Kodikologie und Paläographie im digitalen Zeitalter 4

Codicology and Palaeography in the Digital Age 4

herausgegeben von | edited by

Hannah Busch, Franz Fischer, Patrick Sahle

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Philipp Hegel, Celia Krause

2017

BoD, Norderstedt

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 4. September 2017.

Some Roads to Script Classification: Via Taxonomy and Other Ways

Torsten Schaßan

Abstract

In codicology, the features of a script play an important role for dating and localising the manuscript. There are other questions that can be dealt with by examining these features, e.g. questions of intellectual history, influences of literary genres, or influences of organisational aspects of scriptoria on the shape of a script. But especially in the context of manuscript cataloguing the classification of script is of highest importance if other evidence such as a colophon or references like the naming of celebrations for local saints cannot be found. In order to contextualise the features of a script, palaeography has always striven for inference of a taxonomy from visual properties. Like in other disciplines, the community was not successful in achieving one common naming schema but constituted concurring taxonomies. Thus, the question arises what to do with these in times of the need to search huge amounts of manuscript related data in portals? New approaches in standardisation on the one hand, and semantic technologies and methods for image processing on the other hand, offer new possibilities to access to the manuscripts.

Zusammenfassung

In der Kodikologie spielt die Merkmale einer Schrift für die Datierung und Lokalisierung der Handschrift eine wichtige Rolle. Zwar lassen sich auch andere, geistesgeschichtliche Fragestellungen an diese Merkmale anknüpfen, wie etwa der Einfluss der Textsorte oder die Organisationsform eines Skriptoriums auf die Schriftgestalt, aber insbesondere im Kontext der Handschriftenkatalogisierung dient die Schrift dort, wo Kolophon, Nennung lokaler Heiliger oder andere inhaltliche Bezüge fehlen, der Ermittlung dieser wichtigen Information. Um die Merkmale einer Schrift in größere Bezüge einordnen zu können hat die Paläographie immer schon versucht, aus den visuellen Eigenschaften eine Klassifikation abzuleiten. Wie in anderen wissenschaftlichen Zweigen auch hat sich die Zunft aber nicht auf ein Benennungsschema einigen können, sondern konkurrierende Klassifikationen ausgebildet. Wie soll aber in einer Zeit, da verstärkt Handschriftenkataloge und andere handschriftenbezogene Informationen in Portalen durchsucht werden können und aufgrund der überwältigenden Menge auch durchsucht werden müssen, mit diesem Problem umgegangen werden?

Neue Standardisierungsversuche auf der einen Seite, semantische Technologien und Bildverarbeitungsmethoden auf der anderen Seite bieten Möglichkeiten, Zugänge zu Handschriften zu ermöglichen.

1 Introduction

Today, a large amount of manuscript data is available from various digitisation efforts. How can these data be accessed? How do we find the way to a single manuscript or to a defined group of manuscripts? Besides other means such as the subjects covered, the works contained in, or persons related to, the manuscript, the classification of scripts may give researchers a tool to find the needle in the haystack. The main questions addressed in this paper are: How can such a classification be established? How has it been done in the past? What are the challenges and how can they be overcome with the possibilities of modern technologies and algorithms? While addressing these questions it shall be made clear that this article is not written from a palaeographer's point of view or claims to be a quest for some 'truth' about scripts and the names assigned, but from the perpective of someone who strives to support research by publishing manuscript data and who has to pave paths through masses of data, images, and descriptions alike. Thus, this paper focuses more on information theory and the usefulness or power of algorithms. Additionally, palaeography will be examined only in the context of script description and classification and not according to its possible other functions such as its relation to society and language or as an art (Cf. Stutzmann 2005, 16f.; Castro Correa 2014, 248).

