Kodikologie und Paläographie im digitalen Zeitalter 3

Codicology and Palaeography in the Digital Age 3

Schriften des
Instituts für Dokumentologie und Editorik

herausgegeben von:

Band 10

Schriften des Instituts für Dokumentologie und Editorik — Band 10

Kodikologie und Paläographie im digitalen Zeitalter 3

Codicology and Palaeography in the Digital Age 3

herausgegeben von | edited by

Oliver Duntze, Torsten Schaßan, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Johanna Puhl, Patrick Sahle

2015
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.

Herstellung und Verlag: Books on Demand GmbH, Norderstedt
ISBN: 978-3-7347-9899-3
Einbandgestaltung: Johanna Puhl, basierend auf dem Entwurf von Katharina Weber
Satz: LaTeX, Bernhard Assmann und Torsten Schaßan

Inhaltsverzeichnis

Appendices

Vorwort

Das Institut für Dokumentologie und Editorik (IDE) präsentiert hiermit den dritten Band zur *Kodikologie und Paläographie im Digitalen Zeitalter*, veröffentlicht als zehnter Band in der Schriftenreihe des IDE. Der Band versammelt ein weiteres Mal aktuelle Berichte aus dem Bereich computergestützter Handschriftenforschung. Sie spiegeln den aktuellen Stand digitaler Forschung an mittelalterlichen Handschriften wider und bilden damit einen »Schnappschuss« einer sich rasant weiterentwickelnden Forschungslandschaft.

Das IDE hatte durch einen Call for Papers zur Einsendung von Beiträgen aufgerufen. Die eingereichten Beiträge wurden einer internen Begutachtung im Herausgebergremium und einem anonymisierten Peer-reviewing externer Fachgutachter unterzogen.

Verweise zu Webseiten und Online-Resourcen in den bibliographischen Anhängen schließen nach Möglichkeit Angaben zu Publikationsort und -zeitraum mit ein. URL-Adressen wurden generell Ende Mai 2015 überprüft.

Unser Dank gebührt allen beitragenden Autorinnen und Autoren dafür, dass sie diesen Band überhaupt erst ermöglicht haben. Darüber hinaus ist einer Reihe von unentbehrlichen Helferinnen und Helfern herzlich zu danken: Stefan Dumont (Berlin) für französische Korrekturen; Johanna Puhl (Köln) für die Einbandgestaltung; Bernhard Assmann (Köln) bewältigte erneut alle technischen Finessen der Drucklegung.

Berlin, Graz und Wolfenbüttel im Juni 2015, die Herausgeber

Preface

The Institute for Documentology and Scholarly Editing (IDE) hereby presents the third volume on *Codicology and Palaeography in the Digital Age*, published as the tenth volume in the IDE series. The volume again acquires papers from the area of computer-aided manuscript research. They document the current state of research on medieval manuscripts and form a "snapshot" of a fast changing field.

The IDE published a call for papers. The papers sent in have been reviewed both internally by the board of editors as well as by external experts in an anonymised peer-review process.

References to web sites and online resources in the bibliographies include, as far as possible, information on place and date of the publication. URLs were checked in late May 2015.

We are grateful to all contributors who made this volume possible. In addition, we have to thank for indispensable support at various stages in preparing this publication: Stefan Dumont (Berlin) for remarks and corrections on French texts; Johanna Puhl (Cologne) for cover design; Bernhard Assmann (Cologne) once again smoothly created the print version.

Berlin, Graz, and Wolfenbüttel, June 2015, the editors

Einleitung

Oliver Duntze

Das diesen Band einleitende »Manifest« ist das Resultat des durch einen der Herausgeber des letzten Bandes zur *Kodikologie und Paläographie im Digitalen Zeitalter* (KPDZ) in der Schriftenreihe des IDE, Malte Rehbein, mitorganisierten Dagstuhl-Seminars 12382 (Perspectives Workshop: Computation and Palaeography: Potentials and Limits).[1] Der Beitrag ist wegen seiner grundlegenden Ausrichtung als Zweitveröffentlichung, gewissermaßen als erweiterte Einleitung, in den vorliegenden Band aufgenommen worden. *Tal Hassner*, *Malte Rehbein*, *Peter A. Stokes* und *Lior Wolf* zeigen hier den Stand der computergestützten Paläographie in ihren verschiedenen Ausprägungen auf. Als Problemfelder stellen sich dabei die weniger durch technische als durch rechtliche Rahmenbedingungen erschwerte Datenerhebung dar, die Notwendigkeit einer interdisziplinären Zusammenarbeit von WissenschaftlerInnen sehr unterschiedlich geprägter Wissenschaftskulturen und das Problem technischer »black boxes«, deren Arbeitsweise für die paläographischen Fachwissenschaftler häufig nicht nachvollziehbar ist und deren Ergebnisse daher schwer zu interpretieren sind. Die Autoren schlagen verschiedene Ansätze zur Lösung dieser Probleme vor, u.a. die Ausbildung von »in-betweenern« - WissenschaftlerInnen, die sowohl eine geisteswissenschaftliche als auch eine informationstechnische Expertise besitzen, die Offenheit computergestützter Systeme, die den AnwenderInnen die Möglichkeit manueller Eingriffe ermöglichen, oder auch die Erarbeitung einer interdisziplinären Ontologie, die sowohl informationstechnologische als auch paläographische Sachverhalte abbilden kann.

I. Digitale Reproduktion als paläographisches Werkzeug

Die Beiträge dieser Sektion fallen in den von *Hassner et. al.* als »Data Acquisition« bezeichneten Bereich. Der Ausgangspunkt einer genuin ›digitalen‹ Untersuchung von Manuskripten ist in den meisten Fällen die digitale Reproduktion einer Handschrift. Neben herkömmlichen Reproduktionsverfahren – hochauflösende digitale Fotos oder Scans – existieren inzwischen verschiedene spezialisierte Methoden, die auch für Handschriften zur Anwendung kommen.

---

[1] *Dagstuhl Manifestos*, Volume 2, Issue 1, pp. 14–35. <doi:10.4230/DagMan.2.1.14>

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. IX–XII.

Multispektalfotografie (MultiSpectral Imaging, MSI) findet bereits häufiger Anwendung bei der Lesbarmachung schlecht lesbarer oder beschädigter Manuskripte. *Fabian Hollaus*, *Melanie Gau*, *Robert Slabatnig*, *William A. Christens-Barry* und *Heinz Miklas* stellen in ihrem Beitrag verschiedene Techniken vor, die zur Verbesserung der Lesbarkeit von MSI-Reproduktionen angewendet werden können und bewerten diese. Der zweite Teil des Beitrags schließt an die in KPDZ II publizierte Projektvorstellung »The Enigma of the Sinaitic Glagolitic Tradition« an und stellt neueste Erkenntnisse vor. Hier wurden die beschriebenen Techniken zur Lesbarkeitsverbesserung von MSI-Aufnahmen erfolgreich eingesetzt sowie Verfahren für Bildbearbeitung und Layoutanalyse weiterentwickelt.

Reproduktionstechniken wie die Multispektralfotografie zeigen deutlich die von *Hassner et. al.* hervorgehobene Notwendigkeit interdisziplinärer Zusammenarbeit im Bereich der digitalen Paläographie. Derartige Techniken bedürfen stets einer vertieften technischen Expertise, die in der akademischen paläographischen Lehre wohl kaum vermittelt werden kann. Die Zusammenarbeit geisteswissenschaftlicher und technischer Forschung ist hier unerlässlich.

Eine wichtige Voraussetzung für eine breitere Nutzung elaborierter Reproduktionstechniken wie der Multispektralfotografie wäre eine leichtere Zugänglichkeit und Portabilität der eingesetzten Geräte und – damit verbunden – eine Reduktion der Kosten für die Herstellung der Reproduktionen. Derzeit können ›neue‹ Reproduktionstechniken nur in wenigen Projekten eingesetzt werden, und paläographische Untersuchungen sind im Kontext der noch vorwiegenden Forschung als Leistung individueller Forscher, wie z.B. bei Dissertationen oder kleineren Forschungsprojekten mit geringem Fördervolumen, auf herkömmliche Reproduktionen angewiesen. In diesem Kontext ist *Christine Voths* Beitrag als praxisorientierter Beitrag zu verstehen, der zeigt, dass auch ohne spezialisierte Reproduktionsverfahren hergestellte Scans oder Fotografien als Ausgangsbasis für die Entzifferung schwer lesbarer und beschädigter Manuskripte dienen können. Am Beispiel einer medizinischen Handschrift kann *Voth* zeigen, dass in vielen Fällen auch die Bildbearbeitungsalgorithmen herkömmlicher Grafiksoftware – die richtige Einstellung der vielfältigen Parameter vorausgesetzt – zu einer nennenswerten Verbesserung der Lesbarkeit auch bei problematischem Ausgangsmaterial führen können.

II. Verwaltung von Erschließungsdaten

Die zweite Sektion versammelt Fallstudien, die sich auf die Organisation und Auswertung von Menschen erzeugter Daten über Schriften und Handschriften konzentrieren.

Den Nutzen quantifizierender Methoden in der Paläographie demonstriert *Rombert Stapel* am Beispiel des spätmittelalterlichen Schreibers Hendrik van Vianen. Die statis-

tische Auswertung verschiedener paläographischer Merkmale in der Transkription der Handschriften Henrik van Vianens ermöglicht es, einen skripturalen »Fingerabdruck« des Schreibers zu erstellen.[2] Dieser erlaubt nicht nur eine Schreiberidentifikation, sondern gibt auch Hinweise auf die chronologische Entwicklung der Schreibpraxis eines individuellen Schreibers. Anders als Beiträge aus dem Kern der Paläographie kann Stapel dabei auf eine graphematische Annotation weitgehend verzichten und seine Beobachtungen auf orthographische Phänomene stützen.

Matthieu Bonicel und *Dominique Stutzmann* stellen in ihrem Beitrag den Prototypen einer für das iPad entwickelten Applikation vor, die eine kollaborative Annotation von Digitalisaten mittelalterlicher Handschriften ermöglicht und für die Speicherung der Annotationen das SharedCanvas-Datenmodell einsetzt. Als Prototyp zeigt die Anwendung, auf welche Weise technische Neuerungen wie Touchscreens und mobile Endgeräte für die mediävistische Forschung eingesetzt werden können. Dabei rückt auch das in den Digital Humanities häufig vernachlässigte Problem der Ergonomie von Softwareanwendungen in den Fokus. Der Touchscreen ist als haptisch-visuelles mediales Interface dem Untersuchungsgegenstand, der mittelalterlichen Handschrift, deutlich näher als es bei herkömmlichen Desktop-Computern der Fall ist.

Für die Datierung von Papierhandschriften sind Wasserzeichen eines der wichtigsten Hilfsmittel. An verschiedenen Orten sind deshalb Dokumentationen entstanden, welche es erlauben sollen, Papiere gleicher Herstellungstranchen zu identifizieren. Der Beitrag von *Erwin Frauenknecht* und *Maria Stieglecker* berichtet von den Ergebnissen des Projekts, das *Christina Wolf* in KPDZ I[3] vorgestellt hat. Es ist dabei eine Web-Anwendung entstanden, die Bearbeitern und Benutzern umfangreiche Funktionalität zur Verfügung stellt. Ein solches Portal demonstriert eindrucksvoll, wie bei ihrer Entstehung akzidentielle Kulturprodukte durch moderne Informationstechnologien zu wichtigen Forschungsinstrumenten werden.

Der abschließende Beitrag verlässt den Bereich der Handschriftenforschung im engeren Sinne und widmet sich der Paläographie von Inschriften. *Elisa Pallottini*

---

[2] Vgl. dazu auch folgende, einschlägige Arbeiten: Fiebig, Annegret. *Urkundentext. Computergestützte Auswertung deutschsprachiger Urkunden der Kuenringer auf Basis der eXtensible Markup Language (XML).* Leinfelden-Echterdingen: Thorbecke, 2000. Schriften zur südwestdeutschen Landeskunde 33; Hofmeister, Wernfried, Andrea Hofmeister-Winter, Georg Thallinger. »Forschung am Rande des paläographischen Zweifels. Die EDV-basierte Erfassung individueller Schriftzüge im Projekt DamalS.« *Kodikologie und Paläographie im digitalen Zeitalter.* Hrsg Malte Rehbein, Patrick Sahle u. Torsten Schaßan. Norderstedt: BoD, 2009. Schriften des Instituts für Dokumentologie und Editorik 2. S. 261-292. <urn:nbn:de:hbz:38-29748>; Stutzmann, Dominique. »Paléographie statistique pour décrire, identifier, dater... Normaliser pour coopérer et aller plus loin?« *Kodikologie und Paläographie im Digitalen Zeitalter 2.* Hrsg. Franz Fischer, Christiane Fritze u. Georg Vogeler. Norderstedt: BoD, 2010. Schriften des Instituts für Dokumentologie und Editorik 3, S. 247-277. <urn:nbn:de:hbz:38-43535>

[3] Wolf, Christina. »Aufbau eines Informationssystems für Wasserzeichen in den DFG-Handschriftenzentren.« *Kodikologie und Paläographie im Digitalen Zeitalter.* Hrsg. Malte Rehbein, Patrick Sahle und Torsten Schaßan. Norderstedt: BoD, 2009, 97–107. <urn:nbn:de:hbz:38-29639>

beschreibt das von ihr zusammengestellte Corpus mittelalterlicher Inschriften in Viterbo und damit den Aufbau einer spezialisierten Datenbank mit einer lokal und zeitlich eingegrenzten Erfassungsbasis. Im Rahmen eines PhD-Projekts entstanden, kann sich der Umfang der Datenbank nicht mit dem Datenvolumen langfristiger Forschungsprojekte messen, doch sind die Aspekte der Corpuszusammenstellung, die *Pallottini* beschreibt, exemplarisch für viele kleiner dimensionierte Projekte. Der Übergang in strukturierte Dokumentationsformen macht die Ambiguität verbaler Beschreibungsformen, aber auch mit ihnen erfasster historischer Realität deutlich.

Die hier versammelten Studien zeigen die Vielfalt computergestützter paläographischer, kodikologischer und epigraphischer Forschung, doch wird auch deutlich, dass die von *Hassner et. al.* eingeforderte Standardisierung im Bereich der computergestützten Manuskriptforschung ein dringend zu bearbeitendes Desiderat ist.

Computation and Palaeography: Potentials and Limits[1]

Tal Hassner, Malte Rehbein, Peter A. Stokes, Lior Wolf (Eds.)

Abstract

This manifesto documents the program and outcomes of Dagstuhl Seminar 12382 'Perspectives Workshop: Computation and Palaeography: Potentials and Limits'. The workshop focused on the interaction of palaeography, the study of ancient and medieval documents, with computerised tools, particularly those developed for analysis of digital images and text mining. The goal of this marriage of disciplines is to provide efficient solutions to time and labor consuming palaeographic tasks. It furthermore attempts to provide scholars with quantitative evidence to palaeographical arguments, consequently facilitating a better understanding of our cultural heritage through the unique perspective of ancient and medieval documents. The workshop provided a vital opportunity for palaeographers to interact and discuss the potential of digital methods with computer scientists specialising in machine vision and statistical data analysis. This was essential not only in suggesting new directions and ideas for improving palaeographic research, but also in identifying questions which scholars working individually, in their respective fields, would not have asked without directly communicating with colleagues from outside their research community.

Zusammenfassung

Dieses Manifest dokumentiert das Programm and Resultate des Dagstuhl Seminars 12382 'Perspectives Workshop: Computation and Palaeography: Potentials and Limits'. Der Gegenstand des Workshops war die Interaktion von Paläographie und dem Studium antiker und mittelalterlicher Dokumente mit computergestützten Werkzeugen, insbesondere solchen für die Analyse digitaler Bilder und für Text-Mining. Mit dieser interdisziplinären Zusammenarbeit sollen effiziente Lösungen für zeit- und arbeitsintensive paläographische Aufgaben bereitgestellt werden. Darüber hinaus sollen Forscher quantitative Argumente für paläographische Diskussionen an die Hand gegeben werden, welche in der Folge ein besseres Verständnis unseres

[1] Author names in alphabetical order. This paper was first published in *Dagstuhl Manifestos*, Volume 2, Issue 1, pp.14–35. doi.:10.4230/DagMan.2.1.14.

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 1–27.

kulturellen Erbes durch die einmalige Perspektive antiker und mittelalterlicher Dokumente ermöglichen. Der Workshop bot eine wichtige Gelegenheit für Paläographen, mit Spezialisten für maschinelles Sehen und statistischer Datenanalyse ins Gespräch zu kommen und das Potential digitaler Methoden zu diskutieren. Dies war nicht nur essentiell, um neue Forschungsrichtungen und Ideen für verbesserte paläographische Forschung zu entwickeln, sondern auch, um Forschungsfragen zu identifizieren, welche die einzeln in ihren Feldern arbeitenden Forscher nicht ohne die direkte Kommunikation mit Kollegen außerhalb ihrer Forschungscommunities gestellt hätten.

Executive Summary

Perspectives Workshop 18.–21. Sept., 2012 – http://www.dagstuhl.de/12382
1998 ACM Subject Classification I.5.4 Applications (Text processing, Computer vision), I.7 Document and Text Processing, H.3.7 Digital Libraries, J.5 Arts and Humanities (Literature)

The Schloss Dagstuhl Perspectives Workshop on 'Computation and Palaeography: Potentials and Limits' focused on the interaction of palaeography, the study of ancient and medieval documents, and computerised tools developed for the analysis of digital images in computer vision (a full report of which is available in [18]). During the workshop, the interaction between domain experts from palaeography and computer scientists with computer vision backgrounds has yielded several very clear themes for the future of computerised tools in palaeographic research. Namely,

- difficulties in communication between palaeographers and computer scientists is a prevailing problem. This is often reflected not only in computerised tools failing to meet the requirements of palaeography practitioners but also in the terminology used by the two disciplines. Better communication should be fostered by joint events and long-term collaborations.
- computerised palaeographic tools are often black boxes which put the palaeographer on one end of the system, only receiving a systems output, with little opportunity to directly influence how the system performs or to communicate with it using natural palaeographic terminology. The long-term desire is to have the scholar at the center of the computerised system, allowing interaction and feedback in order to both fine-tune performance and better interpret and communicate results. This is crucial if palaeography is to become a truly evidence based discipline. To this end the use of high-level terminology, natural to palaeography, should be integrated into computerised palaeographic systems.
- palaeographic data, scarce to begin with, is even more restricted by accessibility and indexing problems, non-standard benchmarking techniques and the lack

of accurate meta-data and ground truth information. Multiple opportunities were identified for acquiring data and disseminating it both in the palaeographic research community and outside to the general public.
- palaeographic research is largely restricted to the domain of experts. Making palaeography accessible to non-experts by using computerised tools has been identified as an effective means of disseminating valuable cultural heritage information while at the same time potentially giving rise to crowdsourcing opportunities, such as those proved successful in other domains.

In addition to these themes, several specific recommendations regarding research infrastructure and support were made. These include:

1. A clear articulation of standards for digital image acquisition followed by all digital imaging projects when possible.
2. EU-wide harmonisation of copyright and licensing practices. Copyright or contractual use restrictions on photographs of cultural heritage items create many barriers for researchers. In many cases, tax-funded or state-supported research projects must expend significant financial and human resources on negotiating and paying for reproduction rights, even if those rights are being obtained from state repositories.
3. Ideally, set copyright appropriately to allow for large-scale studies of collections of manuscript images. Making large sets of images more easily available at an international scale would greatly facilitate the pursuit of significant new research questions.
4. Encouraging an interdisciplinary research agenda including disciplines dealing with computable images from various perspectives such as medical imaging, cognitive sciences, Cultural Heritage Imaging (CHI), or Natural Language Processing (NLP).

This manifesto elaborates on the existing challenges and limitations of the field and details the long-term recommendations that have emerged in the workshop.

1. Introduction

Manuscripts are the most important witnesses to and artefacts from our shared cultural heritage of the European Middle Ages. Current estimates are that close to one million manuscript books survived along with countless archival documents from a period stretching across more than a millennium. Cumulatively, these documents are the chief sources of history, history of science, literature, and art history (due to the presence of manuscript decoration) from that period. Moreover, these manuscripts are important subjects of scientific enquiry in their own right, as they bear witness to the

history of the book, to scribal and monastic culture, the history of the development of handwriting systems, languages and dialects, the history and genealogy of texts over time, and the evolution of strategies for organising texts and knowledge.

Although often taken more broadly, palaeography is in essence the study of old handwriting from manuscripts. As such, palaeographers are often asked one of four questions regarding manuscript documents from the past: what was written? when was this written? where was it written? and by whom? Answering these questions, and indeed reading the text itself, are basic prerequisites for any kind of work with primary sources, and the study of almost all fields relevant to the ancient and medieval past therefore depends on them. In this respect palaeography is sometimes regarded as a "mere" auxiliary discipline. However, palaeography also extends beyond this: it encompasses the history of one of humanity's most pervasive technologies – writing – and therefore raises questions of cultural history, the development and spread of ideas, and so on, along with the deep understanding of the transmission and use of texts which it brings. Misunderstandings here can lead to significant errors in scholarship, such as basing historical arguments on charters which prove to be late forgeries [47], or conducting studies of spelling and automatic authorship attribution without considering the effects of textual transmission, both scribal and editorial, and the changes that this brings [49].

Palaeography as a discipline typically involves difficult, complex, and time-consuming tasks, often involving reference to a variety of linguistic and archaeological data sets, and the invocation of previous knowledge of similar documentary material. Due to the involved reading process, it is difficult to record how the final interpretation of the document was reached, and which competing hypotheses were presented, adopted, or discarded in the process. It is also difficult to acknowledge and present the probabilities and uncertainties which were called on to resolve a final reading of a text. As a result, palaeographical discussion tends towards assertions based on experience with little supporting evidence – sometimes none at all – and this has led to an allegedly "authoritarian" discipline which depends on "faith" [10] or "dogma" [16] and is based on "informed guesswork" [16]. It is perhaps no surprise that the discipline itself suffers as a result [10, 3].

Palaeography as a discipline is, however, of high relevance for society and economy. All of the world's written heritage was written by hand until the invention of printing, and texts written by hand have remained important ever since. Manuscripts are hence one of the major sources of knowledge of human culture and society, crossing the borders of modern nations, for most of what we call history. However, unlike printed texts which are distributed through libraries, handwritten sources are often accessible only to a very small and highly trained group of experts, and hundreds of thousands if not millions of manuscripts are scattered around the world. They can be difficult to find and difficult to read, are often written in an old language, and

frequently deal with a subject matter that can be understood only by experts. On the other hand, however, they can be a valuable resource also for public interest such as regional economies and tourism, as demonstrated by highly successful exhibitions which charge for entry such as the Book of Kells at Trinity College in Dublin, a book which was also an inspiration for creativity and the generation of further derived art. There are relatively few examples of manuscripts exploited in this way, but this material remains important for connecting people with their heritage and fostering identity, be it local, regional, national or pan-national.

Research can enhance and popularise the access to this largely untapped resource and can increase the number of beneficiaries of the documents. This is an investment that may bring large returns in the long term. In addition, the area of digital palaeography which is examined in this manifesto promotes technical research in challenging problems, such as processing of ancient documents, and can help develop techniques that may be helpful in other areas.

2. Computation in Palaeography

2.1. State-of-the-Art

Partly in response to the perception of palaeography as "dogma", scholars worldwide have been developing and employing new technologies and computer-based methods for palaeographic research. This approach, often referred to as Digital Palaeography [7] and situated in the wider field of Digital Humanities, aims to improve and enhance the traditional methods. Its goal is to help efficiently solve palaeographic issues and/or provide more quantitative evidence to palaeographical arguments, and in consequence to cater for a better understanding of our cultural heritage.

As of today, there are numerous projects concerned with developing such methodologies. These encompass a wide range of scientific, interdisciplinary approaches such as forensic document analysis, optical character recognition, quantification of "scribal fingerprints", metric analysis, quantitative methods, advanced manuscript analyses such as DNA and imaging techniques such as multi-spectral digitisation, classification systems and databases. Although some achievements have been made already, much research is still required. For instance, something as seemingly fundamental as the automated recognition of characters in handwritten texts has proven extremely complex, due largely to the very wide variation in styles of handwriting, the often poor quality of surviving manuscripts, the lack of standard orthographies which complicates prediction, and so on.

Such computational methods as proposed by digital palaeography have been the subject of research in the last few years, but most of this has been theoretical or applied only to small cases, partly because of the very high degree of labour that is

typically involved [7, 2, 48, 40]. The applications to date have also focused almost exclusively on the question of scribal identity, ignoring other aspects of palaeography. Furthermore, they tend to view letter-forms as objects outside the manuscript or documentary context in which they were written, but palaeographers have long understood that handwriting depends heavily on the context in which it is produced ([4, 53], among many). Much more significantly, these methods tend to make the computer a "black box" which receives images of manuscripts at one end and returns a classification of the handwriting at the other (for examples see [40]). However, they are normally heavily dependent on very subtle and often unstated assumptions about the underlying data [44], but it is difficult or impossible for "traditional" palaeographers to evaluate these, so that usually scholars cannot evaluate the "black box" and so are rightly reluctant to accept its results [49, 58, 9, 43]. The major challenge for computational approaches is to provide a system which presents palaeographical data quickly and easily in a way which scholars can understand, evaluate, and trust. The success and impact of research and initiatives in computational methods so far ([54] with publications [40, 15, 13, 30, 5]) has shown the strong need to combine scientific computing and palaeography in order to further investigate the interdisciplinary methods and scientific fields. It is also apparent that no institution – let alone a single scholar – is capable of undertaking comprehensive research that encompasses all those methods (and potentially more). Thus, a joint effort is required, preferably on an international level.

As became very evident during scholarly meetings on this topic [54, 62, 5], palaeographers and computer scientists tend to think in different terms and tend not to agree even on very basic notions such as "evidence" or "meaning". Successful collaboration between researchers in humanities and in computer science is not nearly as simple as "define a computational problem and find an algorithm to solve it." The input is often loosely defined, and the output needs to be more than just a score on some abstract scale. It is therefore crucial to identify a common level at which effective communication can be established.

2.2. Challenges

During the Dagstuhl Perspectives Workshop, the unmediated interaction between palaeographers and computer scientists yielded several very clear questions and themes for the future of research in Digital Palaeography. These include the following four challenges:

1. *How to optimise collaboration between all the different domain experts involved in Digital Palaeography*?
   Barriers in communication between palaeographers and computer scientists are a prevailing problem. This is often reflected not only in computerised tools failing

to meet the requirements of palaeographers but also in the different terminologies used by the two disciplines. It was recommended that better communication should be fostered by joint events and long-term collaborations.

2. *How to ensure that palaeographers remain in control of their research, whilst taking advantage of the possibilities of computerised approaches*?
   Computerised palaeographic tools are often "black boxes" putting palaeographers on one end of the system, only receiving a system's output, with little opportunity to directly influence how the system performs, or to communicate with it by using natural palaeographic terminology. The long-term desire is to have the scholar at the centre of the computerised system, allowing interaction and feedback in order both to fine-tune performance and to interpret and communicate results more effectively. This is crucial if palaeography is to become a truly evidence-based discipline. To this end the use of high-level terminology, natural to palaeography, should be integrated into computerised palaeographic systems.
3. *How to facilitate sharing, not only of palaeographical data and results, but also of the methodologies involved in palaeography generally*?
   Palaeographic data is scarce and access to it is restricted by copyright and indexing problems, non-standard benchmarking techniques, and the lack of accurate meta-data and ground-truth information. During the workshop, multiple opportunities were identified for the acquisition of data and for its dissemination in the palaeographic research community and to the wider public.
4. *How to use the outreach potential offered by computerised technologies to enrich palaeographical knowledge*?
   Palaeographic research is an expert domain. Making palaeography accessible to nonexperts by using computerised tools has been identified as an effective means of disseminating valuable cultural heritage information while at the same time potentially giving rise to other opportunities, such as crowd sourcing and others which have proved successful in other domains.

2.3. Needs

In this manifesto we address both the technical aspects of the collaboration between computer scientists and humanists as well as conceptual tools such as "mid level features" and "ontologies" (discussed below) that can serve as means for effective communication among practitioners. The emphasis of this discussion is not on the most efficient algorithm, producing the most accurate results. It is also not on the least ambiguous and most meaningful definitions. Instead, the emphasis is on the most effective and fruitful communication.

