsgeschichte der Edition des *Walt Whitman Archive*. Das Archiv, das Aufzeichnungen und Werke von und über den US-amerikanischen Literaten Walt Whitman sammelt, existiert seit 1996. Während die frühe Arbeit des Archivs Whitmans Gedichtmanuskripte in den Vordergrund stellte, hat es im Laufe der Zeit auch Materialien wie seine Korrespondenz, Notizbücher und Prosamanuskripte in seine Sammlung aufgenommen.

Die Herausgeber Ed Folsom und Kenneth M. Price begründen den Zweck des Archivs folgendermaßen:

Our aim is to produce a scholarly edition of Whitman on the web.[...] Whitman's work was always being revised, was always in flux, and fixed forms of print do not adequately capture his incessant revisions. Moreover, the economics of print publication have led previous editors to privilege one edition or another of Whitman's writings [...]. Our goal is to create a dynamic site that will grow and change over the years. (Ken Price 2002, „About the Archive“)

Diese Aussage wird auf nachhaltige Weise umgesetzt, denn neben einer aktuellen Version, sind noch eine Archivversion bis zum Jahr 2000 und eine weitere bis 2007 vollständig verfügbar. Alle Versionen haben zudem Änderungsbeschreibungen im *WWA Changelog*. Dort werden alle laufenden Änderungen zur Nachverfolgung für Benutzerinnen und Benutzer beschrieben. Erst bei großen Umstellungen der Edition (nach einigen Jahren), wird das Gesamtsystem versioniert. Dann ersetzt eine neue Version das Gesamtbild der alten Version, die weiterhin adressierbar und verfügbar bleibt. Der technische Aufwand einer Versionierung des Gesamtsystems ist groß, denn schließlich müssen die vergangenen Versionen volumnfassend weiterbetrieben werden. Das zu gewährleisten gelingt nur, wenn zum einen sämtliche Ressourcen in einem gemeinsamen System verfügbar sind, also keine Daten über externe Services miteingebracht werden. Zum anderen, wenn beim Softwareentwurf der Edition, was Layout und Funktionalitäten betrifft, auf langlebige, gut zu wartende Software geachtet wird.

### 3 Ergebnis

Die Untersuchung digitaler Editionen hat gezeigt, dass Versionierung in der wissenschaftlichen Praxis von Editorinnen und Editoren sehr oft noch ein Randthema darstellt. Aber zumindest jedes vierte Editionsprojekt aus dem evaluierten Datensatz setzt Formen der Versionierung bereits um und das aus unterschiedlichen Gründen. Editionen werden versioniert, weil sie kollaborativ erarbeitet und editorische Entscheidungen auf diese Art nachvollziehbar werden. Editionen werden aber auch versioniert, weil Zwischenstände von Projekten mit langer Laufzeit veröffentlicht werden sollen, oder weil die Edition in Etappen veröffentlicht werden muss, weil das zu edierende Material zu umfangreich oder die Quellen zu verstreut sind.

Ganz allgemein steht hinter einer Versionierung der verantwortungsbewusste Umgang mit Änderungen. Einen unmittelbaren Mehrwert haben Benutzerinnen und

Benutzer, weil sie sich auf Versionen bzw. mit Datum zitierbare Zustände der Edition beziehen und Prozesse des wissenschaftlichen Arbeitens nachvollziehen können. Vollkommene Transparenz bei der Nachverfolgung von Arbeitsprozessen bieten Serviceplattformen mit Versionskontrollsysteem, die über einen Versionsvergleich auf Zeilenebene den editorischen Prozess im Text freilegen können.

Was die technische Umsetzbarkeit von Versionierung anlangt, existiert in der Form der Änderungsbeschreibung, eine einfache Lösung. Einfach ist es auch, Daten von einem Versionskontrollsysteem auf einer Plattform verwalten zu lassen. Technisch aufwendig wird es erst dann, wenn die Daten und der Code zusammenspielen müssen, also wenn vergangene Versionen des Gesamtsystems wiederherstellbar sein sollen. Wenn zudem externe Daten und Softwarekomponenten Teil der Edition sind, wird der Aufwand, die Edition zu warten und eine lauffähige Altversion davon zu archivieren, sehr groß. Es wäre sinnvoll, in Zukunft Methoden zu entwickeln, die diesen Aufwand erleichtern, allerdings fehlt es momentan noch an *best practice* Modellen in den DH.

Aus den Versionierungstypen lässt sich auch erschließen, für wen vordergründig versioniert wird. Spielt sich Versionierung ausschließlich in den Daten im Backend ab, dann ist das vorrangig für die Workflows der involvierten Editorinnen und Editoren gewinnbringend. Liegen die Daten in einem öffentlichen Projektrepository auf Versionskontrollplattformen, dann hat jeder Datensatz eine Versionsgeschichte und kann maschinell verarbeitet und ausgewertet werden, aber zugleich fällt es externen Benutzerinnen und Benutzern schwer, sich in den Daten der Edition zurechtzufinden. In diesem Fall wird die digitale Edition zu einer abstrakten Informationsressource.<sup>17</sup>

Alle hier besprochenen Editionen präsentieren ihr Wissen über Webinterfaces. Das ist auch der Ort, wo nach Veröffentlichung der Edition, Änderungen kommuniziert werden sollen. Oft sind diese Aufzeichnungen schon im Backend vorhanden und müssten, um breiten wissenschaftlichen Nutzen zu haben, ins Frontend geholt werden. Leitlinien für die Vermittlung dieser Informationen müssen erst entstehen.

Als Ergebnis dieses Beitrags steht keine Bewertung der präsentierten Versionierungstypen, sondern eine Bestandsaufnahme, die offenlegt, wie verbreitet es ist, Editionen zu versionieren. Transparenz und Nachvollziehbarkeit von Änderungen schaffen Vertrauen und das wiederum intensiviert den wissenschaftlichen Austausch, weil über Versionen der Edition stabile Diskussionsgrundlagen geschaffen werden, die der digitalen Edition ihre autoritative Stellung zurückgibt.

---

<sup>17</sup> Patrick Sahle spricht in diesem Zusammenhang von Transmedialisierung der Edition. Die Edition wird unabhängig von ihrer medialen Realisierung (Sahle 2010).

## Literatur

- Ainsworth, Peter und Godfried Croenen, Hrsg. 2013. „The Online Froissart. Version 1.5.“ Zugriff: 1. September 2019. <https://www.dhi.ac.uk/onlinefroissart/index.jsp>.
- Baillot, Anne, Hrsg. „Briefe und Texte aus dem intellektuellen Berlin um 1800.“ Berlin. Humboldt-Universität zu Berlin. Zugriff: 2. April 2020. <http://www.berliner-intellektuelle.eu>.
- Bleier, Roman. 2021. „How to cite this digital edition?“ *Digital Humanities Quarterly* 15 (3). Zugriff: 1. Dezember 2021. <http://digitalhumanities.org/dhq/vol/15/3/000561/000561.html>.
- Bleier, Roman und Sean M. Winslow, Hrsg. 2019. *Versioning Cultural Objects: Digital Approaches*. Norderstedt: Books on Demand.
- Burnley, David und Alison Wiggins, Hrsg. 2004. „Auchinleck Manuscript. Version 1.1.“ Zugriff: 1. September 2019. <https:// auchinleck.nls.uk/>.
- Bürgermeister, Martina. 2019. „Extending Versioning in Collaborative Research.“ In *Versioning Cultural Objects: Digital Approaches*, hg. v. Roman Bleier und Sean M. Winslow. 171-190. Norderstedt: Books on Demand.
- Carvais, Robert, Hrsg. 2014. „Cours de Descodets.“ Zugriff: 2. Mai 2020. <http://www.desgodets.net>.
- . 2014. „L’Edition des Cours de Descodets.“ In *Cours de Descodets*, hg. v. Robert Carvais. Zugriff: 2. April 2020. <http://www.desgodets.net/edition-des-cours>.
- Chacon, Scott und Ben Straub. „Pro Git.“ Zugriff: 2. April 2020. [git-scm.com/book/en/v2/Getting-Started-About-Version-Control](http://git-scm.com/book/en/v2/Getting-Started-About-Version-Control).
- Collins-Sussman, Ben, Brian W. Fitzpatrick und C. Michael Pilato. 2004. *Version Control with Subversion: Next Generation Open Source Version Control*. Sebastopol: O’Reilly.
- Duden 2020. „Versionierung.“ In *Onlinewörterbuch*. Berlin: Dudenverlag. Zugriff: 2. April 2020. <https://www.duden.de/suchen/dudenonline/Versionierung>.
- emchateau. 2015. „desgodetsSources.“ *Github*: <https://github.com/desgodets/desgodetsSources>.
- Fraistat, Neil, Elizabeth Denlinger und Raffaele Viglianti, Hrsg. „The Shelley-Godwin Archive.“ Zugriff: 2. April 2020. <http://shelleygodwinarchive.org>.
- Franzini, Greta, Peter Andorfer und Ksenia Zaytseva. 2016–2020. „Catalogue of Digital Editions: The Web Application.“ Zugriff: 10. März 2020. <https://dig-ed-cat.acdh.oeaw.ac.at>.
- Folsom, Ed und Kenneth M. Price, Hrsg. 1995–2020. „The Walt Whitman Archive.“ Zugriff: 2. April 2020. <https://whitmanarchive.org>.
- Iliffe, Rob und Scott Mandelbrote, Hrsg. „The Newton Project.“ Zugriff: 2. April 2020. <http://www.newtonproject.ox.ac.uk>.
- Kassell, Lauren, Michael Hawkins, Robert Ralley, John Young, Joanne Edge, Janet Yvonne Martin-Portugues. Hrsg. 2019. „‘Casebooks’, The casebooks of Simon Forman and Richard Napier, 1596–1634: a digital edition.“ Zugriff: 2. April 2020. <https://casebooks.lib.cam.ac.uk>.
- Milonas, Ellie, Christopher B. Zeichmann, Birkin James Diana, GaiaLembi, Carlos Rotger, AllenZYoung, mlsatlow, and Adam Bradley. „iip-texts“ In *Github*. Zugriff: 2. April 2020. <https://github.com/Brown-University-Library/iip-texts>.
- Pierazzo, Elena. 2014. „Digital Scholarly Editing: Theorie, Models and Methods.“ *HAL*. Zugriff: 2. April 2020. <http://hal.univ-grenoble-alpes.fr/hal-01182162>.
- Sahle, Patrick. 2016. „What is a Scholarly Digital Edition?“ In *Digital Scholarly Editing: Theories*

- and Practices*, hg. v. Matthew James Driscoll und Elena Pierazzo, 19–39. Cambridge, UK: Open Book Publishers. doi:10.11647/OBP.0095.02.
- . 2010. „Zwischen Mediengebundenheit und Transmedialisierung. Anmerkungen zum Verhältnis von Edition und Medien“ In *Editio* 24: 23–36. doi:10.1515/edit.2010.004.
  - , Hrsg. 2008–2020. „A Catalog of Digital Scholarly Editions. Version 3.0.“ Zugriff: 2. April 2020. <http://www.digitale-edition.de>.
- Santacroce, Ferdinando, Rasmus Askl-Olsson und Jakub Narebski. 2016. *Git Mastering Version Control*. Birmingham: Packt Publishing.
- Satlow, Michael L. Hrsg. 2002–2020. *Inscriptions of Israel/Palestine*. Zugriff: 15. Mai 2020. <http://www.brown.edu/iip>.
- , Hrsg. 2008–2020. „Documentation.“ In *Inscriptions of Israel/Palestine*. Zugriff: 15. Mai 2020. <https://library.brown.edu/iip/about/documentation>.
- Siemens, Ray, the Electronic Textual Cultures Lab Research Group et al. 2012. „Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media“ *Humanities Commons*: 445–61. Zugriff: 15. Mai 2020. <https://hcommons.org/deposits/item/mla:83/>; doi:10.17613/M6KS3G.
- Shelley-Godwin Archive. Datensatz. Zugriff: 15. Mai 2020. <https://github.com/umd-mith/sga>.
- TEI Consortium, Hrsg. 2020. „TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.0.0.“ Zugriff: 2. April 2020. <http://www.tei-c.org/Guidelines/P5/>.
- „2 The TEI Header.“ In *TEI P5. Version 4.0.0*. Zugriff: 2. April 2020. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>.
  - „<revisionsDesc>.“ In *TEI P5. Version 4.0.0*. Zugriff: 2. April 2020. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-revisionDesc.html>.
  - „2.6 The Revision Description.“ In *TEI P5 Version 4.0.0..* Zugriff: 2. April 2020. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD6>.
- Viglianti, Raffaele. 2014. „index.html“ In *Github*. Zugriff: 2. April 2020. <https://github.com/umd-mith/sga/blob/master/docs/guidelines/docs/index.html>.
- Wiggins, Alison. 2004. „Archive of Site updates.“ In: *Auchinleck Manuscript. Version 1.1*, hg. v. David Burnley und Alison Wiggins. Zugriff: 2. April 2020. <https://auchinleck.nls.uk/editorial/archive.html>.
- Wikibooks contributors. 2018. *The Devonshire Manuscript*. Wikibooks. Zugriff: 20. Mai 2020. [https://en.wikibooks.org/w/index.php?title=The\\_Devonshire\\_Manuscript&oldid=3469218](https://en.wikibooks.org/w/index.php?title=The_Devonshire_Manuscript&oldid=3469218).
- . 2014. „The Devonshire Manuscript/A Note on this Edition.“ In *The Devonshire Manuscript*. Wikibooks, the Free Textbook Project. Zugriff: 15. Mai 2020. [https://en.wikibooks.org/w/index.php?title=The\\_Devonshire\\_Manuscript/A\\_Note\\_on\\_this\\_Edition&oldid=2659306](https://en.wikibooks.org/w/index.php?title=The_Devonshire_Manuscript/A_Note_on_this_Edition&oldid=2659306).
- Young, John T. 2019. „Newton Project: Transcription and Tagging Guidelines (Version 7)“ In *The Newton Project* hg. v. Rob Iliffe und Scott Mandelbrote. Zugriff: 2. April 2020. <http://www.newtonproject.ox.ac.uk/about-us/tagging-and-transcription-guidelines>.
- Young, John, Michael Hawkins und Robert Ralley. 2019. „Editorial and tagging guidelines. A Critical Introduction to the Casebooks of Simon Forman and Richard Napier, 1596–1634.“ In ‘Casebooks’, hg. v. Lauren Kassell, Michael Hawkins, Robert Ralley, et al. Zugriff: 2. April 2020. <https://casebooks.lib.cam.ac.uk./about-us/editorial-and-tagging-guidelines>.

## 4 Anhang: Übersichtstabelle Editionen und Versionierungsstrategien

	Edition	Änderungsbeschreibung	Revision Description	Versionskontrollsysteme	Versionierung des Gesamtsystems
1	The Thomas Gray Archive				
2	Bérardier de Bataut, Essai sur le récit, Entretiens sur la manière de raconter				
3	Becerro Galicano of San Millán de la Cogolla				
4	Euripides Scholia				
5	The Hooke Folio Online				
6	The Workdiaries of Robert Boyle				
7	The Holinshed Project				
8	Corpus of the Inscriptions of Campā				
9	Henrik Ibsens Skrifter				
10	Mark Twain Project				
11	The Auchinleck manuscript				
12	The Casebooks Project				
13	The Gospel according to St. John				
14	The Proceedings of the Old Bailey				
15	Digital Mishnah				
16	The Journals of the Lewis and Clark Expedition				
17	The Vergil project. Resources for Students, Teachers, and Readers of Vergil				
18	Inscriptiones Graecae in Croatia Repertae (IGCR)				
19	Carl-Maria-von-Weber-Gesamtausgabe				
20	Livingstone's 1871 Field Diary				
21	Inscriptions of Aphrodisias Project				
22	Inscriptions of Roman Tripolitania				
23	Edition der falschen Kapitularien des Benedictus Levita				
24	William of Ockham: Dialogus				
25	The Walt Whitman Archive				

	Edition	Änderungsbeschreibung	Revision Description	Versionskontrollsysteme	Versionierung des Gesamtsystems
26	Schenker Documents Online				
27	Inscriptions of Israel/Palestine				
28	The Shelley-Godwin Archive				
29	The Online Froissart				
30	The Newton Project				
31	Livingstone Online				
32	A Social Edition of the Devonshire MS (BL Add 17,492)				
33	Les Bibliothèques Virtuelles Humanistes				
34	Søren Kierkegaards Skrifter				
35	The Planctus for William Longsword				
36	William Golder Electronic Edition				
37	Les Cours d'Antoine Desgodets				
38	Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg				
39	Digital Schomburg Editions: African American Women Writers of the 19th century				
40	Faust Edition				
41	Theodor Fontane: Notizbücher				
42	Schopenhauer's Library: Annotations and marks in his Spanish books				
43	Project Tarsian: First Corinthians in Vaticanus Arabicus 13				
44	The Wandering Jew's Chronicle				
45	Darwin Online: Online Variorum of Darwin's Origin of Species				
46	Dámos Database of Mycenaeans at Oslo				
47	Streuvels en zijn uitgevers – brieven				
48	Briefwechsel Sauer-Seuffert				
49	William F. Cody Archive				
50	Digital Athenaeus				

	Edition	Änderungsbeschreibung	Revision Description	Versionskontrollsysteme	Versionierung des Gesamtsystems
51	Magistri Guillelmi Autissiodorensis Summa de officiis ecclesiasticis				
52	Edition Humboldt Digital				
53	Digitale Edition der Briefe Erdmuthe Benignas von Reuß-Ebersdorf (1670-1732)				
54	Kazinczy Ferenc összes művei				
55	Csokonai Vitéz Mihály összes művei				
56	Gyöngyössi János művei				
57	Berzenyi Dániel levelezése				
58	Magyar írók levelezése				
59	A Ráday-család levelezése				
60	DER STURM. Digitale Quellenedition zur Geschichte der internationalen Avantgarde				
61	P.S. Post Scriptum: A Digital Archive of Ordinary Writing				
62	Armer Heinrich Digital				
63	Hermann Bahr – Arthur Schnitzler: Briefwechsel, Aufzeichnungen, Dokumente (1891–1931)				
64	Tagebücher von Andreas Okopenko. Digitale Edition				
65	Schnitzler Critical Digital Edition				
		34	29	16	8

## **Projektberichte**

---

**Project reports**



# Die Chronik des Matthäus von Edessa (Matt<sup>c</sup>ēos Urhayec'i) Online

Tara L. Andrews

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Tara L. Andrews (Projektleitung), Schiwa Aliabadi-Pongratz, Tatev Atayan, Sascha Kaufmann, Anahit Safaryan, Christian Veigl
<b>Institutionen:</b>	Universität Wien – Institut für Geschichte, Universität Bern – Walter-Benjamin-Kolleg
<b>Fördergeber:</b>	Schweizerischer Nationalfonds (#159433)
<b>Website:</b>	<a href="https://editions/byzantini.st/ChronicleME/">https://editions.byzantini.st/ChronicleME/</a>

Die *Chronik des Matthäus von Edessa* wurde im 12. Jahrhundert von einem armenischen Augenzeugen des Ersten Kreuzzugs sowie der Gründung der Grafschaft Edessa verfasst. Trotz ihrer Bedeutung als Quelle für die Geschichte von Byzanz, jener der Kreuzzüge und des christlichen Nahen Ostens (Andrews 2016) wurde die Chronik bis jetzt noch nie zur Gänze kritisch ediert. Die frühere *editio maior* (Vałarşapat 1898) basiert hauptsächlich auf einer einzigen Handschrift. Das derzeitige Projekt läuft seit 2015; Ergebnisse werden sukzessive auf der Projektwebseite veröffentlicht, bis der Text vollständig ediert, übersetzt und annotiert vorliegt.

Unsere digitale Edition der *Chronik* hat zum Ziel, den Lesenden nicht nur den erstellten Text sowie eine Darstellung seiner Varianten zu bieten, sondern auch die Möglichkeit, in den historischen Inhalt des Textes und die Geschichte seiner späteren Rezeption einzusteigen. Dazu verfolgen wir die allgemeinere Absicht, die Möglichkeiten von Textkritik im digitalen Medium auszuloten. Ein zentrales methodologisches Anliegen des Projekts besteht somit in der Auseinandersetzung mit dem Prozess des kritischen Edierens selbst – Entzifferung und Transkription, Kollation, stemmatische Analyse, Erstellung von Text und Kommentar; dies erfolgt durch die Beschreibung der einzelnen Schritte im Rahmen eines computerbasierten Objektmodells.

Unser Modell basiert auf einer Darstellung der unterschiedlichen Textzeugen und ihrer Übereinstimmungen (der *collation graph*; vgl. Schmidt und Colomb 2009; Andrews und Macé 2013). Diese Darstellung ist in einer Graphdatenbank gespeichert. Aufbauend auf diesem Basismodell können wir die Kollation durchsehen und verfeinern; abweichende Lesarten und ihre Beziehungen untereinander erfassen; bestimmte Varianten als *bedeutsam* zum Zweck der stammatischen Analyse markieren; den Editionstext aus den verfügbaren Lesarten zusammenstellen und, wenn nötig, emendieren. Das Datenmodell dient auch als Basis für verschiedene Textansichten – zum

Beispiel können einzelne Zeugen entweder in diplomatischer oder normalisierter Form dargestellt werden; die kritische Edition kann in ein TEI-XML-Format oder in ein tabellarisches Format für weitere stemmatische Analysen exportiert werden; eine Liste der Varianten kann erzeugt werden; der Kollationsgraph selbst kann auf jeder ausgewählten Ebene der Normalisierung betrachtet werden.

Die Graphdatenbank speichert auch Annotationen zum Text. Diese können satzweise Übersetzungen, Kennzeichnungen der im Text genannten Personen bzw. Orte, Informationen über Daten und die relative Reihenfolge der Ereignisse sowie Verweise auf weitere Quellen beinhalten. Die Annotationsstruktur wurde so entwickelt, dass sie mit dem CIDOC-CRM-Standard und ausdrückbaren Linked Open Data-Formaten wie RDF-XML kompatibel ist.

Unser Ziel ist es schließlich, eine Edition des Textes vorzulegen, die gänzlich digital ist: nicht nur dahingehend, dass sie in einem digitalen Medium veröffentlicht wird, sondern auch in dem Sinne, dass sie von Anfang an auf einem computerbasierten Modell beruht, das philologische Expertise inkorporiert; und weiters auch in dem Sinne, dass sie die Editionsdaten als maschinenlesbare APIs sowie für Menschenlesbare Visualisierungen zur Verfügung stellt.

## Literatur

- Andrews, Tara L. 2016. *Matt'ēos Urhayec'i and His Chronicle: History as Apocalypse in a Crossroads of Cultures*. Leiden; Boston: Brill.
- Andrews, Tara L. und Caroline Macé. 2013. „Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata.“ *Literary and Linguistic Computing* 28 (4): 504–21. doi:10.1093/lrc/fqt032.
- Matthew of Edessa (Matt'ēos Urhayec'i). 1898. *Žamanakagrut'iwn*. Hg. v. Mambrē Mēlik'-Adamean und Nersēs Tēr-Mik'ayēlean. Vałarşapat.
- Schmidt, Desmond und Robert Colomb. 2009. „A Data Structure for Representing Multi-Version Texts Online.“ *International Journal of Human-Computer Studies* 67: 497–514. doi:10.1016/j.ijhcs.2009.02.001.