2 What do we need script classification for, and why?

Stokes uses the term *palaeography* in the narrow sense as "the study of (medieval) handwriting with view towards its history and development and the identification, localization, and dating of *scribes*." (Stokes 2012, 137; emphasis by the author.) One would want to add: Palaeography is needed for localising and dating *manuscripts*. This task, the localisation and dating, has to start with 'basic truth', i.e. located and/or dated manuscripts, examine the script — and other, external evidences —, recognise its features and compare undated manuscripts with these examples in order to localise and date them. Now, this is what palaeographers have been doing since ages.

    In order to share this information, e.g. through catalogues, the palaeographer had to describe what he/she saw in the manuscripts. Even today, catalogues describe scribal features. Derolez points out the problem with this approach: "How is it possible to proceed in such a way that the description of a specimen of handwriting is as clear

and convincing to its reader as it is to its author?" (Derolez 2003, 7) Firstly, author and reader have to use a shared language in order to understand each other, and, secondly, the reader has to know what to look for if he wants to recognise what the author has seen in a certain manuscript. And, as Derolez continues: "The method applied hitherto in palaeographical handbooks has produced an authoritative discipline, the pertinence of which depends on the authority of the author and the faith of the reader." (Derolez 2003, 9) This refers to the fact that the description establishes a special relation of belief, that the description in a catalogue is detailed enough to be understood and accurate enough to be true.

A very recent example for this problem is the following: Bernhard Bischoff is an authority if it comes to script, localising, and dating of manuscripts. When Hoffmann reviewed the last volume of Bischoffs *Katalog der karolingischen Handschriften*, edited by Birgit Ebersperger and published post-mortem, he criticised the editor for adding *Gudianus latinus 269* to the catalogue. She interpreted Bischoff and added Corvey as a place of origin. Hoffmann asks: "Und wer kann gar mit Sicherheit sagen, daß es Corveyer Hände waren? " (Hoffmann 2015, 17) It seems that Hoffmann would probably have trusted Bischoff but he scrutinises — and challenges — Ebersperger.[1]

Still, the long(er) descriptions of script have always been assigning a name to the script in question. This name represents the most common features of a certain script and is generalised from distinct hands. The term just offers a general impression of a script and does not allow for describing a certain hand. This term is listed in the indices of catalogues in order to allow for easy access to the manuscripts. The community learned to agree — more or less; we will come back to that — on a common list of terms.[2] The naming convention derived would be the basis for a controlled vocabulary in the first place and could be arranged into a taxonomy or even an ontology. Here, 'Digital Humanities methods' come into play, i.e. the application of technologies such as *RDF*, *TripleStores* and so on.

One remark on the notion of Stokes in his 2012 article on *Palaeography and the 'Virtual Library' of Manuscripts* in which he claims that "[t]he use of verbal description can probably never be avoided, because any use of a manuscript or facsimile is an act of interpretation." He continues: "We must be told which aspects of the letterforms are considered significant, how these significant differences compare between samples,

---

[1] Stutzmann claims that the status of being a connoisseur and, thus, an authority has long since been overcome and replaced by set objective criteria and precise terminology. A difference, however, would remain in the criteria applied. But, as Stutzmann critisised Derolez's system, by introducing 'accuracy' as another aspect to distinguish scripts, some level of subjective interpretation beyond the nomenclature continues to influence the analysis. (Stutzmann 2005, 19f)

[2] While Overgaauw had to conclude in 1994 that huge advances had been made for Carolingian and pre-Carolingian scripts but still no such advances were possible for the Gothic scripts (Overgaauw 1994, 100), Stutzmann reviewed the work of Derolez as "far more than just another palaeographic hand-book which offers a comprehensive nomenclature of gothic scripts." (Stutzmann 2005, 1)

and so on, and if we are not told this we are at sea as is demonstrated by an existing attempt to categorize letterforms by images alone." (Stokes 2012, 141) This quotation will have to be reviewed in the light of the techniques described in the last part of this paper as there are ways of measuring differences with digital technologies.