Data Acquisition

Repositories across the European Union have been engaged in large-scale digitisation efforts in recent years, resulting in collections of hundreds of thousands or even millions of digital images of manuscript books and materials. Digital Palaeography relies on the existence of these digital surrogates of manuscripts. Moreover, some of the most exciting prospects of this field can only be demonstrated on sufficiently large collections. However, enabling this first requires modification of both policies and acquisition practices.

Specifically, from the computer user's perspective, obtaining digital copies calls for suitable procedures and for standardisation. Recently, Shweka et al. have suggested specific practices drawing on their experience in large-scale digitisation [45]. These suggestions range from minimum resolution, to the usage of particular rulers and background, and also include suggested policies regarding availability and manipulability during viewing. It is emphasised that taking into account the potential usage of a computer system to analyse the image does not degrade the experience of the human viewer. For example, while image analysis is much easier on a blue/green background, for human viewing purposes, such a background can be easily replaced. We propose the following:

1. A clear articulation of standards for digital image acquisition followed by all digital imaging projects when possible. Where such standards already exists (e.g. "DFG-Praxisregeln 'Digitalisierung'" of the German Research Foundation (DFG) [17] or JISC Guidelines in the UK [23]), they should be checked against the requirements of palaeographers (see also [56]) and, if necessary, extended to encompass and meet them in full. These include practices such as:
    - Proper use of colour bars and grey cards.
    - Appropriate use and documentation of illumination and equipment (e.g. lighting parameters including positioning, hardware).
    - References to size of original objects using shared standards.
    - Metadata descriptions of digitised objects following internationally accepted standards such as MIX/METS; if one takes several images of the same object (e.g. different lighting, multiple sizes, multispectral), it is important that the corresponding metadata indicates that these are images of the same object, and what the relationship between the images is.
    - Information that links multiple names and catalogue records when original objects have no single identifier (e.g., a manuscript with shelf marks that change over time and that is also referred to by other common names in scholarly literature).
    - File naming conventions in order to facilitate the creation of good metadata

and their proper sequence of images when books or other documents are being digitised.

2. A set of guidelines articulating how to capture digital and analogue images across a wide range of technologies – e.g., scanning objects and photographic negatives, using digital and analogue cameras, digitising microfilm.
3. EU-wide harmonisation of copyright and licensing practices. Copyright or contractual use restrictions on photographs of cultural heritage items create many barriers for researchers. In many cases, tax-funded or state-supported research projects must expend significant financial and human resources on negotiating and paying for reproduction rights, even if those rights are being obtained from state repositories (cf. [33] and [59]).
4. Furthermore, rights tend to be granted only to scholars or research groups on a one-by-one basis, which frustrates large-scale studies of collections of manuscript images [42]. It might be useful to call attention to libraries and museums with progressive policies that help researchers, such as the Austrian State Library, which makes images paid for by one project freely available to subsequent researchers needing those images. Making large sets of images more easily available at an international scale would greatly facilitate the pursuit of significant new research questions (e.g., large-scale comparative studies of handwriting that map regional and national developments of hands across time).
5. Freedom of resources produced by cultural institutions must be actively encouraged because it benefits the owners and enables research. The more it generates connections the more it becomes valuable: as well as research connections, it also generates connections back to the institutions themselves, bringing value to those institutions (as demonstrated by examples such as [11], for which see further below).
6. Encouraging an interdisciplinary research agenda including disciplines dealing with computable images from various perspectives such as medical imaging, cognitive sciences, Cultural Heritage Imaging (CHI), or Natural Language Processing (NLP).

Tools, Libraries and Resources

The overall objective of tools, software libraries, and resources to be developed in the context of palaeography is to provide support in establishing the correlation between text as shape and text as meaning; which, in the most general of senses, can also be understood as one of the aims of palaeography as a subject.

The starting point is to firmly acknowledge and map out the domains of expertise of the agents involved in the process, namely, humans and computer-based tools. On the one hand, computers excel at dealing with "big data", namely at tasks ranging

from holding large amounts of data in memory to carrying out process-intensive computations such as the identification of fine differences and rare occurrences within large datasets. On the other hand, humans (including palaeographers) excel at dealing with data which is ambiguous, complex, or broad, in the sense that the datasets are made of heterogeneous pieces of data. Humans also excel at making sense of the data, at expressing its gestalt in the sense that the whole of the data expresses more than the sum of its parts.

Taking these distinct sets of skills into account, the highest priority in developing computational resources for palaeography is the production of semi-automatic and interactive tools, where palaeographers can continually intervene, inform, correct, understand, use, and reuse results produced by and processes implemented by these tools. Only in this manner will palaeography benefit optimally from the respective strengths of the human and computational agents. Ideally, developing such semi-automatic and interactive tools will stimulate the establishment of a mutually beneficial continuous feedback loop between human and machine, whereby humans will be involved at all levels of reasoning, machines will be able to learn from human input, and palaeographers and others will learn and create new knowledge more effectively through the use of machines [22, 25].

We recognise that a critical mass of data is required for performing research, and the preattentive perception of the data by researchers is a major factor in building new hypotheses. This critical mass of data can, on the other hand, only be obtained through usable and ergonomic tools. Hence, in tool development for Digital Palaeography, focus groups, user testing and proper user interface design is needed in consultation with humanities scholars as end-users (for which see also [24] and [26]). A further requirement that emerges from here is the recognition of tool-development as academic research to encourage Digital Humanities scholars to publish their work and make it usable by a broader audience.

In the following, we outline the specific levels at which helpful computational tools can be developed as well as possible ways of keeping the humans in the loop. All tools developed should be compatible with one another and combinable at will (or, more precisely, as long as the notions involved are compatible, the tools should be). They might be used sequentially, or contribute to one another. We have identified the following categories of multi-level computational tools for Digital Palaeography:

1. Low-level tools:
    - Binarisation
    - Segmentation
    - Alignment, matching and registration of features (for similarity measures) including expert features of handwriting extraction (e.g. angles, curvatures, strokes)

- Physical feature extraction
- Similarity measures (for comparison between characters, words, texts, fragments, documents, corpora)

2. Mid-level tools:
    - Clustering
    - Classification
    - Character recognition
    - Word spotting
    - Cross-modality search engines, where the input for the searches might not be in the same form as the dataset that is searched, e.g.:
        - Search for a string in a text / corpus
        - Search for an image in a text / corpus
        - Search for a string in an image / a set of images
        - Search for an image in an image / a set of images
        - Search for a shape (shape would here be a hand-drawn input e.g. SVG, as opposed to an image that would be in a rasterised format)
    - Image-text (shape-meaning) correlation
3. Databases, where the data is organised in a way that allows fast queries of (for example):
    - Metadata
    - Transcripts
    - Images
    - Properties of the text (author, genre, date etc.)
    - Scripts and scribal features
4. Higher-level tools:
    - Interfaces, ergonomics, user experience ("UI"/"UX")
    - Searches of combinations of characters/words (bigrams, trigrams, possibly of shapes and/or images)
    - Correspondences in expert vocabularies
    - Inferences of paraphrases and synonyms for searches through metadata (widening searches by applying fuzzy techniques on search terms, by proceeding by analogy, etc.)
    - Web services
    - Web-based research environments for online collaboration and benchmarking within a global community.

Approaches and tools that keep humans in the loop can further be classified along

two main lines: data acquisition/exchange, and cognitive triggers/feedback loops. These include:

1. Data acquisition and exchange:
    - Provision of training data / annotated data
    - Online training / expert-in-the-loop
    - Crowd-sourcing
2. Feedback loops and cognitive triggers:
    - Drawing / touch screen technologies
    - Simple interactive image enhancements
    - Visualisation aspects of interactions with all the tools listed above (of results, of databases), interactive visualisations – e.g., time varying graphs – with customisability as a priority [22, 26]
    - Rationale building support, tracking of expert hypotheses in interpretation building
    - Statistical tools – with tests of significance
    - Information sharing systems
    - Transcription tools linking text and image.

3. Towards a Research Agenda for Computation and Palaeography

3.1. Challenges

This section is focused more on challenges than on constraints. We use the term "challenge" because it seems that, although the hurdles presented below do constitute some forms of limitations, we do not believe them to be insurmountable.

Context and Meaning

The first observation is that something is generally excluded from systematic analysis, namely the interpretation of data. Contextual knowledge and meaning, which are required for interpretation, are both concepts that are usually best handled by humans because they involve unstructured and non-formalised knowledge. This type of knowledge is often implicit in the natural scholarly environment, and although attempts can be made to structure and formalise contextual knowledge and sense-making processes, their continual evolution dooms the task to permanent incompleteness – which justifies the palaeographers' wish for more involvement and interactivity at various levels of computational formulation and formalisation.

Access to Data

The second observation is that, beyond context and meaning, current computational constraints are usually related either to access to data or to data retrieval. The problem of access to data is largely political in nature, therefore differing between countries and bylaws involved (e.g. in the UK, access and use of images can be drastically restricted due to copyright and licensing issues: see "Data Acquisition" above). In case of data retrieval, it is largely the degree of flexibility of the search tools that limits their usability, their usefulness and thereby their use. Search tools often present difficulties of precision and recall, and this is usually due to parameters being either too inflexible or, paradoxically, too flexible. This mismatch between the flexibility needed and the flexibility provided by the search tools is in fact a good illustration of what we have identified as the major bottleneck in the collaboration between computational and palaeographical research, and that bottleneck once again is down to communication.

It might seem at first that problems in communication are easy to solve, and that it is "just" a matter of listening and understanding, a matter of ironing out differences. However, even in our group of twenty people at Dagstuhl from different backgrounds, where all were accustomed to collaborative scholarship, a striking recurring difficulty in understanding each other was apparent – a trait that would most definitely be accentuated in a larger group and in a group where cross-disciplinary communication is not a current practice. Some of the symptoms of this problem are enumerated below, as well as some examples, and possible measures to treat them. Unfortunately, these hints for solutions will only treat the symptoms; the roots of the problem run much deeper and are mostly cultural, originating in the traditions of each and every discipline.

Interdisciplinary Approaches to Research

Scholarly endeavours are all rooted in their own traditions. In spite of our non-subscription to Snow's Two Cultures theory [46] – the next section on terminology will illustrate how the business of creating knowledge is a Many-Cultures system rather than a Two-Cultures system – one of the high-level observations about research methodologies in palaeography and in computer science is that they differ widely. As noted above, computer scientists tend to be problem solvers. Their approach to knowledge creation is typically to break down a large task into smaller tasks and then to solve these tasks, iteratively, until a satisfactory solution of the initial large task is found (where "satisfactory" is often left to their own discretion). In the tradition of computer sciences, there is a further convention of not deriving natural interpretation from the methodology. In other words, the output needs additional cognitive processing to be interpreted, and computer sciences do not traditionally have ways of doing so. In contrast, palaeographers tend to approach knowledge creation in a different way. Their method is typically to derive questions from questions, where

a new question often has the value of answering the preceding question (see further "Exploring and Questioning, not Answering", below).

Communicating between these two approaches can evoke situations in which what may first seem to be a misunderstanding or misconception turns out, finally, to release synergies. Take, for instance, a question in palaeography for which a corresponding computational solution has been developed. The discussion between palaeographers and computer scientists might lead to an emphasis on the weaknesses or incompleteness of the proposed computational solution. But it might also reveal the need to reformulate the original question, or might open up the potential for new, related questions. In any case, this interdisciplinary communication helps to augment research on either side, and ideally on both sides.

It seems, therefore, that practices exist by which scholars operate at different levels of abstraction and explicitness; for example, palaeographers' relatively abstract way of formulating problems might not translate well into formal computer language. Being aware of these different modes of communication might help to smooth out some of the difficulties and minimize possible frustrations, but the differences in traditions are not likely to change much, and the problems of terminology remain to be addressed. Indeed, these differences are strengths insofar as they allow approaches to different types of questions, and so they should be embraced rather than ignored or suppressed.

Terminology

As hinted above, the differences in research cultures are deeper than different methodological approaches to research (e.g. questioning versus problem-solving). For example, the use of specialised terminology in each domain, where words can coincide but carry different meanings, presents a much greater challenge than is apparent at first sight. A telling example is that of the word "feature". In image processing, "feature" has a very specific meaning: it describes a defined behaviour in terms of signal, an idealised profile such as a step, a ridge, a trough. In palaeography, too, the word "feature" is used with a very specific meaning; it describes the aspects of a stroke that make it characteristic of a certain hand, a certain scribal school, a certain area, or a certain type of document (e.g. its ductus, or the variation in its width). The two domains have therefore their own typical – i.e. accepted and shared – use for the word within their community, but this usage does not translate smoothly from one community to the other. This example is only one of the many that illustrate the terminological challenges that might be encountered (some others are "ontology" and "pattern" which are discussed further below).

It is also worth noting that this issue with the uses of specific terms in various disciplines constitutes a bottleneck in communication not only between computer scientists and palaeographers. Within the computer sciences themselves, communities

such the data mining community and the image processing community also share some words, but not necessarily the meaning attached to them ("feature" is an example once again). Similarly, palaeography has long been troubled by differences in terminology, despite the best efforts of the Comité international de paléographie latine and others to standardise them. The differences run deeper than simple choice of words: expert vocabularies in each discipline and in each sub-domain carry their own implicit contexts and assumptions that can prevent people from understanding each other across and even within fields (cf. [38] and [10]).

The Problem of the Black Box

The last type of bottleneck for communication and mutual understanding across scholarly disciplines resides in the fact that expertise implies tacit knowledge, and tacit knowledge tends to produce "black boxes", namely systems – whether human or machine – which take inputs and produce results without giving any indication of how those results were obtained. Computational algorithms are often perceived as black boxes by palaeographers, and palaeographical expertise is also seen as a kind of black box by computer scientists and indeed by other experts in the Humanities. The main issue here is to not concentrate exclusively on "cracking open" the black boxes to understand all the internal nuts and bolts that power them. Rather what is required is the establishment of trust between the communities. This trust might best be created by communicating an understanding of the principles and assumptions behind the inner working of the black boxes and not of the details of the methods and their implementation. Establishing that trust will alleviate the anxieties that black boxes tend to generate; it will thereby ease communication and collaboration.

Two (non-exclusive) natural solutions to such bottlenecks and lack of trust can be summarised as the introduction of an "in-betweener" and communication of "mid-level features"; both of these are discussed further below.

3.2. Directions

It is worth noting that the technical limitations outlined above are not reviewed in more detail here because, in the light of the potential problems in communication already discussed, they seem largely surmountable. In fact, through the discussions, round tables, and Q&A-sessions during the Dagstuhl workshop, it often emerged that computational approaches offer a lot more possibilities than single experts might have predicted. As a result, any prognosis of technical limitations voiced here would carry the inherent risk of outlining pre-emptive delimitations.

Interdisciplinarity and the "In-Betweener"

The Dagstuhl workshop can serve as a best-practice or "template" for future interdisciplinary communication. Further joint sessions at conferences and similar events need

to be held. But communication between computer scientists on the one hand and Humanities scholars on the other is only a starting point. Interdisciplinary projects between the fields need to be strengthened, and all participating disciplines will draw significant benefits from them. Experts in scientific computing should not merely implement requirements formulated by the Humanities, but should also suggest ideas based on their excellence and expertise. At the same time, scholars in Computer Science should acknowledge the relevance of research questions and methods from the Humanities. Although the disciplines have different semiotics and separate proof systems, interdisciplinary communication and cooperation leads to better understanding and consequently to new knowledge.

Interdisciplinary workshops are invaluable, but also necessary is the interdisciplinary individual: the "in-betweener" introduced above. This is a middle-person, a translator: a person who is versed enough in each of the collaborating fields to understand enough of each of the discipline-specific lexical fields to foster good communication and fruitful exchanges. Dedicated specialised "in-betweeners" have already been used very successfully in some Digital Humanities contexts, such as the positions of "project analysts" at the Department of Digital Humanities in King's College London, and their application to palaeography is to be encouraged.

Communication, Intelligibility and the "Black Box": Evidence-based Palaeography

Given the task of classifying a written fragment, an authoritative palaeographer might examine the page and simply state his or her classification of it, typically providing little evidence for how this conclusion was reached [16, 10, 9]. Somewhat analogously, given an image of a fragment, a computerised system might output the class of script for which the fragment scored the highest, along with the score itself. The mathematical procedures and formulas that led to this conclusion would remain inaccessible inside the "black boxes." Both the authoritative palaeographer and the computer leave little room for further discussion or debate on the results, and their work is therefore somewhat limited in expanding the science of palaeography, even though the answer might still be of a great help to a historian, for example, working on that specific manuscript.

Here, we suggest establishing a system for palaeographic representation which is accessible to both scholars and machines and can serve as the foundation of an evidence-based palaeography.

This representation system would rely on "mid-level" features or descriptors as introduced above. The mid-level features seek to define a shared vocabulary between disciplines, a shared meeting ground where each field can intervene with its own perspective. The term "mid-level" here means that these descriptors require visual

identification, unlike low-level features that are extracted computationally from the images and which cannot readily be verified by a human observer. This identification is meant to be as unambiguous as possible, such that if one researcher or computer system identifies or detects that such a feature exists in the test, other researchers or systems can verify this claim. In other words, one should be able to dispute almost completely on a factual basis any evidence that is structured according to these mid-level features.

Determining the mid-level qualifier is crucial: communication needs to be more finegrained than any abstract conceptual discussion around principles would be; and it must not become bogged down in the (sometimes murky) fine details. Specifically, these features must hold high-level meanings to the palaeographers on one hand, but must still be concrete enough to be definable in terms of a computerised system on the other. From a computational point of view, this is similar to the way by which facial features are used to identify faces in photos [61]. An example candidate for mid-level features are palaeographic "letter features", used to describe and identify handwriting. These are amenable to computerised analysis [29]. This approach has the inherent risk of systematising and formulating the field-specific strategies, thereby possibly compromising the potential for creativity as well as the integrity of each discipline, slowing down progress and over-constraining the problem spaces. However, this seems to be a more than acceptable compromise compared to the risks carried by repeated breaks in communication and failed exchanges. The approach therefore warrants much more investigation, and as a starting-point mid-level features and their application to the "black box" problem are discussed further below.

The use of mid-level features requires both the authoritative palaeographers and some of the computer systems to adapt. Palaeographers, like experts in other domains, often cannot provide the rationale that led to their decisions. In data analysis, systems that are built for maximal accuracy are built to solve the specific task at hand, and not necessarily to rely on rules that are interpretable; adding to the requirement of accuracy the requirement of interpretability would typically hurt the performance of the system. Despite these adaptations, the potential benefits are very significant. While scholars are reluctant to use the output of black-box systems, we expect much easier adaptation to computer systems which provide clear evidence for their classification. As was discussed during the Dagstuhl Workshop, the choice is between having an accurate system that lies unused and having a somewhat less accurate system that scholars are happy to employ.

Ontologies instead of Terminologies

Difficulties in communication have arisen several times in the discussion already, including not only communication between disciplines but even within them. In

particular, many efforts have been devoted to creating a unified terminology in palaeography [10]. These efforts have met with great challenges and were not able to achieve their goals. We suggest embracing the differences in terminologies and the complex relations that exist between them and focussing instead on developing an ontology.

An ontology, in this understanding, is a representation of a knowledge domain which is based upon well defined entities, each having a unique meaning. Various structural links are then used to define relations, such as "subst of" (also known as "is a"), "related to", and so on. Each term can also contain a list of synonyms and translations, a definition, references to other terminologies, and remarks. Instead of traditional classification systems, ontologies are being used more and more widely in Humanities scholarship, especially in cultural heritage documentation, because of the much greater flexibility that they allow. Examples of ontologies that are widely used in the Humanities include the EDM model of the Europeana library of digital objects [14]; and the CIDOC Conceptual Reference Model [20], which has become an international standard [21].

An example of part of the envisioned ontology for palaeography could be:

[TERM]
id: PCO0000345
name: triangular ascender
def: triangular decoration at the tip of an ascender
synonym: wedged ascender EXACT
related to:
is_a: PCO0000221 ascender decoration

The top level of the ontology could contain the terms "Manuscript Studies" and underneath "Palaeography" and "Codicology". The latter could be based on an effort recently envisioned ([28], and compare also [31]). We note that the field of palaeography is much less ordered than codicology, and this can lead to challenges in representing it. Nevertheless, the top distinctions under palaeography could be "Allographs" and "Graphemes", or something similar. Some of the terms would be descendants of terms from both these branches, e.g., "Caroline a" (cf. [51]).

Ontologies allow for unified treatment of metadata associated with documents as well as to mining of such resources. For example, projects like the "Medieval Electronic Scholarly Alliance" [34] and the "Manuscripts Online" project [32] aim to provide federated searches which span multiple resources. However, each resource might use a different name for exactly the same term, or the same name for different things. For example, English vernacular script of the eleventh century has been labelled "Caroline minuscule" [27], "Anglo-Saxon Round minuscule" [6], "Anglo-Saxon Vernacular minuscule" [12], and "English Vernacular minuscule" [52]. Given an

ontology, it is fairly straightforward to come up with reasonable methods to expand the search to include all these terms, and then rank the combined results together, and this is an approach which those projects will use, although its application in practice is far from trivial.

In the past, some projects on building ontologies (in general) were less successful than others. One of the authors of this manifesto (TH) has participated, as a student, in the construction of an ontology for representing 3D shapes. He describes a very frustrating process in which committee members debated many minute details and which ultimately led to a tool that no researcher uses. In contrast, another author (LW) is an avid user of the Human-Phenotype-Ontology [19], which is used by clinical geneticists to describe phenotypes, many of which are visual. He has witnessed the power of the ontology in facilitating the merging of disperse terminologies and the extremely useful data mining and classification tools that that it has entailed. Interestingly, this ontology has originated from a collection of medical data called OMIM, which was written by thousands of authors, each using their own terminology [36].

Based on discussing these cases we came to the conclusion that ontologies should rely at first on the expertise of specific authoritative palaeographers rather than on the community at large, and that they must be accompanied by datasets and computational tools that employ the ontology in question.

Exploring and Questioning, not Answering

It is increasingly being recognised in related fields of Digital Humanities that the "black and white" response often given by computational methods is incompatible with the approaches and interests of Humanities scholars. Furthermore, as already discussed here, it is very difficult to move from computational results to "real world" meaning, but for most Humanities scholars this "real world" meaning is the principal or only point of interest. It is therefore becoming increasingly evident that palaeographers prefer to harness computational methods not to provide answers to real-world questions, but rather to manage large amounts of data in ways that allow them to draw their own conclusions. Furthermore, it has been observed that cognitive processes in academic research can be enhanced through visualisation, particularly when applied to material which is inherently visual such as palaeography and manuscript studies [22, 50].

Some research questions of historical content have results which Humanities scholars can verify with a relatively high degree of confidence. One example is joins, that is, identifying pages or fragments of pages from now dismembered books. For problems like these, computational methods can usefully propose "real-world" answers, for example by providing a set of images of pages which are likely to be from the same book, and which the scholar can then check. In contrast, other problems are difficult

or impossible to verify against the historical "truth," and computational methods which attempt to answer these have not been accepted because of this difficulty of verification. Here, Humanities scholars need to be able to "cross-examine" the results, including also the method and the assumptions which underlie them [9, 43]: if they cannot verify them then they cannot have any confidence in the results. This is closely related to the "black box" problem discussed above and, as already noted, it is a significant challenge for future work.

However, an alternative approach is rather to develop computational methods that allow researchers to manipulate and visualise the content on their own terms, and to communicate this data as evidence to a broader audience. Scholars in Digital Humanities have referred to the "virtue of automated analysis... not [as] the ready delivery of objective truth, but instead the more profound virtue of bringing us up short, of disturbing us in our preconceptions" [44]. Therefore, once a critical mass of data has been reached, problems of this sort should be treated with computational methods designed to aid discovery, exchange, interpretation, and presentation of knowledge, not providing answers to historical or other "real world" questions ([22]; cf. also [8]). This has important implications for collaborative work with computer scientists, since it is fundamentally different from the "algorithms to solve problems" approach which is more typical of the latter's field (see "Interdisciplinary Approaches to Research", above).

The "80/20" Problem: Working with Imperfect Results

Given the inherently ambiguous nature of data in the Humanities, it is unrealistic to expect or even aim for perfect results in the medium term. However, "imperfect" or incomplete results can still be of significant value, as there is a bottleneck of turning digitised manuscripts into texts which can be processed by a machine. Rather than attempting to "solve" this problem in the short term, further consideration is encouraged as to what can be done with computational results that are accurate to (for example) 80%, 60% and so on. Current success rates for handwriting recognition are still extremely low (as low as 30%), however, and research that promises to increase that rate should be encouraged and funded. A success rate of 80% text recognition is still bad (every fifth word would contain an error), but if it is clear which 20% are inaccurate, the 80% of data becomes usable, and following the Pareto phenomena [37], achieving these 80% becomes cheaper than focussing on the expensive remaining 20%. Furthermore, as just discussed, scholars in the Humanities do not typically expect or even desire a final, "correct" answer, but rather want tools to help them process large quantities of material. In circumstances like this, simply reducing the size of a search-space by 80% may be a very significant improvement.

This demands several prerequisites: first, computer scientists must have verifiable ways of establishing confidence in their results matching the "ground truth": as

discussed, this is often challenging or even impossible, but in some specific cases is generally achievable (e.g. text recognition and word spotting). Second, Humanities scholars must learn to understand the implications of the inaccuracies: a given type of inaccuracy will not be significant for some research questions but will be highly significant for others. Close consideration must also be given to the role of false positives versus false negatives: for example, if a computer is being used to reduce a search space which a human researcher then examines, false positives are probably expected and tolerable, but false negatives are not. These considerations again require close communication between the disciplines. Third, investment should be made in identifying new research which can be enabled by computational methods which are largely but not entirely accurate. This may include manually correcting the inaccuracies (which could still save substantially on research time), or in designing new research which is not affected by the types of inaccuracies. Close parallels already exist in fields such as computational linguistics, distant reading, and "big data" research, and lessons learned there can also be applied here.

Outreach and Dissemination

Looking beyond the academic and research audience, very significant potentials exist for outreach and dissemination of work in cultural heritage. As noted above, handwritten manuscripts and documents form a very large part of the world's cultural heritage, with prominent examples including the Book of Kells and Lindisfarne Gospels, the Dead Sea Scrolls, through more recent examples such as Abraham Lincoln's handwritten copy of the Gettysburg Address, or Michel Proust's draft manuscript of À la recherche du temps perdu [39]. This rich cultural heritage has proven to be of great interest to a wide public, and can also help to empower minority or other disenfranchised groups and regions through informing them better of their history, heritage, language, and so on (one example of this is the Lindisfarne Gospels, which recently toured in exhibitions in North-East England). This "virtual repatriation" of cultural heritage represents a promising area of further development. More generally, however, both repositories and research institutions are frequently criticised for spending public money on material that is not accessible to those who provided the funds, and online resources can help to overcome this. Indeed, this increased access and "democratisation" is a frequent promise of Digital Humanities, although it has not necessarily been fulfilled in practice [41].

The introduction not only of digitised images but also of computerised techniques opens up new ways of sharing this information with the broad population. One particularly effective example of this is the Walters Art Museum, whose policy of releasing digital images of manuscripts using Creative Commons licensing, and of distributing these images through a range of social and other media, has lead directly to very wide public recognition of their holdings, so much so that a search for "koran"

in Google Images returns a highly disproportionate number of results from that museum – far more than from much larger and better-known institutions such as the British Library or the Bibliothèque nationale de France [35]. Even more exciting is the prospect of people conducting their own research, or tapping into non-expert traditions as a way of enriching scholarly knowledge. Although the process of opening up "virtual" manuscript archives to the public has already begun, these projects are still in their infancy. Reaching out, collecting, and processing the knowledge that may be available in regional traditions, on the other hand, has not been sufficiently explored. Doing so by using "crowd sourcing" techniques is an exciting new research direction and has already been applied to transcription and identification of manuscripts and musical scores, among others (e.g. [57, 60, 1]).