# Der Regensburger Reichstag 1576 – digital

Roman Bleier, Eva Ortlieb, Florian Zeilinger

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Roman Bleier, Elisabeth Brantner, Magdalena Ebner, Gabriele Haug-Moritz (Projektleitung), Josef Leeb, Christiane Neerfeld, Eva Ortlieb, Constanze Rammer, Georg Vogeler (Projektleitung), Florian Zeilinger
<b>Institutionen:</b>	Historische Kommission bei der Bayerischen Akademie der Wissenschaften (HiKo), Universität Graz – Zentrum für Informationsmodellierung (ZIM)
<b>Fördergeber:</b>	DFG (386773508), FWF Der Wissenschaftsfonds (I 3446)
<b>Website:</b>	<a href="http://gams.uni-graz.at/context:rta1576">http://gams.uni-graz.at/context:rta1576</a>

Im Frühsommer 1576 versammelten sich in Regensburg Kaiser Maximilian II. und seine Räte sowie rund 200 Stände des sog. Heiligen Römischen Reichs Deutscher Nation bzw. deren Vertreter. Bis Mitte Oktober desselben Jahres verhandelten sie über reichspolitische Fragen (Steuer, Religion, Justiz) und Außenpolitik (Frankreich, Polen, Russland). Ebenfalls präsent waren Gesandte auswärtiger Mächte sowie zahlreiche Bittsteller.

Mit dem Reichstag (Kurzüberblick: Liebmann 2009) hatte sich das Reich ein Forum geschaffen, um mittels kommunikativer Prozesse zu gemeinsamen Entscheidungen zu gelangen. Die Versammlungen stehen damit im Kontext europaweiter vormoderner Ständeversammlungen bzw. Parlamente (Hébert 2014).

Bereits im 19. Jahrhundert erkannte man die Bedeutung der Überlieferung dieser Versammlungen. In den vier Abteilungen der Editionsreihe “Deutsche Reichstagsakten” wurden inzwischen zahlreiche Bände veröffentlicht, von denen einige auch digitalisiert vorliegen.<sup>1</sup> Zwei Einschränkungen lassen sich dennoch nicht völlig ausgleichen. Die auf das gedruckte Buch zugeschnittene Erschließung erweist sich als zu wenig flexibel, um das Quellenmaterial für vielfältige Fragestellungen verwendbar zu machen. Die weitgehende Beschränkung des editorischen Blicks auf die Inhalte der Beratungen ließ andere Aspekte in den Hintergrund treten. Beides betrifft insbesondere das in den letzten Jahren verstärkte Interesse an Interaktion, Kommunikation und Verfahren auf frühneuzeitlichen Ständeversammlungen.

<sup>1</sup> Bandübersicht: <https://www.historischekommission-muenchen.de/publikationen.html>. Digitalisiert: <https://www.historischekommission-muenchen.de/digitale-publikationen/reichsversammlungen-und-reichstage-1376-1662.html>.

Die bisher geltenden Editionsrichtlinien für die Reichstage ab der zweiten Hälfte des 16. Jahrhunderts wurden 1988 festgelegt (Lanzinner 1988, 65–80). Sie trugen dem Umfang des Aktenmaterials Rechnung, indem sie das Editionskonzept formal, zeitlich und inhaltlich so eng wie möglich definierten (Haug-Moritz 2020, 20; Lanzinner 1988, 66–69). Die einzelnen Editionen listeten die eingesehenen Archivbestände und Archivalien auf; nur wenige, nämlich maximal fünf der potenziell mehr Abschriften einzelner Stücke wurden explizit genannt. Dem seinerzeit im Zentrum stehenden Erkenntnisinteresse an ständischen Versammlungen Rechnung tragend dokumentieren sie, was den Reichstag in seiner Gesamtheit betrifft (v.a. proponierte Verhandlungspunkte und -ergebnisse bzw. vor der Mainzer Kanzlei Präsentiertes), dazu die Protokolle einzelner Kurien über deren interne Beratungen. Während vieles teils im Volltext transkribiert und kollationiert, teils rejestriert wurde, bot man die Supplikationen einzelner Bittsteller lediglich im Aktenreferat dar (Lanzinner 1988, 27–32; 86–87; 110; 113).

Mit der gegenständlichen digitalen Edition, finanziert durch die *DFG* und den *FWF*, beschreitet die *HiKo* in Kooperation mit dem *ZIM* an der Universität Graz nun neue Wege, um den Benutzer/innen vielfältigere und zugleich einfachere Arbeiten mit den Reichstagsakten zu ermöglichen. Die Edition ruht dabei auf drei Säulen:

(1) Erstmals wird eine Datenbank mit der archivalischen Dokumentation der Überlieferung (AD) angelegt, welche die entsprechenden zentralen, zu dieser Zeit bereits stark angewachsenen Aktenbestände auf Stückebene verzeichnet und die Auswahl der Quellen für die Transkription transparenter macht. Verzeichnet werden dabei rund 10.000 Stücke aus 34 Archiven. Zu einigen Archivalien (Supplikationen) werden in Kooperation mit dem Österreichischen Staatsarchiv auch Bilder zur Verfügung gestellt.

(2) Daneben wird eine digitale Edition erstellt, wodurch weit mehr Volltexte dargeboten werden können als zuvor. Insgesamt wird die Edition Material im Umfang von rund 4.000 Seiten enthalten. Protokolle werden dabei integral im Volltext transkribiert und in der AD tagesgenau erfasst. Auch Berichte einzelner Reichstagsgesandten an ihre Herren werden erstmals transkribiert.

(3) Verbunden werden die ersten beiden Säulen der Edition durch eine Datenbank, die auf Grundlage eines projektspezifischen Modells mit Daten aus der AD und den edierten Texten gefüllt wird. Dabei liegt der editorische Fokus, Forschungen zum vormodernen Parlamentarismus folgend, auf dem für Reichstage zentralen Aspekt des Beratens bzw. der Interaktion und Kommunikation zwischen den politischen Akteuren (Haug-Moritz 2021).

Die Langzeitarchivierung der Projektdaten und die Umsetzung der digitale Edition erfolgt im Geisteswissenschaftlichen Assetmanagement System (GAMS) der Universität Graz. Die AD und die edierten Texte werden in TEI repräsentiert, wobei beim Datenmodell der AD eine Anbindung an die Encoded Archival Description (EAD)

geschaffen wurde und beim Datenmodell für die Edition die bisherigen editorischen Richtlinien und die Praxis der Editionsbände der Reichstagsakten berücksichtigt wurden. Das kommunikations- und interaktionszentrierte Datenmodell ist einerseits in der Auszeichnung von Personen, Orten, Daten und Themen und der Anbindung an entsprechende Normdaten (GND, Geonames) realisiert, andererseits durch die Zuordnung der Entitäten zu einer im Projekt entwickelten Domänenontologie von Kommunikation in vormodernen parlamentarischen Versammlungen, die wiederum an die Top-Level-Ontology CIDOC CRM (E5 Event) anknüpft. Die zusätzliche TEI-Anreicherung erfolgt mithilfe eines kontrollierten Vokabulars im TEI Attribut @ana. Basierend auf der Ontologie werden Daten aus dem TEI extrahiert und als RDF in einen Triplestore gespeichert. Der Triplestore ist die bereits oben erwähnte Datenbank, die, als dritte Säule der Edition, eine wichtige Erschließungsfunktion hat. Dieses Verfahren kommt auch bei anderen Projekten am ZIM zum Einsatz (Pollin und Vogeler 2014; Pollin 2019) und realisiert, durch die Verwendung von RDF, die Anknüpfung an CIDOC CRM und entsprechende Normdaten, die Anbindung der Editionsdaten an das Web of Data. Georg Vogeler hat diese Art von Editionen, die sich auf die Inhaltsebene der Texte konzentrieren, diese über eine Datenbank erschließen und für komplexe Fragestellungen aufbereiten, Assertive Editions genannt (Vogeler 2019). Das Projekt *Regensburger Reichstag 1576 – digital* strebt an, eine solche Edition umzusetzen und damit neue Forschungsperspektiven zu eröffnen.

## Literatur

- Angermeier, Heinz. 1988. Vorwort des Abteilungsleiters. In *Der Reichstag zu Speyer 1570*, hg. v. Maximilian Lanzinner, 19–25. Göttingen: Vandenhoeck & Ruprecht.
- Bleier, Roman, Florian Zeilinger und Georg Vogeler. 2022. „From Early Modern Deliberation to the Semantic Web: Annotating Communications in the Records of the Imperial Diet of 1576.“ In *Digital Parliamentary Data in Action 2022. Proceedings of the Digital Parliamentary Data in Action (DiPaDA 2022) Workshop co-located with 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHB 2022)*, hg. v. Matti La Mela, Fredrik Norén, Eero Hyvönen, 86–100. Zugriff: 17. Juni 2022. <http://ceur-ws.org/Vol-3133/paper06.pdf>.
- Encoded Archival Description. Zugriff: 17. Januar 2022. <https://www.loc.gov/ead>.
- Haug-Moritz, Gabriele. 2020. *Ständeversammlungen digital edieren. Ein neues Editionskonzept für den Reichstag (RT) 1576(–1662): Grundlagen, editorische Konsequenzen, praktische Umsetzung*. (Arbeitspapier als Diskussionsgrundlage). Zugriff: 17. Januar 2022. [https://static.uni-graz.at/fileadmin/projekte/reichstagsakten-1576/Homepage/Materialien\\_Workshop\\_2\\_Editionskonzept.pdf](https://static.uni-graz.at/fileadmin/projekte/reichstagsakten-1576/Homepage/Materialien_Workshop_2_Editionskonzept.pdf).
- . 2021. „Deliberieren. Zur ständisch-parlamentarischen Beratungskultur im Lateineuropa des 16. Jahrhunderts.“ *Historisches Jahrbuch* 141: 114–55.
- Hébert, Michel. 2014. *Parlementer. Assemblées représentatives et échange politique en Europe occidentale à la fin du Moyen Age*. Paris: Éd. deBoccard.

- Lanzinner, Maximilian, Hg. 1988. *Der Reichstag zu Speyer 1570*. Göttingen: Vandenhoeck & Ruprecht.
- Liebmann, Edgar. 2009. „Reichstag.“ In *Enzyklopädie der Neuzeit Bd. 10*, hg. v. Friedrich Jaeger, 948–53. Stuttgart: Metzler.
- Pollin, Christopher. 2019. „Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology.“ In *RODBH 2019, Proceedings of the Doctoral Symposium on Research on Online Databases in History*, hg. v. Thomas Riechert, Francesco Beretta und George Bruseker, 7–14. Zugriff: 17. Januar 2022. <http://ceur-ws.org/Vol-2532>.
- Vogeler, Georg. 2019. „The ‚assertive edition‘. On the consequences of digital methods in scholarly editing for historians.“ *International Journal of Digital Humanities* 1: 309–22. Zugriff: 17. Januar 2022. <https://link.springer.com/article/10.1007/s42803-019-00025-5>.

# CoReMA – Cooking Recipes of the Middle Ages

Astrid Böhm, Julia Eibinger, Helmut W. Klug, Christian Steiner

<b>Team members:</b>	Helmut Klug (Projektleitung AT), Astrid Böhm, Julia Eibinger, Elisabeth Raunig, Christian Steiner; Bruno Laurioux (Projektleitung FR), Denise Ardesi, Corentin Poirier
<b>Institutions:</b>	University of Graz – Institute Centre for Information Modelling – Austrian Centre for Digital Humanities, Institute for German Studies; University of Tours – Département d’histoire, Laboratoire CESR (Centre d’Etudes Supérieures de la Renaissance)
<b>Funding body:</b>	FWF Der Wissenschaftsfonds (I 3614), ANR Agence Nationale Recherche (ANR-17-CE27-0019-01)
<b>Website:</b>	<a href="https://gams.uni-graz.at/corema">https://gams.uni-graz.at/corema</a>

The international project CoReMa seeks to conduct intercultural research into medieval recipes and their interrelations using an interdisciplinary approach. The hand-written recipe traditions of France and the German-speaking countries up to 1500 AD are examined in terms of their origin, their relationship to each other and their migration through Europe. This corpus includes more than 80 manuscripts and about 800 recipes. The texts in question have been handed down in Early New High German, Medieval Latin and Middle French. These multilingual texts are prepared in line with modern standards and analysed using current quantitative and qualitative research methods. To enable computer-aided analysis, the recipe collections, their contained texts and metadata are modelled in TEI/XML (digital transcription and edition) and examined via Semantic Web technologies (digital annotation and data visualisation). The data are fed into a long-term archiving infrastructure (GAMS, Institute Centre for Information Modelling) in which it can be further researched. All recipes are enriched using vocabularies for ingredients, cooking processes and kitchen tools as well as for metadata relevant to history and culture (e.g. religious, cultural or medicinal aspects). Based on this information the project is able to unearth competing or diverging eating habits across language borders, text migration as well as mutual influence of neighbouring countries on their respective cuisines. NLP methods are used for analysis of the German texts regarding historical language stages to examine text relations in the written records.

Workflows and data acquired during the project will be made available for further use in line with the Open Science concepts and FAIR-principles: transcription workflow and transcription principles (theory and practice, in cooperation with KONDE)

can be fully re-used. The annotation vocabularies (ingredients, dishes, kitchen tools, instructions) will be made available including all assigned semantic *Wikidata*-concepts and thus form an essential base for further research in the field of culinary history. Available concepts in *Wikidata*<sup>1</sup> are checked and enriched with additional data (e.g. links of relevant ontologies, for example FoodOn<sup>2</sup> or SNOMED<sup>3</sup>) where applicable. Concepts not yet available are created. In addition to practical considerations, the main aim of using *Wikidata* is to make the data acquired in the project easily available to the community and to enable further use of this data. In addition, the annotation scripts (Python and XSLT) for the transfer of annotation vocabularies to TEI/XML will be available for download. Text data will be offered in plain text for further use by NLP tools.

The recipe texts are uniformly recorded by a hyperdiplomatic re-transcription of the historical sources and will also be available for further use as TEI/XML. The transcription of the source text not only notes the difference in character inventory, but all text-structuring elements. The entire character inventory is recorded in a character description in line with TEI guidelines. The description is based on the theoretical results for describing characters from the *DigiPal* project<sup>4</sup> and also uses the character identifiers of the Medieval Unicode Font Initiative.<sup>5</sup>

The produced data are not only the starting point for research questions of the project, but form a solid base for further research in the fields of German Studies/Linguistics, Palaeography, etc. Furthermore, the CoReMA project will provide a model for integration of additional texts in the research environment. The project aims to provide (technical) impulses for all disciplines concerned with medieval and early modern history, culinary history, Digital Scholarly Edition and Digital Humanities.

---

<sup>1</sup> <https://www.wikidata.org>.

<sup>2</sup> <http://foodon.org>.

<sup>3</sup> <https://browser.ihtsdotools.org>.

<sup>4</sup> Describing Handwriting I-VII; <http://www.digipal.eu/blog>.

<sup>5</sup> <https://folk.uib.no/hnooh/mufi>.

# MUSIL ONLINE – interdiskursiver Kommentar

Artur R. Boelerl

**Mitarbeiterinnen und Mitarbeiter:** Anke Bosse, Artur R. Boelerl, Walter Fanta, Franziska Mader

**Institutionen:** Universität Klagenfurt – Robert-Musil-Institut für Literaturforschung, Kärntner Literaturarchiv (– RMI/KLA), Österreichische Nationalbibliothek

**Fördergeber:** FWF Der Wissenschaftsfonds (P 30028-G24)

**Website:** <http://musilonline.at/>

Das Projekt sondiert Möglichkeiten zur Einbettung eines interdiskursiven Kommentars ins Internetportal *MUSIL ONLINE* auf der Basis von XML/TEI. Dazu werden sowohl die Strukturen eines solchen Kommentars konzipiert als auch Modellkommentare erarbeitet und die Umsetzung an ausgewählten Texten aus dem umfangreichen Werk Musils erprobt. Die so gewonnenen Erkenntnisse werden in der Folge auf das gesamte Portal ausgedehnt und die Online-Kommentierung vorbereitet.

Mit diesem doppelten Ziel – einerseits generell eine interdiskursive Kommentarstruktur zu entwickeln und andererseits dieser Struktur entsprechende Modellkommentare zu den Werken Musils zu bieten – verfolgt das Projekt nicht nur das große Desiderat der gegenwärtigen Musil-Forschung, sondern auch ein brennendes Desiderat der Editionswissenschaft, die im Bereich der Kommentarkonzeption Orientierung sucht. Das Werk Robert Musils bietet dafür den perfekten Lackmustest. Seine Schriften gelten als eine der wichtigsten literarischen Interpretationen der europäischen Gesellschaft in der ersten Hälfte des 20. Jahrhunderts, der verschiedenen Prozesse ihres Fortschritts, ihrer Krise und ihres Niedergangs. Als Teil des immateriellen Kulturerbes in Österreich und damit auch des kollektiven globalen Gedächtnisses wird es auf *MUSIL ONLINE* offen zugänglich gemacht.

Dieser Rahmen erfordert eine kohärente Kommentierung, die den vielfältigen Bedeutungsebenen des Werkes Aufmerksamkeit zollt und diese in einen Dialog miteinander bringt, indem er ihre jeweiligen wissenschaftlichen, philosophischen, kulturellen und politischen Hintergründe aufzeigt. Dazu bedarf es einer neuen Form des interdiskursiven Kommentars in einer auf Interaktivität angelegten, dynamischen digitalen Umgebung, so dass die Benutzerinnen und Benutzer ihren je individuellen Zugang zu Musil finden können, indem sie die verschiedenen Aspekte der Intertextualität des Werks online nach ihrem Erkenntnisinteresse abrufen können. Deshalb werden

zunächst die vorhandenen Anmerkungen und Kommentare zu Musils verschiedenen Schriften, ihr einschlägiger Beitrag zum Verständnis der Absichten des Autors und seiner Rolle im Kontext seiner Zeit untersucht. Daneben wird die textgenetisch-biographisch orientierte Kommentarstruktur der Klagenfurter Ausgabe der Werke Musils (2009) digitalisiert und in Form eines *Textgenetischen Dossiers* (TGD) in die Online-Präsentation eingebettet. Gleichzeitig wird der Charakter eines Kommentars zu Musil von Grund auf neu und unter Berücksichtigung der technischen Möglichkeiten einer Online-Präsentation entwickelt: So entsteht auf der konzeptuellen Grundlage der von Dieter Henrich et al. entwickelten Konstellationsforschung, als Verfahren angepasst an die spezifischen Anforderungen eines literarischen Textkorpus und einer so vielschichtigen und komplexen Epoche wie der sog. Musil-Zeit zwischen 1900 und 1940, in formaler Analogie zum TGD ein *Interdiskursives Dossier* (IDD).

Das IDD bildet die Grundlage des Bereichs *Verstehen* (=interdiskursiver Kommentar) und stellt ein Scharnier zwischen den Bereichen *Lesen* (=Text) und *Schauen* (=Dokument) dar, fungiert aber zugleich als Einstieg *sui generis* in das Onlineportal. Erreicht wird diese Funktionalität durch die Auszeichnung nach XML/TEI aus der Referenzliteratur (in den zunächst zeitlichen, damit aber indirekt auch thematisch verschiedenen Perspektiven *bei* Musil, die den Diskursraum des Autors umfasst und dessen Referenztexte erschließt, sowie *über* Musil, die den Diskurshorizont der Forschungsliteratur abbildet) gewonnenen Elementen als Termini (*terms*) mit unterschiedlicher Gewichtung (*term types*, z. B. *catchword*, *subject*, *discipline*, *discourse*). So erwächst ein stets textbasiertes Bezugsgeflecht, das sich nutzerspezifisch verschiedenen Bedürfnissen intuitiv anpasst.

## Literatur

- Boelderl, Artur R. 2018. „Vom Livre irréalisé zum Texte hyperréalisé? Ein Abriss der Fragestellungen, Problemfelder und Lösungsansätze im Zusammenhang mit der Entwicklung von Kommentarstrukturen und Modellkommentaren für eine interaktive Kommentierung der Schriften Robert Musils im Internetportal MUSIL ONLINE.“ *Zeitschrift für digitale Geisteswissenschaften. Sonderband 2: Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, hg. v. Roland S. Kamzelak und Timo Steyer. Zugriff: 17. Januar 2020. [http://www.zfdg.de/sb002\\_010](http://www.zfdg.de/sb002_010).
- Fanta, Walter, und Artur R. Boelderl. 2020. „MUSIL ONLINE – Vorüberlegungen zur Kommentierung.“ In *Annotieren, Kommentieren, Erläutern. Aspekte des Medienwandels*, hg. v. Wolfgang Lukas und Elke Richter. Berlin-Boston: de Gruyter, 147–57.
- Fanta, Walter. 2019. „Musil online total.“ In *Forschungsdesign 4.0*, hg. v. Jens Klingner, et al. Open-Access-Publikation des Instituts für Sächsische Geschichte und Volkskunde Dresden, 149–79. Dresden: Institut für Sächsische Geschichte und Volkskunde. doi:10.25366/2019.04.

# Intertextuality in the Legal Papers of Karl Kraus: A Scholarly Digital Edition

Ingo Börner, Vanessa Hanneschläger, Isabel Langkabel, Katharina Prager

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Katharina Prager (Projektleitung), Ingo Börner, Vanessa Hanneschläger, Isabel Langkabel
<b>Institutionen:</b>	Ludwig Boltzmann Institut für Digital History, ACDH-CH ÖAW, Wienbibliothek im Rathaus
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds (P 31138-G30)

Angeregt durch die Rechtsreformen der österreichischen Ersten Republik – insbesondere des Presserechts – trug Karl Kraus (1874–1936) seinen seit 1899 über seine Zeitschrift *Die Fackel* ausgefochtenen ‚Kampf ums Recht‘ nun auch regelmäßig in die Gerichtssäle. Der Wiener Rechtsanwalt Oskar Samek (1889–1959) wurde hierbei ab 1922 sein wichtigster Verbündeter und vertrat Kraus in der (außer-)gerichtlichen Regelung von Medien-, Ehrenbeleidigungs-, Urheberrechtsdelikten u. v. m.

Samek ist es zu verdanken, dass sich die Akten, die Kraus’ Prozesse dokumentieren, erhalten haben. Er nahm diese Unterlagen auf seiner Flucht vor den Nationalsozialisten mit sich ins Exil nach New York, von wo sie nach Sameks Tod 1959 an die Wienbibliothek im Rathaus zurück kamen. Dort werden sie seither verwahrt und wurden im Zuge der Reorganisation des Karl Kraus-Archivs (seit 2012) neu geordnet, verlistet und gescannt. Die Digitalisate des etwa 4000 Blatt starken Bestands sind seither über die Plattform *Karl Kraus Online* und in der *digitalen Wienbibliothek* verfügbar.

Bereits in den 1990er-Jahren wurden die Rechtsakten der Kanzlei Oskar Samek von Hermann Böhm ediert. Diese vier Bände dienen der digitalen Edition als Textgrundlage. Dazu wurde die Vorlage mit der Texterkennungssoftware *Tesseract* eingelesen und entsprechend den *TEI Guidelines* codiert.

Zur weiteren Bearbeitung der Dateien wird eine Applikation für die XML-Datenbank *eXist-db* entwickelt, mit der die Dateien für die eigentliche digitale Edition erzeugt und weiter angereichert werden. In diesem Bearbeitungsschritt erfolgt zunächst die Verknüpfung der Dokumente mit den über eine IIIF-Schnittstelle seitens der Wienbibliothek bereitgestellten digitalen Faksimiles. In Böhms Edition nicht oder nur teilweise aufgenommene Dokumente werden über die Plattform *Transkribus* erschlossen und anschließend in die Datenbank zur Weiterbearbeitung eingespielt. Die Einbindung weiterer Materialien aus der Plattform *AustriaN Newspapers Online* (ANNO) mittels IIIF-Schnittstelle ist ebenfalls vorgesehen.