What one can find in all the data about scripts in the databases is the result of the reduction of specimen to simple, short terms. The collection of terms from indices forms the first approach to a controlled vocabulary. Yet, the community of cataloguers tends not to be satisfied with just a small number of terms and names for scripts but always strives for a better distinction between the scripts. A brief survey of recent cataloguing and digitisation projects highlights the problem to find a balance between the advantages of a very short list of terms versus a longer list of (probably) more accurate terms. Only such projects have been chosen which offer the cataloguer a predefined list of terms.[3] Table 1 lists the terms provided by the projects *ENRICH*, *Europeana Regia*, and the Swiss manuscript portal *e-codices*: The terms have been defined in the TEI schema for manuscript description. They are used at the elements `<handNote>` and `<scriptNote>`, more specifically on the attribute `@script` on these elements.

While there has only been defined a relatively small number of terms in the *ENRICH* project, the other projects add numerous terms to the list. All of the terms added are specific to the experiences made by the project partners respectively, representing the scripts that are common to the manuscripts in the collections or of that geographic region. During cataloguing those manuscripts, the participating institutions and heads of the projects must have felt the need to use these terms. The rather short list of the *ENRICH* project was meant to allow for searching and grouping the manuscripts by script in the first place. However, the other institutions and projects must have thought about a better representation of the heterogeneity of scripts. It is clear that too large a number of terms will serve none of the needs one might have: a list of too many entries will neither allow for faster access to the manuscripts, nor be able to describe the world of scripts and differences between hands and scripts in enough detail to replace imaging and the experience of the palaeographer. It is clear that, the closer one looks at scripts, the more differences one will recognise until not even two hands or scripts have the same properties in order to be called 'one script' or hand.[4]

---

[3] Not included are databases such as Manuscripta Mediaevalia which allow cataloguers to review the terms other catalologuers have used before them and just choose from them or enter any term they want. For that practice see Riecke 2009, 225: "Die Ansetzung der Eintragung sollte sich an den bereits vorhandenen Termini orientieren [...]. Bislang wurden beispielsweise eingegeben: Buchschrift, gotische; Capitalis; Geheimschrift; Humanistica; Kanzleischrift; Kurrentschrift; Kursive; Majuskel, angelsächsische; Minuskel, karolingische; Perlschrift; Rotunda; Textualis formata; Unziale; Vortragsakzent."

[4] Cf. for that idea the 'Coastline paradox', which Peter Robinson applied to textual scholarship problems of textual variation in his paper (Robinson 1996. Original by Mandelbrot 1983).

ENRICH	Europeana Regia (added values)	e-codices (added values)
carolmin	capquad	antiqua
textualis	caprust	precar
cursiva	uncialis	spaetcar
hybrida	semiunc	praegot
humbook	benevent	gotica
humcursiva	luxeuil	semicursiva
kanzlei	corbie	greek[5]
kurrent	insulmin	hebrew
	alemmin	
	raetmin	
	carolgot	
	textura	
	rotunda	
	cancell	
	bastarda	
	cursant	
	cursrec	

Table 1: The terms provided by the projects *ENRICH*, *Europeana Regia* and *e-codices*

The general question would then be whether a community could agree upon a single list of terms for script classification at all? There are two kinds of problems connected to this: On the one hand, there is the difficulty to agree upon proper names for scripts that are similar to each other and might be distinguished only by minute characteristics. On the other hand, scripts might have been given different names over time although the visual features of those scripts would suggest likeness.

3 Traditional approaches

In order to examine the features of a script in larger contexts, palaeography has always striven to establish a taxonomy with inferences of the visual properties. Similar to other disciplines, the community was only partially successful to achieve one common naming schema, but constituted concurring taxonomies. These inherited different types of problems such as having different names for similar scripts, entities

---

[5] The inclusion of the terms 'greek' and 'hebrew' would add to the issues of classification of script discussed so far other aspects such as 'script and language'. As the topic of this paper is to examine the possibilities that certain technologies offer in order to overcome some problems, the inclusion of non-Latin scripts shall not be addressed any further.

overlapping geographically, or names of entities changing meaning over time causing ambiguity as one cannot be sure whether the term still covers the same entity, e.g. an area.