In order to realise this potential fully certain requirements remain. As the Digital Walters project clearly demonstrates, one requirement is again that of sufficiently permissive copyright and licensing conditions: if people are not allowed to use images in ways that they wish, or if it is unclear whether they may so use them or not, then they typically will not use them at all [35]. The material must also be free not only of licensing restrictions but also of technical ones: again, if the images are available only in proprietary viewers or other limiting formats then access to them diminishes accordingly. Furthermore, the difficulties in communication which have been discussed between palaeographers and computer scientists become even more pronounced when moving beyond the professional researcher to the wider public. However, the same principles advocated here, such as mid-level features and "in-betweener" specialists, are also relevant to this broader challenge. These principles need to be extended to other areas both of academic but also of public interest such as local history, genealogy, art history, language (including regional dialects), name studies, calligraphy, arts and crafts, and so on. As researchers are increasingly pressured to demonstrate the "impact" and value to society of their work, and as they discuss how best to measure and achieve it [55], digital palaeography is already addressing these concerns and also has an ideal scope of study which already has demonstrable public interest. Extending these concerns and combining the pre-existing interest presents an outstanding opportunity for taking this new and relatively marginalised field of study and bringing it to the forefront of public and academic awareness.

Acknowledgements

The authors wish to thank Schloss Dagstuhl for their support of this workshop. PAS also thanks the European Research Council (ERC): some of the results presented here are funded by the European Union Seventh Framework Programme (FP7) under

grant agreement no 263751. This text has received substantial contributions from all members of the Workshop, for which see Participants list below.

4. Participants

Name	Affiliation
Dimitris Arabadjis	National TU – Athens, GR
Nachum Dershowitz	Tel Aviv University, IL
Matthieu Exbrayat	Universite d'Orleans, FR
Shira Faigenbaum	Tel Aviv University, IL
Melanie Gau	Universität Wien, AT
Tal Hassner	Open University – Israel, IL
R. Manmatha	University of Massachusets – Amherst, US
Ophir Munz-Manor	The Open University of Israel – Raanan, IL
Eyal Ofek	Microsoft Res. – Redmond, US
Micalis Panagopoulos	Ionian University – Corfu, GR
Robert Sablatnig	Universität Wien, AT
Wendy Scase	University of Birmingham, GB
Timothy Stinson	North Carolina State Univ., US
Peter A. Stokes	King's College London, GB
Dominque Stutzmann	Ecole Praqtique des Hautes Etudes – Paris, FR
Segolene Tarte	University of Oxford, GB
Lior Wolf	Tel Aviv University, IL

Bibliography

1. Ancient Lives. <https://www.zooniverse.org/project/ancientlives>.
2. J. F. A. Aussems. *Christine de Pizan and the Scribal Fingerprint: A Quantitative Approach to Manuscript Studies*. PhD thesis, University of Utrecht, Utrecht, 2006. <http://igitur-archive.library.uu.nl/student-theses/2006-0908-200407/UUindex.html>.
3. Mary Beard. A Don's Life: University cuts, redundancies – and bye-bye palaeography at King's College London. *Times Literary Supplement*, January 28, 2010. <http://timesonline.typepad.com/dons_life/2010/01/university-cuts-redundancies-and-byebye-palaeography.html>.
4. Leonard E. Boyle. *Medieval Latin Palaeography: A Bibliographical Introduction*. University of Toronto Press, Toronto, 1984.
5. S.J. Brookes. "Digital Resources for Palaeography" One-Day Symposium. *DigiPal Project Blog*. King's College, London, 2011. <http://www.digipal.eu/blogs/news/%E2%80%98digital-resources-for-palaeography%E2%80%99-one-day-symposium/>.

6. Michelle P. Brown. *A Guide to Western Historical Scripts from Antiquity to 1600*. British Library, London, 1990.
7. Arianna Ciula. Digital palaeography: Using the digital representation of medieval script to support palaeographic analysis. *Digital Medievalist*, 1(1), 2005. <http://www.digitalmedievalist.org/journal/1.1/ciula/>.
8. Tanya Clement, Sara Steger, John Unsworth, and Kirsten Uszkalo. How not to read a million books. *Seminar on the History of the Book*, Rutgers University, New Brunswick NJ, 5 March, 2009. <http://people.brandeis.edu/~unsworth/hownot2read.html>.
9. Tom Davis. The practice of handwriting identification. *The Library (7th series)*, 8(3):251–276, 2007. doi:10.1093/library/8.3.251.
10. Albert Derolez. *The Palaeography of Gothic Manuscript Books from the Twelfth to the Early Sixteenth Century*. Cambridge Studies in Palaeography and Codicology, 9. Cambridge University Press, Cambridge, 2003.
11. Digital Walters. Baltimore: The Walters Art Museum. <http://www.thedigitalwalters.org/>.
12. David N. Dumville. Specimina codicum palaeoanglicorum. In *Collection of Essays in Commemoration of the 50th Anniversary of the Institute of Oriental and Occidental Studies*, pp. 1–24. Kansai University Press, Suita, Osaka, 2001.
13. European Science Foundation. *Humanities Exploratory Workshops: Digital Palaeography*. <http://www.esf.org/activities/exploratory-workshops/humanities-sch/workshops-detail.html?ew=10865>.
14. Europeana library of digital objects. *Europeana Data Model*. <http://pro.europeana.eu/edm-documentation/>.
15. Franz Fischer, Christiane Fritze, and Georg Vogeler, editors. *Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2*. Schriften des Instituts für Dokumentologie und Editorik, 3. Books on Demand, Norderstedt, 2011. <http://kups.ub.uni-koeln.de/4337/>.
16. David Ganz. "Editorial Palaeography": One teacher's suggestions. *Gazette du Livre Médiéval*, 16:17–20, 1990. <http://www.palaeographia.org/glm/glm.htm?art=ganz>.
17. German Research Foundation. *DFG-Praxisregeln. "Digitalisierung"*. 2013. <http://www.dfg.de/formulare/12_151/index.jsp>.
18. Tal Hassner, Malte Rehbein, Peter A. Stokes and Lior Wolf, editors. Computation and palaeography: Potentials and limits (Dagstuhl PerspectivesWorkshop 12382). *Dagstuhl Reports* 2:9 (2012): 184–199. doi:10.4230/DagRep.2.9.184.
19. HPO. *Human-Phenotype-Ontology*. <http://www.human-phenotype-ontology.org/>.
20. International Council of Museums – International Committee for Documentation. *The CIDOC Conceptual Reference Model*. <http://www.cidoc-crm.org/>.
21. International Standards Organization. *Information and Documentation – A Reference Ontology for the Interchange of Cultural Heritage Information*, 2006. <http://www.iso.org/iso/catalogue_detail?csnumber=34424>.
22. Martyn Jessop. Digital visualization. *Literary and Linguistic Computing*, 23(3):281–293, 2008. doi:10.1093/llc/fqn016.

23. JISC. *Basic Guidelines for Image Capture and Optimisation.* <http://www.jiscdigitalmedia.ac.uk/stillimages/advice/basic-guidelines-for-image-captureand-optimisation>.
24. JISC. *Usability Foundation Study and Investigation of Usability in JISC Services*, 2004. <http://www.jisc.ac.uk/whatwedo/programmes/presentation/usability.aspx>.
25. JISC. *Visualisation Foundation Study*, 2004. <http://www.jisc.ac.uk/whatwedo/programmes/presentation/visualization.aspx>.
26. JISC. *Development of Personalisation for the Information Environment 1 (DPIE1)*, 2008. <http://www.jisc.ac.uk/whatwedo/programmes/resourcediscovery/dpie1.aspx>.
27. Neil R. Ker. *Catalogue of Manuscripts containing Anglo-Saxon.* Clarendon Press, Oxford, 1957.
28. Robert Kummer. Semantic technologies for manuscript descriptions-concepts and visions. In Fischer et al. [15], pp. 133–154, 2011. <http://kups.ub.uni-koeln.de/4347/>.
29. N. Levy, Lior Wolf, Nachum Dershowitz, and Peter A. Stokes. Estimating the distinctiveness of graphemes and allographs in palaeographic classification. In *DH2012 Book of Abstracts*, ed. by Jan Christoph Meister et al., pp. 264–267. University of Hamburg, 2012. <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/estimating-the-distinctiveness-of-graphemes-and-allographs-in-palaeographic-classification/>.
30. London: King's College. *DigiPal: Digital Database and Resource of Palaeography, Manuscripts and Diplomatic.* <http://www.digipal.eu/>.
31. London: University of the Arts. *Ligatus: An English/Greek Terminology for the Structures and Materials of Byzantine and Greek Bookbinding.* <http://www.ligatus.org.uk/glossary/>.
32. *Manuscripts Online. Written Culture 1000 to 1500.* <http://www.manuscriptsonline.org/>.
33. Max Planck Institute. *Best Practices for Access to Images: Recommendations for Scholarly Use and Publishing.* Berlin, 2009. <http://www.mpiwg-berlin.mpg.de/PDF/MPIWGBestPracticesRecommendations.pdf>.
34. MESA. *Medieval Electronic Scholarly Alliance.* <http://mesa.performantsoftware.com/>.
35. William Noel. *The Commons and Digital Humanities in Museums.* CUNY, New York, 2012. <http://www.youtube.com/watch?v=XPJ_kciC15I>.
36. OMIM. *Online Mendelian Inheritance in Man.* <http://www.ncbi.nlm.nih.gov/omim/>.
37. Pareto. Pareto Principle. *Wikipedia.* <http://en.wikipedia.org/wiki/Pareto_principle>.
38. Susannah B.F. Paletz, Christian D. Schunn, and Kevin H. Kim. The interplay of conflict and analogy in multidisciplinary teams. *Cognition*, 126(1):1–19, 2013. doi:10.1016/j.cognition.2012.07.020.
39. Elena Pierazzo and J. André. *Autour d'une séquence et des notes du Cahier 46: enjeu du codage dans les brouillons de Proust – Around a Sequence and some Notes of Notebook 46: Encoding Issues about Proust's Drafts.* King's College, London, 2012. <http://research.cch.kcl.ac.uk/proust_prototype/about.html>.
40. Malte Rehbein, Patrick Sahle, and Torsten Schaßan, editors. *Kodikologie und Paläographie im Digitalen Zeitalter - Codicology and Palaeography in the Digital Age.* Schriften des Instituts für Dokumentologie und Editorik. Books on Demand, Norderstedt, 2009. <http://kups.ub.uni-koeln.de/volltexte/2009/2939/>.

41. M. Reisz. Surfdom. *Times Higher Education Supplement*, 2028:34–39, 8 December 2011. <http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=418343&c=2>.
42. *Rights Reserved. – Free Access. Nutzungsrechte für historische Quellen und Handschriften im Internet.* <http://www.infoclio.ch/de/node/127076>.
43. Lambert Schomaker. Advances in writer identification and verification. *Proc. of 9th Int. Conf. on Document Analysis and Recognition (ICDAR 2007), 26 September, Curitiba, Brazil*, 2:769–773, 2007.
44. D. Sculley and Bradley M. Pasanek. Meaning and mining: The impact of implicit assumptions in data mining for the humanities. *Literary and Linguistic Computing*, 23(4):409–424, 2008. doi:10.1093/llc/fqn019.
45. Yaakov Choueka, Lior Wolf, Roni Shweka, and Nachum Dershowitz. Automatic extraction of catalog data from digital images of historical manuscripts. Forthcoming.
46. Charles Percy Snow. *The Two Cultures.* Cambridge University Press, 2012.
47. F. M. Stenton. The supremacy of the Mercian kings. *English Historical Review*, 33(132):433–452, 1918. doi:10.1093/ehr/XXXIII.CXXXII.433.
48. Peter A. Stokes. Palaeography and image processing: Some solutions and problems. *Digital Medievalist*, 3, 2007/8. <http://www.digitalmedievalist.org/journal/3/stokes/>.
49. Peter A. Stokes. Computer-aided palaeography: Present and future. In Rehbein et al. [40], pp. 309–338, 2009. <http://kups.ub.uni-koeln.de/volltexte/2009/2978/pdf/KPDZ_I_Stokes.pdf>.
50. Peter A. Stokes. Palaeography and the "virtual library" of manuscripts. In B. Nelson and M. Terras, *Digitising Medieval and Early Modern Culture*, pp. 137–169. Medieval and Renaissance Texts and Studies, 426. Arizona Center for Medieval and Renaissance Studies, Tempe AZ, 2012.
51. Peter A. Stokes. Modeling medieval handwriting: A new approach to digital palaeography. In *DH2012 Book of Abstracts*, ed. by Jan Christoph Meister et al., pp. 382–385. University of Hamburg, 2012. <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/modeling-medieval-handwriting-a-new-approach-to-digital-palaeography.1.html>.
52. Peter A. Stokes. *English Vernacular Script from Æthelred to Cnut, ca 990 – ca 1035.* Boydell, Woodbridge, Forthcoming.
53. Dominique Stutzman. *AAA - AΔΛ* - Alphabet, Ambiguité et Actualité (paléographique): l'ontologie des formes alphabetiques. *Paléographie Médiévale*, 2012. <http://ephepaleographie.wordpress.com/2012/01/22/aaa-%CE%B1%CE%B4%CE%BB-alphabet-ambiguite-actualites-paleographique-ontologie-formes-alphabetiques/>.
54. Tagung. *Codicology and Palaeography in the Digital Age.* Conference report, July 2009. <http://www.i-d-e.de/aktivitaeten/veranstaltungen/2009-tagung-kpdz/report/>.
55. Simon Tanner. *Measuring the Impact of Digital Resources: The Balanced Value Impact Model.* King's College, London, 2012. <http://www.kdcs.kcl.ac.uk/innovation/impact.html>.
56. Melissa Terras. *Digital Images for the Information Professional.* Ashgate, Aldershot, England, 2008.
57. *Transcribe Bentham.* <http://www.transcribe-bentham.da.ulcc.ac.uk/>.

58. Georg Vogeler. Kodikologie und Paläographie im digitalen Zeitalter / Codicology and palaeography in the digital age. *AHF-Information*, (206), 2009. vgl. <http://www.hsozkult.de/conferencereport/id/tagungsberichte-3025>.
59. C. Von Oertzen and K. Wilder. *Scholarly Publishing and the Issues of Cultural Heritage, Fair Use, Reproduction Fees and Copyrights*. Berlin, Max Planck Institute, 2008. <http://www.mpiwg-berlin.mpg.de/PDF/MPIWGWorkshop1-2008Report.pdf>.
60. *What's the Score*. <https://www.zooniverse.org/lab/wts/>.
61. Lior Wolf, Tal Hassner, and Yaniv Taigman. Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1978–1990, 2011.
62. European Science Foundation. *Exploratory Workshop on Digital Palaeography*. Conference report, 2011. <http://hsozkult.geschichte.hu-berlin.de/tagungsberichte/id=3968>.

Digitale Reproduktion als paläographisches Werkzeug

Digital imaging as a palaeographic tool

Readability Enhancement and Palimpsest Decipherment of Historical Manuscripts

Fabian Hollaus, Melanie Gau, Robert Sablatnig
William A. Christens-Barry, Heinz Miklas

Abstract

This paper presents image acquisition and readability enhancement techniques for historical manuscripts developed in the interdisciplinary project "The Enigma of the Sinaitic Glagolitic Tradition" (Sinai II Project).[1] We are mainly dealing with parchment documents originating from the 10th to the 12th centuries from St. Catherine's Monastery on Mount Sinai. Their contents are being analyzed, fully or partly transcribed and edited in the course of the project. For comparison also other mss. are taken into consideration. The main challenge derives from the fact that some of the manuscripts are in a bad condition due to various damages, e.g. mold, washed out or faded text, etc. or contain palimpsest (=overwritten) parts. Therefore, the manuscripts investigated are imaged with a portable multispectral imaging system. This non-invasive conservation technique has proven extremely useful for the examination and reconstruction of vanished text areas and erased or washed off palimpsest texts. Compared to regular white light, the illumination with specific wavelengths highlights particular details of the documents, i.e. the writing and writing material, ruling, and underwritten text. In order to further enhance the contrast of the degraded writings, several Blind Source Separation techniques are applied onto the multispectral images, including Principal Component Analysis (PCA), Independent Component Analysis (ICA) and others. Furthermore, this paper reports on other latest developments in the Sinai II Project, i.e. Document Image Dewarping, Automatic Layout Analysis, the recent result of another project related to our work: the image processing tool *Paleo Toolbar*, and the launch of the series *Glagolitica Sinaitica*.

Zusammenfassung

In dieser Arbeit werden Bildaufnahmemethoden sowie Methoden zur Verbesserung der Lesbarkeit für historische Manuskripte vorgestellt, die im Rahmen des interdisziplinären Projekts "The Enigma of the Sinaitic Glagolitic Tradition" (Sinai II Projekt)

[1] <http://www.caa.tuwien.ac.at/cvl/research/sinai/>

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 31–46.

angewandt und entwickelt werden. In dem Projekt werden hauptsächlich Pergamentdokumente analysiert, die zwischen dem zehnten und zwölften Jahrhundert im Katherinenkloster am Berg Sinai entstanden. Im Rahmen des Projektes werden die betreffenden Texte analysiert, vollständig oder teilweise transkribiert sowie ediert. Die größte Herausforderung stellt hierbei der schlechte Zustand der Handschriften dar, da diese teilweise von Schimmel befallen sind, verblichenen oder entfernten Text enthalten oder palimpsestiert sind. Deshalb werden die betreffenden Handschriften mithilfe eines portablen multispektralen Aufnahmesystems digitalisiert. Diese nicht invasive Konservierungsmethode hat sich als besonders nützlich für die Untersuchung und Rekonstruktion von solchen schlecht lesbaren Handschriften erwiesen. Verglichen mit normalem Weißlicht, bietet eine Untersuchung in ausgewählten spektralen Bereichen die Möglichkeit gewisse Details eines Dokuments, wie verblasste Schrift oder Palimpsesttext, besser sichtbar zu machen. Um den Kontrast zwischen den verblichenen Schriften und dem restlichen Dokument weiter zu verstärken, werden sogenannte Blind Source Separation Techniken - wie Principal Component Analysis (PCA) oder Independent Component Analysis (ICA) - angewandt. Des Weiteren werden in dieser Arbeit die jüngsten Entwicklungen im Sinai II Projekt beschrieben: Diese beinhalten Techniken für Document Image Dewarping und Automatic Layout Analysis, die Herausgabe der Edition *Glagolitica Sinaitica* sowie das Resultat aus einem anderen Projekt, das sich mit einer ähnlichen Zielsetzung beschäftigt: Das Bildverarbeitungsprogramm *Paleo Toolbar*.

1. Introduction

Apart from the description and classification, the main task of the work with historical manuscripts consists of the decipherment and edition of their contents. This may be difficult due to paleographic or linguistic peculiarities, but also because of the bad preservation state of the relevant source. Thus, it is essential to develop and evaluate methods of image and readability enhancement for this kind of documents.

In our interdisciplinary projects "Critical Edition of the New Sinaitic Glagolitic Euchology (Sacramentary) Fragments with the Aid of Modern Technologies" (P19608, 2007-2010) and "The Enigma of the Sinaitic Glagolitic Tradition" (P23133; 2011-2014), both funded by the Austrian Science Fund (FWF), we have been investigating various approaches of manuscript analyses with scientific and computational methods, among them Image Binarization (cf. Lettner 2009), Character Classification (cf. Vill 2008), Optical Character Recognition (OCR) (cf. Diem 2011), Layout Analysis (cf. Kleber 2009, Garz 2011), and Fragment Puzzling (Kleber 2008). In the scope of both Sinai projects also chemical analyses have been executed on the ancient documents by material chemists from the Vienna Academy of Fine Arts, namely X-Ray Fluorescence

Analyses (XRF) and, in the latter project also InfraRed Spectroscopy (FTIR). The results are being published in the new series *Glagolitica Sinaitica* (GlagSin) by Holzhausen publishing house and will comprise not only facsimile and critical editions, but also a separate technical volume on analysis methods and computational results.

The manuscripts investigated in this study were all written on parchment and originate from the 10^{th} to 14^{th} centuries. They are written in Glagolitic, the oldest Slavic script (cf. Miklas). The documents are in a poor condition and its texts partly barely or not at all discernible, due to partially faded or washed out ink, stains, mold, and other corruption by background clutter. The readability of multiple texts in palimpsests poses additional difficulties. One object in particular, the so called *Missale Sinaiticum* (Cod. Sin. Slav. 5/N), has long been almost impossible to read and transcribe, due to the mysterious discoloration phenomenon of brown or black ink that has turned to white (cf. Miklas). The second object of investigation is one of the Budapest Glagolitic Fragments (Duod. Slav. 2, Nr. 5), which contains characters, that are faded out but still darker than the background.

Both objects were digitized with MultiSpectral Imaging (MSI) technology that has proven a valuable basis for the investigation of historical manuscripts using computational methods. This approach is valued because it is non-invasive and more capable of enhancing the contrast of the degraded writings than methods that rely on normal white light illumination (Easton; Rapantzikos). Using MSI, the transcription of the contents is considerably easier and more comprehensive, as the results of the Archimedes Palimpsest project also have shown (cf. Easton). Our custom built and portable MSI system is described in detail in Section 2 (cf. Lettner; Hollaus).

Successful approaches to contrast enhancement of badly legible historical documents are dimension reduction techniques like Principal Component Analysis (PCA) and Independent Component Analysis (ICA) as demonstrated in Easton, Salerno, and Lettner.

Easton et al. use the PCA approach for the separation and enhancement of the diverse textual layers in the Archimedes palimpsest. They suggest merging those of a set of PCA images that typically emphasize the writing into one pseudocolor image. Salerno et al. when investigating the Archimedes Palimpsest applied the PCA approach and also analyzed the ICA method, which used the FastICA methodology (cf. Hyvärinen). Rapantzikos et al. compared the performance of the PCA technique to the performance of the Linear Spectral Mixture Analysis for the separation of the text layers contained in a palimpsest-ms. and state that the PCA approach is better suited for this task.

Lettner et al. proposed a technique for the enhancement of one or more text layers using not only spectral but also spatial information. In this procedure, a priori knowledge provided by philological experts is incorporated into a text line model used for the generation of a binary mask to encode the text regions. The enhancement

itself is performed by applying the Multivariate Spatial Correlation technique (cf. Wartenberg). The authors were able to show that their technique is superior to the PCA method.

Similar to Lettner we implement spatial correlation, since our approach is also based on text line detection. But in contrast to Lettner we do not require a priori knowledge, as the text lines are automatically found in the investigated regions. To the best of our knowledge this is the first time that the Fisher Linear Discriminate Analysis (LDA) approach is used for the enhancement of historical texts. So far it had been successfully used for face recognition tasks: Belhumeur et al. show that their LDA based dimensionality reduction outperforms the Eigenface method, which utilizes the PCA transformation.

While PCA and ICA are unsupervised dimension reduction tools, we apply the LDA approach, which is a supervised dimension reduction method and hence requires a labelled subset of the multispectral samples. In order to find a training set for the LDA classifier, we apply a semi-automated procedure, in which a subset of multispectral observations is labelled as belonging to text or background regions. The labelling procedure was especially designed for the *Missale Sinaiticum*, where a correct labelling of the fore- and background regions is difficult since in parts characters are barely visible within the multispectral images or best visible in different spectral ranges. Therefore, the labelling is performed on PCA images with their enhanced contrast. Due to the bad condition of the *Missale Sinaiticum* the PCA images are also corrupted by noise; hence labelling by application of a binarization algorithm would have been too error-prone. Instead, the labelling is performed in an iterative manner and is based on the detection of text lines, since the detection of text lines is more robust against noise — e.g. background clutter or lighting variations — compared to binarization methods.

In the last chapter, several further document analysis and image enhancement results will be presented, namely Document Image Dewarping, Automatic Layout Analysis, and two different toolboxes for paleographic document analysis.

The next section introduces a number of image enhancement techniques based on MultiSpectral Imaging (MSI).

2. Image Enhancement Techniques based on MultiSpectral Imaging

2.1. MSI Approach and System Setup

The MSI approach has proven its applicability for the investigation of ancient and problematic sources, since it is a non-invasive analysis tool capable of increasing the legibility of indecipherable texts. Our portable MSI system employs two sets of

LED panels, which provide 11 different narrowband spectral ranges. Compared to an MSI setup that applies optical filters, the LED illumination has two major advantages: First, the heat put on the manuscripts is reduced since broadband illumination is not required. Second, geometrical distortions stemming from optical filters are avoided and an image registration step is not necessary since the optical characteristics are not changed during acquisition. The images are acquired by two cameras, namely a Hamamatsu greyscale camera and a Nikon D4 SLR camera (cf. Figure 1). As a result the multispectral images already provide improved contrast and readability of the ancient writings in several spectral ranges — compared to ordinary white light illumination.

Nevertheless, due to their bad condition, some manuscript portions still remain unreadable. So we have implemented three different enhancement techniques, all of which are based on dimension reduction and applied statistics of the MSI results.

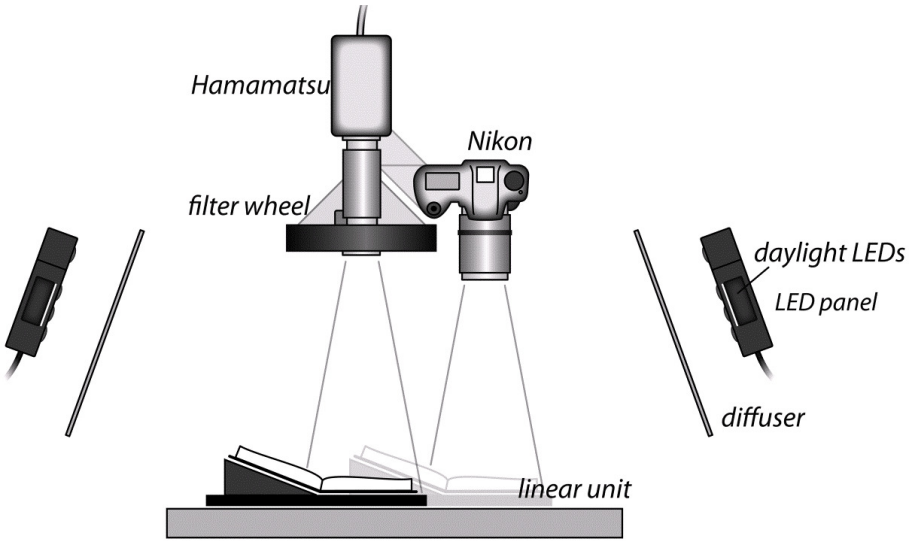

Figure 1. Portable MSI system setup of the Sinai Project.

2.2. Comparison of Image Enhancement Techniques

This section contains a performance comparison between three different dimension reduction techniques: PCA, ICA, and LDA.

These techniques are used to lower the third dimension of a multispectral scan in order to extract the relevant information. For folios containing only one single

textual layer, the MSI scans are reduced to just one image to emphasize the content in question, in our case the text. For palimpsests, on the other hand, the third dimension of the MS scan is reduced to two images, one emphasizing the underwritten, the other the overwritten text. Dimension reduction techniques can be grouped into unsupervised (PCA and ICA) and supervised (LDA) dimension reduction approaches.

Unsupervised Dimension Reduction Techniques

This technique category makes no use of class information. This category of enhancement techniques finds a transformation of the multispectral scan that removes any contained correlation. Such correlation stems mainly from material spectral correlation.

PCA

The PCA transformation finds an orthogonal transformation, on which the data is projected, and removes redundancies contained in the MS scan. The transformation of the multispectral data x is formally defined by:

$$y = Wx$$

where y is the transformed signal, and W is the transformation matrix. The columns of the transformation matrix are filled with the eigenvectors of the covariance matrix of the zero mean normalized data. The eigenvectors are sorted in descending order according to their corresponding eigenvalues.

In this step, the first k PCA images are considered, since one writing is usually emphasized by multiple principal components (cf. Lettner; Easton). For the *Missale Sinaiticum* it was experimentally found that five PCA images have to be considered. For the Budapest Glagolitic Fragment k was set to three, because the spectral signature of the writing in this manuscript is less varying than in the *Missale Sinaiticum*, and the writing is described by the first three eigenvectors. We noticed that in the PCA images, the ruling scheme is better recognizable than the degraded characters. Therefore, in the coarse enhancement stage the samples are labelled as belonging to text lines or belonging to intermediate regions instead of belonging to text or background regions.

We noticed that the PCA output is generally ordered, meaning the first PCA image shows the text layer - if the writing is in relative good, i.e. rich-in-contrast condition. If the document contains a great amount of background clutter or faded-out characters, however, the first image may show the enhanced background clutter instead, whereas the image showing the writing is emphasized by another principal component.