Im Anschluss an die Kollationierung werden die Dokumente auch inhaltlich entsprechend den *TEI Guidelines* ediert. Dabei gibt es zwei große Herausforderungen: Erstens muss ein Weg gefunden werden, bereits bestehende Modellierungen bzw. Ordnungslogiken dieses Bestands (wie in der Böhm-Edition und in der Bestandsliste der Wienbibliothek vorgefunden, aber auch wie in *Karl Kraus Online* als Grundlage eines anti/biographischen Zugangs angelegt) wiederzugeben und in TEI abzubilden. Andererseits ist die Entscheidung zu treffen, welche über den Text hinausgehenden Elemente der Edition (Entitäten wie Personen, Institutionen und Orte, aber auch Klassifikationen von Dokumententypen und juristischen Vorgängen) in TEI abgebildet werden und für welche sich andere Datenorganisationsformate besser eignen. So ist etwa vorgesehen, Taxonomien im Format des *Simple Knowledge Organisation System* (SKOS) abzubilden und in die Edition einzubinden. Ebenso sollen Personen und Institutionen in die auf APIS basierende Datenbank *Personen der Moderne Base* (PMB) eingepflegt und von dort aus eingebunden werden.

Neben diesen bestands- bzw. editionsinhärenten Aspekten ist auch die Verknüpfung bzw. der Vergleich mit externen Textressourcen ein zentrales Anliegen. Um der Frage der Intertextualität nachgehen zu können, werden bis zum Projektende 2021 die edierten Unterlagen einerseits mit den historischen Gesetzestexten, die im *Rechtsinformationssystem des Bundes* (RIS) verfügbar sind, in Beziehung gesetzt und andererseits mit Kraus' eigenem Schreiben über die so dokumentierten Prozesse in seiner Zeitschrift *Die Fackel*, die ebenfalls digital im Volltext vorliegt, verglichen.

## Literatur

- Austrian Academy Corpus AAC. Die Fackel Gate. Zugriff: 17. Januar 2020. <http://corpus1.aac.ac.at/fackel>.
- Böhm, Hermann, Hg. 1995–1997. *Karl Kraus contra ... Die Prozeßakten der Kanzlei Oskar Samek in der Wiener Stadt- und Landesbibliothek*, Band 1–4. Wien: Wiener Stadt- und Landesbibliothek.
- Hannesschläger, Vanessa, und Peter Andorfer. 2019. „I Want it All, I Want it Now. Literature researcher meets programmer.“ In *Twin Talks at DHN2019: Understanding Collaboration in DH. Proceedings*, hg. v. Steven Krauwer und Darja Fišer. Copenhagen. Zugriff: 17. Januar 2020. <https://cst.dk/DHN2019Pro/TwinTalksWorkshopProceedings.pdf>.
- Hannesschläger, Vanessa und Katharina Prager. 2015. „Ernst Jandl und Karl Kraus – Two Lives in Bits and Pieces.“ In *Biographical Data in a Digital World 2015. CEUR workshop proceedings Vol. 1399*, hg. v. Serge ter-Braake et al. Amsterdam. Zugriff: 17. Januar 2020. <http://ceur-ws.org/Vol-1399>.
- Prager, Katharina. 2015. „Einer, der's gut mit mir meint, vermißte meine Biographie.“ Anti/Biographische Affekte um Karl Kraus. In *BIOS* 1+2: 266–80. doi:10.3224/bios.v28i1-2.15.
- Prager, Katharina. 2017. Objekt des Monats April 2017: Karl Kraus und sein „Drang zu activer Klageführung“ Wien. Zugriff: 17. Januar 2020. <https://www.wienbibliothek.at/bestaende>

sammlungen/objekt-monats/objekte-monats-2017/objekt-monats-april-2017-karl-kraus-drang.

Prager, Katharina. 2017. Karl Kraus Online. Wien: Wienbibliothek im Rathaus. Zugriff: 17. Januar 2020. <https://www.kraus.wienbibliothek.at>.

„Sammlung Prozessakten Oskar Samek — Karl Kraus.“ Digitale Wienbibliothek. Zugriff: 17. Januar 2020. <https://www.digital.wienbibliothek.at/wbrobv02/content/titleinfo/2540007>.

# travel!digital

Ulrike Czeitschner

<b>Team members:</b>	Ulrike Czeitschner (Primary Investigator), Victoria Eisenheld, Barbara Krautgartner, Laura Lanig
<b>Institutions:</b>	ACDH-CH ÖAW
<b>Funding body:</b>	Austrian Academy of Sciences (go!digital 1.0), ÖAW0204
<b>Website:</b>	<a href="https://traveldigital.acdh.oeaw.ac.at">https://traveldigital.acdh.oeaw.ac.at</a> , ARCHE data download <a href="https://id.acdh.oeaw.ac.at/traveldigital">https://id.acdh.oeaw.ac.at/traveldigital</a>

Hardly any other medium covers a similarly complex and at the same time exciting variety of topics as the modern travel guide. Since the early 19th century, classics like the *Baedeker* convey travel expertise and profound background knowledge. The selection and assessment of destinations and attractions on site, the composition of facts and figures, but especially the ever changing perspectives and regularly revised references to history and the present, culture and society provide vivid insights into the many facets of travelling culture. Last but not least, the reader learns a lot about the motivation of travellers, their preferences and interests, the current conventions, *zeitgeist* fashions and trends. As researchers are always on the lookout for significant and promising sources, it comes as no surprise that the fascinating guidebooks attract their attention. So it comes that today you can explore these valuable contemporary witnesses online and follow the tracks of former travellers.

The travel!digital collection brings together German *Baedeker* travel guides on non-European countries published between 1875 and 1914. Key components include linguistic and in particular semantic encoding, domain-specific knowledge organization and content contextualization by Linked Open Data. Focusing on people and monuments, two essential elements not only of guidebook texts but cultural narratives in general, the project aims at showing that semantic technologies have great potential to unlock a discourse that goes far beyond travel literature.

## travel!digital WEB APP

Tiles on the start page provide easy access to the *Baedeker* guidebooks. A reader combines richly annotated full-texts and digital facsimiles of the rare historical prints and offers detailed tables of content as well as a page-turning function. A citation suggestion along with a link for each page is included.

<b>TEXT TYPE</b>	1st editions of German Baedeker travel guides
<b>EXTENT</b>	5 volumes = 3.089 pages = 1.51 million tokens / 1.21 million running words
<b>PERIOD</b>	1875–1914
<b>SUBJECT</b>	non-European destinations
<b>ANNOTATION</b>	XML/TEI [P5], Part-of-Speech Tagging, TreeTagger, STTS Stuttgart-Tübingen-TagSet, tokenEditor
<b>SEMANTIC TECHNOLOGIES</b>	SKOS Simple Knowledge Organization System, LOD Linked Open Data, Skosmos

The index browser shows entry frequencies of word forms, lemmas, PoS-information and personal names. Filters and autocomplete facilitate both index and full-text search. In addition, comprehensive metadata of the digital resources as well as the historical prints are available.

The web app accesses ACDH Vocabularies API to retrieve and display information from the travel!digital thesaurus, the core component of this project. Linking included historical as well as contemporary lexicon entries and LOD sources with corpus occurrences, this domain-specific vocabulary reveals the impressive diversity in the *Baedeker* universe and invites explorers on a digital journey through time.

The entire data collection is freely available at ARCHE (A Resource Centre for the HumanitiEs), the ACDH repository. It comprises the full-text corpus including linguistic and semantic mark-up, digital facsimiles of all volumes, a TEI-schema of the applied annotations as well as the extensive thesaurus with more than 4.200 entry terms and over 3.200 links to the LOD cloud. Since this rich resource was created in the spirit of openness, you are invited to (re)use it intensively!

## Bibliography

- Bock, Benedikt. 2010. *Baedeker & Cook – Tourismus am Mittelrhein 1756 bis ca. 1914*. Bern, Berlin, Frankfurt am Main, New York, Paris, Wien: Peter Lang.
- Buzard, James. [1993] 1998. *The Beaten Track. European Tourism, Literature and Ways to “Culture” 1800–1918*. Oxford: Clarendon Press.
- Czeitschner, Ulrike, and Barbara Krautgartner. 2017. “Discursive Constructions of Culture: Semantic Modelling for Historical Travel Guides.” *Sociology and Anthropology* 5: 323–31. doi:10.13189/sa.2017.050406.

# Zweisprachiger Spracherwerb: Longitudinalstudie anhand schriftlicher Texte der Hermagoras-Volksschule – Digitalisierung, Auszeichnung, Auswertung

Ursula Doleschal, Lisa Rieger

<b>Mitarbeiterinnen:</b>	Ursula Doleschal (Projektleitung), Lisa Rieger
<b>Institutionen:</b>	Universität Klagenfurt, Institut für Slawistik
<b>Fördergeber:</b>	Fakultät für Kulturwissenschaften der Alpen-Adria-Universität Klagenfurt
<b>Website:</b>	<a href="https://campus.aau.at/cris/project/0f4de0c9610ea75c01611937ce72037f">https://campus.aau.at/cris/project/ 0f4de0c9610ea75c01611937ce72037f</a>

Dieses Projekt hat zum Ziel, mit Methoden der Digital Humanities den Fortschritt in Sprach- und Textkompetenz von Schülerinnen und Schülern der zweisprachigen *Mohorjeva ljudska šola/Hermagoras Volksschule* in Klagenfurt mit dualem deutsch-slowenischem Immersionsunterricht auf Basis von ihnen verfasster, authentischer Texte zu analysieren. Dafür wurde von 2015 bis 2018 eine Klasse von der zweiten bis zur vierten Schulstufe begleitet. In diesem Zeitraum wurden von 11 Schülerinnen und Schülern 234 Texte, zusammengefasst in 24 Textsammlungen, digitalisiert, transkribiert, ausgezeichnet und mit einem Auswertungsprogramm<sup>1</sup> hinsichtlich quantitativer Parameter analysiert. Nach einer Gesamtauswertung erfolgte eine Auswertung nach Erstsprachen.

Die Transkription erfolgte manuell in *Transkribus*. Dabei wurde nur das, was von den Schülerinnen und Schülern selbst verfasst wurde, im vorliegenden Wortlaut transkribiert. Anhand der Kopien ergaben sich folgende Probleme:

1. An den Seiten abgeschnittene Wörter: Diese wurden nur so weit transkribiert, wie die Buchstaben noch eindeutig erkennbar waren, ansonsten wurde mit dem POS-Tag „gap“ auf die Unvollständigkeit hingewiesen.
2. Schwer leserliche und schlecht erkennbare Wörter: Diese wurden so gut wie möglich rekonstruiert, um das Wort im Weiteren lemmatisieren zu können.

---

<sup>1</sup> Das Auswertungsprogramm wurde nach gemeinsamer Absprache von der Computerlinguistin am Institut für Germanistik der AAU Klagenfurt, Elisabeth Eder, MA, speziell für den Zweck des Projekts erstellt.

3. Abgeschnittene oder unleserliche längere Textpassagen: Diese führten dazu, dass der Text aus der Analyse ausgeschieden wurde.

Die slowenischen Texte wurden mit dem morphosyntaktischen Tagger *Obeliks* (Romih, Krek und Kosem, n.d.), die deutschen mit WebLicht (CLARIN-D, 2017a) ausgezeichnet. Die Zuordnung der Kategorien und Eigenschaften erfolgt bei *Obeliks* anhand der Ergebnisse aus dem Projekt *Jezikoslovno označevanje slovenščine* (JOS) (Inštitut Jožef Stefan, n.d.), in WebLicht basiert sie auf dem Stuttgart-Tübingen Tagset (Schiller et al. 1999). Beim Vergleich der Guidelines (Erjavec et al. 2010) wurde beschlossen, aufgrund folgender Phänomene von einem direkten Vergleich zwischen den beiden Sprachen abzusehen:

1. Kategorien, welche nur in einer der beiden Sprachen vorhanden sind,
2. unterschiedliche Zuordnungen zu den jeweiligen Kategorien und
3. unterschiedlicher Umfang bei der Auszeichnung morphologischer und syntaktischer Eigenschaften.

Trotz der statistisch niedrigen Fehlerquote war es notwendig, die Auszeichnungsergebnisse in folgenden Fällen manuell zu berichtigen:

1. abgeschnittene und unleserliche Wörter,
2. Rechtschreibfehler und
3. abgeteilte Wörter.

Das Auswertungsprogramm errechnete aus den XML-Dateien für jede Textsammlung, aufgeteilt nach Schülerin bzw. Schüler, folgende Parameter: durchschnittliche Satzanzahl und Satzlänge, durchschnittliche Wortanzahl sowie die durchschnittliche Anzahl von Verben und Konjunktionen. Nach einer Gesamtauswertung wurden die Ergebnisse nach den Erstsprachengruppen Deutsch, Slowenisch, Deutsch und Slowenisch sowie Slowenisch und BKS-Sprachen zusammengefasst. Die Entwicklung der slowenischen und deutschen Texte wurde zuerst getrennt betrachtet, bevor am Ende eines jeden Unterkapitels die prozentuellen Veränderungen einander gegenübergestellt wurden. Totale Zahlen wurden nicht berücksichtigt, um einen Vergleich erschwerende Faktoren, wie unterschiedliche Sprachstruktur und Auszeichnungsstandards, möglichst zu minimieren.

Durch die quantitative computergestützte Analyse konnten objektive Daten für die Beurteilung der Sprachkompetenz der Schülerinnen und Schüler gewonnen werden. Allerdings stellte sich auch heraus, dass reine Zahlen Spielraum für Interpretationen

lassen und somit in Zukunft nicht als alleiniges Mittel zur Beurteilung von Sprachkompetenz eingesetzt werden sollten, sondern vielmehr als objektive Ergänzung zu einer qualitativen Analyse. Aus technischer Sicht wäre es für die weitere Forschung wünschenswert, durch interdisziplinäre Zusammenarbeit, aufbauend auf den aufgezeigten Problemen einen einheitlichen Leitfaden für die Digitalisierung und Auszeichnung von Texten zu erarbeiten und somit einen Vergleich zwischen ähnlich gelagerten Projekten zu ermöglichen.

## Literatur

- CLARIN-D. 2017a. „WebLicht. Main Page.“ Zugriff: 5. April 2019. [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page).
- CLARIN-D. 2017b. „WebLicht. Tools in Detail.“ Zugriff: 5. April 2019. [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Tools\\_in\\_Detail](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Tools_in_Detail).
- Erjavec, Tomaž, et al. 2010. „Oblikoskladenske specifikacije JOS.“ Zugriff: 5. April 2019. <http://nl.ijs.si/jos/msd/html-sl/josMSD-sl.html>.
- Institut Jožef Stefan. o.J. „Projekt JOS: jezikoslovno označevanje slovenskega jezika.“ Zugriff: 5. April 2019. <http://nl.ijs.si/jos/index-sl.html>.
- Romih, Miro, Simon Krek und Iztok Kosem. o.J. „Sporazumevanje v slovenskem jeziku. Označevalnik.“ Zugriff: 5. April 2019. <http://oznacevalnik.slovenscina.eu/Vsebine/Sl/SpletniServis/SpletniServis.aspx>.
- Schiller, Anne, et al. 1999. „Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset).“ Zugriff: 5. April 2019. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.

# Kommentierte Werkausgabe Werner Kofler (Prosa, Lyrik, Radio, Film, Theater)

Claudia Dürr, Wolfgang Straub

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Wolfgang Straub (Projektleitung), Claudia Dürr, Helmut W. Klug, Carina Koch, Elisabeth Raunig
<b>Institutionen:</b>	Universität Wien, Universität Klagenfurt – Robert-Musil-Institut für Literaturforschung, Kärntner Literaturarchiv, Universität Graz – Zentrum für Informationsmodellierung
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds
<b>Website:</b>	<a href="http://www.wernerkoefler.at">http://www.wernerkoefler.at</a>

Digitale Editionen von Gegenwartsliteratur stehen aufgrund urheberrechtlicher Restriktionen vor spezifischen Herausforderungen. Die Werkausgabe des österreichischen Schriftstellers Werner Kofler (1947–2011) setzt daher auf die Vorteile einer Hybridedition: Die gesammelte Prosa ist vollständig kommentiert Ende 2018 in drei Bänden zu insgesamt 1700 Seiten erschienen (FWF-Projekt *Kommentierte Werkausgabe Werner Kofler: Prosa*, 2015–18). Die Forschungsplattform <http://www.wernerkoefler.at> bietet den umfangreichen Kommentar auch digital an. Der Stellenkommentar umfasst rund 5500 Einträge, die durch Primärtextzitate ergänzt werden, und ist nach Kategorien durchsuchbar, mittels derer Koflers Denkräume erschlossen werden können (z. B. Politikerinnen und Politiker, Philosophen, Literaturzitate). Der Onlinekommentar bietet darüber hinaus die Möglichkeit, die schriftlichen Einträge mit Fotos und Faksimiles von Archivmaterial aus dem Kärntner Literaturarchiv, wo der Nachlass Koflers aufbewahrt wird, zu ergänzen.

Das 2018 bis 2022 laufende Projekt *Kofler intermedial* erarbeitete in einem ersten Schritt Stellenkommentare zu den dramatischen Texten, zur Lyrik und zu versprengter Prosa Koflers, die in den Bänden IV (Hörspiele) und V (Lyrik, Kurzprosa, Dramatisches) der Werkausgabe im Frühjahr 2023 erschienen. Zugleich wurde im Projekt mit neuen Wegen der Hörspielen und der Weiterverarbeitung der Kommentardaten experimentiert. Dazu wurden 15 Radioarbeiten Koflers auf der Website zugänglich gemacht, eine Novität, bedenkt man die zugrundeliegende rechtliche Situation. Die Edition von Hörspielen hat weder in der etablierten Editionspraxis noch in der Digitalen Edition eine nennenswerte Vorgeschichte, sodass das Projekt mit der Digitalen Edition zweier Hörspiele (Typoskript, Lesetext, Audio, Renotat) theoretisch, methodisch und

praktisch Neuland betrat. Im Zentrum steht dabei wohl die Präsentation der primären und sekundären Daten und ihre Zugänglichmachung für ein Fachpublikum. Die Forschungsplattform versteht sich auch als eine Materialsammlung zur Ergründung intermedialer Phänomene etabliert werden. Das Werk Koflers ist prädestiniert wie wenige für einen solchen Zugang, da die Grenzen zwischen Text und audiovisuellen Arbeiten fließen. Das Projekt verknüpft Ausschnitte aus den dramatischen Arbeiten direkt mit den jeweiligen Textpassagen der Prosawerke und macht Transformationen sicht- und hörbar.

Ziel des Projekts war es, über die fachwissenschaftlichen Fragestellungen hinaus anhand eines konkreten Text- und Medienkorpus die Möglichkeiten der Zusammenarbeit von Philologie und Digital Humanities abseits ‚herkömmlicher‘ (mit urheberrechtsfreien Textkorpora arbeitender) digitaler Editionen auszuloten und geeignete Instrumentarien der Analyse und der Präsentation visueller und auditiver Inhalte zu entwickeln.

# Kommentierte Online-Edition des Gesamtbriefwechsels Ludwig von Ficker

Markus Ender

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Eberhard Sauermann (Projektleitung 2012–2015), Ulrike Tanzer (Projektleitung 2016–2022), Ingrid Fürhapter, Markus Ender, Joseph Wang
<b>Institutionen:</b>	Universität Innsbruck – Forschungsinstitut Brenner-Archiv, University of Bergen – Wittgenstein Archives
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds (P P24283-G23; P 29070-G23)
<b>Website:</b>	<a href="https://edition.ficker-gesamtbriefwechsel.net">https://edition.ficker-gesamtbriefwechsel.net</a>

Die *Kommentierte Online-Edition des Gesamtbriefwechsels Ludwig von Ficker* entsteht seit 2012 im Rahmen zweier Forschungsprojekte am Forschungsinstitut Brenner-Archiv der Universität Innsbruck. Mit der umfangreichen Korrespondenz Ludwig von Fickers wird einer der Kernbestände des Tiroler Literaturarchivs textkritisch aufbereitet, in eine digitale Edition überführt und der Wissenschaft zur Verfügung gestellt.

Die Edition kann auf eine lange Vorgeschichte verweisen, deren Grundlage in der editorischen Arbeit liegt, die 1964 begann und der eine vierbändige gedruckte *Auswahlausgabe der Briefe Ludwig von Fickers* erwuchs (erschienen 1988 bis 1996 im Salzburger Otto-Müller-Verlag sowie im Innsbrucker Haymon-Verlag). In diese Edition wurden 1300 ausgewählte Briefdokumente aufgenommen; parallel dazu wurden im Brenner-Archiv kontinuierlich die Briefe Ludwig von Fickers an diverse Adressatinnen und Adressaten eingeholt und transkribiert. Zunächst wurden Briefabschriften in Form von Typoskripten erstellt, seit den 1980er-Jahren wird die Transkription in elektronischer Form mittels *MS-Word* durchgeführt. Dieser Datenbestand bildet das Fundament der kommentierten Online-Edition.

Aufgrund der erfolgten Vorarbeiten, der Heterogenität des Quellenmaterials und der großen Textmengen mussten in den Projekten die textkritischen Arbeitsabläufe nach ökonomischen Gesichtspunkten bzw. nach Effizienzkriterien ausgerichtet und abgestimmt werden. Im Gegensatz zur gedruckten Briefausgabe stellt die digitale Edition keine Auswahl dar, sondern kompiliert alle vorhandenen Korrespondenzstücke des Gesamtbriefwechsels. Insgesamt umfasst sie mehr als 17.300 Einzeldokumente (hauptsächlich Briefe und Karten, daneben Kondolenzschreiben, Einladungen etc.), davon

entfallen 4.000 auf die Briefe von Ludwig Ficker an diverse Adressatinnen und Adresaten und 13.300 Briefe von über 2.200 Korrespondenzpartnerinnen und -partnern an Ludwig von Ficker. Von sämtlichen Dokumenten wird ein edierter Text erstellt; lediglich im Fall von Beilagen ohne handschriftliche Vermerke (Konzerteinladungen, Fotos o. Ä.) wird der Inhalt als Regest wiedergegeben.

Als Grundbedingung wurde festgelegt, dass zum Einen die Transkriptionen von Hilfskräften erstellt werden konnten und zum Anderen die textphilologische Bearbeitung der Transkripte (Textkritik, Annotation, Kommentierung) auch ohne Kenntnis von XML möglich sein musste. Folglich werden für die Edition DOCX-Dateien mittels auf die Edition optimierter XSLT-Stylesheets automatisiert in TEI-P5-konforme XML-Dateien transformiert. Zu diesem Zweck kommt auch eine Programmroutine auf der Basis des *Apache Camel Frameworks* zur Anwendung.

Das Editionskonzept ist einem strikt textbasierten Ansatz verpflichtet, was zur Folge hat, dass die Reproduktion der Materialität der archivalischen Quellen kein primäres Ziel von Ontologie und Modellierung darstellt. Der *textus constitutus* ist zitierfähig und entspricht zeichengetreu dem Brieftext des Originals; der Fokus auf die Textebene bedingt jedoch, dass zwar auf formaler Ebene die Absatz- und Zeilenumbrüche nachgebildet werden, auf eine vollständige diplomatische Umschrift hingegen verzichtet wird. Die für jedes Korrespondenzstück erfassten strukturellen Metadaten umfassen Sender, Empfänger, Provenienznachweis sowie Archivsignatur. Annotiert werden Personen- und Ortsnamen, Werk- und Aufsatztitel; dies geschieht weitgehend händisch. Um die Anschlussfähigkeit an andere Editionsprojekte, Interoperabilität, Retrieval sowie erweiterte Bearbeitungs- und Explorationsfunktionen zu ermöglichen, werden die Annotationen mit GND-Normdaten verknüpft. Kommentiert werden in der Regel Einzelstellen, die Flächenkommentare werden weitgehend ausgelagert.