An example for the first issue, having different names for similar scripts, is the problem with terminology for Bastarda scripts as well as for Insular Carolingian minuscle. Stokes mentions this example of overlapping respectively divergent terminology. Alexander Rumble's guidelines include, among other terms, the 'round Anglo-Saxon minuscle'. This script is called 'Caroline minuscle' by Ker, 'Anglo-Insular minuscle' by Boyle, 'Anglo-Saxon round minuscle' by Brown, 'Anglo-Saxon vernacular minuscle' by Dumville, 'English Caroline minuscle' by Roberts, and, finally, 'English vernacular minuscle' by Stokes himself. (Stokes 2012, 147) One — traditional — way to deal with this issue would be to define a concordance and mention divergent names together, as e.g. Derolez does. (Derolez 2003; a concordance Stutzmann 2005, 63)

An example for the second problem, entities overlapping geographically and being not defined clearly, can be found in related subjects which are relevant for palaeography: the names of places and regions as used for localising script and manuscripts. When we find entries such as 'Südostdeutschland', 'Österreich', or 'Bayern' in a catalogue like Bischoff's catalogue of ninth centuries manuscripts, which entities do these refer to?

An example for the last issue, the change of names of entities over time, may be the distinction between 'Niederdeutschland' and 'Norddeutschland'.[6] Here, the same methodologies have to be applied.

4 'Healing' concurrence

At this point, it is necessary to repeat the definitions of some terms that are regularly — but sometimes perplexingly or wrongly — in the discussion of characteristics:

- A *controlled vocabulary* is just a collection of terms describing one aspect or feature. If the controlled vocabulary covers all aspects and is therefore 'complete', it is called a *nomenclature*.

- A *taxonomy* is an ordered, mono-hierarchical classification of the terms of a nomenclature.

- An *ontology* adds the relations between the terms to the mono-hierarchical classification.

- A *folksonomy* is a 'democratised' version of a taxonomy, derived from collaborative, social tagging.

---

[6] Examples taken from Hoffmann 2015, 45.

But how are any of these technologies applied? Cataloguers who write their documents in TEI-XML directly are offered lists of values for the classification of script during the cataloguing, as shown in table 1. The list is provided as part of a schema which defines the 'grammar' of a document. Cataloguers are provided with a template file that contains all the necessary structures and serves as a spreadsheet.[7] The template file references the schema file. The schema file contains the information about the scripts we want the encoders to specify and, respectively, the values of the attributes @script which can be used to name the script used in the manuscript. The list has been defined using the TEI *ODD* document type. *ODD* is short for 'one document does it all'. The *ODD* allows to define a schema from within the TEI. In this *ODD* file, the list of values is supplied. The definition of the list of values looks like this:

```
<classSpec ident="att.handFeatures" type="atts" mode="change" module="tei">
 <attList>
  <attDef ident="script" mode="change" usage="rec">
   <defaultVal>other</defaultVal>
   <valList type="semi" mode="replace">
    <valItem ident="carolmin">
     <desc xml:lang="de">Karolingische Minuskel</desc>
    </valItem>
    <valItem ident="textualis">
     <desc xml:lang="de">Textualis</desc>
    </valItem>
    <valItem ident="cursiva">
     <desc xml:lang="de">Kursive</desc>
    </valItem>
    <valItem ident="hybrida">
     <desc xml:lang="de">Hybrida</desc>
    </valItem>

    <!-- more values might be defined here -->
    <valItem ident="other">
     <desc>any other type of script</desc>
    </valItem>
    <valItem ident="unknown">
     <desc>script information not available</desc>
    </valItem>
   </valList>
  </attDef>

    <!-- more attributes might be dealt with here -->
 </attList>
</classSpec>
```

A list like this represents a controlled vocabulary. The cataloguers will be able to choose from this list of terms during their cataloguing. Managing the list helps to foster the interoperability of data and allows for faceted browsing of the catalogue entries if implemented. Lists like these are helpful especially if the list of terms included is rather short.