ICA

In contrast to the PCA transformation, the ICA approach finds a projection matrix W that is not necessarily orthogonal. PCA follows the assumption that sources within the data (e.g. the text(s)) are uncorrelated, whereas ICA assumes that the sources are

statistically independent. Those sources that carry any specific content (e.g. text or palimpsest text) are detected by maximizing their independence. Similar to Salerno we applied this method in order to separate the two layers of a palimpsest. We applied the FastICA algorithm (for more details on ICA cf. also Hyvärinen; Salerno; Hollaus).

Supervised Dimension Reduction Techniques

LDA

PCA and ICA as unsupervised dimension reduction techniques find the relevant information within the multispectral scans by themselves. LDA as a supervised dimension reduction algorithm requires a prepared subset of the samples with labelled fore- and background pixels as a training set for the classifier. Such labelling could be achieved by applying a simple document binarization technique (cf. Su), but since historical documents can be corrupted with background clutter, simply applying a binarization algorithm is error-prone. That is why we suggest a more careful procedure instead: In a first step, the areas in an input image with the most visible content, e.g. text lines, are determined. Text line regions are found with a text line detection algorithm similar to the one in Yosef et.al. Depending on the document, both UV fluorescence images and the output of the PCA transformation can provide suitable input images. The pixels of this input image are labelled as belonging to text line or intermediate, i.e. background, regions. Then the LDA-based dimension reduction is applied in order to generate an enhanced image by bringing out the degraded text and enhancing the contrast.

The PCA transformation maximizes the so-called scatter of the transformed feature vector y. In contrast, the LDA method as a supervised dimension reduction tool seeks a projection W that maximizes the ratio of the between-class scatter to the within-class scatter, i.e. it exploits the class information in order to select discriminant features, e.g. for text and background (for more details on LDA cf. also Duda).

If the manuscript is a palimpsest, a three-class problem is considered, where overwriting, underwriting, and background classes are considered. Since the overwriting of the manuscript investigated is typically most visibly separable from other content under Near Infrared (NIR) illumination, whereas the underwriting is not, the labelling of the overwriting is performed on NIR images by applying a binarization algorithm (cf. Su). Afterwards, the underwriting is found by applying the procedure explained above.

2.3. Results

For image enhancement of degraded documents, dimension reduction techniques have yielded two advantages: On the one hand, an investigation of the entire scan

is avoided, while on the other hand, the contrast and visibility of the degraded characters is enhanced and the resulting images show a considerable increase in legibility compared to the images contained in the multispectral scan.

In this study we compared three different dimension reduction techniques applied on the entire multispectral scan on two different manuscripts: PCA, ICA, and LDA. The results show that all three approaches are capable of considerably enhancing the contrast compared to the unprocessed multispectral images. In the best case the output is one single image showing the relevant image content, e.g. the text.

We noticed that the PCA output is generally ordered, meaning the first PCA image shows the writing — if the writing is in relative good, i.e. rich-in-contrast condition. However, if the document contains a great amount of background clutter or faded-out characters it is not guaranteed that the first image shows the degraded writing, since the background clutter may be enhanced in the first image, whereas the image showing the writing is emphasized by another principal component.

The ICA approach does not order the resulting images; hence the ICA image showing the respective content most clearly must be selected manually.

In contrast, the LDA approach considers a two class problem and thus its output is a single image which enhances the writing, given a previously applied line detection algorithm separated the classes correctly. A qualitative analysis showed that the LDA approach achieves better performance in the case of background clutter and faded-out characters.

The results of the image enhancement techniques were evaluated by using a qualitative legibility assessment, conducted by a philologist experienced in reading the Glagolitic script. In order to avoid bias the scholar did not evaluate entire image patches, but instead single characters. We extracted 212 single characters from 7 test panels, each containing several hard to read characters. The test panels were normalized between the minimum and maximum intensity value. A grouping into categories of visual quality was not used, due to high variety of the enhancement results. Instead, the scholar was asked to pairwise compare two single characters, which were produced by different techniques. The philologist was not told which method was used for the enhancement of a particular character. For each compared letter pair, the scholar decided which letter was found more legible and assigned a 1 to the superior result image and a 0 to the inferior counterpart. Afterwards, the sum of the assigned scores was calculated.

The performance of the LDA based technique improves if it is solely applied to a region in which the characters have a similar contrast. Hence, we applied the algorithm exclusively to the test panels. For the PCA and ICA approaches it cannot be assumed that a transformation based on local statistics gains a better enhancement result than a transformation which has been calculated on the basis of the entire folio. Therefore, the PCA and ICA techniques have been applied in a local and global manner.

Since the text layer is emphasized by several principal components, we manually selected the best PCA outcome. The output of the (Fast)ICA algorithm depends on a random initialization and on several parameters we applied on the MSI data and selected the best results for the evaluation task. Additionally, the multispectral image with the best visible text was added to the test set. Two examples of Missale Sinaiticum can be seen in Figure 2 and Figure 3.

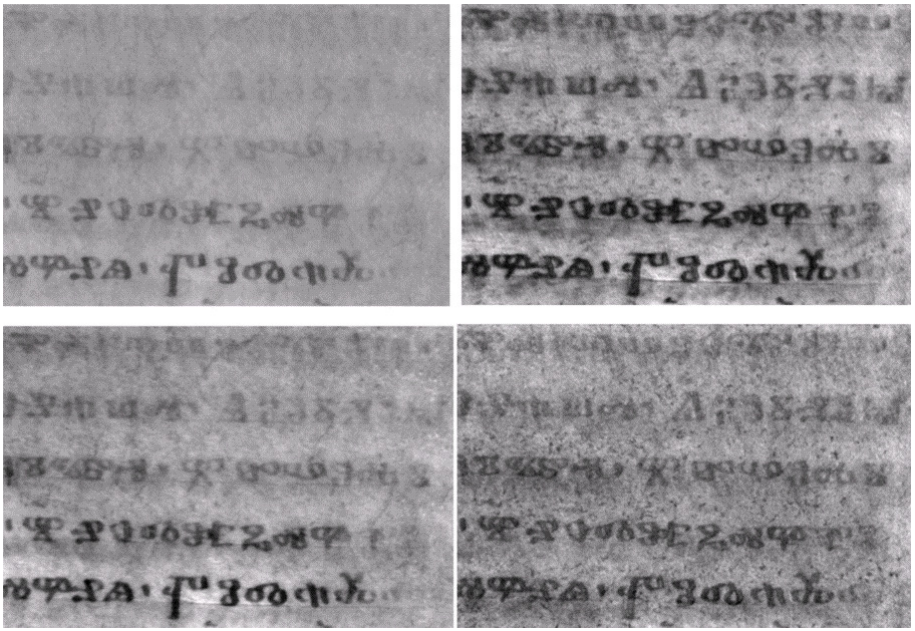

Figure 2. Missale Sinaiticum: top left: white light image. Top right: LDA, bottom left: PCA, bottom right: ICA output.

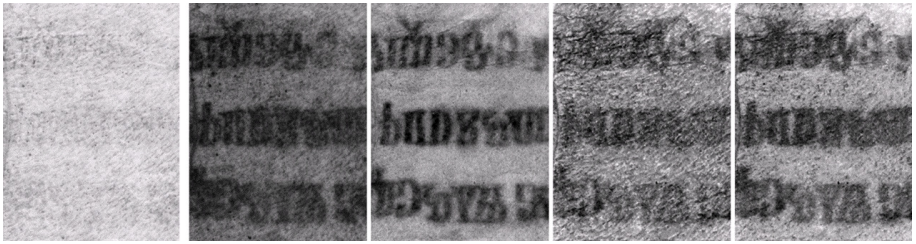

Figure 3. Budapest Glagolitic Fragment: from left to right: white light image. UV fluorescence image. LDA, PCA, ICA results.

In both cases the LDA output is superior to the other enhancement techniques, since it is capable of restoring the degraded text parts. In Figure 3 the PCA performs poorly, with lower readability than in the UV fluorescence image.

3. The Sinai Project II – Related Results

Since our last report on the Sinai Project several new developments have been made. They will be described shortly in the following.

3.1. Document Image Dewarping

The other readability enhancement method presented in this work, namely document image dewarping, is concerned with the rectification and smoothing of images of documents that are deformed due to age and bad storage conditions. The proposed algorithm has been adapted to artificially warped documents showing a conventional Latin computer font, e.g. *Times New Roman*, as well as images of Glagolitic script in two variants, a) artificially warped documents written in a standardized computer font, e.g. *Glagolica Bulgarian*, for experimental purposes and the binarized images of the ancient Sinai manuscripts (cf. Figure 4). The warping of all three types of documents could be straitened successfully and the readability thus improved.

The evaluation of the dewarping of the Glagolitic documents showed a particular improvement by reducing the Straightness Standard Error. An evaluation of the results on the artificially warped Latin documents by means of applying standard OCR on the original as well as on the dewarped text shows that the straightened text had a higher recognition rate by 51% on average (cf. Steinböck).

To further enhance the dewarping results on the Glagolitic document images, we plan to also take into account alignment and orientation of the characters, as the Glagolitic script can be not only a standing script, but also a hanging one.

3.2. Document Layout Analysis

The layout entities considered in this approach include body text, embellished initials, plain initials and headings. These textual elements are disassembled into segments, and a part-based detection is done which employs local gradient features known from the field of object recognition, the Scale Invariant Feature Transform (Sift).

These features describe the structures in a scale-, rotation- and illumination invariant manner. Hence, this approach does not rely on a binarization step, but is directly applied to the gray scale image, and it is robust to variations in shape, illumination, writing orientation, and (background) noise, too. Thus, it is suitable for ancient handwritten documents with varying layouts and degradation effects.

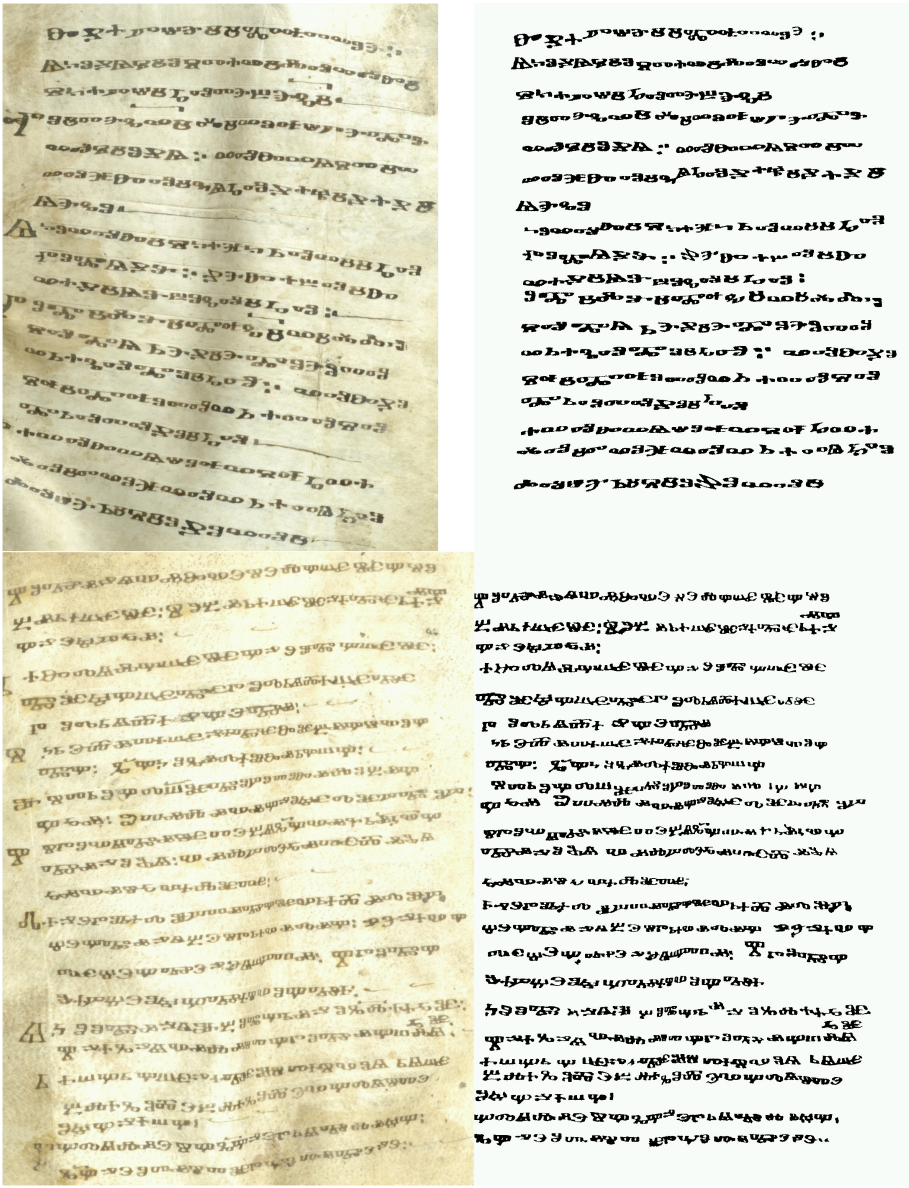

Figure 4. Results of the dewarping algorithm on 2 folia of the Psalter of Demetrius (Cod. Sin. Slav. 3/N): original images (left), rectified images (right).

As the whole entity cannot directly be inferred from the mere positions of the interest points, a localization algorithm is needed that expands the interest points according to their scales and the classification score to regions that encapsulate the whole entity. Therefore, a cascading algorithm is proposed that successively rejects weak candidates applying voting schemes (cf. Garz).

The evaluation shows that the method is able to locate main body text in ancient manuscripts. The detection rate of decorative entities is not as high as for main body text but already yields to promising results (cf. Figure 5).

3.3. *nomacs image lounge* and Toolbox for Manuscript Analysis

The image viewer *nomacs image lounge* (<http://www.nomacs.org>) developed by image processing experts at the Computer Vision Lab at Vienna UT already features several useful functions for multispectral image analysis, e.g. image panning (also between connected computers), and manuscript analysis, e.g. a false color tool for enhancing the visual contrast between e.g. text and palimpsest. The latest development is a toolbox for manuscript analysis. It was developed in collaboration with manuscript researchers in order to make certain processing and reading steps better and easier. The toolbox includes an automatic line detection tool based on the text line detection technique similar to Yosef et.al. The calculated lines, depending on the script either below or above the text, can be visualized directly on the shown manuscript image. It can also measure the factual distance between two points of a loaded page in relation to the image resolution. Another tool makes cutting out characters for the creation of character comparison tables very fast by simply clicking into the according letter and automatically extracting it in a bounding box. The toolbox will be included in the next release of *nomacs*.

3.4. External Project: Paleo Toolbox

Recently, one of the authors (Christens-Barry) has developed an imaging and processing software tool, *Paleo Toolbox*, to aid researchers having varying levels of computer skill. *Paleo Toolbox* implements workflows developed for spectral imaging projects, including the Archimedes Palimpsest Project and the St. Catherine's Palimpsests Project being conducted by EMEL[2], with particular focus on the identification and increased legibility of undertext. It uses a simple tool palette interface that builds upon the plugin architecture of ImageJ, a widely used open source tool for all major platforms[3].

---

[2] <http://rsbweb.nih.gov/ij/>.

[3] Available at <http://rsb.info.nih.gov/ij/>.

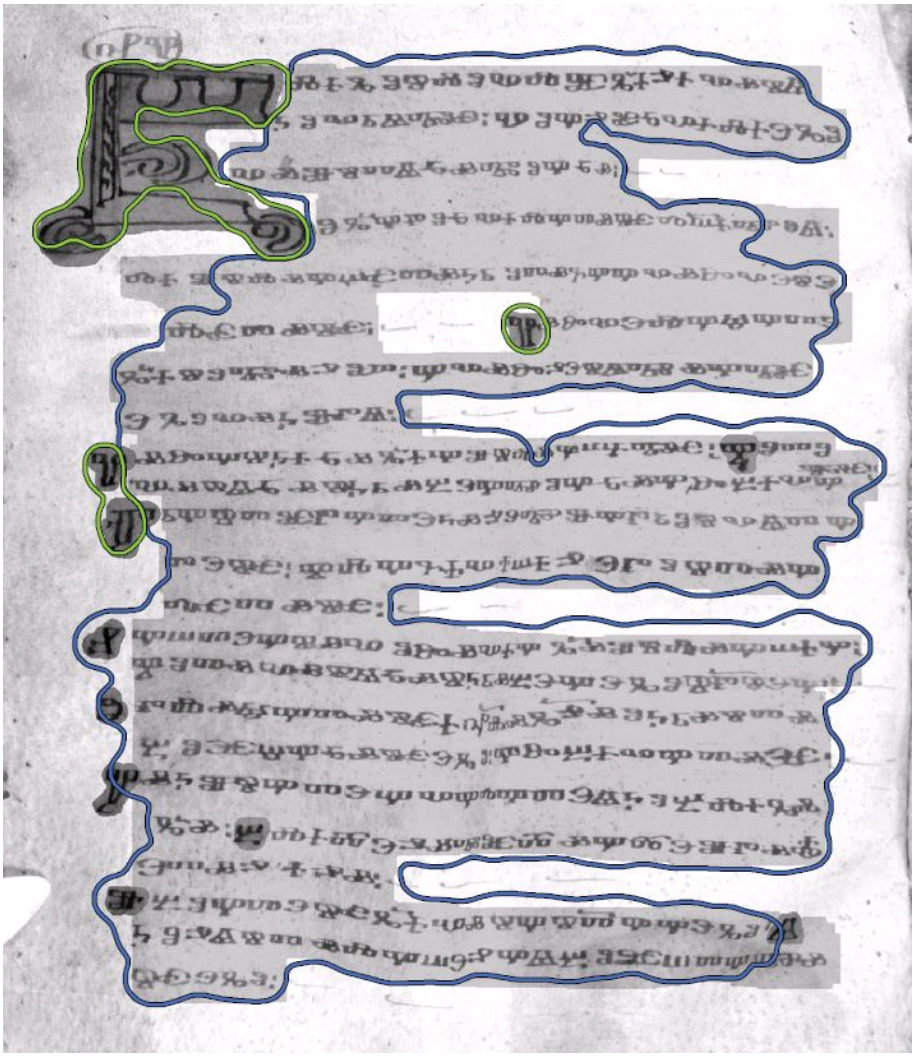

Figure 5. Automatic Layout Detection on Psalterium Demetrii Sinaitici fol. 43v.

Paleo Toolbox operations offer the large number of image viewing and image processing operations that are available in the standard ImageJ libraries and plugins, without requiring great expertise of the user. These rely largely on quantitative manipulation, measurement, and visualization of features of interest. It includes a set of image analysis (statistical approaches such as PCA), visualization (pseudocolor display and color channel manipulation) and markup tools (textual annotation; region-of-interest recording and replay). It more generally allows users to perform operations that assist in codicological and paleolographic investigations. *Paleo Toolbox* can be adapted to the needs of individual scholars: ImageJ's plugin architecture facilitates the development or adaptation of processing routines to match the goals of particular studies or the content of particular texts.

Its design is driven by the recognition that a linear model of image analysis, in which images are initially processed solely by computer scientists and are then subsequently viewed solely by humanities scholars, is both static and inadequate. *Paleo Toolbox* is designed to allow scholars with little expertise in image manipulation to dynamically explore images, to adjust display features (contrast, zoom, pan, color scheme etc.), but also to create annotations and markup. These can be used to guide computer scientists in adaptively choosing particular processing procedures based on the content.

Paleo Toolbox is based on an iterative model that envisions feedback from scholars that informs further processing. This model is more similar to the interplay between the human eye and the human brain than conventional approaches that conceive vision as a unidirectional flow of information from image processors (technologists) to visual interpreters (scholars). A further, central goal at the heart of *Paleo Toolbox* is to assist the collaboration between technologists and scholars, while adopting standard image and metadata formats that facilitate exchange, storage, and retrieval of images, annotations, and working materials.

Based on the insights of one of the authors (Miklas) at a recent working conference on the St. Catherine's Palimpsests Project, *Paleo Toolbox* is presently being configured to allow users to choose between a basic viewing environment, suited to the needs of non-experts to view and annotate images in a simplified setting, and a more sophisticated processing environment encompassing the full range of processing operations. This is intended to ease the joint use of a common tool by technologists and scholars, at their various levels of comfort in using software tools, during both individual and collaborative projects. (Please contact Bill Christens-Barry to obtain a copy).

3.5. Series Glagolitica Sinaitica

Finally, we would like to announce that the first edition of our new series *Glagolitica Sinaitica* (GlagSin) was issued by Holzhausen publishing house in August 2012

containing the facsimile edition of the Psalterium Demetrii Sinaitici and the inserted Medical Folia (Sin. Slav. 3/N); the critical edition of both manuscripts is well under way. The following editions will comprise not only other facsimile and critical editions of Sinai manuscripts, but also a separate volume on analysis methods, material chemistry, and computational results.

Acknowledgments

The authors would like to thank the Austrian Science Fund (FWF) for funding the project under grant P23133.

Bibliography

Belhumeur, P. N., J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection." In: *ECCV* 1 (1996). 45–58.

Diem, M., R. Sablatnig, M. Gau, and H. Miklas. "Recognizing Degraded Handwritten Characters." *Codicology and Palaeography in the Digital Age 2.* Ed. F. Fischer, C. Fritze, G. Vogeler. Norderstedt: BoD, 2011. 295–306. <http://kups.ub.uni-koeln.de/4337/>.

Duda, R.; Hart, P.; Stork, D. (ed.) (2001): *Pattern Classification*^{2}. New York: Wiley-Interscience, 2001.

Easton, R., W. Christens-Barry, and K. Knox. "Spectral Image Processing and Analysis of the Archimedes Palimpsest." *Proceedings of the 19th European Signal Processing Conference, EUSIPCO. 2011.*

Garz, A., M. Diem, and R. Sablatnig. "Layout Analysis of Ancient Manuscripts Using Local Features." Holappa, M. (Ed.): *Eikonopoiia. Digital Imaging of Ancient Textual Heritage 2010: Technological Challenges and Solutions.* (Commentationes Humanarum Litterarum. Vol. 129). Helsinki: Societas Scientiarum Fennica, 2011. 150–162.

Glagolitica Sinaitica. Wien: Holzhausen Verlag, 2012ff. <https://shop.verlagholzhausen.at/hhshop/Psalterium-Demetrii-Sinaitici.-Monasterii-s.-Catharinae-cod.-Slav.-3/n-adiectis-foliis-medicinalibus.htm?x3ce42=f72b414588eef8293a09877c464a828a>.

Hollaus, F., M. Gau, and R. Sablatnig. "Multispectral Image Acquisition of Ancient Manuscripts." *EuroMed* 2012. 30–39.

Hyvärinen, A.; Oja, E. "Independent Component Analysis: Algorithms and Applications." *Neural Networks.* Vol. 13/4-5 (2000). 411–430.

ImageJ. Image Procession and Analysis with Java. <http://rsb.info.nih.gov/ij/>.

Kleber, F., and R. Sablatnig. "Scientific Puzzle Solving: Current Techniques and Applications." *Computer Applications and Quantitative Methods in Archaeology (CAA). Making History Interactive.* Williamsburg (VA), 2009. (CD publication).

Kleber, F., R. Sablatnig, M. Gau, and H. Miklas. "Ruling Estimation for Degraded Ancient Documents Based on Text Line Extraction." In: *2nd Int. Conf. on Electronic Visualisation and the Arts (EVA). Digital Cultural Heritage – Essential for Tourism.* Vienna. 79–86.

Lettner, M.; Diem, M.; Sablatnig, R.; Miklas, H. (2008): "Registration and Enhancing of Multispectral Manuscript Images." *16th European Signal Processing Conf. (EUSIPCO)*. Lausanne. (CD publication).

Lettner, M.; Sablatnig, R. (2009): "Document Image Binarization in Multispectral Images using Markov Random Fields." *33rd Workshop of the Austrian Association for Pattern Recognition (AAPR): Challenges Visual Learning*. Ed. Roth, P.M., T. Mauthner, and T. Pock. (Schriftenreihe der OCG. Vol. 254). Stainz, 2008. 85–96.

Miklas, H., M. Gau, F. Kleber, M. Diem, M. Lettner, M.C. Vill, R. Sablatnig, M. Schreiner, M. Melcher, and E.-G. Hammerschmid. "St. Catherine's Monastery on Mount Sinai and the Balkan-Slavic Manuscript-Tradition." *Slovo. Towards a Digital Library of South Slavic Manuscripts*. Ed. H. Miklas and A. Miltenova. Sofia: Bulgarian Academy of Science, Institute of Literature, 2008. 13–36, 286 (Res.).

Rapantzikos, K., and C. Balas. "Hyperspectral Imaging: Potential in Non-Destructive Analysis of Palimpsests." *Int. Conf. on Image Processing (ICIP)* 2 (2005). 618–621.

Salerno, E., A. Tonazzini, and L. Bedini. "Digital Image Analysis to Enhance Underwritten Text in the Archimedes Palimpsest." *Int. Journal on Document Analysis and Recognition*. Vol. 92/4 (2007). 79–87.

Steinböck, D.: *Document Image Dewarping*. Bachelor Thesis. Vienna Univ. of Technology. 2012.

Su, B.; S. Lu, and C.L. Tan. "Binarization of Historical Document Images Using the Local Maximum and Minimum." *Document Analysis Systems*. (2010) 159–166.

Vill, M.C., and R. Sablatnig. "Static Stroke Decomposition of Glagolitic Characters." *2nd Int. Conf. on Electronic Visualisation and the Arts (EVA). Digital Cultural Heritage – Essential for Tourism*. Ed. R. Sablatnig, J. Hemsley, P. Kammerer, E. Zolda, and J. Stockinger. Vienna, 2008. 95–102.

Wartenberg, D. "Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis." In: *Geographical Analysis*. Vol. 17/4 (1985), 263–283.

Yosef, I. B., N. Hagbi, K. Kedem, and I. Dinstein. "Line Segmentation for Degraded Handwritten Historical Documents." In: *ICDAR* (2009). 1161–1165.

What lies beneath: The application of digital technology to uncover writing obscured by a chemical reagent

Christine Voth

Abstract

The Anglo-Saxon manuscript, London, British Library, Royal 12. D. XVII, is the oldest extant manuscript of Old English medical remedies (s. x^{med}). The manuscript ends incompletely on fol. 127v. On this same folio, a chemical reagent has been applied to several lines of writing along the long margin, obscuring the underlying text. This study adapts the digital technology established by Peter Stokes for uncovering palimpsests (2011) to uncover the writing hidden underneath the reagent. The methods are set out step by step in order that they may be repeated by others, or transferred for use on other damaged manuscripts. This study also reveals the underlying text as a possibly corrupted Hiberno-Latin charm, and provides an analysis and discussion of the text of the charm and its relation to the text of the manuscript. Overall, the results of this study further our understanding of the uses to which medical manuscripts were put during the Middle Ages, as well as providing a new means by which we can access text in damaged manuscripts.

Zusammenfassung

Die anglelsächsische Handschrift, London, British Library, Royal 12. D. XVII, ist die älteste existierende Handschrift altenglischer medizinischer Heilmittel (Mitte 10. Jh.). Die Handschrift endet unvollständig auf fol. 127v. Auf derselben Seite wurde eine chemische Reagenz auf mehrere Zeilen Schrift am Rand entlang der langen Seite angewendet, welche die Schrift unlesbar gemacht hat. Diese Studie wendet die von Peter Stokes (2011) zum Sichtbarmachen von Palimpsesten etablierte digitale Technologie an, um die Schrift unter der Reagenz lesbar zu machen. Die Methoden sind Schritt für Schritt erklärt, damit sie von anderen wiederholt oder auf andere beschädigte Handschriften übertragen werden können. Darüber hinaus deckt diese Studie auf, dass es sich bei dem unterliegenden Text womöglich um einen fehlerhaften Hiberno-Lateinischen Zauber handelt und analysiert und diskutiert den Text des Zaubers und seine Beziehung zum Text der Handschrift. Insgesamt erweitern die Ergebnisse dieser Studie sowohl unser Verständnis von der Anwendung

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 47–64.

medizinischer Handschriften im Mittelalter als auch von neuen Methoden den Text beschädigter Handschriften zugänglich zu machen.