Diese Editions- und Erschließungsstrategien gewährleisten einen zuverlässigen und für derivative Forschung relevanten kulturwissenschaftlichen Quellenfundus sowie die zeitgemäße Repräsentation von Archivalien. Darüber hinaus vermag die Edition dem Bedarf des Literaturarchivs nach digitaler Langzeitarchivierung und Bestandsschonung Rechnung zu tragen.

## Literatur

- Ender, Markus. 2015. „Zum Mehrwert der kommentierten Online-Edition des Gesamtbriefwechsels Ludwig von Fickers.“ In *Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte*, hg. v. Thomas Bein. Berlin, Boston: De Gruyter, pp. 35–46.
- Wang, Joseph. 2016. „From DOCX to TEI to Literature Map.“ Poster. Zugriff: 30. Sept. 2019. <http://tei2016.acdh.oeaw.ac.at/sites/default/files/Joseph%20Wang%20From%20DOCX%20via%20TEI%20to%20Literature%20Map.pdf>.

# Arthur Schnitzler – Kritische Edition (Frühwerk) III

Konstanze Fiedl, Ingo Börner, Anna Lindner, Marina Rauchenbacher, Isabella Schwentner

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Konstanze Fiedl (Projektleitung), Ingo Börner, Anna Lindner, Marina Rauchenbacher, Isabella Schwentner
<b>Institutionen:</b>	ACDH-CH ÖAW; Universität Wien – Institut für Germanistik
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds (P 30513)
<b>Website:</b>	<a href="https://www.univie.ac.at/germanistik/projekt/arthur-schnitzler-fruehwerk-3">https://www.univie.ac.at/germanistik/projekt/arthur-schnitzler-fruehwerk-3</a> , <a href="https://www.oewa.ac.at/de/acdh/projects/arthur-schnitzler-fruehwerk">https://www.oewa.ac.at/de/acdh/projects/arthur-schnitzler-fruehwerk</a> , <a href="https://schnitzler-werke.acdh-dev.oewa.ac.at">https://schnitzler-werke.acdh-dev.oewa.ac.at</a>

Bis 2011 existierte von den Werken Arthur Schnitzlers keine historisch-kritische Ausgabe. Dies liegt einerseits an seiner unleserlichen Handschrift (vgl. Fiedl 2017, S. 144) und andererseits an der komplizierten Nachlass-Situation: 1938, beim ‚Anschluss‘ an Hitlerdeutschland, waren seine hinterlassenen Papiere nach Cambridge in Sicherheit gebracht worden. Während die privaten Teile des Nachlasses (v. a. das Tagebuch und die Korrespondenzen) im Deutschen Literaturarchiv in Marbach aufbewahrt werden, befinden sich die Werkmanuskripte nach wie vor großteils an der *University Library* in Cambridge. Andere Teile des Nachlasses liegen verstreut in Institutionen, wie etwa in Exeter, Genf und Jerusalem. Eine historisch-kritische Ausgabe blieb daher lange ein Desiderat, zumal die im Fischer-Verlag erschienenen Lesetexte aufgrund verschiedener Korrektur- und Bearbeitungsvorgänge äußerst unzuverlässig sind.

In zwei FWF-Vorgängerprojekten *Arthur Schnitzler – Kritische Edition (Frühwerk) I* und *II* (2010–2018) erschienen mittlerweile elf Ausgaben im de Gruyter Verlag: *Lieutenant Gustl* (2011), *Anatol und Sterben* (2012), *Liebelei* (2014), *Frau Bertha Garlan* (2015), *Die Frau des Weisen*, *Die Toten schweigen und Ein Abschied* (jeweils 2016), *Ein Ehrentag* (2017), *Blumen* (2018) sowie *Reigen* (2019). Die beiden Einakter *Der grüne Kakadu* und *Paracelsus* sind derzeit in Vorbereitung zum Druck.

Bereits ab der zweiten Projektphase wurden die Daten entsprechend den TEI-Richtlinien bearbeitet; der Satz des Drucktextes erfolgte mittels LaTeX. Für die Darstellung der Transkriptionen der Handschriften wurde ein XML-Zwischenformat entwickelt, das einen Import nach Adobe Indesign ermöglichte. Das derzeit laufende Projekt *Arthur Schnitzler – Kritische Edition (Frühwerk) III* zielt auf eine Umstellung zur Hybridausgabe, welche anhand des Bandes *Blumen* erprobt wurde (vgl. Börner u. Schwentner 2019).

*Blumen* weist eine komplexe Entstehungsgeschichte auf, die sich vor allem durch eine weitreichende Umordnung des Textmaterials auszeichnet. Die genetische Abfolge der Niederschrift wurde rekonstruiert und in der gedruckten Edition als fortlaufender Text zusätzlich wiedergegeben. Um die spezifische Entstehungsgeschichte von *Blumen* und damit die makrogenetischen Prozesse besser nachvollziehbar zu machen, bot sich eine digitale Edition an. Daher wurde zunächst das Datenmodell für die Edition in Hinblick auf eine digitale Darstellung der Textgenese überarbeitet und erlaubt nun eine Verknüpfung von handschriftlich vorliegenden Textstufen und Drucktext auf Segment-Ebene.

Für die Transkription des handschriftlichen Materials kommt die Software *Transkribus* zum Einsatz, die es ermöglicht, die Transkriptionen in der Webdarstellung mit den digitalen Faksimiles auf Basis von ermittelten Bildkoordinaten zu verknüpfen und für eine Web-Darstellung basierend auf der XML-Datenbank *eXist-db* aufzubereiten. Darüber hinaus wird in der digitalen Komponente eine Möglichkeit zur semantischen Annotation, vor allem der im Kommentar vorkommenden Personen, Orte und Texte, ausgeschöpft. Diese sollen mit entsprechenden Normdaten und Daten aus anderen Schnitzler-Projekten verknüpft werden. Dementsprechend ist im Rahmen der derzeit in Arbeit befindlichen Edition zum Stück *Der einsame Weg* ein Datenaustausch mit der in Cambridge entstehenden Edition von *Professor Bernhardi* vorgesehen, da sich die Entstehungsgeschichte der beiden Texte anfänglich überschneidet.

## Literatur

- Börner, Ingo und Isabella Schwentner. 2019. „Blumen, neu arrangiert. Die Historisch-kritische Ausgabe von Arthur Schnitzlers Frühwerk auf dem Weg von der Buch- zur Hybridedition. Am Beispiel des Bandes Blumen.“ In *Textgenese in der digitalen Edition*, hg. v. Anke Bosse und Walter Fanta. Berlin und Boston: de Gruyter.
- Fliedl, Konstanze. 2017. „Arthur Schnitzler. Schrift und Schreiben.“ In *Die Werkstatt des Dichters. Imaginationsräume literarischer Produktion*, hg. v. Klaus Kastberger und Stefan Maurer, 139–62. Berlin, Boston [u.a.]: de Gruyter. doi:10.1515/9783110466850-010.

# **Andreas Okopenko: Tagebücher aus dem Nachlass (Hybridedition)**

Desiree Hebenstreit, Laura Tezarek, Christiane Fritze, Christoph Steindl

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Roland Innerhofer (Projektleitung), Holger Englerth, Bernhard Fetz, Christiane Fritze, Desiree Hebenstreit, Arno Herberth, Stefan Karner, Christoph Steindl, Laura Tezarek, Stefan Winterstein, Christian Zolles
<b>Institutionen:</b>	Universität Wien – Institut für Germanistik, Österreichische Nationalbibliothek – Literaturarchiv
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds (P 28344)
<b>Website:</b>	<a href="https://edition.onb.ac.at/okopenko">https://edition.onb.ac.at/okopenko</a>

Das FWF-Projekt *Andreas Okopenko: Tagebücher aus dem Nachlass (Hybridedition)* (P 28344 Einzelprojekte) wurde 2015 bis 2018 als Kooperation zwischen dem Institut für Germanistik der Universität Wien und dem Literaturarchiv der Österreichischen Nationalbibliothek (ÖNB) durchgeführt. Grundlage des Projekts war der 2012 vom Literaturarchiv erworbene Nachlass des österreichischen Schriftstellers Andreas Okopenko (1930–2010), der umfangreiche Tagebuchaufzeichnungen Okopenkos von der Jugend an bis kurz vor seinem Tod enthält. Für die Edition wurden die Tagebücher aus dem Zeitraum 1949 bis 1954 ausgewählt, die über 3.000 Seiten umfassen und aus Heften, Beilagen und Briefkonvoluten bestehen.

Ziel des Projektes war die Erstellung einer Edition, die aktuellen wissenschaftlichen Standards entspricht und eine inhaltliche Aufarbeitung bietet, um neue Erkenntnisse zu Okopenko selbst, zu seinem Werk und zur österreichischen Literaturgeschichte der 1950er-Jahre zu gewinnen.

Die ausgewählten Tagebücher wurde von Projektmitarbeiterinnen und Projektmitarbeitern an der Österreichischen Nationalbibliothek im Format JPEG 2000 digitalisiert und dort zur Langzeitarchivierung gespeichert. Die Transkription der handschriftlich verfassten Tagebücher erfolgte manuell, für die maschinenschriftlichen Textteile wurde die an der Universität Innsbruck entwickelte Software *Transkribus* eingesetzt.

Ausgehend von den Richtlinien der *Text Encoding Initiative* für Handschriftenbeschreibungen wurde für die Transkription des Quellenmaterials ein auf die Tagebücher abgestimmtes Datenmodell entwickelt. Jedes Tagebuchkonvolut wurde in einem separaten XML-Dokument gespeichert. Alle Projektdaten wurden beim Bearbeiten in einem git-Repositorium zur Dokumentation abgelegt und fortlaufend versioniert.

Die inhaltliche Erschließung des Materials erfolgte über die Erstellung von Registerdateien, auf die von in den Tagebüchern genannten Personen, Orten, Institutionen und Werken referenziert wird. Neben einem umfangreichen biografischen Abriss zum Leben Okopenkos wurden drei inhaltliche Themenkommentare verfasst (literarische Netzwerke, Medien, zeithistorische Diskurse). Seit Version 2.0 sind Zusatzmaterialien aus dem Vor- und Nachlass Okopenkos als Ergänzung zur Edition verfügbar, die Studierende der Universität Wien ediert haben. Ebenso bietet die Version 2.0 einen Einstieg in das Material über eine Objektgalerie und einen Eintrag des Tages.

Die digitale Edition wurde innerhalb der nachhaltigen Infrastruktur für digitale Editionen an der Österreichischen Nationalbibliothek umgesetzt, welche auf dem *Geisteswissenschaftlichen Asset Management System* (GAMS) der Universität Graz basiert. Die TEI-XML-Dokumente werden mit XSLT-Skripten transformiert und zur Anzeige gebracht. Das Design wurde mit einem externen Dienstleister entwickelt und im Frontend-Framework Bootstrap umgesetzt, wodurch unter anderem eine optimale Darstellung auf kleineren Bildschirmgrößen gewährleistet wird. Hauptelement der Edition ist eine synoptische Ansicht der Tagebücher, in welcher die hochauflösenden Faksimiles entweder zusammen mit einer Lesefassung, einer diplomatischen Umschrift oder der TEI-XML-Version betrachtet werden können.

## Literatur

- Fritze, Christiane und Christoph Steindl. „Digitale Editionen an der Österreichischen Nationalbibliothek – eine Infrastruktur.“ Forschungsblog der Österreichischen Nationalbibliothek, 06.06.2019. Zugriff: 17. Januar 2020. <https://www.onb.ac.at/forschung/forschungsblog/artikel/digitale-editionen-an-der-oesterreichischen-nationalbibliothek-eine-infrastruktur>.
- Zolles, Christian. „Hypertext-Pionier nun im Hypertext.“ ORF science, 17.12.2018. Zugriff: 17. Januar 2020. <https://science.orf.at/stories/2952576>.

# Mittelhochdeutsche Begriffsdatenbank (MHDBDB)

Peter Hinkelmanns, Katharina Zeppezauer-Wachauer

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Katharina Zeppezauer-Wachauer, Peter Hinkelmanns, Daniel Schläger, Klaus M. Schmidt
<b>Institutionen:</b>	Universität Salzburg – Interdisziplinäres Zentrum für Mittelalter und Frühneuzeit (IZMF)
<b>Fördergeber:</b>	Universität Salzburg
<b>Website:</b>	<a href="http://mhdbdb.sbg.ac.at">http://mhdbdb.sbg.ac.at</a>

Auf der Grundlage digitaler Editionen wird der mhd. und fnhd. Wortschatz seit den 1970er-Jahren in der *Mittelhochdeutschen Begriffsdatenbank* (MHDBDB) onomasiologisch, also über die Wortbedeutung, erschlossen (begründet durch Klaus M. Schmidt, Bowling Green State University). In mehreren technischen Stufen und einer 1992 durchgeführten Zusammenschließung mit dem Projekt *Namen in deutschen literarischen Texten des Mittelalters* von Horst Pütz (Christian-Albrechts-Universität zu Kiel) ist die MHDBDB, die seit 2002 an der Universität Salzburg installiert ist, mittlerweile auf über 10 Millionen Tokens, verteilt auf über 650 Texteditionen unterschiedlichster Textsorten und -gattungen, angewachsen. Die Editionen sind zum Großteil vom Projekt nachdigitalisierte Printeditionen, mittlerweile ergänzen aber auch erste *Born-Digital*-Editionen das Korpus.

Seit 2016 wird der Relaunch der MHDBDB vorbereitet, dessen Ziel neben einer gestalterischen ‚Rundumerneuerung‘ in puncto Bedienung, Abfragemöglichkeiten und Webdesign vor allem die Standardisierung der Daten ist. Die bestehenden Texte der Datenbank werden nach TEI-XML kodiert. Zukünftig sollen überwiegend genuin digitale TEI-Editionen aufgenommen und auf eine Retrodigitalisierung weitestgehend verzichtet werden.

Bereits heute sind komplexe Suchanfragen zur Literatur und zum Wortschatz des Mittelhochdeutschen möglich. Es können Textstellen etwa über die Suche nach dem Vorkommen eines oder mehrerer zueinander in Beziehung stehender lemmatisierter Tokens, aber auch über die onomasiologischen Begriffe des Begriffssystems gesucht werden. Auch kombinierte Suchen können durchgeführt werden, um etwa zwischen Wortbedeutungen zu unterscheiden. Beispielsweise führt der Wortartikel *druc* mehrere Bedeutungen an: 1) *Schlacht*, 2) *Willensausübung auf andere*, 3) *Zweikampf*, 4) *Physikal. chem. Eigenschaften*, 5) *Buchherstellung/Buchdruck*. Auch Informationen zur

Wortbildung und zu Wortfamilien listet der Wortartikel – zu *druc* etwa die Nomina *druckerie* (Druckerei) und *drucker* (Buchdrucker).

Suchanfragen können in beliebiger Kombination durchgeführt werden, beispielsweise in welchen Texten *druc* in der Bedeutung 5) *Buchherstellung/Buchdruck* gebraucht wird oder ob und wie die lateinische Wortvariante *editum* in der Bedeutung *edieren, herausgeben* verwendet wird (einmal übrigens in Heinrich Münsingers *Regimen sanitatis*, einmal bei Priester Konrad).

Um beliebig viele linguistische Annotationsebenen wie etwa *Part-Of-Speech* (POS) oder Phrasen- und Satzstrukturen annotieren zu können, werden zukünftig die linguistischen Annotationen im *Stand-Off*-Verfahren mit *Linked Open Data* (LOD) auf die Tokens der Texte bezogen und nicht wie bisher auf die graphematischen *Types* im Wortartikel. Dieses Verfahren ermöglicht weiters das Hinzufügen von Metadaten zu jeder Annotation, etwa um deren Autorschaft zu kennzeichnen.

Sämtliche Daten werden im Zuge des Relaunches als LOD in das *Semantic Web* eingebunden und (abseits der TEI-Editionen) nach LOD-Standards codiert. Die Vernetzung mit Metadatenrepositorien wie *Wikidata* oder der *Gemeinsamen Normdatei* (GND) ermöglicht das wechselseitige Anreichern der Daten und wird wesentlich zu einer besseren Abfragbarkeit des MHDBDB-Korpus beitragen. Normdaten zu Personen, Orten, Ereignissen und Werken werden des Weiteren auch die Auffindbarkeit und Filterung der Texte des Korpus vereinfachen.

Der Relaunch der MHDBDB wird nicht nur zu erheblichen Vorteilen für diejenigen führen, die selbst nach Wörtern, Zeichenketten, Wortarten und Begriffen suchen möchten. Zusätzlich können auch Editorinnen und Editoren ihre digitalen Editionen als externe Korpora in die Begriffsdatenbank einbinden und so automatisch POS annotieren und lemmatisieren, mit einer onomasiologischen Systematik (dem Begriffssystem) erschließen und über Normdaten anreichern. Die Benützbarkeit und Sichtbarkeit der Editionen wird damit maßgeblich – ohne besonderen Aufwand für die Herausgeberinnen und Herausgeber – verbessert und ausgeweitet werden.

## Literatur

MHDBDB 3.0: Hinkelmanns, Peter und Katharina Zeppezauer-Wachauer. 2020. „ez ist ein wârheit, niht ein spel, daz netze was sinewel. Die MHDBDB im Semantic Web.“ In *Bamberger interdisziplinäre Mittelalterstudien*, hg. v. Ingrid Bennewitz und Martin Fischer. doi:10.20378/irb-48993.

MHDBDB 3.0: Hinkelmanns, Peter. 2019. „Mittelhochdeutsche Lexikographie und Semantic Web: Die Anbindung der ‚Mittelhochdeutschen Begriffsdatenbank‘ an Linked Open Data.“ In *Digitale Mediävistik*, hg. v. Roman Bleier, Franz Fischer, Torsten Hiltmann, Gabriel Viehhäuser und Georg Vogeler. *Das Mittelalter. Perspektiven mediävistischer Forschung* 24, no. 1: 129–41. doi:10.1515/mial-2019-0009.

- MHDBDB 2.0: Schmidt, Klaus M. und Horst P. Pütz. 2001. „Die mittelhochdeutsche Begriffsdatenbank.“ *Zeitschrift für deutsches Altertum und deutsche Literatur* 130: 493–5.
- MHDBDB 1.0: Schmidt, Klaus M. 1978. „Errungenschaften, Zukunftsmusik und Holzwege auf dem Gebiet der inhaltlichen Textanalyse mit Hilfe des Elektronenrechners.“ In *Maschinelle Verarbeitung altdeutscher Texte III. Beiträge zum Symposium Tübingen 17. - 19. Februar 1977*, hg. v. Paul Sappler und Erich Straßner, 101–11. Tübingen: Niemeyer.

# Ambraser Heldenbuch: Transcription and Scientific Dataset

Mario Klarer, Aaron Tratter, Hubert Alisade

<b>Team members:</b>	Mario Klarer (principle investigator), Hubert Alisade, Veronika Führer, David Messner, Markus Saurwein, Claudia Sojer, Aaron Tratter
<b>Institutions:</b>	University of Innsbruck – Department of American Studies
<b>Funding body:</b>	Austrian Academy of Sciences (go!digital-2.0)
<b>Website:</b>	<a href="https://www.uibk.ac.at/projects/ahb/">https://www.uibk.ac.at/projects/ahb/</a>

Between January 2017 and December 2019, a research group led by Univ.-Prof. Dr. Mario Klarer at the University of Innsbruck worked on the go!digital 2.0 project *Ambraser Heldenbuch: Transcription and Scientific Dataset*, funded by the Austrian Academy of Sciences. The project's aim is to transcribe the entire *Ambraser Heldenbuch* (Vienna, Austrian National Library, Cod. ser. nova 2663) on the occasion of the 500th anniversary of Emperor Maximilian's death.

At the beginning of the sixteenth century, Emperor Maximilian I commissioned Hans Ried, a scribe from Bolzano, to copy texts for the *Ambraser Heldenbuch*.<sup>1</sup> The *Ambraser Heldenbuch* consists of twenty-five important texts of German medieval literature on approximately 5 + 238 parchment folios. Fifteen of these texts have survived solely in the *Ambraser Heldenbuch* as Early New High German versions, including *Erec* by Hartmann von Aue, the anonymous *Kudrun* and *Mauritius von Craûn*. Most modern editions of these uniquely transmitted texts are re-translations into standardized Middle High German that therefore do not represent the original textual source. For many years, scholars and editors have been demanding a transcription of the entire *Ambraser Heldenbuch* that is faithful to the manuscript text.<sup>2</sup>

The project team transcribed the *Ambraser Heldenbuch* with the software Transkribus. The generated data were saved as XML files, which contain the transcribed text as well as the coordinates of the respective lines in the manuscript. The three text columns of each page of the *Ambraser Heldenbuch* were labeled as *text regions*, which were, in turn, divided into *lines*. The codex's unique status as the only source for a number of canonical medieval literary works required an allographic

---

<sup>1</sup> See Alisade (2019).

<sup>2</sup> E.g. Leitzmann (1935); Gärtner (2006); Homeyer and Knor (2015).

transcription that refrained from any kind of standardization or normalization. For example, variants of letters, such as long ‹s› or rounded ‹r›, were maintained in order to be as faithful as possible to Hans Ried's use of characters and diacritical signs in the *Ambraser Heldenbuch*.

Another important aspect of the project was the tagging of, e.g., verses, stanzas, and caesuras, as well as abbreviations, initials, and rubrications, all of which provide added value to the transcription proper. The meticulous processing of the texts of the *Ambraser Heldenbuch* paves the way for a variety of potential applications in further research and editorial endeavors.

The consistent allographic transcription of the entire *Ambraser Heldenbuch* will serve as a reference and basis for linguistic investigations of Early New High German and will facilitate editorial projects on individual texts from the *Ambraser Heldenbuch*.

In early 2022, the transcription of the entire *Ambraser Heldenbuch* was published as an eleven volume print and open access edition by De Gruyter (Klarer 2022a), which, in addition to the allographic and the diplomatic transcriptions of the text, also synoptically juxtaposes the manuscript scans on each page. Furthermore, the data set, together with the transcription guidelines, were published on GitLab (Klarer 2022b) and on Zenodo (Klarer 2022c).

## Bibliography

- Alisade, Hubert. 2019. "Zur Entstehungsgeschichte des Ambraser Heldenbuchs. Die Beauftragung Hans Rieds." In *Kaiser Maximilian I. und das Ambraser Heldenbuch*, ed. by Mario Klarer, 27–35. Vienna: Böhlau.
- Gärtner, Kurt. 2006. Introduction to *Erec: Mit einem Abdruck der neuen Wolfenbütteler und Zwettler Erec-Fragmente*, by Hartmann von Aue, XI–XLIII. 7th ed., ed. by Albert Leitzmann, Ludwig Wolff, and Kurt Gärtner. Tübingen: Niemeyer. Altdeutsche Textbibliothek 39.
- Homeyer, Susanne, and Inta Knor. 2015. "Zu einer umfassenden Untersuchung der Schreibsprache Hans Rieds im Ambraser Heldenbuch." *Zeitschrift für Deutsche Philologie* 134 (1): 97–103.
- Klarer, Mario, ed. 2022a. *Ambraser Heldenbuch: Gesamttranskription mit Manuscriptbild*. 11 vols. Berlin: De Gruyter. Transcriptiones 1.
- . 2022b. "Ambraser Heldenbuch: Transkription und wissenschaftliches Datenset." Accessed February 9, 2022. [https://gitlab.com/mario\\_klarer/ambraser-heldenbuch-transcription-and-scientific-dataset](https://gitlab.com/mario_klarer/ambraser-heldenbuch-transcription-and-scientific-dataset).
- . 2022c. "Ambraser Heldenbuch: Transkription und wissenschaftliches Datenset." Accessed February 9, 2022. doi:10.5281/zenodo.6025611.
- Leitzmann, Albert. 1935. "Die Ambraser Erecüberlieferung." *Beiträge zur Geschichte der deutschen Sprache und Literatur* 59: 143–234.