---

[7] <https://github.com/schassan/cataloguing/blob/master/tei-msDesc_template.xml>.

If one wanted to allow for a better search even with a large number of entries, one could arrange the terms in a mono-hierarchical classification, a taxonomy. In the resulting hierarchy, one would distinguish broader and narrower terms, e.g. *Cursiva* as broader and *Cursiva antiquior* or *Cursiva recentior* as more specialised terms. Furthermore, concurring terms like the ones mentioned above could remain as they are whilst their relation with others can be expressed. To express a hierarchy of or relations between terms, one could apply several semantic web technolgies: one of these is the usage of the *Web Ontology Language* (OWL).[8] The technical realisation the hierarchy for the example in *OWL* might look like that:

```
<rdf:RDF>
 <owl:Ontology rdf:about=""/>
 <owl:Class rdf:ID="script"/>
 <owl:Class rdf:ID="cursiva" rdf:about="http://anyuri.com/scripts#cursiva">
  <rdfs:subClassOf rdf:resource="#script"/>
  <rdfs:label>Cursiva</rdfs:label>
  <rdfs:comment>This class covers all types of cursive script.</rdfs:comment>
 </owl:Class>
 <owl:Class rdf:ID="cursant">
  <rdfs:subClassOf rdf:resource="#cursiva"/>
  <rdfs:label>Cursiva antiquior</rdfs:label>
  <rdfs:comment>This class covers antique cursive script.</rdfs:comment>
 </owl:Class>
 <owl:Class rdf:ID="cursrec">
  <rdfs:subClassOf rdf:resource="#cursiva"/>
  <rdfs:label>Cursiva recentior</rdfs:label>
  <rdfs:comment>This class covers recent cursive script.</rdfs:comment>
 </owl:Class>
</rdf:RDF>
```

This example implements the ontology in *RDF* syntax and therefore uses elements from the *RDF Schema* (RDFS) namespace. RDFS offers elements to define sub- and super-classes as well as relations between classes such as 'sameAs', 'similarTo', or 'relatedTo'. With the means of semantic web technologies and *RDF* it is possible to enhance a controlled vocabulary respectively a taxonomy to an ontology.

5 Machine-aided approaches

Already in 1979, Bernhard Bischoff recognised that palaeography, which used to be an art of vision and empathy, becomes an art of measuring by technical means.[9]

---

[8] Although the abbreviation for the Web Ontology Language should be *WOL*, the reason for choosing *OWL* is not entirely clear. Possible explanations include that the inventor of that language chose to introduce a more interesting one, stating that "Why not be inconsistent in at least one aspect of the language which is all about consistency?" (Schreiber) Another one is that this acronym has been chosen as a tribute to William A. Martin's *One World Language* knowledge representation project from the 1970s.

[9] "Mit technischen Mitteln ist die Paläographie, die eine Kunst des Sehens und der Einfühlung ist, auf dem Wege, eine Kunst des Messens zu werden." (Bischoff 1986, 19)

In a 'machine-aided approach', the traditional palaeographic method is enhanced by the aid of some automated methods, e.g. measuring.[10] Already in 1977, Gilissen pioneered with the statistical analysis of quantitative measurements such as pen-angle, pen-width, etc. (Gilissen 1977, cited in Stokes 2012, 145)

An example for a machine-aided approach to palaeographical research is the project DigiPal. DigiPal has been developed at the University College London by Peter Stokes et al. The basis of DigiPal is a database to which researchers can add images of manuscripts and detailed descriptions such as own characterisations of scripts and classifications and other metadata. Users can cut out single letters which then will be displayed side-by-side with other occurrences of the same letter. This generates collections which can be grouped, compared, and searched for. All this work has been and is to be done manually. Moreover, as there are only single letter-forms stored in this database, they lack the context of the word, line, and entire page. This is exactly what the algorithms presented afterwards are going to look at. The machine-aided part of the project consists of a set of search algorithms which are able to search both in the descriptions of script as well as in the descriptions of characteristics. The characteristics of script can be detailed enough to serve as a finger-print of a script. These fingerprints are compared by the algorithms.