1. Introduction

The application of chemical reagents came into vogue during the nineteenth century when scholars and antiquarians alike applied the liquid solutions to portions of medieval manuscripts where the writing had faded or had been erased (as well as some palimpsests) in order to make the faint writing more visible.[1] A number of Anglo-Saxon manuscripts were treated in this manner: the F-recension manuscript of the Anglo-Saxon Chronicle (London, British Library, Cotton Domitian A. VIII, fols. 30–70) shows evidence of reagent having been applied to the margins and text on fols. 58r, 60r, 61r and 67r (Baker 2000: XVI). Sometime after Grímur Jónsson Thorkelín made his transcription in 1787, a chemical reagent was applied to faded portions of the text of the Beowulf manuscript (London, British Library, Cotton Vitellius A. XV) on fols. 179r and 198v (Wrenn 1953: 12). When working on his edition of the Leofric Missal (Oxford, Bodleian Library, Bodley 579) in 1883, F. E. Warren received permission to apply a solution of hydrophosphate of ammonia to portions of the texts which were no longer readable.[2] The application of a reagent, typically similar in chemical makeup to the oak galls used in medieval ink, was intended to darken erased or faded scripts, and in many cases, the reagent temporarily darkened the text. But in all cases, irreparable damage from the chemicals occurred, changing the surface of the parchment in such a way as to make reading the underlying text nearly impossible for later use.

Until recently, scholars were restricted to the use of ultraviolet lamp, fortuitous natural light (Parkes 1991: 263), or a magnifying glass to aid in reading through the damage left by the reagent, often with limited results. However, the availability of high quality colour images of medieval manuscripts and the advent of digital technology provide new access to texts underlying a reagent. Adapting the image manipulation methods developed by Peter Stokes (2011) for use with palimpsests, I was able to digitally enhance a folio from the tenth-century Anglo-Saxon manuscript London, British Library, Royal 12. D. XVII that had been treated with a reagent. Although the procedures described below derive from Stokes's methods, their importance lies in making accessible the application of digital enhancement to texts that have been damaged by reagents or have darkened due to age or environment.

---

[1] In 1912, palaeographer E. M. Thompson recommended certain 'least harmful' reagents for use on palimpsests (p. 65).

[2] No scholarly work has been undertaken recording the number of Anglo-Saxon manuscripts having been treated with a reagent; these are just a few examples of the more well-known manuscripts.

Digitally uncovering palimpsests involves creating a colour difference between the parchment and the erased writing where little to none can be distinguished by the naked eye (Stokes 2011: 42). The same principles apply to working with a reagent because the ink and the reagent are often indistinguishable. This paper presents the digital manipulation in three sections: in the First, I will share the methodology of uncovering obscured writing; next, I will discuss what was discovered under the chemical reagent and my preliminary analysis; and finally, I will relate these results back to the manuscript itself, addressing the results of my findings in relation to both, the codicology of the manuscript and the content of the page upon which the text was written.

1.1. The Manuscript

Royal 12. D. XVII is an Old English manuscript of medical remedies, copied in a single hand of s. x^{med}. The manuscript is made up of three books, the first two of which are known collectively by the name 'Bald's Leechbook' because of the colophon naming Bald as the owner of the book, copied at the end of book II. The third book, which is now referred to as 'Leechbook III', is not believed to have been part of the original 'Bald's Leechbook' compilation, but copied into the current manuscript collection probably because of similar medical content. In its entirety, Royal 12. D. XVII comprises some of the earliest extant vernacular medical texts in Europe and is important to our understanding of early medieval medicine, as well as the production and use of medical texts in Anglo-Saxon England.

The manuscript ends incompletely on fol. 127v. Comparing the text of the final folio with the table of contents of book three (on fol. 111r, see Table 1), three remedies are apparently missing. The last few lines of chapter 71 (on the treatment of ulcers) are written at the top of fol. 127v, followed by the chapter on the treatment of 'yellow disease', which is numbered as chapter 72 in the text, but 73 in the table of contents. Chapter 73, providing remedies for when one's bowels are on the outside, is actually listed as chapter 74 in the table of contents and ends incompletely after only five lines. To sum up, chapter 72 on venom is missing entirely as are chapters 75 on 'inner sickness' and 76 for the making of a holy salve. This would seem to indicate that there is a missing gathering of unknown size at the end of the manuscript.

A chemical reagent has been applied to several lines of text written along the long margin of 127v (see figure 1). Only a few letters of the text are visible to the naked eye as the reagent has caused significant damage to the parchment – beyond the nature of the dark brown after-effects of the reagent – as the organic material of the page is now very delicate with a glossy appearance. In addition, several scrape marks are visible across the top of the reagent, indicating some attempt had been made in the past to try to remove it. Because of the incomplete nature of the third book, it seemed

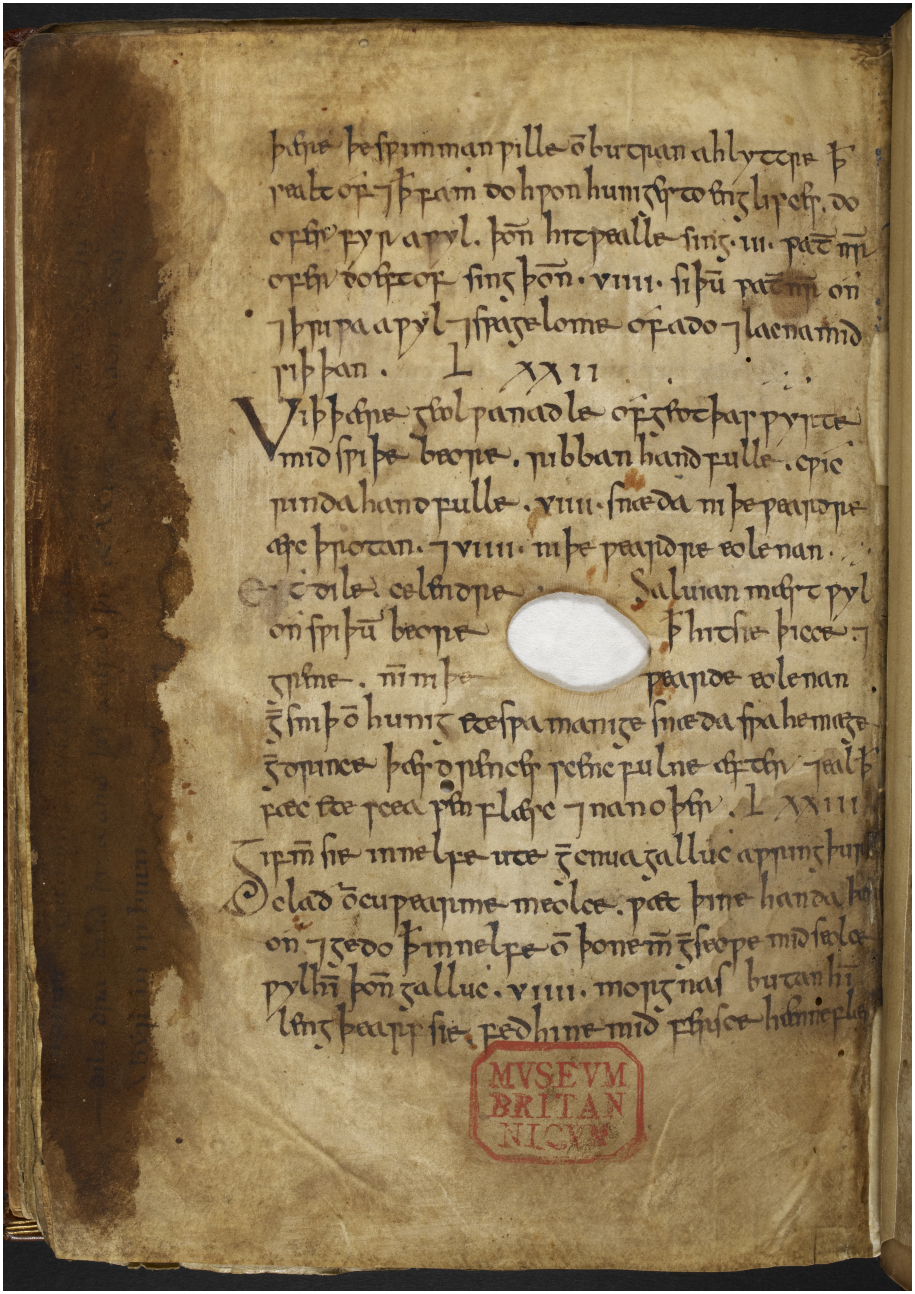

Figure 1. ©British Library Board: London, British Library, Royal 12. D. XVII, fol. 127v before any adjustment.

Chapter	Heading	
lxxi	*Wiþ springe smiring 7 sealf*	For ulcer, an ointment and a salve
lxxii	*Wiþ attre drenc 7 smiring*	For venom, a draught and a salve
lxxiii	*Wiþ þære geolwan adl*	Against the yellow disease
lxxiiii	*Wiþ þam gif innelfe si ute*	For when the bowels are on the outside
lxxv	*Wiþ ælcre innan untrymnesse 7 wiþ hefignesse 7 wiþ hleor blæce*	For all inner sickness, and for heaviness, and for a rash of the cheek
lxxvi	*Be þam hu man scyle halige sealfe wyrcean*	How one can make a holy salve

Table 1. Book III, Table of Contents.

important to conduct a full codicological analysis of the manuscript to uncover the writing under the reagent and to see if what is written might include or shed any light on any of the missing remedies.

Like the rest of the manuscripts of the Royal library that survived the fire at Ashburnham House, Royal 12. D. XVII was given to the care of the British Museum in 1757. When Royal 12. D. XVII was edited in 1865 by Oswald Cockayne as part of the Roll Series, no mention was made of the use of a chemical reagent on the final folio. In regard to the text of the long margin, Cockayne transcribed the following (p. 360):

> [line 1] (*completely obscured*)
>
> [line 2] *dila dra bind þ.....wod þi..... A Byr in ir bren.*[3]

If the reagent was not applied by Cockayne in an attempt to clarify the text – and he makes no mention of taking that effort in his edition – it would appear the reagent was then applied at some point after 1865. No record exists from the British Museum of permission being given to anyone for the application of the reagent.

2. Methodology

Scholars have had some success in viewing faint writing or erasures by using ultraviolet light, particularly in the case of palimpsests, because the 'organic material in the parchment fluoresces under such lighting' (Easton 2004: 3) and the underlying ink is then more visible to the naked eye. But given the similar chemical nature of reagents

[3] In Cockayne's transcription he only notes two lines, including words from line 3 on line 2 (see transcription below). Leonhardi (1905) makes no mention of the reagent in his diplomatic edition of the manuscript.

and ink, UV light may be of limited help in viewing writing covered by a reagent. I had the occasion to view 127v under both ultraviolet and white light to see if either environment aided in making the obscured portions of the text more visible. However, due to damage to the parchment from the reagent, neither method was successful. By far the best means of uncovering obscured writing has been Multispectral Imaging, as demonstrated with great success in the Archimedes Palimpsest Project. However, while extremely effective, MSI is also incredibly expensive, and thus not a feasible option for projects of limited size.

The success in uncovering palimpsest text using digital manipulation provided a more economic option for uncovering the underwriting. Although most graphics editing programmes can be used for the purposes, I used a freeware programme known as GIMP: the GNU Image Manipulation Program (version 2.6),[4] which offers the same tools for digital image manipulation as the commercial Adobe Photoshop. There is also a freeware programme developed for Macintosh by Peter Stokes called Image Viewer. Using GIMP, I adapted the methods established by Stokes for palimpsests and created a colour distribution histogram and then adjusted the colours of the image in an attempt to clarify the underlying text. To these I added a step of inverting the image so that all the text underlying would appear lighter than the reagent-covered parchment. This produced the best results.

2.1. Histogram[5]

A colour distribution histogram is a 'statistical summary of [...] the distribution of colour' in an image (Stokes 44), with the peaks of a histogram indicating the highest concentration of colour in the image. The arrows underneath the histogram are positioned to indicate the lightest colours of the image (to the left) and the darkest colours of the image (on the right). It is possible to take that concentration of colours and 'spread' it out to improve the readability of the obscured writing by narrowing the available colour range to either side of the peak of the histogram. This can be done on an entire folio or a very small portion of text. The best results will be to focus the histogram on the portion of the text covered by the reagent. As seen in figure 2, I selected a square of text to show the distribution, and adjusted the histogram to increase the differences between the lightest and the darkest colours. Moving the right-most arrow to the very edge of the peak lightens up the background, and moving the left-most arrow to the other edge of the peak, makes the darker portions (that is, the writing) appear darker. The levels that worked best were (47, 1.00, 112).

---

[4] Different versions of software may have different menu options than those indicated here from v. 2.6.

[5] In GIMP, this is found under Menu: colors > levels.

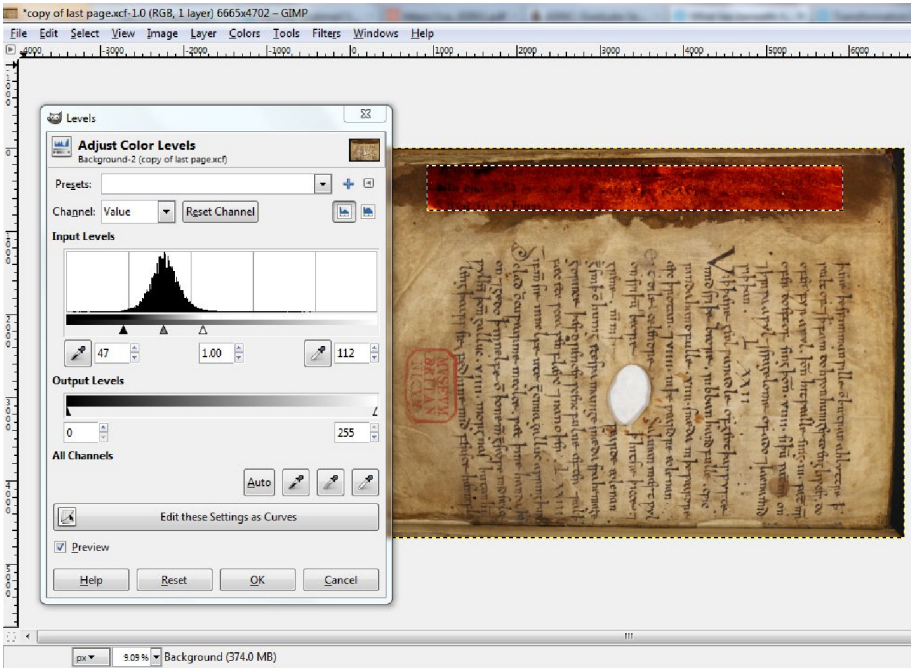

Figure 2. Histogram on 127v.

2.2. Colour Adjust[6]

The next step was to take the same portion of the image and make a further distinction between the background and the foreground. One of the easiest ways to do this is to use the 'monochrome' function. This will present the image in greyscale format, in which it is much easier to distinguish letter forms. However, in most cases, and particularly in the case of reagents and faded writing, an additional step is needed before changing the image to greyscale. This involves a process known as channel manipulation where essentially the colours of the image can be altered individually or in tandem in order to clarify the distinction between the parchment and the ink. Since computers represent colours through different intensities of red, green and blue, channel manipulation involves changing the intensities and allocation of one or more of these three colours. In the GIMP programme, this is done in 'channel mixer'. The primary background colour of the image after adjusting the histogram was red; this is

6 Menu: colors > component > channel mixer

indicated as percentages in the channel mixer function by settings of red channel 100, green channel 0, blue channel 0. Increasing the amount in the red channel column – that is, increasing the percentage of red colour in the image – eliminated much of the background noise and made the forms of the letters more apparent. After making these adjustments, I used the monochrome function to greyscale the image (this is also found within channel mixer) (see figure 3).

Figure 3. Colour adjustment (histogram) plus monochrome.

Because a high resolution photograph of a manuscript is quite large, there is often the problem of pixilation, 'noise' or artefacts which can distort the viewer's perception. Blurring the image or using a despeckling filter often presents a clearer image.[7] As this can be done at any time in the process, I was able to use this filter when it was most needed. For example, when the writing was very faint, I found it better to blur the image immediately after adjusting the histogram. But in the portions of the text where the writing or the reagent was particularly dark, it was better to use despeckle or blur after adjusting the colours, but before changing the image to greyscale.

[7] Menu: filters > blur OR filters > enhance > despeckle.

These steps can be taken on large portions of text, as I've shown in the previous images, but it may be necessary to highlight and manipulate a single word or segment of the image, particularly if the reagent covers just a single word or small portion of the text. In the case of 127v, the greatest amount of damage to the parchment has occurred in the upper left corner (when the image is in landscape orientation) where the pigment has darkened significantly and the surface of the parchment is thin and reflective. As someone has tried to remove the reagent with a sharp instrument (thus removing some of the underlying text as well), the sections on the very right are lighter, and thus better results were achieved by performing different adjustments on different portions of the folio (see figures 4 and 5).

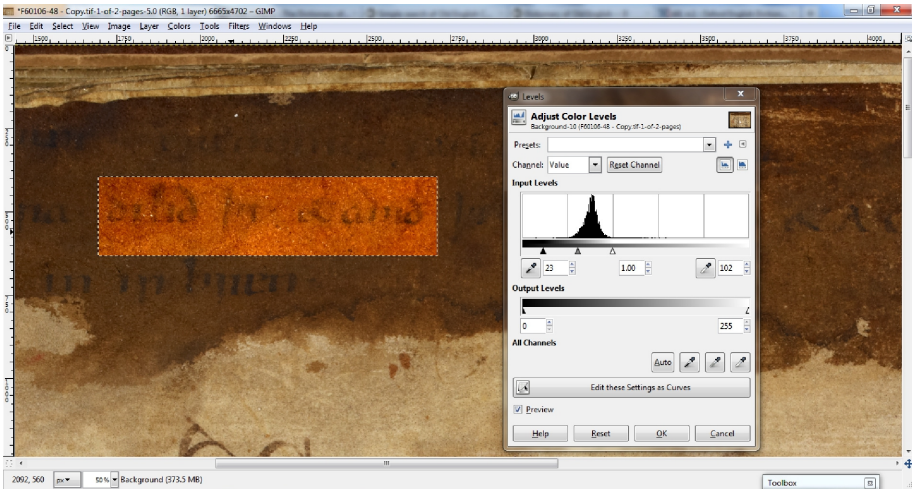

Figure 4. Using histogram to colour adjust individual words.

2.3. Inverting the Image

Experimenting with the different functions within GIMP, I came across a way to manipulate the colours of the entire folio by inverting the image so that the darkest portions of the page, the reagent and the ink of the main text, became the lightest colours. This presented a clearer picture of the underlying text as well as the damage the reagent had caused to the parchment. A histogram of the image showed two sub-peaks which represented the colours of the parchment and the main ink and a main peak, representing the darkest colour of the image, which would be the reagent. I narrowed the histogram (50, 1.00, 205) to either side of the two largest peaks, in an

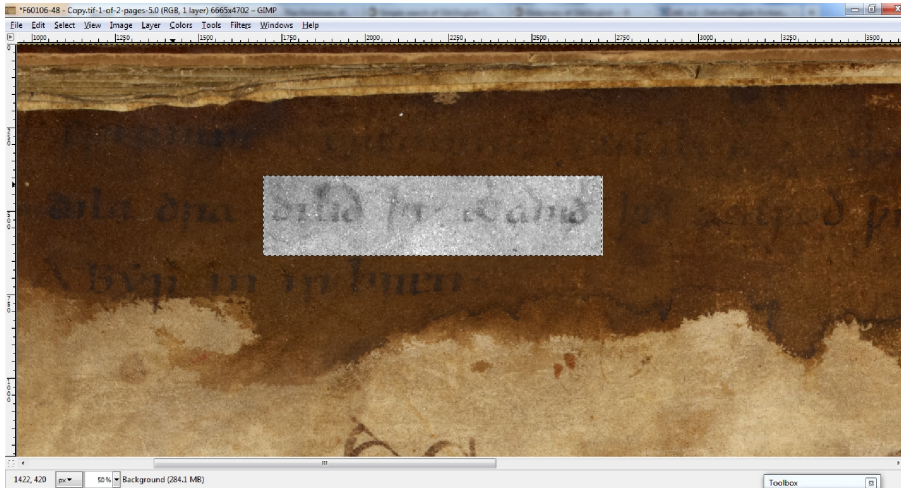

Figure 5. Monochrome plus despeckle on small portions of text after histogram.

attempt to narrow the range between the darkest and lightest colours in the image. Then, using the channel mixer, I altered the colours of the parchment entirely so that all the areas not covered by the reagent were tinted green, the ink and the reagent appeared red (200 red channel, -200 green channel, -200 blue channel) (see figure 6). The final step was to change the image to monochrome as seen in figures 7 and 8, giving us a better idea of the distribution of the writing under the reagent.

3. The Readable Text

It is possible to make out several words with only one or two letter forms showing through with some clarity, which I've indicated in the transcription below in Table 2. The square brackets indicate where I had the impression of word divisions but the letter forms are too severely damaged to make that out for certain. I've also indicated possible word/letter alternatives in the footnotes.

In comparing the above results to what was visible to the naked eye in 1865, it is possible to see that most of the top line had faded entirely already by 1865, whether from damage, erasure or time. The application of the chemical reagent and the subsequent attempt at removal may also have caused further obscuring of the writing. The marks below the marginal commentary annotation which look like water damage when seen in the inverted image are where the chemical reagent bled through to the

recto. It would appear that at least two coats of reagent were applied, although it is uncertain whether they were applied in the same sitting or at different times.

Figure 6. The first step in inverting an image.

[line 1]	† *swernum*[8] *Tremme*[.....][9] *lara* [....] [*...a*] [*d..*] *In*[10] *Iren*
[line 2]	*dila dra dilið þy & anið þri _eawoð*[11] *þy & A Cin & Adon*[12] *[..æs]*[13] [..]
[line 3]	*A Byr in ir bren.*

Table 2. Transcription of the underlying text.

The underlying text is in a hand of Vernacular Minuscule c. 1000, and does not appear elsewhere in the later annotations or additions to the manuscript. While at a glance the text does not represent recognisable Old English, the orthography, such as

[8] or spernum

[9] At least one, possibly two words between *Tremme* and *lara*

[10] **t** is subscripted directly below the **n**, so it may be long to In\t/ or [..æs\t/]

[11] or *_eawod*; the formation of Insular d and eth are nearly indistinguishable in vernacular script.

[12] or *Adam*

[13] This word begins with a majuscule letter, looks to be an *R*, *B* or a *K* with at least one, possibly two letters between it and *æ*

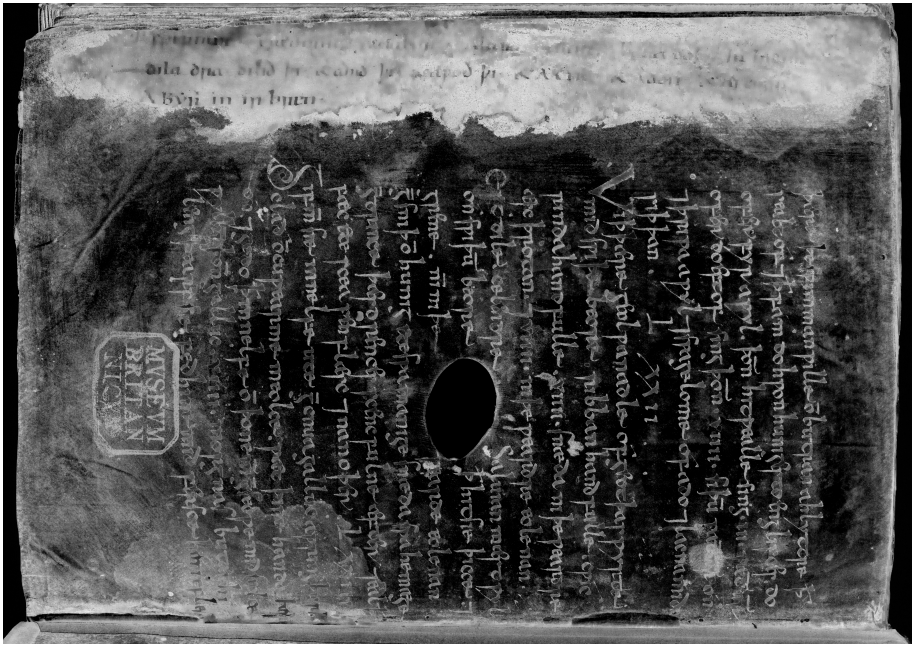

Figure 7. The inverted image of 127v.

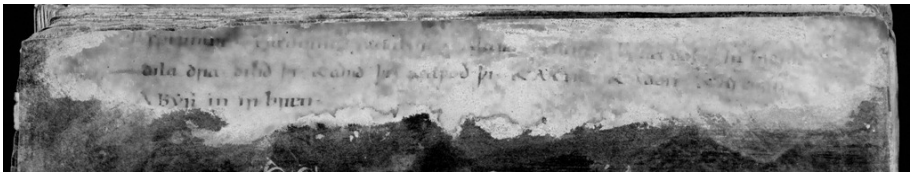

Figure 8. A closer view at the underlying text with the image inverted.

dotting the **y** of *Byr* (the second word on the third line of text), appears to be Late West Saxon.

Despite the loss of part of the text, the marginal annotation resembles an incantation, or charm. At first glance, it has only the vaguest resemblance to Old English in the first line, along with many unintelligible words, and some alliteration like *dila dra dilið* in the second and third lines. Felix Grendon classified this type of charm or incantation as 'gibberish' or nonsensical in 1904, yet the work of Heinrich Zimmer in 1895 and Howard Meroney in 1945 has shown that the original language of the

majority of extant incantations are a combination of Old Irish and Latin. Recent advances have also been made by Professor David Stifter in the study of the Old Irish origin of medical charms and incantations in the Stowe Missal.[14] Additionally, Lea Olsan has shown that the forms of Anglo-Saxon vernacular and Latin charms were fluid, affected both by transmission and oral tradition (1999:407–8).

4. History of Incantations and their Medical Use

While most charms or incantations have come to the Anglo-Saxon medical corpus through the writings of Classical and Late Antique authors including Alexander of Tralles, Theodorus Priscian, Pliny the Elder, and Marcellus Empiricus transmitted through the medium of *florilegia* texts (Cameron 1982: 135–55; Wallis 1995: 101–26), Anglo-Saxon nonsensical charms most typically derive from a Hiberno-Latin tradition of prayers and incantations that were borrowed into the Anglo-Saxon medical practice. Many of the unfamiliar Old Irish words in these charms have been altered: truncated, divided or even combined with other Old Irish or Latin words, which does not make it easy to recover their original forms. The use of unintelligible language in charms, incantations or amuletic remedies is believed by some scholars to have strengthened the perceived efficacy of the remedy by 'virtue of being unintelligible' (Pollington 2000: 422). Corrupted Hiberno-Latin in Anglo-Saxon charms occurs with some frequency in the eleventh century medical manuscript known as the *Lacnunga* (London, British Library, Harley 585).[15]

These incantations are now considered gibberish because the language has been corrupted in transmission, possibly through a combination of phonetic and orthographic adaptation, resulting in a loss of the original word forms over time. This may have been, in part, due to Anglo-Saxon practitioners or scribes attempting to regulate unfamiliar words into something with a perceived familiarity, or possibly the corruption of the words may have occurred in the process of writing down phonetically the recitation of an unfamiliar language. Alternatively, the transmission of these incantations in this altered state may have been purposeful, so that the true meaning of the incantation would only be known to the practitioner, and the esoteric nature of the incantation preserved (Olsan 2003:357–60).

Within Royal 12. D. XVII there is further evidence for the transmission of a language corrupted over time in an amuletic charm in Book 1 (chapter 65, fol. 53r) that requires the medical practitioner to write down a series of names in Greek as part of a prayer.

---

[14] Professor Stifter discussed the results of his continuing work with the Stowe Missal at his inaugural lecture on the study of Old Irish at the National University of Ireland, Maynooth (2 February 2012).

[15] The most recent edition of the *Lacnunga* by E. Pettit (2001) includes commentary on these types of incantations.

What is written is HΛMMΛNϒEL. BPONIce. NOϒeꝒTΛϒEPΓ.[16] Somewhere in the written transmission of this remedy, some of the original Greek letter forms have been lost, some replaced with Latin letter forms and runes, and the meaning and form of the final word has been lost entirely.