# **Digitale Edition und Analyse der Medialität diplomatischer Kommunikation: Kaiserliche Gesandte in Konstantinopel in der Mitte des 17. Jahrhunderts**

Carina Koch, Lisa Brunner, Anna Huemer, Christoph Würflinger

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Arno Strohmeyer (Projektleitung), Lisa Brunner, Anna Huemer, Philipp Humer, Carina Koch, Jakob Sonnberger, Georg Vogeler, Christoph Würflinger
<b>Institutionen:</b>	Universität Salzburg – Fachbereich Geschichte, Universität Graz – Zentrum für Informationsmodellierung
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds (P30091-G28)
<b>Website:</b>	<a href="http://gams.uni-graz.at/dipko">http://gams.uni-graz.at/dipko</a> , <a href="http://diploko.at">http://diploko.at</a>

Das FWF-Projekt *Die Medialität diplomatischer Kommunikation* untersucht den Informationstransfer zwischen Konstantinopel und dem Kaiserhof in Wien anhand schriftlicher Zeugnisse diplomatischer Missionen (Briefe, Reisebericht, Abschluss- und Geheimbericht, Instruktionen) aus der Mitte des 17. Jahrhunderts mit digitalen Methoden. Die Quellen geben Auskunft über Probleme interkultureller und interreligiöser Koexistenz, daraus resultierender Konflikte und Lösungsstrategien (Meienberger 1973, Droste 2006). Sie vermitteln jedoch keine objektiven Informationen, da sie von Diskurstraditionen, den Interessen und der Selbstdarstellung der Akteure sowie von den Bedingungen der Übermittlung (Postwege, Informationssicherheit) beeinflusst waren. Die Medien hatten also Einfluss auf den Wissenstransfer zwischen den habsburgischen Diplomaten und dem Kaiserhof und bestimmten die öffentliche Wahrnehmung des Osmanischen Reiches (Faulstich et al. 2007). Die ausgewählten Quellen werden deshalb aus medienwissenschaftlicher Perspektive computergestützt analysiert, wodurch a) ein besseres Verständnis der Konstruktion des in (Mittel-)Europa vorhandenen Wissens über die Osmanen und den Islam, das mitunter bis heute nachwirkt, ermöglicht wird und b) umfangreiches Wissen über die Nutzung von *Natural-Language-Processing* (NLP) mit deutschen historischen Texten gesammelt wird.

Ein Ergebnis des Projekts ist die Digitale Edition der Quellen mit verschiedenen Perspektiven auf die Texte.<sup>1</sup> Neben digitalen Faksimiles und Transkriptionen enthält sie semantische Anreicherungen und Visualisierungen, um Wege des Wissenstransfers und Diskurstraditionen aufzudecken. Historische Hintergrundinformationen werden durch die Kontextualisierung der Quellen bereitgestellt.

Die historisch-kritische Ausgabe wird nach den TEI-P5-Richtlinien<sup>2</sup> modelliert und folgt (wenn möglich) dem DTA-Basisformat.<sup>3</sup> Textstrukturen (Absätze, Überschriften) und textkritische Befunde (Korrekturen, Randbemerkungen) bilden die Basiskodierung. Zudem werden historische Interpretationen (z.B.: Personen, Ereignisse, Korrespondenzformeln oder Plagiate) systematisch ausgezeichnet und mit Identifikatoren versehen. Für die Datenanreicherung werden *Semantic Web*-Technologien herangezogen: SKOS<sup>4</sup> als Vokabular für selbst erstellte Register sowie kontrollierte Vokabulareien (z.B.: GND, VIAF, *GeoNames*)<sup>5</sup> zur Datenverknüpfung nach außen. Beschreibende Informationen zum Reisebericht folgen den Richtlinien des TEI-Moduls *msdescription*.<sup>6</sup> Die CMIF-TEI-Erweiterung<sup>7</sup> der Briefkodierung ermöglicht den Datenaustausch mit dem Webservice CorrespSearch.<sup>8</sup> Basierend auf der Kodierung werden die Inhalte für den Vergleich der beiden Medientypen aufbereitet.

Ein Teil der Daten wird experimentell mit NLP-Tools verarbeitet. Dies stellt eine besondere Herausforderung dar, da sich die Werkzeuge für historische Texte erst in der Entwicklung befinden (Piotrowski 2012). Grundlegende Analysen können mit N-Grammen durchgeführt werden, um wiederkehrende Textmuster zu erkennen. Mit Hilfe von DTA-CAB<sup>9</sup> wird mit Lemmatisierung, *Named-Entity-Recognition* zur Identifizierung von räumlichen Informationen oder Personen und *Part-of-Speech-Tagging* zur Analyse der Verwendung von Stil experimentiert. Die Ergebnisse sollen dazu beitragen, Plagiate aufzudecken oder die Frage zu beantworten, wie Informationen vermittelt werden und welche Meinungen (positiv, negativ, neutral), insbesondere im Zusammenhang mit der Charakterisierung von Personen oder der Selbstdarstellung

<sup>1</sup> Digitale Verfahren erzeugen eine transmediale Darstellung von Text. Zur Transmedialität von Text, s. Sahle 2016 und Sahle 2010.

<sup>2</sup> TEI-Consortium: TEI P5-Guidelines for Electronic Text Encoding and Interchange. 2018. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.

<sup>3</sup> Das DTA-Basisformat (DTABF) ist das von Clarin-D und der DFG empfohlene TEI-Format für historische Texte. <http://www.deutschestextarchiv.de/doku/basisformat>.

<sup>4</sup> SKOS Reference W3C Recommendation 18 August 2009, edited by Alistair Miles and Sean Bechhofer, <http://www.w3.org/TR/skos-reference>.

<sup>5</sup> Gemeinsame Normdatei (GND) [https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html), Virtual Authority File (VIAF) <http://viaf.org>, GeoNames <https://www.geonames.org>.

<sup>6</sup> TEI Manuscript Description, <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html#msov>.

<sup>7</sup> Das Correspondence Metadata Interchange Format wird von der TEI Correspondence SIG entwickelt, [https://correspsearch.net/index.xql?id=participate\\_cmi-format&l=de](https://correspsearch.net/index.xql?id=participate_cmi-format&l=de).

<sup>8</sup> correspSearch. Search scholarly editions of letters, <https://correspsearch.net>, s. Dumont 2016.

<sup>9</sup> <https://kaskade.dwdts.de/demo/cab>.

der Schreiber, geäußert wurden. Für die Sentimentanalyse wird semiautomatisch eine Liste stimmungstragender Ausdrücke als Sentimentlexikon herangezogen, da für den betroffenen Zeitraum kein Sentimentlexikon existiert.

## Literatur

- Droste, Heiko. 2006. *Im Dienst der Krone: Schwedische Diplomaten im 17. Jahrhundert*. Berlin: LIT.
- Faulstich, Werner et al. 2007. „Medialität.“ In *Handbuch Literaturwissenschaft, Bd. 1: Gegenstände und Grundbegriffe*, hg. v. Thomas Anz, 203–64. Stuttgart: J.B. Metzler.
- Meienberger, Peter. 1973. *Johann Rudolf Schmid zum Schwarzenhorn als kaiserlicher Resident in Konstantinopel in den Jahren 1629–1643: Ein Beitrag zur Geschichte der diplomatischen Beziehungen zwischen Österreich und der Türkei in der ersten Hälfte des 17. Jahrhunderts*. Bern: Lang.
- Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. San Rafael, Calif.: Morgan & Claypool.
- Sahle, Patrick. 2010. „Zwischen Mediengebundenheit und Transmedialisierung: Anmerkungen zum Verhältnis von Edition und Medien.“ *editio* 24: 23–36.
- . 2016. „What is a scholarly digital edition (SDE)?“ In *Digital Scholarly Editing: Theory, Practice and Future Perspectives*, hg. v. Matthew James Driscoll und Elena Pierazzo, 19–39. Cambridge, UK: Open Book Publishers.

# InCritApp – Interactive Critical Apparatus

Philipp Koncar, Roman Bleier

<b>Team members:</b>	Roman Bleier, Philipp Koncar
<b>Institutions:</b>	Universität Graz – Zentrum für Informationsmodellierung, Technische Universität Graz – Institute of Interactive Systems and Data Science
<b>Funding body:</b>	BMBWF (HRSM)
<b>Website:</b>	<a href="https://github.com/philkon/InCritApp">https://github.com/philkon/InCritApp</a> , <a href="https://gams.uni-graz.at/context:epistles">https://gams.uni-graz.at/context:epistles</a>

The digital edition *St Patrick's epistles. Transcriptions of the seven medieval manuscript witnesses* is based on diplomatic transcriptions of the manuscript witnesses of the two surviving epistles of the fifth-century bishop Patrick of Ireland. The electronic transcriptions follow a document-centred editing approach trying to represent the text in its original manuscript layout (Bleier 2017), but still in the edition it was desired to allow comparison and analysis across the transcriptions and existing editions. This short paper describes a tool that uses distance and close reading methods to achieve such a cross-comparison of versions of a text.<sup>1</sup>

The tool is implemented as a client-side web application, making it usable on any system supporting modern web browsers. After starting the visualization tool, users must select a file to import. Once the data is loaded and processed, the user sees details, such as the number of texts and subsections (e.g., chapters), an options section where users can select the preferred comparison method (table or graph) and other additional settings, as well as a heatmap providing an overview of the data. *InCritApp* can easily be integrated into an existing website, as, for example, demonstrated by the implementation into the St Patrick's epistles edition.<sup>2</sup>

The tool uses data formatted in JavaScript Object Notation (JSON) and users must follow a predefined data format which is based on the *CollateX* JSON export format (*CollateX* – documentation 5.1 json). This format is essentially an alignment table structured as a two-dimensional array, where rows represent word tokens in a sequential order, and columns represent variant readings. Additional metadata has to be provided which is used by the tool for processing and displaying purposes: the

---

<sup>1</sup> *InCritApp* was inspired by already existing tools, such as the *Versioning Machine* and *CollateX* and recent research in automated collation by Elisa Nury (Nury 2018).

<sup>2</sup> [https://gams.uni-graz.at/o:epistles.1996/sdef:TEI/get?mode=InCritApp\\_Intro](https://gams.uni-graz.at/o:epistles.1996/sdef:TEI/get?mode=InCritApp_Intro) (26.6.2020).

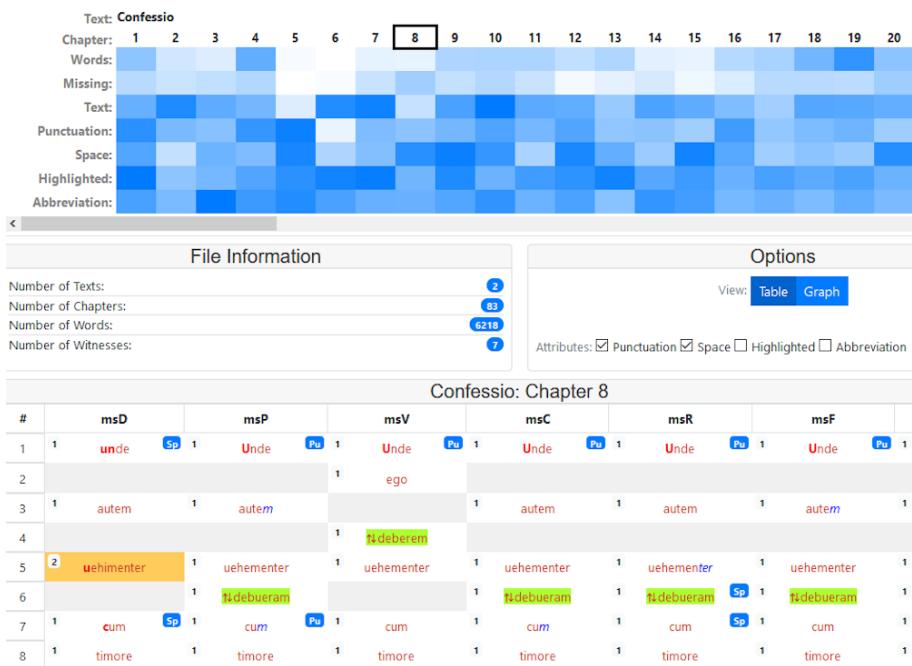


Figure 1: Comparison of the witnesses of St Patrick’s epistles through the table view of *InCritApp*.

version/witness names, a plain text that is used for comparison, a formatted text that is used for display, and optional attributes for visualization can be handed to the application. In case of St Patrick’s epistles, the JSON alignment table is generated from the TEI transcriptions using XSLT. Optional attributes used by this project include information about punctuation, highlighting, spacing between the words and abbreviations used by the medieval scribes.

The heatmap is the point of entry for users, providing them with a distant reading view over all data imported into the tool. It is interpreted as a matrix showing variations (e.g., number of words, number of missing words) across the witnesses and across the subsections of the imported text(s). If the user defined optional attributes in the imported JSON file, they will appear as further rows in the heatmap. Hovering the mouse cursor over a cell will provide details about the respective attribute and subsection. The user can open a detailed comparison view (as table or graph) for individual subsections by clicking on a column in the heatmap. This is the first step into a closer reading of the imported data. To explore the texts further, the user can choose between two different views provided by the *InCritApp* tool: the table and the

graph view. The table view allows for a parallel reading and comparison across all witnesses: each table cell contains a variant reading. The rows follow the sequence of the words as present in the original witnesses and, thus, users can read the text from top to bottom (see Fig. 1). Variations and commonalities between readings are highlighted and indicated by the coloring of table cells and a small number in the top-left corner of a cell. In the graph view, commonalities between the witnesses are combined into single nodes. An edge represents a transition from one word to another and the respective witnesses are shown by a mouse-over effect. For both views the displaying of additional attributes can be controlled in the options section. The presence or absence of punctuation or abbreviations in the witnesses is shown by labels (table view) or by the use of color markers next to a node (graph view).

To conclude, *InCritApp* was developed for a concrete editing project and allows viewing the variant reading of St Patrick’s epistles in different display modes. A user can explore the variations starting from a distant reading view (heatmap) over to a table and graph representation. Each variant in the table view is a hyperlink leading to the word in the diplomatic transcription. Despite being developed in a project context, the tool is open source and can be used with any textual data following our predefined format.

## Bibliography

- Bleier, Roman. 2017. “Digital Documentary Editing of St Patrick’s epistles. Linking the manuscript witnesses to the canonical text.” *Studia Universitatis Babes-Bolyai Digitalia* 62 (1): 9–25. doi:10.24193/subbdigitalia.2017.1.01.
- (ed.). 2019. *St Patrick’s epistles. Transcriptions of the seven medieval manuscript witnesses*. Graz. Accessed December 30, 2020. <https://gams.uni-graz.at/context:epistles>.
- Nury, Elisa. 2018. *Automated collation and digital editions: from theory to practice*. PhD diss., London: King’s College London.
- CollateX. Accessed December 30, 2020. <https://collatex.net>.
- CollateX – documentation. Accessed December 30, 2020. <https://collatex.net/doc>.
- Schreibman, Susan, ed. Versioning Machine 5.0. Accessed September 30, 2020.  
<http://v-machine.org>.

# Grazer Repertorium antiker Fabeln (GRaF, 2017–2019)

Sarah Lang, Ursula Gärtner

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Ursula Gärtner (Projektleitung), Sarah Lang, Alexander Praxmarer, Lukas Spielhofer, Lukas Werzer et al.
<b>Institutionen:</b>	Universität Graz – Institut für Klassische Philologie, Zentrum für Informationsmodellierung
<b>Fördergeber:</b>	BMBWF, Förderreihe Sparkling Science
<b>Website:</b>	<a href="https://gams.uni-graz.at/graf">https://gams.uni-graz.at/graf</a>

Als Output des Sparkling Science-Projektes *Fabula docet – Wer will schon saure Trauben? Grazer Repertorium antiker Fabeln* (GRaF), gefördert vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung (bmbwf), wurde in Zusammenarbeit zwischen dem Institut für Klassische Philologie und dem Zentrum für Informationsmodellierung (ZIM) der Karl-Franzens-Universität Graz ein Repertorium antiker Fabeln entwickelt.

Unter der Leitung von Ursula Gärtner wurde nicht nur eine Edition einer repräsentativen Menge an Primärtexten fachdidaktisch aufbereitet, sondern zur Ergänzung der Primärtexte auch ein fachwissenschaftliches Webportal aufgebaut. Die griechischen und lateinischen Primärtexte selbst werden von einer Übersetzung begleitet und mit Vokabel-, Sach-, Stilistik-, Metrik- und Grammatikangaben angereichert. Fragen zum Text mit dazugehörigen Lösungsvorschlägen werden für jede Fabel im Interpretationsbereich angeboten.

Hierbei wurden jeweils auch textkritische Fragen berücksichtigt. Vergleichstexte in Originalsprache sind mit Übersetzung in einem zusätzlichen Datenstrom verfügbar. Sie können zur Beantwortung der Fragen hinsichtlich intertextueller Bezüge verwendet werden. Hinzu kommt ein kurzes Glossar antiker Autoren mit ihren Hauptwerken. Fachwissenschaftliche Einleitungstexte zur antiken Fabel und zur Textkritik sowie zu den Autoren Aesop, Avian, Babrios und Phaedrus werden neben einer Bibliographie zur antiken Fabel bereitgestellt.

Eine technische Dokumentation sowie Editionsrichtlinien stehen neben den detaillierten Metadaten aus den TEI-Header-Elementen zur Verfügung. Diese vermitteln nicht nur Zusatzinformationen über die Texte selbst, sondern erlauben, die Genese der Ausarbeitungen im Zuge von *Citizen Science* durch am Projekt beteiligte Schulklassen

anhand der Verantwortlichkeiten nachzuvollziehen. Ein Metadaten-Reiter bietet ausführliche Informationen über sowohl Verantwortlichkeiten und Quellentexte als auch weitere für Benutzerinnen und Benutzer möglicherweise relevante Informationen, wie etwa die Zuordnung einzelner Texte zu passenden Modulen des österreichischen Lehrplans. Aufgrund der Lizenzierung als Creative Commons CC-BY-NC kann das Repertorium überdies als *Open Educational Resource* (OER) bezeichnet werden.

Ziel ist, möglichst umfassende wissenschaftliche Ressourcen zum Thema der antiken Fabel für Schülerinnen und Schüler, Lehrerinnen und Lehrer, Studierende und ein wissenschaftlich interessiertes Publikum gleichermaßen bereitzustellen. Die angebotenen Texte sind in TEI/XML kodiert und in der GAMS-Infrastruktur des Zentrums für Informationsmodellierung langzeitarchiviert. Die auf einem *Bootstrap*-Template basierende Webrepräsentation und die LaTeX-basierten PDF-Druckversionen der Texte werden mithilfe von XSLT-Stylesheets nach dem *Single Source*-Prinzip on the fly dynamisch generiert. In der Webrepräsentation bietet das Repertorium eine gewisse Interaktivität in dem Sinne, als Zusatzinformationen hinzuge- oder weggeschaltet werden können. Aufgrund der Schwierigkeiten, die sich in der Langzeitarchivierung und vor allem der dauerhaften Erhaltung der Funktionalität einer im strengen Sinne ‚interaktiven Lernplattform‘ ergeben, wurde hier auf weitere Interaktivität bewusst verzichtet. Das Repertorium ist damit eine lernendenorientierte Digitale Edition,<sup>1</sup> deren Inhalt eine *born digital*-Textausgabe zum Thema der antiken Fabel darstellt.

Es handelt sich beim GRaF-Repertorium weder um eine ‚Digitale Lernplattform‘ noch ein ‚Learning Management System‘. Als ‚digitale Textausgabe‘, die zwar schülerinnen- und schülergerecht aufbereitet ist, aber doch wissenschaftlichen Anspruch hat, erlaubt GRaF einem weiten Publikum von SchülerInnen bis hin zu BachelorstudentInnen erste Erfahrungen im wissenschaftlichen Umgang mit Fabeltexten zu machen. Das Angebot von PDF-Druckversionen richtet sich primär an Lehrende, die durch dieses Angebot an Arbeitsblättern digitale mit analogen Lehrmethoden verbinden können (sog. Blended Learning).

Die Edition wurde mit 30.09.2019 nach einer Projektlaufzeit von 2 Jahren abgeschlossen, eine zukünftige Erweiterung mit neuen Texten im Zuge eines Folgeantrags ist allerdings nicht ausgeschlossen.

## Literatur

- Lang, Sarah und Lukas Spielhofer. 2020. „Digitale Lernplattformen und Open Educational Resources im altsprachlichen Unterricht. Spielräume und Grenzen am Beispiel ‚Grazer Repertorium antiker Fabeln (GRaF)‘.“ *Zeitschrift für digitale Geisteswissenschaften* 5. doi:10.17175/2020\_004.

---

<sup>1</sup> Vgl. auch: Schwinghammer und Schneider 2017.

- Lang, Sarah. 2020. „Die Fabel lehrt – Fabula docet. Das Grazer Repertorium antiker Fabeln als fachdidaktische digitale Ressource.“ In *Funktion und Aufgabe digitaler Medien in Geschichtswissenschaft und Geschichtsunterricht*, hg. v. Krešimir Matijević. Computus Druck: Gutenberg.
- Schwinghammer, Ylva und Gerlinde Schneider. 2017. „sWer so(e) gehoer gelese daz puech. Die deutschsprachige Marginalüberlieferung der Seckauer Margaretenlegende aus der Grazer Handschrift UB, Ms. 781 als Grundlage einer revisionssensiblen, lernerorientierten Digitalen Edition.“ In *Textrevisionen. Beiträge der Internationalen Fachtagung der Arbeitsgemeinschaft für germanistische Edition, Graz, 17. bis 20. Februar 2016*, hg. v. W. Hofmeister und A. Hofmeister-Winter, 19–32. Berlin: deGruyter.

# Franz und Franziska Jägerstätter Edition

Verena Lorber, Joseph Wang-Kathrein

**Mitarbeiterinnen  
und Mitarbeiter:**

Ulrich Lobis, Verena Lorber, Andreas Schmoller, Joseph Wang-Kathrein

**Institutionen:**

Franz und Franziska Jägerstätter Institut, KU-Linz; Forschungsinstitut Brenner-Archiv, Universität Innsbruck

**Website:**

[https://ku-linz.at/forschung/franz\\_und\\_franziska\\_jaegerstaetter\\_institut/forschungsblog/artikel/jaegerstaetter-digital-projekt-zur-erstellung-einer-digitalen-edition](https://ku-linz.at/forschung/franz_und_franziska_jaegerstaetter_institut/forschungsblog/artikel/jaegerstaetter-digital-projekt-zur-erstellung-einer-digitalen-edition)

Im Rahmen des Projektes *Jägerstätter digital* wird an der Katholischen-Privat-Universität Linz eine historisch-kritische digitale Edition des gesamten Nachlasses von Franz Jägerstätter (1907-1943) erstellt. Der Familienvater und Bauer aus St. Radegund (Oberösterreich) zählt zu den bekanntesten Personen, die während des Nationalsozialismus den Wehrdienst aus Gewissensgründen verweigerten und dafür wegen „Zersetzung der Wehrkraft“ zum Tode verurteilt und hingerichtet wurden. Die digitale Edition wird vom Franz und Franziska Jägerstätter Institut umgesetzt und herausgegeben, welches zum zehnjährigen Jubiläum der Seligsprechung des Wehrdienstverweigerers im Oktober 2017 gegründet wurde und im Mai 2018 seine Forschungstätigkeit aufnahm.