Aussems and Brink presented another possibility by looking at a "writer-specific variation in the width of the ink trace" and measuring "the relation between the local direction and width of the ink traces." (Aussems and Brink, 298)

As Stokes points out, "[…] none of these projects accounts for page curvature, image distortion, or the natural expansion and contraction of parchment […] " (Stokes 2012, 145). I think he is exagerating here as the human eye is subject to the same challenges and the palaeographer's decisions have to be questioned as well. The algorithms mentioned above do not account for that either and may have to be adjusted in order to do so. On the other hand, as measurements are summarised over many pages, sometimes entire manuscripts, the deviation may be of little significance.

6 Machine-driven approaches

To overcome the burden to describe every detail in one's own words, only some printed catalogues supply the reader with a series of images from the manuscripts. This is especially true for catalogues of dated manuscripts which contain both images of pages that have a colophon or other means used for dating, and images of sample pages representative of the script used throughout the manuscript. Catalogues of illuminated manuscripts supply images for art historical means. But even 'normal'

---

[10] Stokes called this approach 'computer-aided' and gave an overview of the questions and methods in his 2009 publication.

catalogues sometimes supply a number of images.[11] With these images, the reader can make up his mind and compare the description of script with its actual image. Still, *an image says more than a thousands words.*

But even today, as more and more images of manuscripts are available online for consultation and reference, one would want to access this huge amount of data with the help of standardised terminology or via a pattern that can be found in all these images. This time, we do not need the terminology for summarising the long feature descriptions but in order to subdue the sheer mass of information available to us. To supply a term or a reference to a pattern for every image available will have to be the task of (automated) algorithms, in the best of all worlds implemented as services.

Such algorithms have been proposed e.g. by Bulacu and Schomaker (2007a and 2007b), others built on top of these basic algorithms. (Cf. Fecker et al. 2015) Basically, these algorithms are based on the idea that a script can be described as a multidimensional matrix of attributes such as stroke-width, slant, etc. Once all of these characteristics have been recognised, measured, and assessed, the algorithm is (or should be) able to distinguish between different scripts. Although the algorithms mentioned above have been used for scribal identification and are, thus, intended to find differences in what are supposed to be similar or homogenous measurements, one would think that the difference between scripts — in order to arrive at a classification — e.g. between Caroline minuscule and Gothic scripts, should be greater than the differences between two hands writing both a Gothic minuscule? Another aspect would be that the proposed algorithms strive both for a script identification as well as for a script verification. The identification will separate the characteristics of a script from possibly all other scripts. This does not necessarily imply a writer identification which would mean to assign an identified script to an identified scribe. But once the script has been identified by recognising its features, it should be possible to look for the same features in other manuscripts or on other pages and, thus, verify whether a page or a manuscript has been written by the same 'hand'.

It has to be stressed that importance of certain attributes of script varies if one examins scribal hands or scripts. Finding and defining attributes that scripts have in common and then have an algorithm to process the image data might be as complicated as it is for palaeographers to agree on a common terminology.[12]

---

[11] Whether (text-)catalogues contain images or not seems to depend more on money than on a theory behind their establishment. Thus, catalogues without images are more frequent but there are a number of catalogues that contain sample images, cf. the catalogues of Jena."
[12] For further discussion of the issues cf. Stutzmann 2015.