Because an incantation, by its nature, is spoken aloud, the written form must be such that it can be read or performed as part of a healing ritual; many remedies prescribe that charms be carried or used in rituals as inscriptions. The original language may have been lost, but the written transmission has at least maintained clarity and preserved some rhythmic qualities, albeit at the expense of understanding the actual content. What appears in the incantations in the Anglo-Saxon medical corpus is an attempt to re-create word structure which would seem familiar to someone with knowledge of Latin and Old English. There are some Hiberno-Latin charms found in Anglo-Saxon manuscripts that have survived in forms which are more easily reconstructed to the original language,[17] whereas other incantations provide much more of a challenge in recovery of the original language and sense of the charm.

I have laid out some possible interpretations of the recovered words in the text from 127v in Table 3.[18]

The first line has a sense of Old English, and perhaps comprised the Old English introduction to the incantation, though the inability to recover many words and letter forms at this time makes it difficult to judge this with any level of certainty. If this line is Old English, then the general impression is that the incantation is for the use of iron (*Iren*) in a ritual to strengthen or aid something (*getremme*); possibly alluding to a teaching tradition (*lara*). The emphasis should be placed on **if** in this situation, however, since there are other possible readings for Tremme.

The second and third lines provide more challenge, despite the fact that there are more words recovered here than in the first line. The first word on line two may be translated as dill (*dila*), which is a herb found in over 30 remedies in Royal 12. D. XVII. *Anið* may be a truncated or corrupted version of the Latin word for dill: *anethum*. In which case, there may be an importance being placed on this particular herb as a medical ingredient or cure-all. *Cin* and *Adon* may be Cain and Adam or Adonai as many of these incantations call upon a saint or have Biblical or Christian references.

---

[16] The first word is most typically written in medieval Greek as EMMANOϒHΛ (Emmanuel), the second BEPENIKHE (Veronica), so it is easy to see where the corruption has occurred in these words. Veronica is also called upon as protection against evil in chapter 64, 52v, written BEPPNNIKNE. For more on the transmission of Greek in Anglo-Saxon England, see Bierbaumer, Peter, "Small Latin and Less Greek? Zur Kenntnis der Klassischen Sprachen in altenglischer Zeit", in *The History and the Dialects of English. Festschrift for Eduard Kolb.* ed. Andreas Fischer, Heidelberg: Carle Winter, 1989. 79–90.

[17] See Merony, 'Old English Charms' (1945) and Pettit, *Lacnunga* (2001).

[18] Many thanks to Dr. Elizabeth Boyle, Prof. Paul Russell and Silva Nurmio for help with the philological reconstruction of the Irish portions of this text.

Tremme [Old English: *getremman/trymman* 'to strengthen, confirm']
lara [OE: lar 'teaching, instruction', genitive plural]
In Iren [OE: 'with iron']
dila [OE: *dile* 'dill'; Old Irish: *do lega* 'disolve, digest'; Lat: dilaudo 'distinguish with praise']
dra [Latin: *tria* 'three']
dilið/dilid [Lat: *dilato* 'to dilate, spread out'; OE: *dile*]
þy [OE: demonst. pron. instr.]
anið (þri?) [Lat: *anethum* 'dill, anise']
Cin [OIr: *cenn* 'head' /proper name?/Cain?]
Adon [OE: *adon* 'to take away, banish'/proper name?/Adam? Adonai?]
A [OIr vocative particle]
Byr [OIr *bir* 'spear'/proper name??]
bren [OIr: *brén* 'putrid, rotten']

Table 3. Preliminary linguistic breakdown of the incantation on 127v.

The reference to Cain may be in relation to the slaying of his brother Abel, and a reference to a form of wound by iron. However, the reconstruction of this charm is still in the working stages.

The distribution of majuscule letters in the text should also be noted. *Tremme* is the second word in the first line, but begins with a majuscule **T**. *In Iren* at the end of line one both begin with a majuscule **I**. *Cin*, *Adon*, *Bren* also begin with majuscule letters, although they may also represent proper names. While attention to style and character formation is not always a priority in marginal commentaries, it is interesting to see this kind of specific detail used. This leads me to believe that these particular words, proper names or otherwise, are important to the perceived efficacy of the remedy and were copied from the original.

The hand dates to the end of the tenth/beginning of the eleventh century and the letters are more indicative of Vernacular Minuscule than Caroline Minuscule, thus the scribe's use of the Caroline *et* (the ampersand) rather than the more common Tyronian *nota* is unusual. It is uncommon to see the use of an ampersand in a vernacular text, unless there is an interjection of Latin, and in which case the ampersand is only ever really used within the Latin portion of the text. The presence here may be an indicator of a remnant of the Latin origins of the incantation.

5. Discussion

The incantation begins with a *signe de renvoi*, which is intended as a cross-reference for the reader, directing him/her to a particular place in the manuscript where a notation or correction has been made. However, there is no corresponding *signe* on 127v. As the location of an incantation is rarely coincidental, it is important to establish to which remedies the incantation might correspond. The three remedies on 127v include the end of chapter 71 on ulcers, the entirety of chapter 73 (mis-numbered 72) on yellow disease, probably jaundice, and part of chapter 74 (mis-numbered 73) for when one's insides are on the outside. Ker (1957:332) suggested that the underlying text may provide information regarding the missing remedy, which was my initial impression as well. However, it appears that the chapter on venom was lost before the manuscript was copied, which makes it unlikely the incantation refers to that particular remedy. It is possible that the incantation relates either to the remedy for jaundice or for the inappropriately placed intestines. An examination of the uses of iron and dill in medical remedies in Royal 12. D. XVII shows that iron is most likely to be used in remedies for wounds and internal disorders such as 'broken intestines' (*tobrocenum innoðum*) and 'swollen spleen' (*aswollenum milte*), whereas dill is found in over 30 remedies for various ailments and wounds of the stomach and intestines. Therefore, it seems likely that the incantation pertains to the remedy for internal disorders such as chapter 74.

However, going back to the word that follows the *signe* at the beginning of the first line: *swernum* [*spernum*], it should be noted that it begins with a minuscule long-s, and is followed by a word which begins with a majuscule letter, *Tremme*. This combination is troubling, but it could be explained if that first word is not part of the incantation itself, but is used together with the *signe* as a second point of reference. As there is neither a corresponding *signe* on 127v nor is there a word that resembles *swernum* in the text, this incantation may not refer to a passage on 127v at all, but to one on the facing page of the manuscript, the now missing folio 128r.

6. Conclusion

Uncovering the passage written on the long margin of 127v was attempted not only to identify the text itself, but also to see if that text might shed light on the missing remedies at the end of the manuscript. The marginal notation found was written almost fifty years after Royal 12. D. XVII was copied, and furthers our understanding of the use and readership of medical manuscripts. Medical texts would appear to be a logical vehicle for incantations, yet none were copied into the Royal manuscript before the eleventh century; the other medical compendia where incantations of

this type have been recorded date no earlier than the eleventh century.[19] This alone may tell us something about the earliest transmission of Hiberno-Latin charms in Anglo-Saxon England and their incorporation into the vernacular medical tradition at a time when the production of medical codices was flourishing.

In regard to the codicology of the final quire of the manuscript, the presence of the incantation raises more questions than it answers. The darkened and worn parchment found in the first and final extant quires of Royal 12. D. XVII are consistent with what we know of medieval manuscripts that were kept in a soft binding, and do not suggest the presence of additional quires at the end of the manuscript. Yet the Royal scribe did not note an incomplete exemplar, which would suggest that perhaps single sheet(s) were added to the final quire to complete the text of book III. A missing leaf would supply sufficient parchment for the final remedies which may have included the corresponding *signe* and reference point for the incantation. Furthermore, a single leaf at the end of a quire could easily have come loose and been lost at any point before the manuscript was rebound in 1751.

The methods addressed in this paper provide a useful and economical technique for finding what lies beneath reagents. They could easily be adapted for examining other types of obscured portions of manuscripts. Certain environmental conditions or insufficient storage can result in permanent damage to a manuscript. Codices exposed to a lack of humidity or excessive heat have darkened and become brittle. The techniques discussed here could provide aid in seeing texts obscured by a change in the organic nature of the folio.

Bibliography

Archimedes Palimpsest Project. Baltimore: The Walters Art Museum. <http://archimedespalimpsest.org/>.

Baker, P. *The Anglo-Saxon Chronicle: A Collaborative Edition*, vol. 8, MS F. Woodbridge: Brewer, 2000.

Bosworth, J., and T. N. Toller, *An Anglo-Saxon Dictionary Based on the Manuscript Collections of the Late Joseph Bosworth.* Oxford: Clarendon Press, 1882.

Bosworth-Toller Anglo-Saxon Dictionary. Eds. S. Crist, O. Tichy. 2001–2010. <http://bosworth.ff.cuni.cz/>.

Cameron, M. L. "The Sources of Medical Knowledge in Anglo-Saxon England." *Anglo-Saxon England* 11 (1982): 135–55.

Clark Hall, J. R. *A Concise Anglo-Saxon Dictionary.* 1894. Cambridge: Cambridge UP, 1994.

Cockayne, T. O. *Leechdoms, Wortcunning and Starcraft of Early England.* Vol. 2. London: Longman, 1865. Vol. II.

[19] See reference to the *Lacnunga* above.

Dictionary of Old English Plant Names. Eds. P. Bierbaumer and H. Sauer with H. W. Klug and U. Krischke. 2007-2009. <http://oldenglish-plantnames.org/>.

Easton, R. "Digital Restoration of Erased and Damaged Manuscripts." *Proceedings of the 39th Annual Convention of the Association of Jewish Libraries.* Brooklyn, 2004. <http://databases.jewishlibraries.org/sites/default/files/proceedings/proceedings2004/easton.pdf>.

Grendon, F. "The Anglo-Saxon Charms." *The Journal of American Folklore* 22 (1909): 105–237.

GNU Public License. GNU Operating System, Free Software Foundation. <http://www.gnu.org/licenses/gpl.html>.

Ker, N. *Catalogue of Manuscripts Containing Anglo-Saxon.* Oxford: Clarendon Press, 1957.

Meroney, H. "Irish in the Old English Charms." *Speculum* 20 (1945): 172–82.

Olsan, L. "Charms and Prayers in Medieval Medical Theory and Practice." *Social History of Medicine* 16 (2003): 343–66.

Olsan, L. "The Inscription of Charms in Anglo-Saxon Manuscripts." *Oral Tradition* 14 (1999): 401–19.

Pollington, St. *Leechcraft: Early English Charms, Plant Lore and Healing*. Norfolk: Anglo-Saxon Books, 2000.

Parkes, M. B. "An Anglo-Saxon Text at Fleury: The Manuscript of the Leiden Riddle." in his *Scripts, Scribes and Readers: Studies in the Communication, Presentation and Dissemination of Medieval Texts.* London: Hambledon Press, 1991. 263–74.

Pettit, E. *Anglo-Saxon Remedies, Charms, and Prayers from British Library MS Harley 585: The Lacnunga: Introduction, Text, Translation, and Appendices*, Mellen Critical Editions and Translations. Lewiston: Mellen Press, 2001.

Stokes, P. "Recovering Anglo-Saxon Erasures: Some Questions, Tools and Techniques." *Palimpsests and the Literary Imagination of Medieval England.* Eds. L. Carruthers, R. Chai-Elsholz and T. Silec. New York: Palgrave MacMillan, 2011. 35–60.

Thompson, E. M. *An Introduction to Greek and Latin Palaeography.* Oxford: Clarendon Press, 1912.

Wallis, F. "The Experience of the Book: Manuscripts, Texts and the Role of Epistemology in Early Medieval Medicine." *Knowledge and the Scholarly Medical Tradition.* Ed. D. Bates. Cambridge: Cambridge UP, 1995. 101–26.

Warren, F. E. *The Leofric Missal: As Used in the Cathedral of Exeter During the Episcopate of its First Bishop, A.D. 1050–1072.* Oxford: Clarendon Press, 1883.

Wrenn, C. L. "Introduction: 1. The Manuscript." *Beowulf, with the Finnesburg Fragment*, Ed. W. F. Bolton. London: Harrap, 1953.

Wright, C. *Bald's Leechbook: British Museum Royal manuscript 12. D. XVII*, Early English Manuscripts in Facsimile v. Copenhagen: Rosenkilde-Bagger, 1955.

Zimmer, H. "Ein altirischer Zauberspruch aus der Vikingerzeit." *Zeitschrift für vergleichende Sprachforschung* 33 (1895): 141–53.

Verwaltung von Erschließungsdaten

Organizing descriptive information

The development of a medieval scribe

Rombert Stapel

Abstract

Every individual has a set of traits unique for that person. These include biometric identifiers such as DNA, but the same principal applies to the notion of a scribal fingerprint or human stylome. In contrast to the innate nature of a real fingerprint, such features have been acquired over time and, by definition, are therefore subject to change. Knowledge of the (lack of) consistency of such linguistic or palaeographic identifiers over time is essential in constructing unique personal identifiers for scribes. The present article examines the case of one scribe, working as a secretary for the Teutonic Order in Utrecht and as notary public. His corpus of texts, which includes an important author's copy of the late fifteenth century *Jüngere Hochmeisterchronik*, covers a period of thirty years. By quantifying spelling preferences, character sizes, letter-forms and the use of abbreviations it is possible to monitor the development of his writing through time. It turns out that spelling preferences and the use of abbreviations show remarkably little consistency over a longer period. Only changing patterns in the use of certain letter-forms can be used to create a more stable timeline in Hendrik van Vianen's writings. Furthermore, abrupt changes in the patterns have been used to indicate a phased genesis of the manuscript of the *Jüngere Hochmeisterchronik*.

Zusammenfassung

Jedes Individuum hat ein eine Reihe von Eigenschaften, die einmalig für diese Person sind. Dazu gehören biometrische Merkmale wie die DNA. Die gleichen Prinzipien lassen sich auf die Idee des Fingerabdrucks eines Schreibers oder des menschlichen Stils anwenden. Im Gegensatz zur angeborenen Natur des eigentlichen Fingerabdrucks sind hier die Eigenschaften aber über die Zeit angeeignet und daher per definitionem Gegenstand der Veränderung. Das Wissen um (das Fehlen von) Beständigkeit solcher linguistischer oder paläographischer Eigenschaften ist wesentlich bei der Konstruktion eindeutiger persönlicher Identifikatoren für Schreiber. Der vorliegende Artikel untersucht den Fall eines Schreibers, der als Sekretär für den Deutschen Orden in Utrecht und als öffentlicher Notar tätig war. Sein Textkorpus, das eine wichtige Autorkopie der *Jüngeren Hochmeisterchronik* des späten 15. Jahrhunderts einschließt, umfasst einen Zeitraum von 30 Jahren. Durch die Quantifizierung von

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 67–86.

Präferenzen in der Rechtschreibung, von Buchstabengrößen und -formen und den Gebrauch von Abkürzungen ist es möglich, Entwicklungen in seinen Werken durch die Zeit zu beobachten. Es zeigt sich, dass Rechtschreibpräferenzen und der Gebrauch von Abkürzungen über einen längeren Zeitraum bemerkenswert geringe Konstanz aufweisen und daher nur wechselnde Muster im Gebrauch bestimmter Buchstabenformen genutzt werden können, um eine stabilere zeitliche Einordnung der Schriften Hendrik von Vianens zu erschaffen. Darüber hinaus wurden plötzliche Wechsel in den Mustern bisher genutzt, um die phasenweise Entstehung der *Jüngeren Hochmeisterchronik* zu kennzeichnen.

1. Introduction

One of the aims of palaeographers has traditionally been the discrimination of scribes. Together with dating and localizing a hand it remains one of the key questions posed in the field. When identifying scribes, either manually or by computer, one automatically assumes the existence of a set of features that are unique for this person, this scribe. An almost limitless variety of these traits in medieval manuscripts can be quantified, such as letter angles, orthography, abbreviations, letter-forms and spacing (compare Stokes, particularly 313-314), and a wide variety of studies have approached these quantitative data in different ways (take for instance from the previous volumes: Stokes; Hofmeister, Hofmeister-Winter, and Thallinger; Aussems and Brink; Stutzmann). From a selection of these aspects, Mark Aussems developed a template for scribal discrimination that he dubbed the *scribal fingerprint* (Aussems). Just as with a human fingerprint, or DNA, there is the implication that we are dealing with a unique individual marker.

The same principle applies to the field of stylometry. Here too, there is a silent assumption that an individual employs a distinct – and therefore identifiable – "set of measurable traits of language products", referred to as the *human stylome* (Van Halteren et al.). Stylometry approaches the text primarily as a linguistic object, measuring linguistic features by means of statistical procedures. Applied to medieval texts, stylometry always faces the issue of the dynamic nature of these texts. Scribes often change the orthography or even the text itself, causing a strict distinction between scribe and author to be problematic. In the last couple of years however a number of studies using stylometric methods have tackled this challenge head-on, with promising results, also when distinguishing scribes in a text (Van Dalen-Oskam and Van Zundert; Van Dalen-Oskam; Kestemont and Van Dalen-Oskam; Kestemont).

In their respective fields, both the linguistic approach of a medieval text as well as the palaeographic or codicological studies – unfortunately as of yet without much interaction – are able to provide valuable insight into the workings of a medieval

writer. However, in order to distinguish scribes and/or authors from another, it is essential to know how consistent the expressions of scribes remain during their career. Writing in the Middle Ages remains an inherently 'human' process and it is therefore not surprising that the writing preferences of a medieval scribe vary. Note for instance how some scribes start their manuscripts using a formal script, that gradually degenerates and becomes more current towards the end (compare Parkes 21; McGillivray 55–56). One may also find examples of the opposite, with scribes slowly finding their rhythm in the opening stages of a manuscript before reaching consistency in their handwriting.

Indeed, quantitative case studies by Jacob Thaisen and John McGillivray have shown that the expressions of a single scribe could evolve significantly within a single manuscript (Thaisen, "Probabilistic Analysis"; Thaisen, "Overlooked Variants"; McGillivray). Both examine their manuscripts on a fairly granular level – that of the quire or tale – and also point at the need for more comparative material provided on a grander scale. Until then, it remains difficult to ascertain what parts of the writing of scribes is consistent and what parts could develop over time – both within the confinements of a single manuscript as well as during the career of a scribe. Whereas the material presented below is hardly the grand scale perhaps envisaged, it can showcase the possibilities of such an approach. Moreover, we have chosen to examine the text on a lower level – folio or chapter/paragraph – in order to examine the development of the writings more precisely. Using a unique corpus of writings by a single scribe the development of the scribe's hand can be followed over time. Furthermore, the quantified scribal features of his writings will be used to define the phased genesis of his most notable work, the oldest manuscript of the so-called *Croniken van der Duytscher Oirden* (Chronicle of the Teutonic Order) or *Jüngere Hochmeisterchronik* (e.g. Stapel and Vollmann-Profe).

2. The corpus

The *Croniken van der Duytscher Oirden* concerns the history of the Teutonic Order, a military order that originated in the Holy Land during the Third Crusade. From the thirteenth century onwards the Order became increasingly active in the Baltic region. The *Croniken* was written originally in Middle Dutch, probably in the Utrecht bailiwick of the Teutonic Order. It is generally regarded as the last testimony of a long line of historiographical works produced by members of the Teutonic Order that include well-known works such as the *Livländische Reimchronik*, the chronicles by Peter von Dusburg and Nikolaus von Jeroschin and the *Ältere Hochmeisterchronik*. The *Croniken* became one of the most influential of these texts in the sixteenth century, not only within the Order, but also or perhaps even primarily in upcoming urban

circles in both Prussia and Livonia and beyond. Numerous manuscripts, translated into both Low German and High German were dispersed across Central and Northeast Europe.[1]

The oldest manuscript of the *Croniken* is now held in the *Deutschordenszentralarchiv* in Vienna (Hs. 392).[2] Watermark evidence suggests that the manuscript (in its present state containing 201 folios and 774 chapters)[3] was produced in at least three phases. In a first phase, the quires 3 to 9 (of 19) were assembled using paper dated around 1480. Around 1491 the second half of the chronicle, including a part at the end that described the history of the local Utrecht bailiwick and Land Commanders, was added (quire 10-19). The text on the final page, containing the life of Land Commander Johan van Drongelen (1469-1492), was finished later by the scribe of the manuscript. It suggests a *terminus ante quem* of the rest of the chronicle of 15 August 1492, the day Van Drongelen had died – which is mentioned in the text.[4] Using the same paper a table of content was placed in front of the chronicle (quire 2). Finally, three single bifolia that could, perhaps, be dated around 1496 can be found in three different places, among which quire 1. Only two of these bifolia contain medieval text, the other was originally left blank and now contains seventeenth century notes. All medieval text in the manuscript, thus including those two bifolia, is written in one single hand. And although the Vienna manuscript is in no sense a working copy, for one it has a neat and finished appearance, there is indeed evidence of editorial amendments that one could only associate with the author. The Vienna manuscript is therefore either an autograph, or an author's copy written in close collaboration with the author – the latter being the most probable for various reasons.[5]

The hand that wrote the *Croniken* manuscript belonged to a professional writer named Hendrik Gerardsz van Vianen. Van Vianen was probably secretary of the Land Commander of the Utrecht bailiwick Johan van Drongelen (1469-1492) and later in life became active as notary public. Between 1479 and 1509 he wrote, apart from

---

1 A complete list and description of all the extant manuscripts – including a detailed codicological and palaeographical examination of the Vienna manuscript discussed below – will be part of the dissertation in preparation by the author. The most recent manuscript description of the Vienna manuscript was published in 2000 (Lackner Kat.Nr. 62).

2 Hereafter manuscript We1 (the signature used in the dissertation) or simply the 'Vienna manuscript'.

3 The (modern) chapter numbers used in the new edition that is in preparation as part of the dissertation do not correspond to the ones added by Theodor Hirsch in his edition of the text printed in 1874 (Hirsch) or those present in the edition by Antonius Matthaeus from 1710 (Matthaeus 1–284; 343–360).

4 In the Vienna manuscript the text stops abruptly in the middle of a sentence on the last folio. At least one folio is missing, as well as two sentences that can be reconstructed using the other manuscript copies of the *Croniken*. These two sentences also provide the date of Van Drongelen's death and strictly speaking therefore, the date is not part of the present state of the Vienna manuscript. There is little doubt that it would have been included in the original manuscript though.

5 For the argumentation, again, reference is made to the dissertation currently in preparation. A selection of arguments can also be found in Stapel 345.

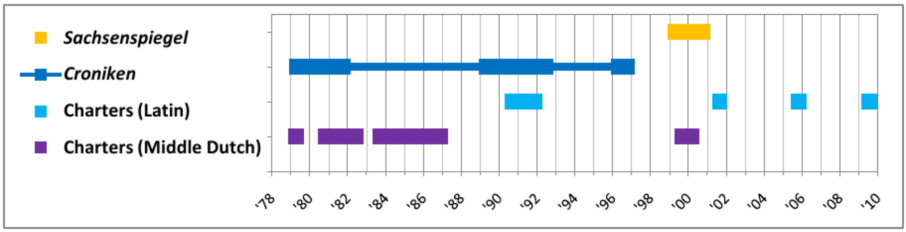

Figure 1. Writing activities of Hendrik van Vianen (1479-1509), as are presently known

the *Croniken* manuscript, a number of land charters. Most of them were written in Middle Dutch and intended for the Teutonic Order. In a couple of these land charters the name and notarial sign of Hendrik van Vianen is revealed. His hand can also be recognized in a manuscript copy of a Middle Dutch Sachsenspiegel (around 1499-1500) that contains owner marks of two sixteenth century Utrecht Land Commanders (The Hague, *Koninklijke Bibliotheek*, 133 H 4). All of his Middle Dutch writings have been transcribed in XML following the guidelines of the Text Encoding Initiative (P5). The corpus comprises more than 130.000 words. Using this corpus we can follow the development of a scribe's hand and scribal preferences.

3. The analysis

With his extant writings ranging from 1479 to 1509, the opening stages of the *Croniken*, written on paper from around 1480, will have been written early on in Hendrik van Vianen's career. By 1491, the approximate date of the paper on which the second half of the *Croniken* is written, Van Vianen had written at the very least numerous charters and over seventy to eighty folios of the *Croniken*. In theory, Van Vianen will have become a more experienced scribe. In the first quire of the *Croniken* manuscript Van Vianen still showed signs of inexperience in his writing – or inconsistency at least. Here the picture emerges of a scribe that needed to get into gear and had not yet developed a persistent writing mode. To the naked eye, the characters appear smaller than elsewhere in the manuscript. This is clearly supported by the average width of the characters calculated for the entire manuscript (Figure 2). The numbers of characters on each line including spaces were extracted using a simple XSLT script. Lines that were not fully written were discarded. This was divided by the width of the text block (135 millimetres). An average was calculated for each folio or – in this particular case – chapter. It turns out that the average width of the characters in the first few folios is much lower (around 2.3 millimetres) than the average of the entire manuscript (2.5 millimetres). Around chapter 125, coinciding with the transition to

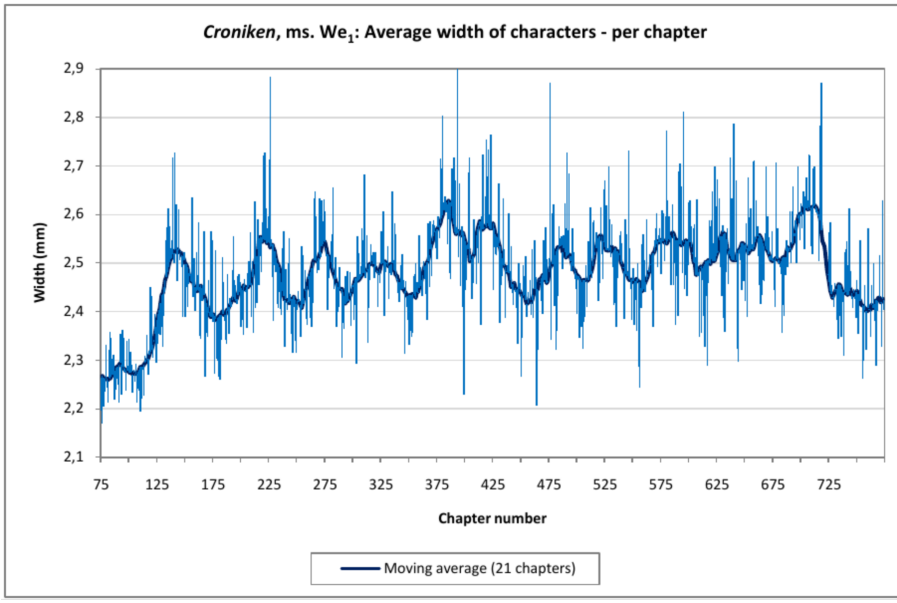

Figure 2. *Croniken*, ms. We1: Average width of characters, calculated per chapter. The table of content (c.1-74) is excluded.

the second quire, the width of the characters quickly increases to a size more in line with the rest of the manuscript. What is striking, is that the width of the characters increases for most of the manuscript. Apparently, Hendrik van Vianen is using more and more space for his characters – 0.1 millimetre more width equals roughly seventy to eighty less characters per page. Only in the so-called Utrecht bailiwick chronicle, placed separately from the rest of the *Croniken* at the end of the manuscript, Van Vianen uses a more compact script again.

Not only the size of the characters in the first few folios differs from the rest of the manuscript. Some of the letter-forms are also aberrant. This is especially the case with the letter w. It is the only letter-form that was quantified for Hendrik van Vianen's writings, for it is such a characteristic feature of his script. Three (or four) different graphic forms of the letter w can be distinguished throughout the writings of Hendrik van Vianen (Table 1). However, only on the first two folios the 'disjointed' form is predominant, not to return elsewhere in the manuscript (Figure 5; compare also ff. 1-4 of the *Sachsenspiegel* in Figure 11). In the remainder of the first quire (ff. 9-21) two other forms subsequently fight for dominance – at first the 'closed' w, then the 'open' variant. Only from the second quire onwards, some sort of – temporary –

Name	Description	Example
‘**Disjointed**’	Two loose strokes of the pen, positioned diagonal alongside each other.	‘want’ (Figure 3)
‘**Closed**’	Two connected and inward facing, curl shaped pen strokes.	‘wert’ (Figure 3); ‘wael’, ‘wt’ (Figure 4)
‘**Open**’	Two parts that open at the top, pointing away from each other, and slightly touch at the bottom.	‘wijsen’ (Figure 4)
‘**Mixtures**’	Various intermediate forms, mixtures of the above categories.	