Eine der ersten Forschungsaufgaben des Institutes stellte die Digitalisierung und Archivierung sowie Reorganisation (nach RNA – Regeln zur Erschließung von Nachlässen und Autographen) des gesamten Nachlasses dar, der seit 2018 im Besitz der Diözese Linz ist und die Grundlage der Edition bildet. Dieser umfasst 94 von Franz Jägerstätter verfasste, 181 an ihn gerichtete Briefe sowie 101 Schreiben von Personen aus Jägerstättlers Umfeld an vorwiegend Franziska Jägerstätter. Zudem beinhaltet der Bestand vier Hefte mit insgesamt rund 200 Seiten, zahlreiche lose Blätter, ein Notizbuch und Aufzeichnungen aus dem Wehrmachtsuntersuchungsgefängnis Berlin-Tegel mit Gedanken von Franz Jägerstätter. Diesen Bestand gilt es, im digitalen Editionsprojekt dauerhaft für die Nachwelt zu sichern und einer breiten Öffentlichkeit zugänglich zu machen.

Der Fokus liegt auf der Erstellung einer historisch-kritischen Edition, die Annotierungen zu Personen, Orten, Heiligen, Organisationen und Bibelstellen sowie Kommentierungen an Einzelstellen bzw. Flächenkommentare enthält, die mit den entsprechenden

Registern verknüpft sind. Dabei werden – sofern vorhanden – GND-Normdaten und *GeoNames* verwendet.

Im Unterschied zu bisherigen Jägerstätter Editionen werden erstmals alle Schritte, die zur Lesefassung führen, sichtbar. Neben den Faksimiles besteht die Franz und Franziska Jägerstätter Edition aus der diplomatischen Umschrift der in Kurrentschrift verfassten Briefe und Lebensdokumente sowie der Lesefassung, die diese Schriften ins „Heute“ transferiert und verständlich macht. Aufgrund bereits erfolgter buchstaben-, aber nicht zeilengetreuer Transkriptionen eines Großteils der Briefe und Lebensdokumente durch die damalige Leiterin des Diözesanarchivs Linz (Dr.in Monika Würthinger), was bereits im Zuge des Seligsprechungsprozesses passiert ist, ist es möglich, alle Repräsentationsstufen einer Quelle sowie die unterschiedlichen Textformen nebeneinander abzubilden.

Das Besondere des Editionsprojektes stellt die Verbindung zweier zumeist getrennt voneinander durchgeführter Arbeitsschritte dar: die Erstellung einer digitalen Edition und die archivarische Tiefenerschließung des Quellenbestandes. Dazu erfolgte die Datenmodellierung nach ISADG (International Standard Archival Description), um die in XML-TEI Dokumenten erfassten Metadaten in einem Folgeprojekt in das Archivinformationssystem des Linzer Diözesanarchivs einzuspielen. Dort wird der Bestand nach der Erschließung und wissenschaftlichen Bearbeitung durch das Franz und Franziska Jägerstätter Institut dauerhaft verwahrt und zugänglich sein. Zu jedem Brief werden folgende Kernelemente der ISADG erhoben: Signatur, Titel, Sender und Senderinnen und Empfänger und Empfängerinnen, Entstehungszeitpunkt, Ort, Umfang, Form, Sprache, Inhalt, Geschichte/Provenienz, physische Beschreibung der Quelle und Kategorisierung (Korrespondenzen oder Lebensdokumente).

Zur Kodierung der Inhalte und Strukturen sowie zum Austausch der Textdaten wird das XML konforme Textauszeichnungssystem TEI verwendet. Die Datenmodellierung und Auszeichnung der Briefe als XML-TEI Dokumente werden vom Franz und Franziska Jägerstätter Institut durchgeführt. Für die technische Umsetzung und Betreuung des Projektes konnte eine Kooperation mit dem Forschungsinstitut Brenner-Archiv der Universität Innsbruck geschlossen werden, das dem Projekt einerseits für technische Fragestellungen beratend zur Seite steht und andererseits Softwarelösungen (beispielsweise für XML-Schema-Adaptierung und Datenanalysen) entwickelt und dem Projekt zur Verfügung stellt.

Der Mehrwert des Projektes liegt darin, dass es sich nicht nur um eine fortschrittliche und bestandsschonende Darstellung von Quellenmaterial handelt, sondern die Daten jederzeit weiterverarbeitet, neu zusammengestellt, ergänzt oder um weitere Bestände, wie den Nachlass von Franziska Jägerstätter oder andere Kooperationspartner und Kooperationspartnerinnen, erweitert werden können. Zudem vereinfachen Such- und Filterfunktionen das Lesen und die Nutzung der Quellen und durch die Annotationen und Hyperlinks ist die Edition komplex, vernetzt und ermöglicht ein

browsen und suchen innerhalb des Textkorpus. Die Quellcodes werden zur Nachnutzung frei verfügbar und die Inhalte der Edition über eine Creative Commons Lizenz nutzbar sein.

# Stefan Zweig digital

Oliver Matuschek, Christopher Pollin, Lina Maria Zangerl

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Manfred Mittermayer (Projektleitung), Oliver Matuschek, Christopher Pollin, Lina Maria Zangerl
<b>Institutionen:</b>	Literaturarchiv Salzburg, Universität Graz – Zentrum für Informationsmodellierung
<b>Fördergeber:</b>	Land und Stadt Salzburg
<b>Website:</b>	<a href="https://www.stefanzweig.digital">https://www.stefanzweig.digital</a>

Ziel des Projekts Stefan Zweig digital (SZD) ist es, den Nachlass des österreichischen Autors Stefan Zweig (1881–1942) online zugänglich zu machen. Die als Kooperation des *Literaturarchivs Salzburg* und des *Zentrums für Informationsmodellierung* entstehende digitale Nachlassrekonstruktion bietet die Möglichkeit, Originalmaterialien des Autors zu verzeichnen und als digitale Faksimiles raum- und zeitübergreifend zu präsentieren. Nicht zuletzt durch Zweigs Exilzeit ergab sich für die von ihm hinterlassenen Dokumente eine Verteilung auf zahlreiche Sammlungen in aller Welt und eine entsprechend heterogene Überlieferungslage. In der Folge standen Originalmaterialien eines der auflagenstärksten und umfassend vernetzten Autoren des 20. Jahrhunderts nur sehr eingeschränkt zur Verfügung.

Im Rahmen der Nachlassrekonstruktion können durch Verlinkung und Markups inhaltliche, personelle, topographische und zeitliche Verbindungen zwischen den erhaltenen Quellen dargestellt werden. Zweigs Werkmanuskripte, die heute im Literaturarchiv Salzburg und in der Stefan Zweig Collection der State University of New York in Fredonia/USA aufbewahrt werden, wurden im Rahmen von SZD erstmals komplett verzeichnet. Neben den Katalogen zu Werkmanuskripten und Lebensdokumenten stehen auf der Plattform eine biographische Übersicht sowie Register zu *Named Entities* (Personen, Standorte, Werke) auf Basis von Normdaten zur Verfügung. Hinzu kommt ein Verzeichnis der erhaltenen Bücher aus Zweigs Bibliothek, das einen wichtigen Einblick in die von ihm wahrgenommene, gelesene und als Quelle für seine Werke genutzte Literatur bietet (Matthias and Matuschek 2018). Der Datenbestand wurde zudem um das Verzeichnis von Zweigs Autographensammlung mit rund 1000 Objekten erweitert und mit den Katalogen der bestandshaltenden Institutionen verknüpft. Das Projekt basiert ausdrücklich darauf, weitere Sammlungen dazu einzuladen, Metadaten und Digitalisate ihrer Originalmaterialien einzubringen. In diesem Zusammenhang sind auch Datensätze und digitale Faksimiles der National Library of

Israel in SZD übernommen worden und nun zusammen mit den übrigen Beständen durchsuchbar.

Die für SZD erschlossenen Daten wurden im Zuge des Projekts nach TEI und RDF überführt und im FEDORA-basierten, digitalen Repository GAMS (Stigler and Steiner 2018) gehostet. Die Digitalisate entsprechen dem IIIF-Standard und werden über den *Mirador*-Viewer zur Verfügung gestellt. Beim *Ingest* der Daten in die GAMS-Infrastruktur werden die Metadaten und Relationen als RDF abgebildet, mit Normdaten verknüpft (z. B. GND, *GeoNames*, *Wikidata*) und semantisch angereichert. Sowohl formale Kriterien der RNAB als auch inhaltliche Aspekte müssen dabei gleichermaßen beschrieben werden, um adäquate Ordnungsstrukturen aufzubauen zu können. Die RDF-Formalisierung der Daten auf Basis einer domänen-spezifischen Nachlass-Ontologie ([stefanzweig.digital/o:szd.ontology](https://stefanzweig.digital/o:szd.ontology)), die auf Wissensbasen bzw. Top-Level-Ontologien (*CIDOC-CRM*, *Records in Context*, Llanes-Padrón and Pastor-Sánchez 2017) referenziert, wird zum einen der Komplexität des Nachlassmaterials und zum anderen den Ansprüchen an interoperable Daten gerecht. Sämtliche aktuellen *Retrieval*- und *Discovery*-Funktionalitäten von SZD basieren auf RDF-Daten und erlauben es, nach semantischen Konzepten und deren Verknüpfungen wie Personen, Standorte, RNA-Kategorien oder Provenienzkriterien zu suchen.

## Literatur

- CIDOC Conceptual Reference Model (CRM)*. 2019. Version 6.2.6. Zugriff: 17. Dezember 2022. <https://cidoc-crm.org>.
- Llanes-Padrón, D. und Pastor-Sánchez, J.A. 2017. „Records in contexts: the road of archives to semantic interoperability.“ *Program* 51 (4): 387–405.
- Matthias, Stephan und Oliver Matuschek. 2018. *Stefan Zweigs Bibliotheken*, Dresden: Sandstein. *Ressourcenerschließung mit Normdaten in Archiven und Bibliotheken (RNAB) für Personen-, Familien-, Körperschaftsarchive und Sammlungen. Richtlinie und Regeln*. 2022. Version 1.1. Zugriff: 17. Dezember 2022. <https://d-nb.info/1271740966/34>.
- Stigler, Johannes und Elisabeth Steiner. 2018. „GAMS – Eine Infrastruktur zur Langzeitarchivierung und Publikation geisteswissenschaftlicher Forschungsdaten.“ *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 71 (1): 207–16.

# Stefan George Digital

Frederike Neuber

**Team members:**

Frederike Neuber

**Institutions:**

Universität Graz – Zentrum für Informationsmodellierung

**Funding body:**

Marie Skłodowska-Curie Action, Initial Training Network ‘Digital Scholarly Editions Initial Training Network (DiXiT)’

**Website:**

<http://gams.uni-graz.at/context:stgd> and

<https://github.com/FrederikeNeuber/stgd-prototype-edition>

The typographical dress of literature shapes our understanding and interpretation of the text, especially when it is such an integral part of poetry as in the literary work of Stefan George (1868–1933). At the end of the 19th century, in times of industrialization of the book market, George put special emphasis on the typographic design of his works, which is either simple or splendid, but always intentional. Thus, design features such as format or layout can have an impact on the literary dimension of the works. In addition, the individualized *St-G-typeface*, a sans serif typeface which is based on George’s handwriting and historical scripts, is an exceptional phenomenon in literary history. Its versions, form repertoire and textual function have not yet been researched enough.

With Stefan George Digital (StGD) a selection of historical prints of George’s poetical works has been edited for the first time and is now available online for further exploration. StGD has been developed as a prototype of a ‘typography-centered edition’ within the editor’s dissertation at the Centre for Information Modelling at the University of Graz and the *Digital Scholarly Editions Initial Training Network DiXiT* (funded 2013–2017 under Marie Curie Actions within the European Commission’s *7th Framework*).

The focus of StGD is the editorial enrichment of information on typography in general and on microtypography in particular, i.e. of the design and use of typefaces and fonts. Currently, StGD includes 25 printed editions of George’s poetical works published between 1897 and 1922. The single works are represented in the corpus variable numbers of times according to their typographical variation. All sources are enriched by bibliographical metadata and digital facsimiles. Four typographical versions of “Der Teppich des Lebens und die Lieder von Traum und Tod mit einem Vorspiel” (1899–1932) are available as edited full texts and with regard to their typographical variance at line level. In addition, all sources are enhanced by descriptions

of the applied fonts based on a description ontology that has been developed for this purpose. This ontology also lays the foundation for the microtypographical inventory, which is part of the edition and a research tool itself. It offers three exploration modes: the type repertoire (Typenrepertoire), e.g. to retrace the genesis of the St-G-typeface, the form repertoire (Formenrepertoire), e.g. to observe the transmission of design features between typefaces, and the typeset repertoire (Satzrepertoire), e.g. to explore the application of fonts in certain text contexts.

The technical ecosystem of StGD is the FEDORA-based asset management system GAMS (*Geisteswissenschaftliches Asset Management System*) that guarantees long-term archiving. The digital images are provided via IIIF in several viewers (*Mirador* and *Open Seadragon*), bibliographical metadata and texts are encoded in XML/TEI. The type descriptions are based on an ontology in RDF/OWL. Since the latter is documented and can be referenced via stable URIs, it can be further used as a Semantic Web resource to be implemented in other projects as well.

## Bibliography

- DiXiT-Website. Accessed February 1, 2020. <http://dixit.uni-koeln.de>.
- Kurz, Stephan. 2007. *Der Teppich der Schrift. Typografie bei Stefan George*. Frankfurt am Main/Basel: Stroemfeld.
- Lucius, Wolf D. von. 2012. “Buchgestaltung und Typographie bei Stefan George.” In *Stefan George und sein Kreis. Ein Handbuch*, ed. by Achim Aurnhammer, Wolfgang Braungart, Stefan Breuer, Ute Oelmann, 467–91. Berlin/Boston: De Gruyter.
- Neuber, Frederike. 2020. *Das Konzept einer typografiezentrierten digitalen Edition der Werke Stefan Georges samt einem Modell zur Beschreibung von Mikrotypografie*. PhD diss, Cologne: University of Cologne.
- Neuber, Frederike 2017. “Typografie und Varianz in Stefan Georges Werk. Konzeptionelle Überlegungen zu einer ‘typografiekritischen’ Edition.” In *Editio* 31: 205-32. Berlin/Boston: De Gruyter. doi:10.1515/editio-2017-0012.

# Karl Wiesinger: Digitale Edition der Tagebücher (1961–1973)

Helmut Neundlinger, Selina Galka

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Georg Hofer (Projektleitung), Selina Galka, Helmut W. Klug, Helmut Neundlinger, Elisabeth Steiner, Michaela Thoma-Stammler
<b>Institutionen:</b>	Adalbert-Stifter-Institut des Landes Oberösterreich, Universität Graz – Zentrum für Informationsmodellierung
<b>Fördergeber:</b>	BMBWF (HRSM)
<b>Website:</b>	<a href="https://gams.uni-graz.at/wiesinger">https://gams.uni-graz.at/wiesinger</a>

Der Linzer Autor und Publizist Karl Wiesinger (1923–1991) war ein Mann der Extreme, nicht nur im politischen Sinn: Bis zuletzt überzeugter Kommunist und zuweilen Verehrer so dunkler Figuren der Geschichte wie Stalin und Mao, führte er über 40 Jahre hinweg ein Tagebuch, dessen Umfang und Radikalität seinesgleichen sucht.

Es ist in vielerlei Hinsicht ein Zeitzeugen-Leben, das sich in der Intimität der persönlichen Niederschrift dokumentiert: Von den Bürgerkriegshandlungen im Februar 1934, die Wiesinger als Kind in Linz miterlebte, über den Austrofaschismus, den Einmarsch der Nazis, ein Soldatenleben zwischen Fronteinsatz und Widerstand, die Nachkriegszeit mit dem alles bestimmenden Kalten Krieg bis zum Zusammenbruch des real existierenden Sozialismus 1989/90 hat Wiesinger in seinen Aufzeichnungen alles beschrieben und kommentiert. Einsetzend mit einer Art Rückblende, die er im Jahr 1950 anlegt und in der er seine persönlichen Erlebnisse und Eindrücke seit dem Jahr 1938 festhält, betreibt er die Arbeit am Tagebuch mit ausdauernder Regelmäßigkeit bis kurz vor seinem Tod im Februar 1991 (Schnalzer-Beiglböck 1995, 45). Das Tagebuch begleitet sein Leben und vor allem sein Schreiben.

Als der Nachlass Wiesingers im Jahr 2012 von einer Nichte des Autors an das Adalbert-Stifter-Institut des Landes Oberösterreich übergeben wurde, fanden sich darin neben Typoskripten und Dokumenten zu Leben und Werk auch die Tagebücher, allerdings nur jene aus den Jahren 1961 bis 1973, der Rest muss als verschollen gelten. In Kooperation mit dem Franz-Nabl-Institut für Literaturforschung und dem Zentrum für Informationsmodellierung der Uni Graz wurde im Herbst 2017 eine umfassende Edition der verbliebenen Tagebücher vereinbart und in Angriff genommen.

Die Tagebücher umfassen im Original rund 750 eng beschriebene Typoskriptblätter. Die mittels Texterkennung zunächst in ein Word-Format umgewandelte Textmenge

belief sich auf ein Dokument von ca. 850 Seiten (ca. 2 Millionen Zeichen). Inhaltlich begleitet das Tagebuch einerseits Wiesingers schriftstellerische Entwicklung vom Dramatiker zum Romancier, andererseits die für diesen Schritt nicht unwesentliche Zuspitzung des Kalten Krieges zwischen West und Ost. Berücksichtigt man die in den Tagebüchern wiederkehrenden Rückblenden zur eigenen Biographie bis in die Kindheitstage zurück, lässt sich dem Dokument eine umfassende Bedeutung hinsichtlich Leben und Werk Wiesingers zuschreiben.

Nicht nur die materielle Form der Überlieferung (Maschintyposkript auf zum Teil qualitativ dürftigem Papier) legte eine Sicherung bzw. Umwandlung der ursprünglichen Form in eine digitale Fassung nahe. Die grundlegenden Problemstellungen für eine Edition ergaben sich sowohl aus der relativen Unbekanntheit bzw. weitgehenden Vergessenheit des Autors als auch aus der starken zeit- und lokalgeschichtlich bedingten Kontextabhängigkeit.

Eine solche Quelle ist zugleich reizvoll und uferlos sowie durchaus problematisch, was die Gestalt und die Umsetzung einer Edition betrifft. Die digitale Form bietet zunächst die Möglichkeit, eine so umfangreiche Quelle nicht nur in ihrer Gesamtheit zugänglich zu machen, sie kann auch die vielfältigen Verknüpfungen, Verbindungen und Entwicklungen lesbar machen, die sich über den Verlauf der Tagebücher zeigen und ausformen. Ein wichtiges Vorbild für die methodische Kommentierung war diesbezüglich das Editionsprojekt der Tagebücher des deutschen Schriftstellers und politischen Aktivisten Erich Mühsam (1878–1934), die den Zeitraum von 1910 bis 1924 umfassen und in einer Hybrid-Edition in Kooperation mit dem Berliner Verbrecher Verlag von Chris Hirte und Conrad Piens herausgegeben worden sind. Die Mühsam-Tagebücher weisen einige Merkmale auf, die signifikante Parallelen zu jenen Wiesingers ergeben: die literarisch-politische Sozialisation in einer bohèmehaften Kneipenkultur, der Zug zum ständigen Agitieren, die dauerhaft prekäre ökonomische Lage, die Einschreibung der Zeitgeschichte in die Biographie sowie die Neigung zur detaillierten Bilanzierung eines ausschweifenden Sexuallebens. Zudem sichert die puristische Umsetzung der digitalen Mühsam-Edition eine hohe Konzentration auf den Primärtext – ein Zugang, der auch die Gestaltung der Wiesinger-Edition von Beginn an leiten sollte.

Ein wesentlicher Unterschied besteht im simplen Faktum des Entstehungszeitraums: Mühsams Tagebücher wurden vor gut hundert Jahren verfasst, die darin erwähnten Personen sind seit langem tot. Bei den Tagebüchern Wiesingers gestaltet sich die Edition hinsichtlich der Persönlichkeitsrechte deutlich problematischer. Wiesingers oft boshafte und beleidigende Bemerkungen über Bekannte, politische Weggefährten oder Schriftstellerkollegen sind an mehr als einer Stelle dazu angetan, zumindest die Gefühle von noch lebenden Familienmitgliedern oder Freunden der im Tagebuch verunglimpften oder attackierten Personen zu verletzen. Eine entscheidende Frage war nun, wie man ein Tagebuch wie jenes von Wiesinger im Sinne der Bereitstellung eines

authentischen Textdokuments überliefert und in welchem Verhältnis man schützend eingreifen muss.

Eine Hauptaufgabe der Arbeit im Rahmen der Edition war es, die schillernden Widersprüchlichkeiten einzufangen, sowohl was das Biographische als auch was das Literarische betrifft. Die digitale Edition der Tagebücher verfügt deshalb über drei Komponenten, die den Text wesentlich ergänzen bzw. kontextualisieren:

1. eine biographische Einführung zu Karl Wiesinger, die seinen komplexen Lebenslauf darstellt und die mit ausgewählten Dokumenten-Scans aus dem Nachlass illustriert ist;
2. die direkte Kommentierung des Textes mittels Register (Personen, Orte, Institutionen als TEI-Dokumente mit Verknüpfungen zu Normdateien wie GND, VIAF und GeoNames sowie weiterführenden Links; Schlagworte modelliert als SKOS-Thesaurus);
3. überblicksartige Kommentare zu jenen thematischen Schwerpunkten, die sich in den Tagebüchern über die Jahre herauskristallisierten und entwickelten. Dies sind – neben den Reiseberichten, die eine Art autonomes Subgenre im Tagebuch darstellen – biographische, werkgeschichtliche sowie politische Kategorien. Diese insgesamt neun Themen werden in den Originaltexten annotiert und in der Webanzeige als eigene Features farblich markiert ausgewiesen. Sichtbar werden dadurch auch zeitlich bedingte Konjunkturen bzw. persönliche Entwicklungen am Korpus der Tagebücher entlang.

Die Digitale Edition wurde mittels des *Geisteswissenschaftlichen Asset Management Systems* (GAMS) am Zentrum für Informationsmodellierung umgesetzt.

Für die digitalisierten Tagebuchtexte wurde gemäß der Richtlinien der *Text Encoding Initiative* ein Datenmodell entwickelt. Die Annotation erfolgte mittels vordefinierter Formatvorlagen für die jeweiligen Entitäten (z.B. Personen, Orte, Institutionen) und anderen Informationen (z.B. Datumsangaben, Anonymisierungen oder Auslassungen durch den Editor) in Word-Dokumenten. Die annotierten Word-Dokumente wurden mit *OxGarage* und XSLT-Stylesheets nach TEI/XML transformiert und letztendlich als TEI/XML-Dokument pro Jahr persistent zitierbar in der GAMS langzeitarchiviert. Mittels XSLT-Stylesheets werden die TEI-Dokumente *on the fly* im Web zur Anzeige gebracht, wobei hier das CSS-Framework Bootstrap genutzt wird.