7 Conclusions, or: Are algorithms the better palaeographers?

The answer to that question depends on what you want the palaeographer to do. The machine is definitely capable of recognising features of scripts. The algorithms can calculate means of pretty much everything: thickness, straightness, or orientation of strokes, height or width of lines, numbers of lines per page, etc. From those general features of script, the algorithm can determine clusters. Depending on thresholds, the machine is able to distinguish individual scripts and maybe even script families. Whether it is possible to attribute terms to these clusters, or if the algorithm ends at the same position as palaeographers did, is an open question.

Anyway, the ultimate question might be: to what end do we use the classification? One option might be to find as many objects as possible that meet one's criteria in the course of a research project in order to analyse them and answer research questions. The other might be to find the 'correct' items in a given set of objects. This would require much higher 'accuracy'. Institutions such as libraries may be more interested in the first option: one needs to find out about similar objects which have to be examined by experts in order to help them with their cataloguing or their research.

Is there anything like being 'right' or 'wrong' when talking about script or, more general, about palaeography?

Most importantly, the history of script(s) is no mono-hierarchical development, emerging from Capitalis via Uncialis to Minuscules, to name but a few. Thus, the technical means to deal with the phaenomenon of scripts could be the implementation and application of a taxonomy, but this would represent a mono-hierarchy. The better way would be to establish an ontology. In order to derive the net of taxonomies, the algorithms presented can be used to generate a basic knowledge.

One has to conclude an overview like the one above with the *almighty Bernhard Bischoff*. Hoffmann cites him as follows: "Berühmt, um nicht zu sagen berüchtigt, ist seine [Bischoffs] Charakterisierung des Reimser Stils: das lange s sei dort stärker geneigt gleich Getreidehalmen, über die der Wind gehe." *Famous, not to say notorius, is his [Bischoff's] characterisation of the Reims style: the long s were more slanted there like the stem of grain in the wind.* (Hoffmann 2015, 40) Not to be left with such wonderfully poetic yet hard to comprehend descriptions may be the task of the new methods.

Bibliography

Aussems, Mark and Axel Brink. "Digital Palaeography." In Rehbein, Malte, Patrick Sahle, and Torsten Schaßan (eds.). *Kodikologie und Paläographie im digitalen Zeitalter - Codicology and Palaeography in the Digital Age.* Norderstedt: Books on Demand, 2009. 293–308. URN: urn:nbn:de:hbz:38-29773.

Bischoff, Bernhard. *Paläographie des römischen Altertums und des abendländischen Mittelalters.* Second revised edition. Berlin: Schmidt, 1986.

Bulacu, Marius L. and Lambert R. B. Schomaker. [2007a.] "Text-independent Writer Identification and Verification Using Textural and Allographic Features." *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Special Issue - Biometrics: Progress and Directions* 29.4 (2007). 701–717.

Bulacu, Marius L. and Lambert R. B.Schomaker. [2007b.] "Automatic handwriting identification on medieval documents." *Proceedings of 14th International Conference on Image Analysis and Processing (ICIAP, Modena, 11–13 September 2007).* Los Alamitos: IEEE, 2007. 279–284.

Castro Correa, Ainoa. "Palaeography, Computer-Aided Palaeography and Digital Palaeography: Digital Tools applied to the Study of Visigothic Script." In Andrews, Tara and Caroline Macé (eds.). *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches.* Turnhout: Brepols, 2014. 247–272.

Derolez, Albert. *The palaeography of Gothic Manuscript Books from the Twelfth to the Early Sixteenth Century.* Cambridge: Cambridge University Press, 2003.

DigiPal: *Digital Resource and Database of Palaeography, Manuscript Studies and Diplomatic.* 2011-2014. <http://digipal.eu>.

e-codices. Virtual Manuscript Library of Switzerland. 2005-. <http://www.e-codices.unifr.ch>.

ENRICH: *ENRICH project.* University of Oxford. 2007-2009. <http://enrich.manuscriptorium.com>.