Table 1. Letter-forms of ‘w’ used by Hendrik van Vianen.

balance is reached between the ‘closed’ and ‘open’ form. Judging from Figure 5 it is difficult to pinpoint a sudden shift in the writing process. Rather, it seems there is a smooth transition exchanging one letter-form for the other. The last instance of a long series of ‘closed’ w’s can be found on f. 78v, followed by a few ‘mixtures’. That is just before the last quire that consists of paper dated around 1480 (ff. 81-92). The use of the ‘open’ w becomes increasingly frequent, only to become the most dominant letter-form in the second half of the *Croniken.* In the table of content (ff. 3-6) too, the ‘open’ w is by far the most frequent – if not the only – form of the letter w.

The preference for the ‘open’ w in both the table of content and the second half of the chronicle runs parallel with the choice of paper, dated around 1491, over a decade later than the paper used for the first part of the chronicle. The fact that this change of paper coincides with a change in writing strongly suggests that the manuscript was produced in several phases and – given the fact that the Vienna manuscript is an author’s copy – so was the text. A scenario in which an old stash of paper was used by Hendrik van Vianen and combined with a more recent selection of paper can almost certainly be excluded: one would not expect a change of script at the same location of the change of paper.

The results achieved with the letter-forms triggered further quantitative analysis of other scribal features. Could it be possible to pinpoint an exact transition in the production of the manuscript? Or do other features strengthen the image of an on-going, sometimes bumpy development of a scribe that is ripening his skills and preferences? In fact, there is evidence of both. Some features appear to have already

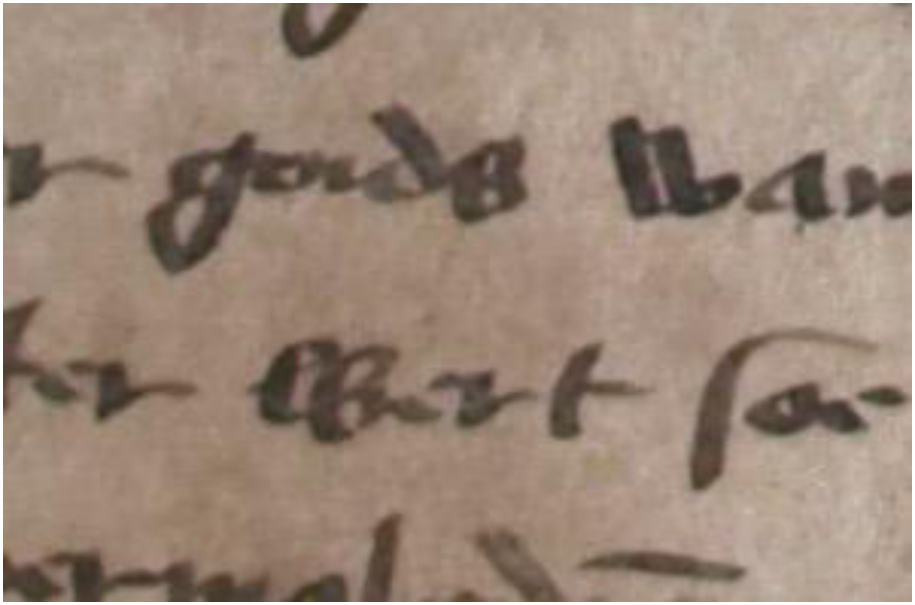

Figure 3. Croniken, ms. We1, f. 9r.

started developing during the first phase of the manuscript production, whereas others appear suddenly. If multiple changes align at a certain point, this might be a good indication that there is some sort of transition in the production process.

Truly abrupt changes can be found in the use of abbreviations in the *Croniken*. The most commonly used abbreviation by Hendrik van Vianen is the horizontal bar representing the letter n. Around twenty-five per cent of all the letters n in the *Croniken* are represented by an abbreviation (13.902 instances). However, the abbreviations are not evenly distributed as one can clearly see in Figure 6. The abbreviations in the first half of the manuscript are much more frequent than in the second. This is also the case for the –much less used– loop representing the letters -er- (126 instances; Figure 6), as well as the abbreviated form of the word 'ende' (English: and; 2.030 instances; Figure 7) or for instance the contraction that refers to the city of Jerusalem ('ihrlm'; not included here). 'Ende' is the most frequent word in the *Croniken* and in fact in most other Middle Dutch texts. The abbreviation consists of a horizontal bar above the letters en. Although the abbreviated form of the word 'ende' also shows a marked decline in popularity roughly between chapter 125 and 225 (immediately after the first quire up until an extensive part of the text that includes

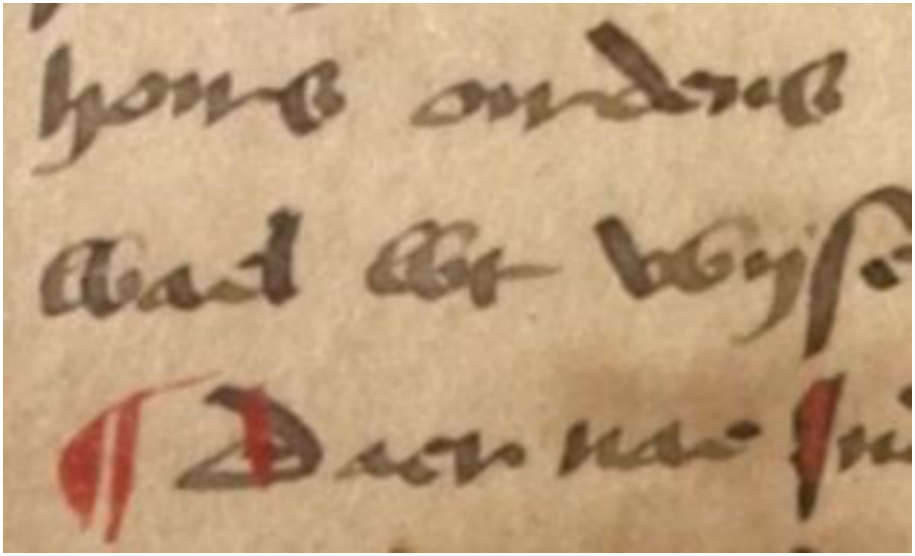

Figure 4. *Croniken*, ms. We1, f. 15v.

Latin privileges), most abbreviations show a change in their use at the same position halfway through the chronicle.

This transition is located at the beginning of the ninth quire – the last quire of paper dated around 1480 (ff. 81-92). A logical conclusion is that somewhere halfway in the ninth quire the writing process had haltered, leaving some of the remaining folios blank for further development. Perhaps we can pinpoint this moment around f. 83v: between chapter 379 and 380 the colour of the ink changes slightly, as does the overall appearance of the script. Compare for instance the larger than average width of the characters of the chapters immediately following this area of the *Croniken* (Figure 2). At the end of chapter 380 it is announced that the Livonian history will be left alone for now, to pursue the history in Prussia again. However, the chapters that follow continue to describe events in Livonia. One folio later – f. 84v (chapter 384) – a new set of sources is introduced to describe the Seventh Crusade. This also means that the chronology of the text is disturbed (chapter 383: 1258, chapter 384: 1245), which seldom occurs in the *Croniken*. Furthermore, all of a sudden stylized small cadels appear at the beginning of many of the chapters 381-393. Last but not least, the hierarchy of the initials adopts a new structure for the chronicle shortly hereafter, from chapter 416 onwards – the beginning of the tenth quire.

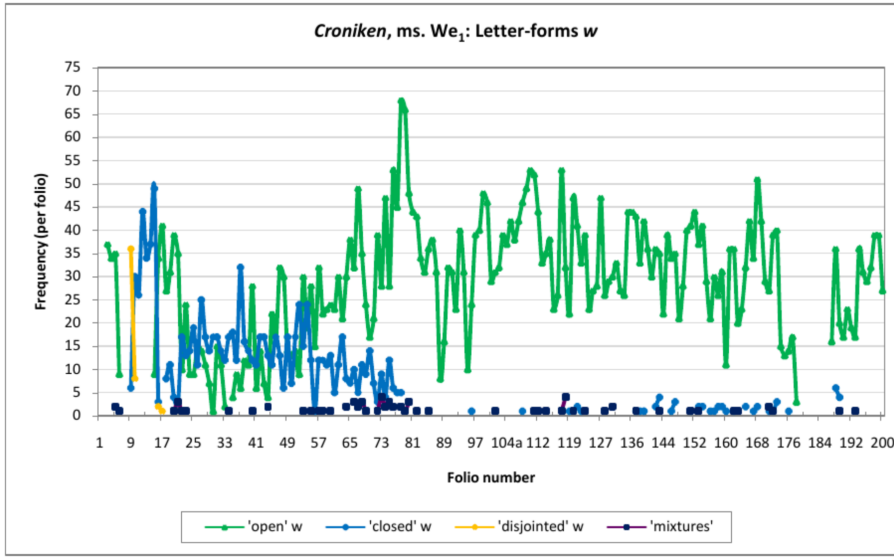

Figure 5. *Croniken*, ms. We1: Letter-forms w

The ninth quire also shows shifts in spelling preferences. However, most of the time these do not correspond completely with the abbreviations. All types of abbreviations in the *Croniken* show a marked fall in their use between chapter 379 and 380. The spelling preferences show either a gradual shift that starts somewhere in the first half of the *Croniken*, or show a more abrupt change elsewhere in or at the end of the ninth quire. A gradual shift is detected in the use of '-ge-' in comparison to '-ghe-' (Figure 8) but also in the earlier mentioned forms of the letter w (Figure 5). A marked change can be observed in the word 'meister' (English: master) in comparison to 'meyster' (or any other word that contains the diphthong '-ei-' or '-ey-') (Figure 9). Here, the shift is located at chapter 416 at the beginning of the tenth quire and the start of the description of a new Grand Master of the Teutonic Order, Poppo von Osterna (1252-1256). Another example is provided by the interchangeable forms of the long vowel '-ae-' and '-ai-' (compare 'daer' (615 instances) and 'dair' (432 instances), English: there). Parallel to the transition from chapter 398 and 400 (chapter 399 is in Latin) the use of the form '-ai-' increases from between ten to twenty per cent to a range of thirty to forty per cent. The surrounding chapters 393 to 415 all contain privileges, but starting from chapter 400 the chronology is rearranged.

It appears that almost all of these script related shifts in the ninth quire can be linked to the content of the *Croniken*. The fact that changes in both script and content

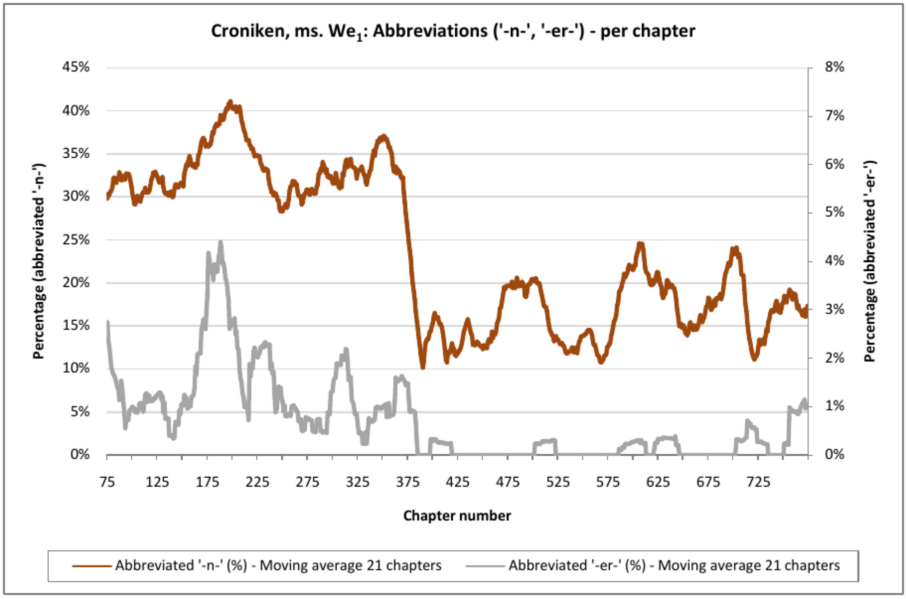

Figure 6. *Croniken*, ms. We1: Abbreviations ('n', 'er')

are aligned, points at delays in the production process of the manuscript. These delays, perhaps short at first, eventually resulted in a ten year period before new paper was added and work on the rest of the chronicle continued. Perhaps one could blame the political situation in the Low Countries for this delay. The struggle for power that started after the death of Duke Charles the Bold of Burgundy in 1477 seriously destabilized the region. In 1483 Emperor Maximilian besieged the city of Utrecht, also damaging the Utrecht commandry of the Teutonic Order situated next to the city walls. From May 1482 to September 1483, Land Commander Johan van Drongelen, the employer of Hendrik van Vianen, even had to leave the convent in Utrecht since it was not safe for him to stay. This is of importance especially because Van Drongelen is known to have enjoyed the company of historiographers in the Low Countries and in all likelihood had at least an active role in the creation of the *Croniken.* Another factor that could have triggered a phased genesis was the collection of relevant sources. The number of sources for the *Croniken* were extensive and evidence shows that they were often collected from various parts of Europe. Such an effort must have taken much time.

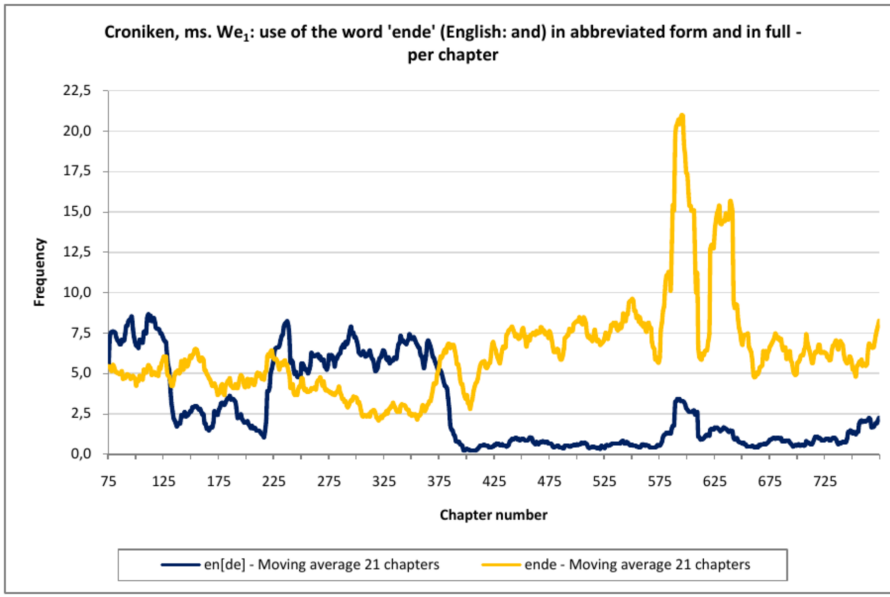

Figure 7. *Croniken*, ms. We1: Abbreviated form of 'ende' and in full (English: and).

4. Development of a scribe

Having outlined the main phases of production of the Vienna manuscript of the *Croniken*, resulting in a first tentative attempt to determine the correct context in which the text was written, it is time to turn our eyes back at the scribe, Hendrik van Vianen. In itself, the *Croniken* is a substantial piece of text written by one scribe that stretches over at least one decade. Some of the scribal features changed gradually within the timeframe of this single text, whereas other changes appeared more suddenly, usually somewhere in or directly after the ninth quire. To determine the consistency of such features during the lifetime of a scribe, we should take a look at the rest of the corpus by Van Vianen.

The first thing that becomes clear is that not all changes in scribal preferences appearing in the *Croniken* remain so during Hendrik van Vianen's career. Note for instance the spelling of the interchangeable combination '-ge-' and '-ghe-'. In the *Croniken* (Figure 8) the dominant form '-ge-' gradually loses terrain to '-ghe-' only to reach a balance in the second half of the manuscript. Throughout the *Sachsenspiegel* (around 1499-1500), too, both combinations appear in roughly equal numbers. In the land charters however, also those written long after the *Croniken*, the combination '-

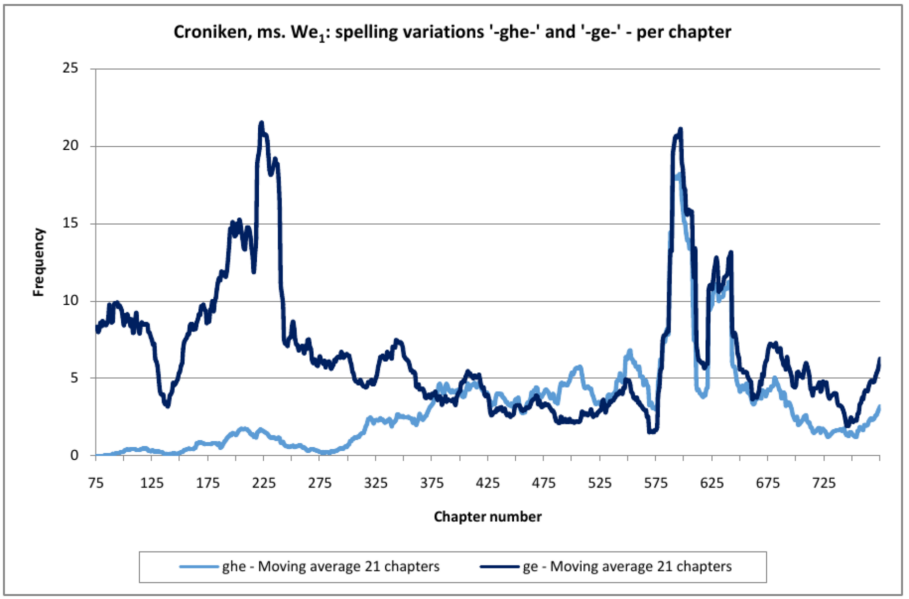

Figure 8. *Croniken*, ms. We1: spelling variations '-ghe-' and '-ge-'.

ghe-' is hardly present, whereas '-ge-' appears regularly, comparable to the beginning of the *Croniken*. For the diphthong '-ei-' or '-ey-', too, changing preferences that were revealed in the *Croniken* (Figure 9) do not correspond to other writings of Hendrik van Vianen. In both the *Sachsenspiegel* as well as all land charters the combination '-ei-' remained the dominant form. Furthermore, the degree of dominance of the non-abbreviated form of the word 'ende' (consistently around 90 per cent in the second half of the *Croniken*) is never equaled in the other writings by Van Vianen. Many of these new scribal preferences that developed during the production of the Vienna manuscript appear to have been short lived or confined to a specific piece of text.

This seems hardly good news to those who wish to use such scribal features to identify scribes. If one individual could change his writing preferences back and forth during his lifetime, one that wishes to discriminate that individual from other scribes would have to overcome serious methodological objections. Admittedly, the cases provided above rely mainly on a few well-chosen examples that are preferably frequent, bear no change in connotation ('meister' vs. 'meyster') and whose results are directly visible. Perhaps a more encompassing, computational approach could

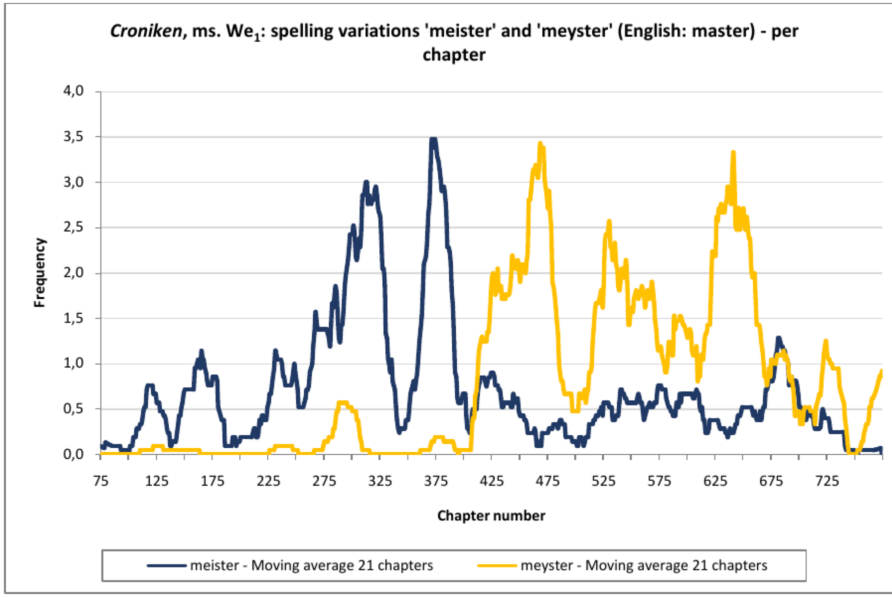

Figure 9. *Croniken*, ms. We1: spelling variations 'meister' and 'meyster' (English: master).

uncover other, less apparent patterns in the writings of Hendrik van Vianen. Some preliminary results from the field of stylometry are encouraging (Kestemont 54–55).

It turns out that the most useful feature to create a timeline of *all* the writings of Hendrik van Vianen is not spelling variation or the use of abbreviations but the letter-form w. In the *Croniken*, the 'open' w gradually becomes more and more dominant. In the second half of the manuscript, the other letter-forms are hardly existent (Figure 5). This is equally the case in the *Sachsenspiegel* written much later (Figure 11), as well as the three land charters that were written around the same time (1499-1500) (Figure 10). Almost all of the land charters that predate the second large phase of the *Croniken* (around 1491) show a clear preference for the 'closed' w that is also present in the first half of the *Croniken*. An important factor in the popularity of the 'closed' w in the charters is that, in contrast to the *littera hybrida* script of the Vienna manuscript and the *Sachsenspiegel*, a mixture was used between a *littera cursiva* (with loops, comparable to the 'closed' w) and a *littera hybrida* (without loops): the so-called *littera cursiva* (C/H).

There is only one marked exception in 1482, in which the 'open' w forms a majority. In the years immediately following, the 'open' w gradually withdraws again. It is

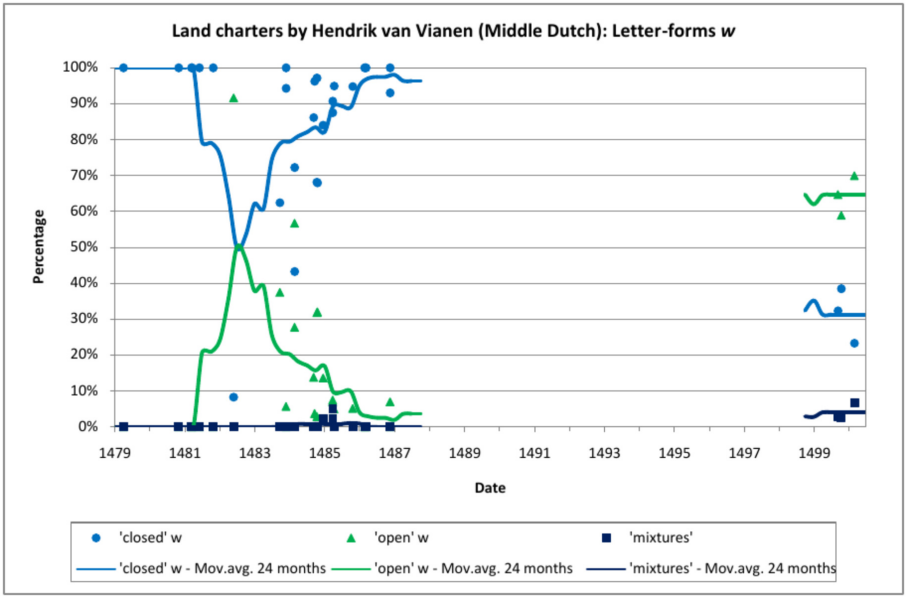

Figure 10. Land charters by Hendrik van Vianen (Middle Dutch): Letter-forms w.

tempting to attribute this practice to the aftereffects of writing the Vienna manuscript, during which process a preference for another letter-form of the w was developed (compare hypothesis 5 of McGillivray 58). After some time, the 'closed' w with its loops, a letter-form that better suits the chosen script of the land charters, would again gain ground. Only years later, in the 1490s, the 'open' letter-form would become and remain the dominant form in all of Hendrik van Vianen's writings.

One other factor strongly influences the consistency and preferences of scribes in their writing: The nature of their exemplar. This is not directly applicable to the Vienna manuscript of the *Croniken*, since it would have had no exemplar except perhaps notes or drafts. However, the characteristics of the text and manuscript of the exemplar could seriously affect the appearance of the scribe's copy (e.g. McGillivray 58). Surely a scribe will implement his own linguistic preferences in his manuscript copy, but he will also tend to take over features of the text as presented in his exemplar. Just recently, Tara Andrews and Caroline Macé have emphasized the importance of small, seemingly trivial textual variants when studying the interdependence of manuscripts and their stemma (Andrews and Macé; see also: Blake and Thaisen). Much though is still unclear in what way scribes were influenced by the appearance of

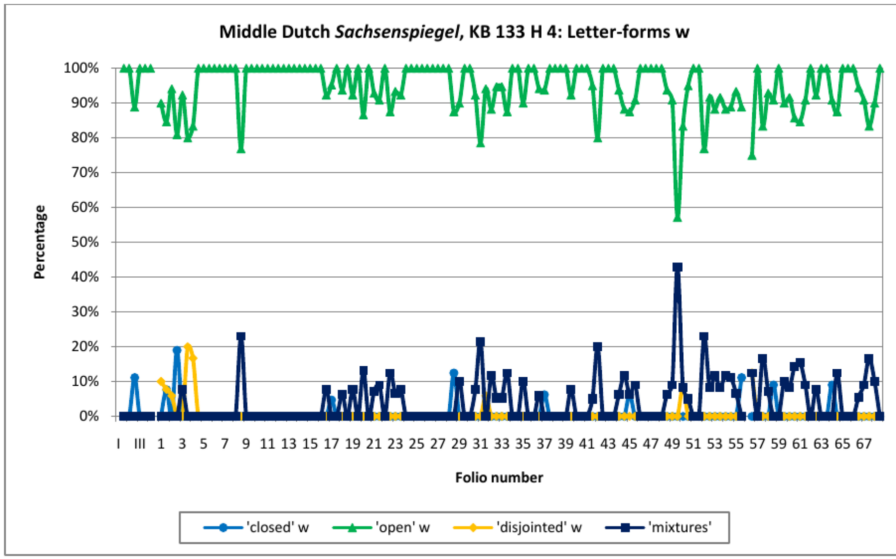

Figure 11. Middle Dutch *Sachsenspiegel*, Koninklijke Bibliotheek, ms. 133 H 4: Letter-forms w.

their exemplars, and to what extent and under which conditions they could implement their own personal writing preferences in the text.[6]

The availability of larger quantities of transcribed text is essential to study these phenomena. For the *Croniken* we not only have digital transcripts of the Vienna manuscript, but also of later Middle Dutch copies from Ghent (Ge), Utrecht (Ut1) and of an eighteenth century edition of an extant manuscript (Ma1). Again, we will turn to the two spelling variations of the word master, 'meister' and 'meyster' (Figures 12 and 13), as displayed earlier (Figure 9). What immediately becomes apparent from the graphs is that all three copies of the *Croniken* copy the orthography of the Vienna manuscript in the first half of the manuscript. However, when the spelling preferences of Hendrik van Vianen had changed dramatically, favouring the spelling 'meyster' over 'meister', most scribes continued the orthography associated with the first half of the Vienna manuscript – either directly using the manuscript Vienna as exemplar or indirectly. The form 'meister' remained the dominant form from beginning to end. Nevertheless, all three copies are also clearly influenced by the changes in spelling preference in the Vienna manuscript. The form 'meyster' was used significantly more in the second half of the text than before, especially in the Ghent manuscript that is

[6] For a brief discussion of the influence of exemplars on scribes and relevant literature see Stokes 315–316.