## Literatur

Schnalzer-Beiglböck, Christiane. 1995. *Karl Wiesinger (1923–1991). eine Monographie unter besonderer Berücksichtigung der Theaterarbeit*. Diplomarbeit, Uni Wien.

# Celtic Divine Names in the Inscriptions of the Roman Province Germania Inferior

Werner Petermandl, Elisabeth Steiner

<b>Team members:</b>	Wolfgang Spickermann (Projektleiter), Werner Petermandl, Verena Reiter, Astrid Schmölzer, Elisabeth Steiner
<b>Institutions:</b>	Universität Graz – Institut für Alte Geschichte und Altertumskunde, Zentrum für Informationsmodellierung
<b>Funding body:</b>	FWF Der Wissenschaftsfonds (P 29274-G25), BMBWF (HRSM)
<b>Website:</b>	<a href="https://gams.uni-graz.at/fercan">https://gams.uni-graz.at/fercan</a>

Within the FWF project *Celtic Divine Names in the Inscriptions of the Roman Province Germania Inferior. A Case Study on Religion in the Context of Cultural Contacts and Cultural Transfer* (P 29274-G25), a digital edition was created that for the first time assembles all Celtic divine names attested in inscriptions of Germania Inferior. The edition further considers the inscribed objects and their iconography. It will, in particular, serve to examine the religious situation in an area which is characterised by the clash of different religious traditions and cultic practices as a result of its incorporation under Roman rule. Epigraphical sources are especially important for the analysis of the relevant phenomena, as in the entire Roman Empire there are hundreds of Celtic divine names recorded in inscriptions. On the other hand, only ten Celtic divine names can be found in literary sources. However, the edition will not only be useful to focus on religious aspects in a narrower sense but also on social issues and corresponding mentalities of the dedicants worshipping divinities addressed by Celtic names. Furthermore, research on the spread and persistence of the Celtic language in the Roman province Germania Inferior will profit from this endeavor. On the whole, the project can be considered a helpful tool to contribute to the overall topic of Romanization.

In addition to a printed volume providing a thorough analysis, the digital edition of the relevant inscriptions offers immediate access to the *EpiDoc* data forming the basis of research. The objective was to create the digital representation from the data already gathered with as little extra effort as possible. This entailed especially the creation of a conversion process from the Leiden Conventions for epigraphic markup to *EpiDoc*. A tailor-made approach was implemented which partly drew on already existing solutions and guidelines (reference for conversion of Leiden markup to *EpiDoc*, *Chapel*

*Hill Electronic Text Converter – CHET-C*). The data were enriched with references to *Pleiades*, *Trismegistos*, Clauss/Slaby, *Epigraphical Database Heidelberg*, *ubi erat lupa* and a project specific thesaurus for the archaeological and iconographic description of the inscriptions and their carriers. Exact geographic location of the modern finding spots was also added where possible. Resulting TEI data and facsimiles are stored in the trusted digital repository GAMS. The publication layer was implemented based on the *Bootstrap* framework, images are provided by a IIIF-compliant image server and viewer. Inscriptions can be browsed via fulltext search, structured indices and map views. All data is published under Creative Commons BY-NC.

The digital edition not only allows for time and place independent access to the inscriptions, but also enables constant updating of the research data. Furthermore, the linked open data approach maximizes re-use by dissemination of the data to platforms like *Pelagios Commons* and *Trismegistos*. The project exemplifies fruitful interdisciplinary cooperation between the fields of Ancient History, Archaeology and Digital Humanities, opening up the data to all research communities.

## Bibliography

- Bodard, Gabriel, Elli Mylonas, Tom Elliott, Simona Stoyanova, Charlotte Tupman, and Irene Vagionakis. “EpiDoc Guidelines: Ancient documents in TEI XML.” Accessed February 1, 2020. <http://www.stoa.org/epidoc/gl/latest/index.html> and <https://sourceforge.net/projects/epidoc>.
- Cayless, Hugh, Martina Filosa, Gabriel Bodard, Tom Elliott, and Scott DiGiulio. 2011. “Chapel Hill Electronic Text Converter (CHET-C).” Accessed February 1, 2020. <http://cds.library.brown.edu/projects/chet-c/chetc.html> and <https://sourceforge.net/p/epidoc/code/HEAD/tree/trunk/chetc-js>.
- “Geisteswissenschaftliches Asset Management System (GAMS).” hdl:11471/521.1.
- “Pelagios Commons.” Accessed February 1, 2020. <http://commons.pelagios.org>.
- “Pleiades.” Accessed February 1, 2020. <https://pleiades.stoa.org>.
- Spickermann, Wolfgang. 2018. “Celtic Divine Names in the Inscriptions of the Roman Province Germania Inferior.” hdl:11471/504.50.
- “Trismegistos.” Accessed February 1, 2020. <https://www.trismegistos.org>.

# ABaC:us – Austrian Baroque Corpus

Claudia Resch

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Claudia Resch (Projektleitung), Ulrike Czeitschner, Matej Ďurčo, Barbara Krautgartner, Daniel Schopper, Eva Wohlfarter
<b>Institutionen:</b>	ACDH-CH ÖAW
<b>Fördergeber:</b>	Kultur-, Wissenschafts- und Forschungsförderung der Stadt Wien (Forschungstipendium <i>Barocke literarische Totentänze von und mit Abraham a Sancta</i> und Projekt <i>Totentanz und Jenseitsvorsorge in Wien: Barocke Bruderschaftsdrucke als Forschungsgegenstand der digitalen Geisteswissenschaften</i> , Nr. LWI 0240 I-III), Österreichische Nationalbank (Projekt <i>Texttechnologische Methoden zur Analyse österreichischer Barockliteratur</i> , Nr. 14738)
<b>Website:</b>	<a href="https://acdh.oeaw.ac.at/abacus/">https://acdh.oeaw.ac.at/abacus/</a>

*Austrian Baroque Corpus (ABaC:us)* ist die Bezeichnung für eine digitale Sammlung von literarischen Texten aus der Barockzeit. Die edierten Texte, die digital gelesen, durchsucht und nachgenutzt werden können, sind Teil einer reichhaltigen und weit verbreiteten Literatur, die sich mit der Vorbereitung auf den Tod beschäftigt und heute beinahe in Vergessenheit geraten ist. Den thematischen Kern der Sammlung bilden Erstdrucke theologisch-erbaulichen Inhalts, insbesondere *Ars moriendi*-, Totentanz- und *Memento mori*-Literatur, die dem bekannten Barockprediger Abraham a Sancta Clara (1644–1709) und seinem Umfeld zugeschrieben werden. Das Kernkorpus im Umfang von etwa 180.000 Token ist seit 2015 online zugänglich und über ein benutzerfreundliches Interface abrufbar.

Zur Erstellung originalnaher Transkriptionen wurden die aus verschiedenen Bibliotheken stammenden Bilddigitalisate der raren Erstdrucke mit XML und verwandten Technologien zu maschinenlesbarem Text verarbeitet und gemäß der Richtlinien der Text Encoding Initiative (TEI) in der Version P5 erschlossen. Die digitalen Texte haben mehrfache Kollationierungsdurchgänge durchlaufen, wurden sorgfältig hinsichtlich ihrer Qualität überprüft und geben den historischen Sprachstand quellengetreu wieder. Offensichtliche Textfehler der Originale wurden in der Transkription zudem ausgewiesen und mit editorischen Anmerkungen versehen. In den einzelnen Werken wurden außerdem historische, biblische und mythologische Personennamen sowie Ortsnamen annotiert, wobei in der letztgenannten Kategorie die Namen von Bergen, Gewässern, Regionen, Ländern und Kontinenten, Städten und Dörfern sowie Straßen und Plätzen in Wien ausgezeichnet wurden.

Im Hinblick auf die Durchsuchbarkeit und Nachnutzung wurden die historischen Texte mit einer manuell korrigierten linguistischen Basisannotation versehen, die folgende Verarbeitungsschritte vorsah: 1. Tokenisierung, 2. Wortartklassifizierung (*Part-of-Speech Tagging*) und 3. Lemmatisierung. Als Klassifikationssystem wurde hierfür das 54-teilige *Stuttgart-Tübingen-TagSet* (STTS) herangezogen und – der Sprache des älteren Neuhochdeutsch entsprechend – geringfügig erweitert. Bei der Lemmatisierung dienten das *Deutsche Wörterbuch* von Jacob und Wilhelm Grimm sowie der *Duden* als Referenzwerke.

Das Ergebnis dieses aufwändigen Annotationsverfahrens sind handverlesende Daten, deren Qualität *ABaC:us* zu einer gefragten Ausgangsbasis für weiterführende Forschungsfragen macht. Sein Wert liegt in der Wiederverwendbarkeit von bereits erarbeitetem Wissen über die Texte, was bedeutet, dass die Projektgruppe selbst, aber auch Nutzerinnen und Nutzer die Annotationen zeitsparend und gewinnbringend für ihre Erkenntnisinteressen einsetzen können.

Die beispielgebende *ABaC:us*-Webapplikation basiert auf dem modularen Publikations-Framework *corpus\_shell* und wurde mit der Intention entwickelt, zeitentfernte, in diversen Bibliotheken beherbergte Drucke aus der Barockzeit für wissenschaftliche Fragestellungen zugänglich zu machen. Das Projektteam hat sich für dieses Interface entschieden, weil es – einem abrahamischen Buchtitel folgend – dazu geeignet ist, *Etwas für alle* (1699) zu bieten, indem es an unterschiedlichen Fragestellungen orientiert ist und mehrere Nutzungsszenarien ermöglicht und unterstützt. So bietet das Interface Lesenden einerseits eine Lektüreansicht mit Faksimiles und stellt Suchenden andererseits Navigationsinstrumente und eine Suchmaschine zur Verfügung, die den direkten Zugang in den Text erlauben – etwa über Inhaltsverzeichnisse, Register und die freie Volltextsuche nach Wort, Lemma oder Wortart. Für weitere wissenschaftliche Nachnutzungsszenarien, die über diese antizipierten Fragestellungen hinausgehen, sind außerdem die Daten im XML/TEI-Format zum Download abrufbar.

Es ist zu wünschen, dass *ABaC:us* weiterhin in Forschung, Lehre und Unterricht rezipiert wird und darüber hinaus auf die Neugier einer interessierten, web-affinen (Fach-)Öffentlichkeit trifft, die Zitate von Abraham a Sancta Clara nicht nur in Anekdoten oder Blütenlesen, sondern auch im unveränderten Wortlaut nachlesen und suchen möchte.

Abbildung 1: *ABaC:us*-Webapplikation.

## Literatur

- Czeitschner, Ulrike und Claudia Resch. 2016. „Repräsentation von | in barocken Buch-Totentänzen im digitalen Medium.“ In *Repräsentation(en). Interdisziplinäre Annäherungen an einen umstrittenen Begriff*, hg. v. Gernot Gruber und Monika Mokre, 35–49. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Resch, Claudia, Ulrike Czeitschner, Eva Wohlfarter und Barbara Krautgartner. 2016. „Introducing the Austrian Baroque Corpus: Annotation and Application of a Thematic Research Collection.“ In *Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age*. Zugriff: 20. September 2019. [http://ceur-ws.org/Vol-1681/Resch\\_et\\_al\\_austrian\\_baroque\\_corpus.pdf](http://ceur-ws.org/Vol-1681/Resch_et_al_austrian_baroque_corpus.pdf).
- und Wolfgang U. Dressler. 2016. „Zur Pragmatik der Diminutive in frühen Erbauungstexten Abraham a Sancta Claras. Eine korpusbasierte Studie.“ In *Linguistische Pragmatik in historischen Bezügen*, hg. v. Peter Ernst und Martina Werner, 235–50. Berlin und Boston: de Gruyter.
- . 2017. „»Etwas für alle« – Ausgewählte Texte von und mit Abraham a Sancta Clara digital.“ *Zeitschrift für digitale Geisteswissenschaften*. doi:10.17175/2016\_005.
- und Ulrike Czeitschner. 2017. „Morphosyntaktische Annotation historischer deutscher Texte: Das Austrian Baroque Corpus.“ In *Digitale Methoden der Korpusforschung in Österreich*, hg. v. Claudia Resch und Wolfgang U. Dressler, 39–62. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- . 2019. „Linguistisch annotierte historische Texte stilistisch auswerten. Musterhaft vorkommende Wortverbindungen im Austrian Baroque Corpus.“ In *Historische Korpuslinguistik*, hg. v. Renata Szczepaniak, Stefan Hartmann und Lisa Dücker, 368–85. Berlin und Boston: de Gruyter.
- . 2022. „Digital Approaches to Analyzing and Understanding Baroque Literature.“ In *New Technologies and Renaissance Studies III*, hg. v. Colin Wilder und Matt Davis, 63–86. New York und Toronto: Iter Press.

# DIGITARIUM – Das Wien[n]erische Diarium digital

Claudia Resch, Nora Fischer, Dario Kampkaspar, Daniel Schopper

**Mitarbeiterinnen und Mitarbeiter:** Claudia Resch (Gesamtprojektleitung), Nora Fischer, Dario Kampkaspar, Michael Pölzl, Nina Claudia Rastinger, Daniel Schopper

**Institutionen:** ACDH-CH ÖAW, Institut für kunst- und musikhistorische Forschungen der Österreichischen Akademie der Wissenschaften (Projektleitung: Anna Mader-Kratky), Forschungszentrum Digital Humanities der Universität Innsbruck (Projektleitung: Günter Mühlberger)

**Fördergeber:** Programm „go!digital 2.0“ der Österreichischen Akademie der Wissenschaften, Kultur-, Wissenschafts- und Forschungsförderung der Stadt Wien (MA7)

**Website:** <https://digitarium.acdh.oeaw.ac.at>

Das *DIGITARIUM* ist eine Digitale Edition von mehreren hundert Ausgaben der Wiener Zeitung (vormals *Wien[n]erisches Diarium*) aus dem 18. Jahrhundert, die im Volltext erschlossen werden. Das *Wien[n]erische Diarium* gilt als das bedeutendste Medium der Habsburgermonarchie des 18. Jahrhunderts. Allein die Tatsache, dass die Zeitung seit 1703 im langen 18. Jahrhundert kontinuierlich zweimal wöchentlich erschien, macht sie zu einer außerordentlich wertvollen Informationsquelle für Wissenschaftlerinnen und Wissenschaftler verschiedenster Disziplinen (u. a. aus der Mediengeschichte, Kunstgeschichte, Musikgeschichte, Literaturwissenschaft, Zeremonialforschung, Namenforschung und der historischen Linguistik) sowie für eine interessierte Öffentlichkeit.

Im Rahmen des Projekts *Das Wien[n]erische Diarium: Digitaler Datenschatz für die geisteswissenschaftlichen Disziplinen* wird eine exemplarische Auswahl von ca. 300 Ausgaben des *Wien[n]erischen Diarium* über den gesamten Verlauf des 18. Jahrhunderts als Trainingskorpus bearbeitet, um einerseits den angesprochenen vielfältigen Forschungs- und Erkenntnisinteressen Rechnung zu tragen und andererseits neue Features des Interface- und Interaktionsdesigns sowie Anforderungen und Ideen für die Entwicklung der Forschungsumgebung zu testen.

Die Erstellung der hochwertigen Volltexte beginnt bei der Bearbeitung der Bilddigitalisate, welche zu einem überwiegenden Teil auf dem Portal *AustriaN Newspapers Online* (ANNO)<sup>1</sup> der Österreichischen Nationalbibliothek zur Verfügung gestellt werden. Da nur eine hohe Bildqualität die korrekte automatische Weiterverarbeitung

---

<sup>1</sup> <http://anno.onb.ac.at/cgi-content/anno?aid=wrz>.



Abbildung 1: Das Wien[n]erische Diarium.

ermöglicht, ist es vereinzelt notwendig, diese Bilder zu verbessern, bevor sie zur automatischen Layoutanalyse (OLR) und Texterkennung (OTR) im Tool *Transkribus*<sup>2</sup> hochgeladen werden können. Nach der erfolgreichen Erkennung werden die generierten Volltexte exportiert und von einer Reihe von XSLT-Skripten zu XML-Dateien im aktuellen Standard der Text Encoding Initiative (TEI P5) nachbearbeitet. Hier hat die Erfahrung gezeigt, dass aufgrund von Problemen mit dem Originaldruck (Flecken, Fältchen, Löcher und Risse des Drucks, undeutliche Buchstaben usw.), dem digitalen Bild (fehlende Schärfe usw.) oder dem Layout (Änderungen innerhalb einer Ausgabe oder des Gesamtlayouts) manuelle Korrekturen erforderlich sind, um die angestrebte hohe Textgenauigkeit zu erreichen. Diese Korrekturprozesse tragen wiederum dazu bei, Vorstufen der automatisierten Verarbeitung iterativ zu verbessern, wodurch das mittlerweile anhand von großen Textmengen trainierte Modell auch zur digitalen Transformation anderer historischer Zeitungen (z. B. der *Zürcher Zeitung*) angewandt werden kann.<sup>3</sup>

<sup>2</sup> <https://transkribus.eu/Transkribus>.

<sup>3</sup> Das Model wird unter <https://readcoop.eu/de/modelle/german-fraktur-18th-century> zur Verfügung gestellt.

Die abschließend korrigierten Transkriptionen werden so mit dem Bild kollationiert, dass die digitalen Volltextversionen das *Wien[n]erische Diarium* seiten-, zeilen-, und zeichengetreu abbilden und damit den historischen Sprachstand der Texte unverändert wiedergeben. Die solcherart mit der Erstellung der Volltexte verbundene Analyse von formalen Charakteristika (Vignetten, Schriften, Rubriken) wird einen wichtigen Beitrag zur Erforschung der Entwicklung des Zeitungslayouts leisten, das generell für das 18. Jahrhundert noch kaum untersucht ist, indem sie die Profilierung und Weiterentwicklung des vertretenen Spektrums an Textsorten zutage fördert und die Untersuchung von Techniken der Leserorientierung und Verständnisförderung ermöglicht.

Das *DIGITARIUM* bietet derzeit (Stand: 2020) eine Übersicht über den Bearbeitungsstand der einzelnen Ausgaben, deren Volltexte und Faksimiles synoptisch dargestellt und durchsucht werden können. Als Forschungsumgebung will das *DIGITARIUM* einer interdisziplinären Community neue Möglichkeiten zur Erschließung des *Wien[n]erischen Diarii* ermöglichen und innovative Forschung fördern. Die Expertise, welche die Projektgruppe bei der – auch international gesehen – noch wenig ausverhandelten Transformation von historischen Zeitungen erworben hat, teilt sie mit der Fachgemeinschaft in zwei neu gegründeten Arbeitsgruppen – *Newspapers and Periodicals der Text Encoding Initiative (TEI)* und *Zeitungen und Zeitschriften* der Digital Humanities im deutschsprachigen Raum (DHd).

## Literatur

- Fischer, Nora. 2019. „Von Orten im Wien[n]erischen Diarium. Anmerkungen zu den Voraussetzungen einer Annotation von Ortsnamen.“ *Wiener Geschichtsblätter* 74 (2): 137–49.
- Kampkaspar, Dario, Daniel Schopper und Claudia Resch. 2017. „Wiennerisches Diarium Digital: Involving users in scholarly annotation.“ In *TEI conference and members' meeting 2017. Book of Abstracts*. Zugriff: 20. September 2019. [https://hcmc.uvic.ca/tei2017/abstracts/t\\_111\\_kampkasparetal\\_scholarlyannotation.html](https://hcmc.uvic.ca/tei2017/abstracts/t_111_kampkasparetal_scholarlyannotation.html).
- Kampkaspar, Dario. 2019. „Das *DIGITARIUM* – Volltexterstellung und Nutzungsmöglichkeiten.“ *Wiener Geschichtsblätter* 74 (2): 131–5.
- Mader-Kratky, Anna, Claudia Resch und Martin Scheutz. 2019. „Das Wien[n]erische Diarium im 18. Jahrhundert. Neue Sichtweisen auf ein Periodikum im Zeitalter der Digitalisierung.“ *Wiener Geschichtsblätter* 74 (2): 93–113.
- Resch, Claudia. 2022. „Volltextoptimierung für die historische Wiener Zeitung mit einem Anwendungsszenario aus der germanistischen Sprachgeschichte.“ In *Digitised Newspapers – a New Eldorado for Historians? Epistemology, methodology, tools and the changing practice of history in the context of mass digitisation of newspapers*, ed. by Estelle Bunout, Frédéric Clavert und Maud Ehrmann. Berlin, Boston: de Gruyter, 89–112.
- . 2018. „‘Zeitung Lust und Nutz’ im digitalen Zeitalter. Partizipative Ansätze zur Erschließung historischer Ausgaben der Wiener Zeitung.“ *medien & zeit. Kommunikation in Vergangen-*

*heit und Gegenwart* 2: 20–31.

- , Dario Kampkaspar und Daniel Schopper. 2018. „Historische Zeitungen kollaborativ erschließen: Die älteste, noch erscheinende Tageszeitung der Welt „under annotation.“ In *Kritik der digitalen Vernunft. DHd 2018 Köln. Konferenzabstracts*, hg. v. Georg Vogeler, 229–32. Zugriff: 20. September 2019. <http://dh2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>.
- . 2019. „Das Wien[n]erische Diarium und seine digitale Erschließung oder ‚Was Zeitungsleser vor Geräte haben müssen?‘“ *Wiener Geschichtsblätter* 74 (2): 115–29.
- und Dario Kampkaspar. 2019. „DIGITARIUM – Unlocking the Treasure Trove of 18th-Century Newspapers for Digital Times.“ In *Digital Eighteenth Century: Central European Perspectives*, hg. v. Thomas Wallnig, Marion Romberg und Joelle Weis, 49–64. Wien, Köln u. Weimar: Böhlau.

# Virtuelle Benediktinerbibliothek Millstatt

Sabine Seelbach

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Sabine Seelbach (Projektleitung), Petra Fiedler, Christa Herzog, Birgit Müllner-Stieger, Norbert Hunor Orban, Cornelia Ortner, Petra Schebach
<b>Institutionen:</b>	Universität Klagenfurt – Institut für Germanistik, Universitätsbibliothek Klagenfurt
<b>Fördergeber:</b>	BMBWF (HRSM); Forschungsrat der AAUK
<b>Website:</b>	<a href="https://virtbibmillstatt.com/">https://virtbibmillstatt.com/</a>

Virtuelle Bibliotheken sind Webportale, die räumlich entfernte Leseobjekte physischer oder elektronischer Daseinsform in einem digitalen Wissensraum zusammenführen und simultan verfügbar machen. Abseits der seit den 1990er Jahren bereits genutzten Methoden zur Volltextsuche und Literaturrecherche, die die traditionelle Bibliothek als materiellen Ort des Lesens begrenzter, physisch verfügbarer Bestände tendenziell ablösten, war das Konzept der virtuellen Bibliothek in jüngerer Zeit in nachgerade gegenläufiger Bewegung verstärkt auf die Rekonstruktion und Bewahrung temporärer Lesegemeinschaften ausgerichtet (vgl. Rapp und Embach 2008, Philippi und Vanscheidt 2014, Seelbach 2017, Nemes 2020). Innerhalb solcher Bemühungen hat insbesondere die mittelalterliche Bibliothek einen zentralen Stellenwert gewonnen und dies nicht allein wegen des allerorts beobachtbaren Attraktions- und Prestigekapitals des kulturell Ältesten. Durch die Rekonstruktion mittelalterliche Bibliotheken werden vielmehr die Bildungsfundamente für eine Archäologie der Moderne freigelegt (Embach, Moulin und Rapp 2011), aber auch die spezielleren Konturen regionaler und institutioneller Identität sichtbar gemacht. Da mittelalterliche Bibliotheken überwiegend das Schicksal der Zerschlagung und multiplen Dislozierung der Bestände in Folge von Klosterauflösungen am Beginn der Moderne teilen, können sie in besonderem Maße von Vorteilen der Digitalisierung profitieren: Emanzipation von Raum und Zeit, Grenzkostenmarginalisierung, Validierung der Erkenntnisse, Akzeleration der Informationsprozesse (Kirmße und Schneider 2019).