Europeana Regia. 2010–2012. <http://www.europeanaregia.eu/>.

Fecker, Daniel, Volker Märgner, and Torsten Schaßan. "Vom Zeichen zur Schrift: Mit Mustererkennung zur automatisierten Schreiberhanderkennung in mittelalterlichen und frühneuzeitlichen Handschriften." Baum, Constanze, and Thomas Stäcker (eds.). *Grenzen und Möglichkeiten der Digital Humanities.* (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). 2015. DOI: 10.17175/sb001_008.

Gilissen, Léon. *Prolgomènes à la codicologie: Recherches sur la construction des cahiers et la mise en page de manuscrits médiévaux.* Ghent: Éditions scientifiques, 1977.

Die Handschriften der Thüringer Universitäts- und Landesbibliothek Jena. Vol. 1–3. Wiesbaden: Harrasowitz, 2002–2016.

Hoffmann, Hartmut. "Bernhard Bischoffs Katalog der karolingischen Handschriften." *Deutsches Archiv* 71 (2015). 1–56.

Mandelbrot, Benoît. "How Long Is the Coast of Britain." *The Fractal Geometry of Nature.* New York (NY): W.H. Freeman and Co., 1983. 25–33.

ODD: *One document does it all.* Meta-schema file defined by the TEI which allows to generate schema files in multiple schema languages. <http://www.tei-c.org/Guidelines/Customization/odds.xml>.

Overgaauw, Eef. "Die Nomenklatur der gotischen Schriftarten bei der Katalogisierung von spätmittelalterlichen Handschriften." *Codices manuscripti* 15 (1994) 100–106.

RDF: *Resource Description Framework.* World Wide Web Consortium (W3C). 2014. <http://www.w3.org/RDF>.

RDFS: *RDF Schema 1.1.* World Wide Web Consortium (W3C). 2014. <https://www.w3.org/TR/rdf-schema>.

Riecke, Anne-Beate. *MXML-Dokumentation. Die Erstellung von Handschriftenbeschreibungen nach den Richtlinien der DFG mit Hilfe von Manuscriptum XML*. Berlin: Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, 2009.

Robinson, Peter. "Is there a Text in These Variants?" In Richard Finneran (ed.). *The Literary Text in the Digital Age*. Ann Arbor (MI): University of Michigan Press, 1996. 99–115.

Schreiber, Guus, cited by Ivan Herman. *Why OWL and not WOL?*. Tutorial on Semantic Web Technologies. World Wide Web Consortium. <https://www.w3.org/People/Ivan/CorePresentations/RDFTutorial/Slides.html#(114)>.

Stokes, Peter. "Computer-Aided Palaeography, Present and Future." In Rehbein, Malte, Patrick Sahle, and Torsten Schaßan (eds.). *Kodikologie und Paläographie im digitalen Zeitalter - Codicology and Palaeography in the Digital Age*. Norderstedt: Books on Demand, 2009. 309–338.

Stokes, Peter. "Palaeography and the 'Virtual Library' of Manuscripts." In Nelson, Brent, and Melissa Terras (eds.). *Digitizing medieval and early modern material culture*. Toronto: ACMRS, 2012. 137–169.

Stutzmann, Dominique. "Nomenklatur der gotischen Buchschriften: Nennen? Systematisieren? Wie und wozu? Recension of: Albert Derolez: The Palaeography of Gothic Manuscript Books. From the Twelfth to the Early Sixteenth Century. Cambridge et. al.: Cambridge University Press 2003.)." *IASLonline*. 2005. <http://www.iaslonline.de/index.php?vorgang_id=995>.

Stutzmann, Dominique. "Clustering of medieval scripts through computer image analysis: Towards an evaluation protocol". *Digital Medievalist* 10 (2015). DOI: 10.16995/dm.61/.

TEI: *Text Encoding Initiative*. <http://www.tei-c.org>.