Im paradigmatischen Fall der Rekonstruktion der Bibliothek des Benediktinerklosters St. Matthias in Trier wurden beispielsweise 526 Handschriften des ehemaligen Klosterbestands aus 34 heute bekannten Standorten zusammengeführt und somit

erstmals geschlossen sichtbar und beforschbar gemacht.<sup>1</sup> Die Präsentation zielte auf Nachhaltigkeit und ein breites Spektrum an Zugängen (Projekt-Homepage, Bibliotheken der Stadt Trier, Handschriftenportal *Manuscripta mediaevalia*<sup>2</sup>). Dazu profitierte das Projekt von der Verknüpfung mit der virtuellen Forschungsumgebung *TextGrid*<sup>3</sup> (mit integriertem Tool Digilib zur Unterstützung der Bilddatenarbeit) mit ihrer komplexen Metadatenverwaltung und den vielfältigen Möglichkeiten der Verlinkung der Text- und Bilddaten, etwa mit Transkriptionen und Editionen, sowie zur Langzeitarchivierung (Rapp 2006, Raspe und Casties 2006) und nutzte Methodologien des web 2.0 wie z.B. das nutzerorientierte Kommentieren und Annotieren und Ontologien zur interpretatorischen Tiefenerschließung der dokumentierten Quellen (Ciula, Spence und Vieira 2008).

Neben institutionenübergreifenden Projekten wie *Fragmentarium*<sup>4</sup> oder *E-Codices*<sup>5</sup> wird in jüngerer Zeit wird auch wieder auf niederschwelligere, weniger komplexe, gleichwohl Entwicklungsfähige open-source basierte Lösungen gesetzt. Projekte wie z.B. *Johann Fischart kommentieren*,<sup>6</sup> *ArcheoInf*,<sup>7</sup> *SalzWiki*,<sup>8</sup> *forschungsdaten.org*,<sup>9</sup> oder eben die *Virtuelle Benediktiner-Bibliothek Millstatt* verwenden die von den MediaWiki-Entwicklern bereitgestellte freie Software nach dem Muster von Wikipedia und Wiki-Commons, die es ermöglicht, Daten zur Verfügung stellen, die auf stetige Erweiterungen, Korrekturen und zahlreiche bzw. wechselnde Beiträger (Autoren) angewiesen sind. Sie erlaubt Verlinkungen zu den Datenbanken mit den Beschreibungen und Bilddaten der besitzenden Institutionen (Bibliotheken), Verknüpfungen mit Normdaten und eine nachhaltige Nachnutzung.

Die Bibliothek des Benediktinerklosters Millstatt, im Mittelalter einer der umfangreichsten Buchbestände Kärntens, wurde nach der Auflösung des Klosters über weite Teile Europas verstreut. Dabei wurde nicht selten die Herkunft aus Millstatt verschleiert, was für die Identifikation der Bücher detektivischen Spürsinn erforderte. Ziele des Projekts, das sich den Bemühungen um digitale Vermittlung des kulturellen Erbes zuordnet, bestehen darin, die Handschriften dieser Bibliothek vollständig zu digitalisieren, nach modernen Prinzipien zu erschließen, sie in einer virtuellen Bibliothek wieder zusammenzuführen und somit erstmals geschlossen sichtbar zu machen und einer eingehenden Erforschung der frühen Wissenschaftsgeschichte Millstatts zur Verfügung zu stellen. Für die Präsentation der Daten wurde eine Form

<sup>1</sup> <https://stmatthias.uni-trier.de>.

<sup>2</sup> <http://www.manuscripta-mediaevalia.de>.

<sup>3</sup> <https://textgrid.de>.

<sup>4</sup> <https://fragmentarium.ms>.

<sup>5</sup> <https://www.e-codices.unifr.ch/de>.

<sup>6</sup> <https://wiki.uni-bielefeld.de/kommentieren/index.php/Gkl:komm>.

<sup>7</sup> <https://projekte.itmc.tu-dortmund.de/projects/tua-archeoinfdfg>.

<sup>8</sup> <https://www.salzwiki.de/index.php/Startseite>.

<sup>9</sup> <https://www.forschungsdaten.org/index.php/Kategorien:Projekte>.

gewählt, die die Bearbeitung und Erweiterung der Datenbasis jederzeit ermöglicht, weltweit lesbar und mit den relevanten internationalen Handschriftendatenbanken (*Manuscripta mediaevalia, manuscripta.at*) vernetzt ist.

Aktuell können folgende Ergebnisse präsentiert werden:

- Insgesamt konnten im Laufe der Vorarbeiten zum Projekt bislang 153 Handschriften eindeutig dieser Provenienz zugeordnet und in einen Arbeitskatalog aufgenommen werden. Dieser verzeichnet die Standorte, Signaturen und Inhalte.
- Von diesen Handschriften sind bislang ca. 80% in Kooperation der UB Klagenfurt mit dem Handschriftenzentrum der Uni Graz digitalisiert worden. Diese Arbeit wird kontinuierlich und zügig fortgeführt.
- Etwa 25% der identifizierten Handschriften wurden im Rahmen von Projektseminaren unter maßgeblicher Beteiligung von Studierenden bereits neu beschrieben und tiefenverschlossen. Mit Hilfe der vorhandenen Handschriftendatenbanken ist es dabei häufig gelungen, den Katalog von Hermann Menhardt (1927) zu ergänzen und zu berichtigen, d.h. Texte exakt zu identifizieren, Autoren zuzuweisen, Parallelüberlieferungen zu ermitteln und die Standorte der in Millstatt fehlenden Codices (z.B. im Falle mehrbändiger Werke) zu ermitteln.
- Eine Homepage mit assoziiertem Wiki wurde erstellt (<https://virtbibmillstatt.com>).

Eingepflegt wurden bislang die Handschriftenbeschreibungen des Pilotprojekts, das ca. 25% des Gesamtkorpus beinhaltet. Eine solche Pilotphase war notwendig, da für die Findung von praktikablen technischen Lösungen eine Zeit des Experimentierens eingeräumt werden musste.

## Literatur

- Ciula, Adriana, Paul Spence und José Miguel Vieira. 2008. „Expressing complex associations in medieval historical documents.“ *Literary and Linguistic Computing* 23: 311–25.
- Embach, Michael, Claudine Moulin und Andrea Rapp. 2011. „Die mittelalterliche Bibliothek als digitaler Wissensraum. Zur virtuellen Rekonstruktion der Abteibibliothek von Trier-St. Matthias.“ In *Mittelhochdeutsch. Beiträge zur Überlieferung, Sprache und Literatur*, hg. v. R. Plate und M. Schubert, 486–97. Berlin/Boston: de Gruyter.
- Kirmße, Stefan und Helmut Schneider. 2019. „Die acht Gesetze der Digitalisierung.“ *FAZ* 23. 9. 2019, 16.
- Nemes, Balazs und Martina Backes. Buchmeisterinne. *Handschriften und Frühdrucke aus dem Freiburger Dominikanerinnenkloster Adelhausen*. Zugriff: 17. November 2020. <https://2020.freiburg.de/pb/1408020.html>.

- Philippi, Sabine und Philipp Vanscheidt, Hrsg. 2014. *Digitale Rekonstruktionen mittelalterlicher Bibliotheken*. Wiesbaden: Reichert.
- Rapp, Andrea und Michael Embach. 2008. „Die Bibliothek der Benediktinerabtei St. Matthias in Trier – ein europaweites Projekt zur Volltextdigitalisierung und virtuellen Rekonstruktion der Handschriften.“ In *Rekonstruktion und Erschließung mittelalterlicher Bibliotheken*, hg. v. Andrea Rapp und Michael Embach, 147–69. Berlin: de Gruyter.
- Rapp, Andrea. *Das Projekt TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung – ein Community-Grid für die Geisteswissenschaften*. [http://www.ahf-muenchen.de/Forschungsberichte/Jahrbuch2006/AHF\\_Jb2006\\_FB\\_BI\\_Rapp.pdf](http://www.ahf-muenchen.de/Forschungsberichte/Jahrbuch2006/AHF_Jb2006_FB_BI_Rapp.pdf).
- Raspe, Martin und Robert Casties. *Digilib. Wissenschaftliches Bildmaterial studieren und kommentieren im Internet*. Zugriff: 17. November 2020. [https://www.mpg.de/bilderBerichteDokumente/dokumentation/jahrbuch/2006/bibliotheca\\_hertziana/forschungsSchwerpunktI/index.html](https://www.mpg.de/bilderBerichteDokumente/dokumentation/jahrbuch/2006/bibliotheca_hertziana/forschungsSchwerpunktI/index.html).
- Seelbach, Sabine. *Colligit fragmenta ne pereant. Zum Bestand der Millstätter Benediktinerbibliothek*. Zugriff: 17. November 2020. [https://www.aau.at/wp-content/uploads/2017/08/16-12-06\\_Kostbarkeiten-9\\_Einleitung\\_Seelbach.pdf](https://www.aau.at/wp-content/uploads/2017/08/16-12-06_Kostbarkeiten-9_Einleitung_Seelbach.pdf).
- Seelbach, Ulrich. 2019. „Ein moderierter Kommentar zu Fischarts Geschichtklitterung.“ In *Johann Fischart, genannt Mentzer. Frühneuzeitliche Autorschaft im intermedialen Kontext*, hg. v. Tobias Bulang, 273–92. Wiesbaden: Harrassowitz.

# Die gelehrte Korrespondenz der Brüder Pez (Hybridedition)

Thomas Wallnig

<b>Mitarbeiterinnen und Mitarbeiter:</b>	Thomas Wallnig (Projektleitung), Patrick Fiska, Manuela Mayer, Ines Peper, Daniel Schopper, Claudia Sojer, Thomas Stockinger, Victor Wang
<b>Institutionen:</b>	Universität Wien – Institut für Geschichte und Institut für Österreichische Geschichtsforschung, ACDH-CH ÖAW
<b>Fördergeber:</b>	FWF Der Wissenschaftsfonds (P-16940; Y-390; P-28016)
<b>Website:</b>	<a href="http://vemg.at/pez-edition-und-nachlass">http://vemg.at/pez-edition-und-nachlass</a> , <a href="https://pezworkshop.org">https://pezworkshop.org</a>

Die Brüder Bernhard und Hieronymus Pez lebten im frühen 18. Jahrhundert als Mönche im Benediktinerkloster Melk in Niederösterreich. Sie wirkten als Historiker und Herausgeber mittelalterlicher Quellen, und sie zählten zu den frühen Vertretern kritischer Geschichtsforschung in Zentraleuropa.

Die an sie adressierten Briefe sind zum Großteil noch heute in Melk erhalten, einige von ihnen selbst versendete Schreiben auch in mehreren Dutzend anderer europäischer Bibliotheken. Die etwa 1150 erhaltenen, meist lateinisch verfassten Gelehrtenbriefe sind seit 2004 im Rahmen mehrerer FWF-Projekte Gegenstand editorischer Bemühungen; etwa die Hälfte von ihnen liegt seit 2015 im Druck vor. Nun wird für die verbleibenden Briefe eine digitale Lösung angestrebt, die das Layout der bisherigen Bände reproduzieren kann, während zugleich die erschienenen Bände im selben Sinn rückwirkend digital bearbeitet werden. Im Zuge dessen soll auch ein genereller Workflow für Hybrideditionen in der Publikationsreihe *Quelleneditionen des Instituts für Geschichtsforschung* erarbeitet werden, in welcher die Pez-Korrespondenz erscheint.

Die Motivation für diese Art der Bearbeitung erschließt sich aus der Natur und Bedeutung der Pez-Korrespondenz. Sie ist eine der wichtigen Quellen zur Wissenschafts- und Ideengeschichte Zentraleuropas in der Vormoderne. Sie dokumentiert eine spezifisch katholische Variante von religiösem Antiquarianismus und historischer Kritik. Die Briefe sind bedeutsam im Hinblick auf das in ihnen und im (ebenfalls online verfügbaren) Nachlass dokumentierte, meist mittelalterliche Quellenmaterial, aber auch im Hinblick auf die durch sie greifbaren zeitgenössischen Gelehrtennetzwerke. In der Pez-Korrespondenz wird schließlich die spezifische Latinität des katholischen

Späthumanismus ebenso greifbar wie die historiographische Einfassung von ‚österreichischer‘ und ‚deutscher‘ Geschichte.

## Literatur

- Rabl, Irene. 2013. „Der digitalisierte Nachlass der Brüder Bernhard und Hieronymus Pez. Ein Projektbericht.“ *MIÖG* 121: 437–44.
- Stockinger, Thomas, et al. 2015. *Die gelehrt Korrespondenz der Brüder Pez. Text, Regesten, Kommentare, 2: 1716–1718*. Quelleneditionen des Instituts für Österreichische Geschichtsforschung 2/2. Wien. Zugriff: 15. September 2019. <https://www.oapen.org/search?identifier=576951>, <http://www.oapen.org/search?identifier=576952>.
- Wallnig, Thomas und Thomas Stockinger. 2010. *Die gelehrt Korrespondenz der Brüder Pez. Text, Regesten, Kommentare, 1: 1709–1715*. Quelleneditionen des Instituts für Österreichische Geschichtsforschung 2/1. Wien. Zugriff: 15. September 2019. <http://www.oapen.org/search?identifier=445402>.
- Wallnig, Thomas. 2019. *Critical Monks. The German Benedictines, 1680–1740*. Leiden, Boston: Brill.

## **Appendix**



## Biographical Notes

**Tara L. Andrews** ist Universitätsprofessorin für Digital Humanities an der Universität Wien. Ihre Forschungsgebiete umfassen die Geschichte des mittelalterlichen Nahen Osten, Textstemmatologie und digitale Modellierung der Überlieferung antiker und mittelalterlicher Texte.

**Roman Bleier** studied History and Religious Studies at the University of Graz, and completed a Ph.D. in Digital Arts and Humanities at Trinity College Dublin with a dissertation on the editing of St Patrick's epistles (*St Patrick's epistles*, <https://gams.uni-graz.at/context:epistles>). Since 2016 he worked as postdoctoral researcher on different digital scholarly editing projects at the Centre for Information-Modelling, University of Graz. He is currently working on the Imperial Diet records of 1576 (*Der Regensburger Reichstag von 1576 digital*, <http://gams.uni-graz.at/context:rta1576>) and Peter of Poitiers' *Compendium historiae*.

**Martina Bürgermeister** ist Doktorandin im Fachbereich der Digitalen Geisteswissenschaften. Seit 2013 ist sie wissenschaftliche Mitarbeiterin am Zentrum für Informationsmodellierung (ZIM-ACDH) an der Universität Graz. In ihrer Dissertation setzt sie sich mit den Themen Provenienz, Versionierung und kollaborative Entwicklung von dynamischen Ressourcen in den Geisteswissenschaften auseinander.

**Markus Ender:** Studium der Germanistik, Geschichte und Philosophie an der Universität Innsbruck. Seit 2012 wissenschaftlicher Mitarbeiter des Forschungsinstituts Brenner-Archiv. Arbeitsgebiete: Literatur des 19. und 20. Jahrhunderts, Editionsphilologie, Archivtheorie. Vorträge und Publikationen zu Ludwig von Ficker und zum „Brenner“, zur Literatur der Moderne sowie zur digitalen Edition. Seit 2018 Senior Scientist am Forschungsinstitut Brenner-Archiv.

**Peter Hinkelmanns** ist Mediävist, Digital Humanist und Senior Scientist an der Paris-Lodron-Universität Salzburg. Seine InteressenSchwerpunkte sind Lexikographie, Sprachgeschichte und Graphentechnologien. Er ist Mitarbeiter der Mittelhochdeutschen Begriffsdatenbank und an weiteren Projekten aus den Digital Humanities beteiligt.

**Helmut W. Klug** is a medievalist and strongly interested in Digital Scholarly Editions as means for making available historical sources. He was project assistant in the federally funded infrastructure project *Kompetenznetzwerk Digitale Edition (KONDE)*, <http://www.digitale-edition.at>). He researches culinary history with a focus on the Middle Ages and Early New Times and heads the Austrian team

of a French-Austrian cooperation project that aims at providing and analysing German, French and Latin cooking recipe collections (*CoReMA*, <https://gams.uni-graz.at/context:corema>).

**Sarah Lang** is a Digital Humanities scholar with a background in Classics, working on the early modern Neo-Latin work of iatrochymist Michael Maier (1568-1622). In her Digital Humanities PhD thesis at the University of Graz (Austria), she has developed a Machine Reasoning and Semantic Web based analysis tool for the alchemical cryptographic device called Decknamen, using Maier's printed works as an example corpus.

**Bernhard Oberreither:** Studium der Germanistik, Kunstgeschichte und Anglistik in Graz und Wien; von 2010 bis 2014 wissenschaftlicher Mitarbeiter des FWF-Projekts *Das Bildzitat* an der Universität Wien, von 2014 bis 2019 Universitätsassistent am dortigen Institut für Germanistik; seit 2019 Mitarbeiter an der Arbeitsstelle für Corpora und Editionen (ACE) am ACDH-CH der ÖAW. Forschungsschwerpunkte: Editionswissenschaft und digitale Geisteswissenschaften, literarische Bildzitate, Literatur und Fotografie, Literatur und Fetischismus.

**Claudia Resch:** Studium Deutsche Philologie sowie Publizistik- und Kommunikationswissenschaft in Wien. Seit 2003 an der Österreichischen Akademie der Wissenschaften; derzeit am Austrian Centre for Digital Humanities and Cultural Heritage. Leitung von Forschungsprojekten wie *Austrian Baroque Corpus*, *Wienerisches Digitarium* und *Vienna Time Machine*. Von 2012-2017 Lehrbeauftragte an der LMU München, seit 2017 an der Universität Wien am Institut für Geschichte, am Institut für Germanistik sowie im Masterstudium Digital Humanities. Seit 2021 *venia docendi* für das Fach Digitale Textwissenschaften.

**Georg Vogeler** is Professor for Digital Humanities at the University of Graz. He studied Historical Auxiliary Sciences in Freiburg (Brsg.) and Munich. In 2002 he received his PhD with a study on late medieval tax registers of German territories and in 2016 the *venia docendi* for his Habilitationschrift on the use of the charters of Frederic II (1194-1250) by his contemporaries in Italy.

**Joseph Wang-Kathrein:** Studium der Humanmedizin und Philosophie. Seit 2006 Erfahrung im Bereich des digitalen Edierens in unterschiedlichen Forschungsprojekten. Forschungsschwerpunkte: Ludwig Wittgenstein, Digital Humanities, Analytische Substanzontologie. Im Moment arbeitet er am Forschungsinstitut Brenner-Archiv und am Digital Science Center der Universität Innsbruck.

The authors of the article “Don’t worry, we’re also doing a book!” are historian and principal Pez correspondence editor **Thomas Wallnig** (Institute of Austrian

Historical Research at the University of Vienna), **Daniel Schopper**, specialist for digital resources and data modelling at the Austrian Center of Digital Humanities and Cultural Heritage, and **Victor Wang**, who is centrally involved in Böhlau's strategic development of digital matters.



# **Publications of the Institute for Documentology and Scholarly Editing / Schriftenreihe des Instituts für Dokumentologie und Editorik**

- 01** Bernhard Assmann, Patrick Sahle. Digital ist besser. Die Monumenta Germaniae Historica mit den dMGH auf dem Weg in die Zukunft – eine Momentaufnahme. Norderstedt: Books on Demand, 2008. ISBN 978-3-8370-2987-1
- 02** Kodikologie und Paläographie im digitalen Zeitalter / Codicology and Palaeography in the Digital Age. Ed. by Malte Rehbein, Patrick Sahle and Torsten Schaßan in collaboration with Bernhard Assmann, Franz Fischer and Christiane Fritze. Norderstedt: Books on Demand, 2009. ISBN 978-3-8370-9842-6
- 03** Kodikologie und Paläographie im digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2. Ed. by Franz Fischer, Christiane Fritze and Georg Vogeler in collaboration with Bernhard Assmann, Patrick Sahle und Malte Rehbein. Norderstedt: Books on Demand, 2010. ISBN 978-3-8423-5032-8
- 04** Birgit Jooss: Die digitale Edition der Matrikelbücher der Akademie der Bildenden Künste München. Ein Instrument zur Erforschung der Attraktivität einer international ausgerichteten Kunsthochschule (1808 – 1920). Norderstedt: Books on Demand, 2011. ISBN 978-3-8423-1278-4
- 05** Johannes Kepper: Musikedition im Zeichen neuer Medien – Historische Entwicklung und gegenwärtige Perspektiven musikalischer Gesamtausgaben. Norderstedt: Books on Demand, 2011. ISBN 978-3-8448-0076-0
- 06** Digitale Urkundenpräsentationen. Beiträge zum Workshop in München, 16. Juni 2010. Ed. by Georg Vogeler and Joachim Kemper. Norderstedt: Books on Demand, 2011. ISBN 978-3-8423-6184-3
- 07** Patrick Sahle: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 1: Das typografische Erbe. Norderstedt: Books on Demand, 2013. ISBN 978-3-8482-6320-2

- 08** Patrick Sahle: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik. Norderstedt: Books on Demand, 2013. ISBN 978-3-8482-5252-7
- 09** Patrick Sahle: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung. Norderstedt: Books on Demand, 2013. ISBN 978-3-8482-5357-9
- 10** Kodikologie und Paläographie im digitalen Zeitalter 3 / Codicology and Palaeography in the Digital Age 3. Ed. by Oliver Duntze, Torsten Schaßan and Georg Vogeler. Norderstedt: Books on Demand, 2015. ISBN 978-3-7347-9899-3
- 11** Kodikologie und Paläographie im digitalen Zeitalter 4 / Codicology and Palaeography in the Digital Age 4. Ed. by Hannah Busch, Franz Fischer and Patrick Sahle, in collaboration with Bernhard Assmann, Philipp Hegel, and Celia Krause. Norderstedt: Books on Demand, 2017. ISBN 978-3-7448-3877-1
- 12** Digital Scholarly Editions as Interfaces. Ed. by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider. Norderstedt: Books on Demand, 2018. ISBN 978-3-74810-925-9
- 13** Versioning Cultural Objects. Ed. by Roman Bleier and Sean M. Winslow. Norderstedt: Books on Demand, 2019. ISBN 978-3-7504-2702-0
- 14** Rekontextualisierung als Forschungsparadigma des Digitalen. Ed. by Simon Meier, Gabriel Viehhauser, and Patrick Sahle. Norderstedt: Books on Demand, 2020. ISBN 978-3-7519-1531-1
- 15** Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing. Ed. by Elena Spadini, Francesca Tomasi, and Georg Vogeler. Norderstedt: Books on Demand, 2021. ISBN 978-3-7543-4369-2
- 16** Digitale Edition in Österreich / Digital Scholarly Edition in Austria. Ed. by Roman Bleier and Helmut Klug. Norderstedt: Books on Demand, 2023. ISBN 978-3-743-102-842