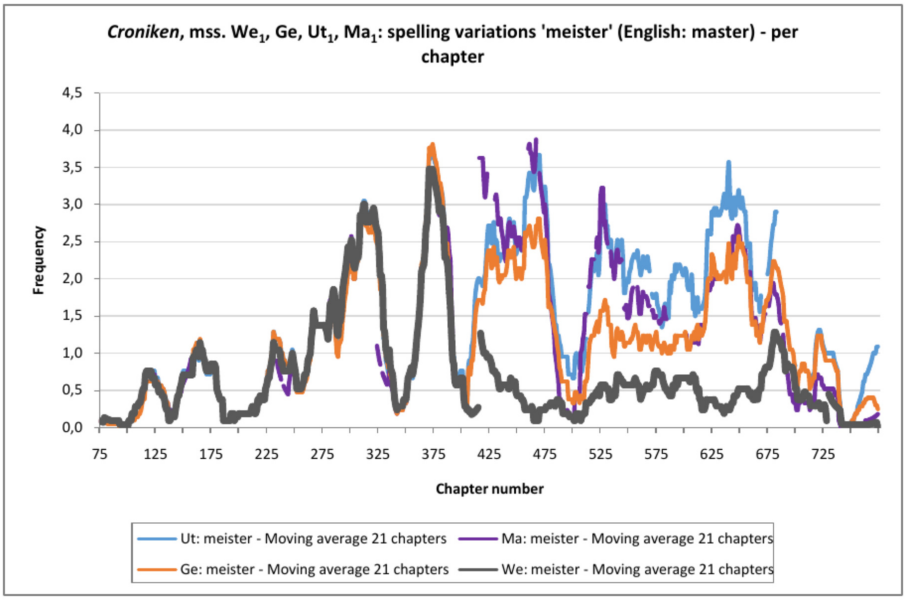

Figure 12. Variation 'meister' in three manuscripts and an edition of an extant manuscript of the *Croniken*.

closely related to the original Vienna manuscript. Did the scribes imitate the spelling preference of Hendrik van Vianen, but refused to imitate his sudden changes in the second half of the text? Or did the form 'meister' used originally by Van Vianen match the scribes' personal spelling preferences and did they continue to use their own personal preference in the rest of their manuscripts? The implications of these effects should be taken into account when drawing up linguistic and palaeographic profiles of medieval scribes.

5. Conclusion

One of the original incentives for quantifying various scribal features of Hendrik van Vianen's writings was to get a grip on the Vienna manuscript's intriguing genesis. Both the manuscript's codicological composition and watermark evidence already pointed at the possibility that the manuscript may have been written in several phases. The investigations laid out above seem to strengthen this assumption. Moreover, the quantified scribal features proved invaluable to pinpoint the exact transitions in the text where the writing process had temporarily haltered. There was a sharp

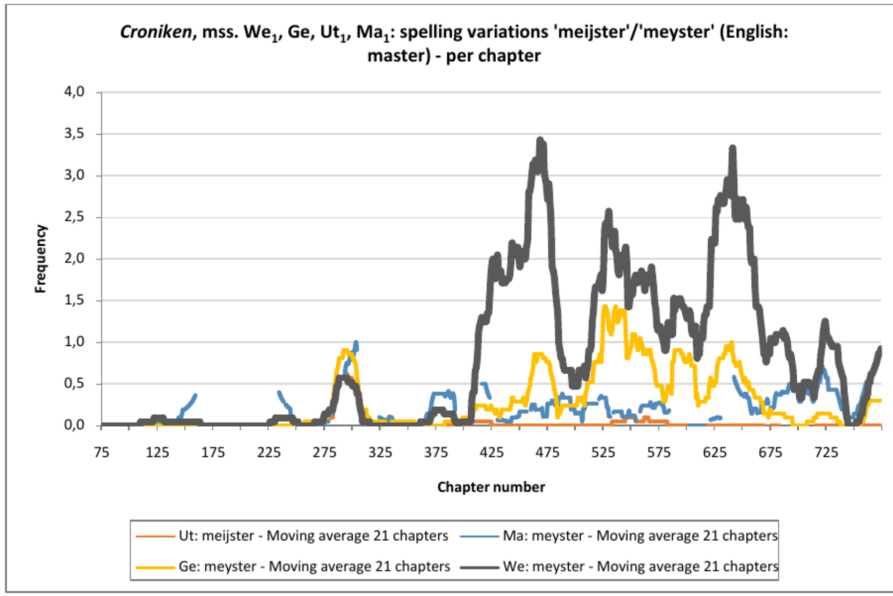

Figure 13. Spelling variation 'meyster' or 'meijster' in three manuscripts and an edition of an extant manuscript of the *Croniken*.

simultaneous decline of all abbreviations halfway through the last quire dated by watermarks around 1480, between chapter 379 and 380. However, changes in other scribal features such as variations in spelling did not correspond to this marked drop. In fact, several changes in spelling that we have examined occurred at various points in the manuscript, sometimes gradually, sometimes more abrupt. However, most abrupt changes befell on different points in or at the end of the last quire dated 1480. As almost every shift could also be linked to changes in textual context– for instance changes of subject, sources or chronology in the text – it is almost certain that these transitions coincided with several shorter or longer periods of inactivity in the manuscript production. Since the Vienna manuscript appears to be an author's copy, these periods of inactivity also apply to the writing process of the *Croniken van der Duytscher Oirden*, also known as the *Jüngere Hochmeisterchronik*.

The results presented above do have a wider importance. Various scholars have been engaged in drawing up linguistic or scribal profiles of individual scribes from quantitative data sets, for instance in order to distinguish scribes from another. The writings of Hendrik van Vianen show that there are some serious difficulties to overcome before a unique personal marker, a *scribal fingerprint* or *stylome*, can be

constructed. Especially spelling preferences and the use of abbreviations showed remarkably little consistency over a longer period of time and spanning different types of writing. Furthermore, the influence of exemplars on a scribe's writing should never be ignored. The letter-form w appeared to be the most constructive feature to create a timeline in Hendrik van Vianen's writings. The 'open' w, without loops, became dominant later in his career even in the charters that were written in a script with looped letters. However, there is also the temporary popularity of the 'open' w earlier on in his career, perhaps following a period in which Van Vianen worked on the Vienna manuscript. Hopefully research on a much grander scale than presented above can help determine which features of Van Vianen's writings should be regarded idiosyncrasies, and what results have more universal implications. Using these data, we can improve the personal profiles of medieval scribes based on quantitative data, increasing their accuracy and avail for both scribal discrimination, dating and perhaps even localizing a hand.

Bibliography

Andrews, T.L., and C. Macé. "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 (2013) 504-521.

Aussems, J.F.A. "Christine de Pizan and the Scribal Fingerprint. A Quantitative Approach to Manuscript Studies." MA Thesis, Utrecht University. <http://igitur-archive.library.uu.nl/student-theses/2006-0908-200407/UUindex.html>.

Aussems, J.F.A., and A. Brink. "Digital Palaeography." *Kodikologie und Paläographie im Digitalen Zeitalter 1 / Codicology and Palaeography in the Digital Age 1.* Ed. M. Rehbein, P. Sahle, & T. Schaßan. Norderstedt: Books on Demand, 2009. 293–308. Print. Schriften des Instituts für Dokumentologie und Editorik 2.

Blake, N., and J. Thaisen. "Spelling's Significance for Textual Studies." *Nordic Journal of English Studies* 3.1 (2004): 93–108. Print.

Hirsch, Th. "Die Jüngere Hochmeisterchronik." *Scriptores Rerum Prussicarum. Die Geschichtsquellen Der Preussischen Vorzeit Bis Zum Untergange Der Ordensherrschaft.* Ed. Th. Hirsch, E. Strehlke, & M. Töppen. V. Leipzig, 1874. 1–172. Print.

Hofmeister, W., A. Hofmeister-Winter, and G. Thallinger. "Forschung Am Rande Des Paläographischen Zweifels: Die EDV-basierte Erfassung Individueller Schriftzüge Im Projekt DAmalS." *Kodikologie Und Paläographie Im Digitalen Zeitalter 1 / Codicology and Palaeography in the Digital Age 1.* Ed. M. Rehbein, P. Sahle, & T. Schaßan. Norderstedt: Books on Demand, 2009. 261–292. Print. Schriften des Instituts Für Dokumentologie und Editorik 2.

Kestemont, M. "What Can Stylometry Learn From Its Application to Middle Dutch Literature?" *Journal of Dutch Literature* 2.2 (2011): 46–65. Print.

Kestemont, M., and K.H. van Dalen-Oskam. "Predicting the Past: Memory Based Copyist and Author Discrimination in Medieval Epics." *Proceedings of the 21st Benelux Conference*

on Artificial Intelligence. Ed. T. Calders, K. Tuyls, & M. Pechenizkiy. Eindhoven, 2009. 121–128.

Lackner, F. *Katalog Der Streubestände in Wien Und Niederösterreich. Nichtarchivalische Mittelalterliche Handschriften Und Fragmente in Korneuburg, Mistelbach, Retz, St. Pölten, Tulln, Waidhofen an Der Thaya, Weitra, Wien, Wiener Neustadt Und Aus Privatbesitz*. I. II. Vienna: Verlag der Österreichischen Akademie der Wissenschaften, 2000. Web. 17 Oct. 2011. Veröffentlichungen Der Kommission Für Schrift- Und Buchwesen Des Mittelalters, Reihe II, Verzeichnisse Der Handschriften Österreichischer Bibliotheken 5/1.

Matthaeus, A. *Veteris Ævi Analecta, Seu Vetera Aliquot Monumenta, Quæ Hactenus Nondum Visa*. 1st ed. X. Leiden, 1710. Print.

McGillivray, M. "Statistical Analysis of Digital Paleographic Data: What Can It Tell Us?" *TEXT Technology* 14.1 (2005): 47–60. Print.

Parkes, M.B. *English Cursive Book Hands, 1250-1500*. 3rd ed. Berkeley/Los Angeles/London: University of California Press, 1980. Print.

Stapel, R.J. "Layer on Layer. 'Computational Archaeology' in 15th-century Middle Dutch Historiography." *Literary and Linguistic Computing* 28.2 (2013): 344–358. Print.

Stapel, R.J., and G. Vollmann-Profe. "Cronike van der Duytscher Oirden (Chronicle of the Teutonic Order) [Jüngere Hochmeisterchronik (Younger Chronicle of the Grand Masters)]." Ed. G. Dunphy. *Encyclopedia of the Medieval Chronicle* 2010: 328–329. Print.

Stokes, P.A. "Computer-Aided Palaeography, Present and Future." *Kodikologie Und Paläographie Im Digitalen Zeitalter 1 / Codicology and Palaeography in the Digital Age 1*. Ed. M. Rehbein, P. Sahle, & T. Schaßan. Norderstedt: Books on Demand, 2009. 309–338. Print. Schriften des Instituts Für Dokumentologie Und Editorik 2.

Stutzmann, D. "Paléographie Statistique Pour Décrire, Identifier, Dater... Normaliser Pour Coopérer et Aller Plus Loin ?" *Kodikologie Und Paläographie Im Digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2*. Ed. F. Fischer, C. Fritze, & G. Vogeler. Norderstedt : Books on Demand, 2010. 247–277. Print. Schriften des Instituts Für Dokumentologie Und Editorik 3.

Thaisen, J. "A Probabilistic Analysis of a Middle English Text." *Digitizing Medieval and Early Modern Material Culture*. Ed. B. Nelson & M. Terras. Arizona: Iter / Arizona Center for Medieval and Renaissance Studies, 2012. 171–200. Print. New Technologies in Medieval and Renaissance Studies Series 1.

—. "Overlooked Variants in the Orthography of British Library, Additional MS 35286." Ed. M.W. Driver. *Journal of the Early Book Society: for the study of manuscripts and printing history* 11 (2008): 121–143. Print.

Van Dalen-Oskam, K.H. "The Secret Life of Scribes. Exploring Fifteen Manuscripts of Jacob van Maerlant's Scolastica (1271)." *Literary and Linguistic Computing* 27.4 (2012): 355–372. Print.

Van Dalen-Oskam, K.H., and J.J. van Zundert. "Delta for Middle Dutch. Author and Copyist Distinction in Walewein." *Literary and Linguistic Computing* 22.3 (2007): 345–362. Print.

Van Halteren, H. et al. "New Machine Learning Methods Demonstrate the Existence of a Human Stylome." *Journal of Quantitative Linguistics* 12.1 (2005): 65–77. Print.

Une application iPad pour l'annotation collaborative des manuscrits médiévaux avec le protocole SharedCanvas : «Formes à toucher»

Matthieu Bonicel, Dominique Stutzmann

Résumé

Formes à toucher est une application pour iPAD développée sous la direction de l'Institut de recherche et d'histoire des textes et la Bibliothèque nationale de France et publiée sous licence GNU-GPL. Conçue en collaboration avec des ergonomes, elle démontre que la technologie n'offre pas simplement une amélioration du confort de l'usager mais modifie profondément les méthodes de travail et les questionnements de la recherche en SHS. Malgré l'accès ubiquitaire aux ressources numérisées, le rapport à l'image reste surtout un rapport d'illustration et non d'exploitation assistée par ordinateur, alors que l'analyse d'images et la prise en compte des formes est au cœur des enjeux actuels en SHS.

Les études d'usage montrent que l'application est utilisable aussi bien dans un contexte de recherche (annotations personnelles ou d'équipe) que pédagogique (paléographie, histoire de l'art). L'utilisation du modèle de données SharedCanvas, fondé sur RDF, permet d'assurer l'interopérabilité avec d'autres bibliothèques numériques et de futures applications similaires dans le domaine du Linked Open Data.

Zusammenfassung

Formes à toucher ist eine iPad-App, die unter der Leitung des Institut de Recherche et d'Histoire des Textes und der französischen Nationalbibliothek entwickelt wurde und unter einer GNU-GPL Lizenz veröffentlicht ist. Die App wurde in Zusammenarbeit mit Usability-Experten entworfen und zeigt so, dass die Technologie nicht nur einfach einen Komfortgewinn darstellt, sondern dass sie die Arbeitsmethoden und geisteswissenschaftlichen Fragestellungen grundlegend ändert.

Bisher hat die Digitalisierung der Quelldokumente nur die Arbeitsbedingungen verändert, jedoch nicht die Methoden. Trotz des allgegenwärtigen Zugangs zu digitalen Quellen bleibt der Bezug zum Bild vor allem ein illustrativer, ohne die Möglichkeiten computergestützter Auswertung zu nutzen, obwohl die Bildanalyse und die Mustererkennung im Herzen der aktuellen Herausforderungen der Geistes- und Kulturwissenschaften stehen.

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 87–104.

Nutzungsstudien haben gezeigt, das die App sowohl in einem wissenschaftlichen Umfeld (durch Annotationen) als auch in der Lehre (Paläographie, Kunstgeschichte) von Nutzen ist. Das System ermöglicht, die auf den Bildschirm gezeichnete Linienführung (als Vektorzeichnung) und ihren Bezug zum Originalbild (via Bildkoordinaten) zu speichern. Die Nutzung des RDF-basierten Datenmodells von SharedCanvas stellt die Interoperabilität mit anderen digitalen Bibliotheken und zukünftigen ähnlichen Anwendungen über Linked Open Data sicher.

Abstract

Formes à toucher is an iPad app, developed under the direction of the Institut de Recherche et d'Histoire des Textes and the Bibliothèque nationale de France and published under a GNU-GPL license. The app has been designed in collaboration with usability experts and proves that this technology does not simply enhance the user comfort, but that it changes fundamentally work methods and scholarly research aims.

Until now digitalisation of source materials only changed the conditions of work but not its methods. Despite the ubiquitous access to digital sources, the relation to the image remains primarily illustrative and does not make use of the possibilities of computer-assisted analysis although image analysis and pattern recognition are core challenges in the humanities and cultural sciences.

Usage analysis have shown that the app is valuable both for scientific use (by annotations) as well as in teaching (palaeography, art history). The system allows to store the (vectorised) lines and their relation to the original image (via image coordinates) as drawn on the touchscreen. The use of the RDF-based data model SharedCanvas ensures the interoperability with other digital libraries and similar future applications via Linked Open Data.

1. Introduction

Le projet *Formes à toucher* a permis de prototyper application pour tablette à écran tactile pour réaliser l'annotation scientifique de documents patrimoniaux et de sources historiques en tenant compte des caractéristiques graphiques et des formes. Cette application doit ouvrir un verrou technologique et permettre, dans un second temps, de formuler un projet de recherche d'analyse d'image et de formes dans le domaine des SHS.

Des axes de recherches essentiels de la recherche pour les sciences de l'érudition portent sur les caractéristiques graphiques : diplomatique et codicologie (mise en page et mise en texte, présence d'éléments de décor), héraldique et iconographie, paléographie (forme des lettres). La révolution numérique n'a pour l'instant pas

encore modifié le rapport à l'image, qui reste illustratif : les bases de données (Initiale, Mandragore, Bibale) sont impuissantes à traiter réellement de l'image et des formes qui y sont représentées : l'indexation est textuelle et linéaire, tandis que l'image dans sa bidimensionnalité reste un tout, ou, au mieux, divisé en formes géométriques (rectangle, cercle, etc.).

L'analyse d'images et la prise en compte des formes est pourtant au cœur des enjeux actuels en SHS : la TEI introduit l'élément <zone>, le projet TILE, Text-Image Linking Environment a publié sa première version en juillet 2011 [1] et trois publications récentes montrent à la fois l'effervescence dans le domaine [2–5] et témoignent d'un état expérimental, sur des objets isolés ou des corpus très limités, de sorte que les recherches, menées par des ingénieurs ne répondent que très partiellement aux questions des sciences humaines et n'en renouvellent pas les questionnements.

La situation actuelle est paradoxale : l'analyse d'image a conquis les espaces du grand public avec des applications alimentées par des données nombreuses (reconnaissance faciale de Google et Facebook) de même que les objets d'annotation, notamment sur les images (Flickr), mais reste cantonnée, pour les usages scientifiques, aux sciences de l'ingénieur avec des projets très spécifiques, sur corpus restreints et homogènes. Or les SHS pourraient profiter de ces avancées technologiques pour mieux répondre aux questions anciennes et poser de nouvelles questions. Les outils informatiques d'aujourd'hui et les humanités numériques ont en effet des enjeux cognitifs forts : il ne s'agit pas seulement de gagner du temps, mais de poser de nouvelles questions et l'ergonomie constitue un défi majeur pour permettre l'appropriation des outils et des apports heuristiques.

C'est à ce verrou technologique et cognitif que s'attaque la réalisation de *Formes à toucher*, **application d'annotation des formes sur support à écran tactile**, permettant de «dessiner» (avec ses doigts ou au moyen d'un stylet) la forme telle qu'elle apparaît sur l'écran et de sauvegarder à la fois la forme du tracé (dessin vectoriel) et son lien avec l'image d'origine (coordonnées sur l'image numérique). Ce système d'annotation doit être fluide et d'usage aisé, de façon à permettre des questionnements tenant compte de cette forme. En particulier, il met à profit les interfaces tactiles qui permettent de tracer directement sur l'image plutôt que de sélectionner une zone géométrique ou de dessiner maladroitement avec une souris. Le système sera souple et permettra soit de dessiner un contour, soit de dessiner la forme en un trait.

Une fois la forme dessinée, le système permettra d'**annoter la forme** et de **partager ses annotations**, de façon à la catégoriser selon les besoins des différents chercheurs. L'application devra aussi permettre ensuite de **visualiser les diverses formes** annotées, de façon à pouvoir valider les rapprochements faits durant l'étape d'annotation (par ex. : positions d'un personnage, motifs iconographiques ou formes des filigranes et lettres).

Le présent outil est une innovation importante pour le travail des chercheurs en SHS. Il met en avant l'ergonomie de l'outil, pour permettre d'ouvrir le verrou technologique de l'appréhension de la forme. Il se fonde sur des réalisations génériques et open source, et permet le partage des annotations et la constitution d'un corpus collectif.

2. L'ergonomie et les humanités numériques

La multiplicité des bases et la diversité des formats, les impératifs de modélisation, les divergences de granularité descriptive, les enjeux de méréologie et granularité physique font des humanités numériques un lieu incantatoire des «bonnes pratiques» et de «l'interopérabilité». Sans contester cette nécessité absolue, la place centrale de «l'outil» dans la capacité des sciences humaines dites «numériques» à poser de nouvelles questions imposent de repenser le rapport du chercheur, à la fois «early adopter» et «end user», à l'interface, à l'ergonomie et aux outils de visualisation (et de validation) des données. Il est nécessaire de replacer l'utilisateur au centre des réflexions, de remettre «l'humain» au cœur des «humanités», fussent-elles numériques. Nous soulignerons ainsi l'importance de l'ergonomie pour les applications, interfaces de saisie et de consultation des données, la visualisation des processus d'analyses comme des résultats, car l'objectif est l'efficacité des hommes et femmes dont l'intelligence fait avancer la recherche. Pour ce faire, il faudra lutter contre deux religions assez répandues : celle des données et celle de l'outil.

Le rôle du chercheur est d'observer, décrypter, interroger ces sources brutes, puis de constituer ce qu'il appellera par transitivité «la source», mais qui est un objet constitué en tant que source. Pour ce faire, le chercheur encode les données observées, s'il utilise un système capable d'ajouter de la valeur sémantique et de l'interprétation au cœur du document tenu pour source (e.g. TEI) ; ou bien il décode les données observées s'il les range dans une base de données. Il faut ensuite traiter les données, ou, autrement dit, poser la question aux données. Or, pour traiter les données, des intermédiaires sont désormais nécessaires : développeurs, outils, écrans, interfaces, claviers. A l'heure actuelle, le chercheur est en périphérie et l'outil au centre. Les «humanités numériques», même quand ce mot ne sert pas de concept marketing, tendent souvent à placer l'outil comme objet au cœur du travail, plutôt que de se focaliser sur ses objectifs. C'est la religion de l'outil.

A notre sens, le rôle du numérique, dans les «humanités numériques» est au moins quadruple, et mieux le comprendre permettra de faire progresser les humanités numériques, dans leur fonctionnement et leur acceptation par la communauté. **I**l y a au moins pour le chercheur quatre étapes différentes : constitution de la source ; validation des données ; visualisation des données traitées ; interaction avec les données traitées. Toutes ces opérations sont effectuées par ordinateur, mais ce n'est pas lui qui est au centre du processus. Ce sont des mécanismes et des processus distincts qui prennent place au même endroit.

C'est ici qu'intervient l'ergonomie, science qui étudie la relation entre l'homme et son milieu de travail, et désigne par extension l'adéquation de l'outil à l'homme et à ses objectifs, en vue d'une utilisation facile, efficace, sûre et adaptée à tous.

Le deuxième mot est celui «d'interface» ou «IHM, interface homme machine». Les développements qui lui sont accordés sont généralement consacrés à l'interface publique, et dans les humanités numériques, à l'interface des éditions et au cas particulier de la mise en forme et à la visualisation des données traitées lors de la publication. Presque jamais on ne parle des interfaces des outils des chercheurs.

Cet enjeu est, selon nous, sous-évalué actuellement dans les humanités numériques. Or, il est important de s'y consacrer, car l'ergonomie accroît l'efficacité. L'idée que la possibilité de recherche est un plus qui justifie tous les sacrifices doit être remise en cause. Il y a certes un avantage évident dans la pérennisation de l'information et son partage, dans la puissance d'analyse, dans la fouille de données aux résultats objectivisables. Mais le gain heuristique n'est réel que si ces méthodes n'empêchent pas les chercheurs de voir et exploiter toutes les sources qu'ils doivent exploiter, sinon il y a un recul caché de la force de la recherche.

Une meilleure ergonomie signifie une meilleure recherche :

- elle **réduit le temps** nécessaire à l'analyse des sources
- elle permet ainsi de constituer une **masse critique** de données
- elle permet d'**affiner** l'analyse ou l'encodage
- elle évite la surabondance d'information et **assure la qualité des données**

Pourtant, quand un budget est amputé, c'est toujours l'ergonomie qui pâtit en premier. Deux préjugés en sont cause, partagés par les chercheurs et développeurs : l'ergonomie et le graphisme, c'est pour faire joli ; les chercheurs sont des gens sérieux, il faut s'adresser à leur raison et non leur offrir des logiciels de jeux.

Du côté des chercheurs, il faut lutter contre une attitude masochiste, qui fait accepter des interfaces peu pratiques et développer des stratégies de contournement : soit en raison de la passion au service de la science, qui permet de passer outre les difficultés quotidiennes, techniques ou administratives, soit parce que les chercheurs sont, par nature, à la pointe : ils sont donc par vocation des «earlyadopters» et trouvent normal d'essuyer les plâtres et d'être des beta-testeurs ; ou bien encore par la croyance erronée

que les questions spécifiques nécessitent des outils spécifiques ou des modèles de données distincts.

Ce besoin ressenti de développement spécifique trouve un écho favorable chez des développeurs et a pour conséquence la multiplication de prototypes et d'applications dédiées, toujours en phase de test et jamais finalisés techniquement. Au final, les chercheurs, victimes de leur *habitus* propre et des circonstances, travaillent souvent avec des outils qui semblent des obstacles à la bonne recherche.

La majorité des débats en humanités numériques s'intéressent à la façon dont la première opération doit être effectuée pour obtenir des données «interopérables» et «pérennes». Ce sont des enjeux majeurs, et nous en parlerons ci-dessous. Pourtant, il faut insister ici sur un point important : le modèle de données n'a pas à se matérialiser dans l'interface, ni lors de la saisie, ni lors de l'interrogation, ni lors de la consultation. Les données sont souples et malléables et doivent être affichées ou exploitées en tant que de besoin.

En effet, la perte de temps n'est pas neutre scientifiquement. L'outil ergonomique permet la bonne recherche. En bonne logique, l'outil non ergonomique empêche la bonne recherche. Non seulement il est lent, nuit à la granularité descriptive, réduit la masse de données produites et trouble la qualité des données. Mais en outre il crée des problèmes humains. Le cerveau est sous-stimulé, frustré par la répétition de gestes mécaniques visant à expliciter l'évidence, c'est-à-dire gaspiller des efforts cognitifs, alors même que l'objectif final est précisément l'intelligence fine des processus culturels, humains et sociaux, qui requiert un esprit entier. Cette frustration intellectuelle naît du reste de la perception confuse que ce travail pénible est inutile, et elle engendre, à son tour, des problèmes de gestion d'équipe.

Il faut se garder d'adorer l'ordinateur comme le veau d'or et rappeler que le premier ouvrage de paléographie statistique date d'avant l'ordinateur, c'est l'essai de W. Meyer sur l'écriture gothique en 1897[4]. Il est donc possible de faire sans ; il faut mesurer la plus-value.

Le futur des humanités numériques, c'est l'ergonomie et la création d'une "expérience utilisateur" fluide et limpide, ou, pour reprendre les termes de la définition, confortable, efficace, sûre et de prise en main immédiate. Les questions sont et doivent être complexes, les calculs de même, mais pas les outils. C'est dans le dialogue avec les chercheurs et en remettant ceux-ci au centre du processus de la recherche que l'on pourra créer des outils paramétrables et adaptables aux questions de chaque équipe. Les conditions de ce dialogue peuvent être améliorées par des formations transdisciplinaires, notamment pour que les chercheurs SHS identifient ce qui est faisable (convertir un ficher EAD en base de données plutôt que de le recopier) et à qui s'adresser ; mais aussi et surtout en accroissant la formation des chercheurs non au développement, mais d'une part aux outils et logiciels, et d'autre part aussi au graphisme pour les décomplexer face à leurs besoins, puis en délimitant bien les

rôles, en acceptant que le chercheur exige des développements adaptés à ses besoins scientifiques, en reconnaissant que l'outil doit être ergonomique. En revanche, le chercheur doit s'immiscer le moins possible dans le développement de l'outil ; le développeur doit oublier qu'il a affaire à un public à la fois captif et passionné, qui passera au-dessus de tous les obstacles formels, mais n'en reste pas moins constitué d'êtres humains.

Un autre axe majeur est le traitement et la visualisation des données selon leur nature. Une information graphique doit être visualisée de façon graphique, une information temporelle de façon temporelle, car c'est ainsi que l'on tire le mieux parti du fonctionnement du cerveau humain. Ce qui est difficile pour l'ordinateur peut être aisé pour le cerveau humain (et inversement), à condition qu'on lui donne les moyens de valider l'information avec ses capacités les plus immédiates, éventuellement en cours de traitement. Ici aussi, la modélisation complexe et fine ne doit pas aboutir à des interfaces complexes. Non seulement l'interface n'a pas à être le décalque du modèle d'information car le cerveau humain est capable de gérer l'implicite, ce que l'ordinateur ne sait pas faire, mais en outre le système cognitif humain réagit plus vite à une organisation spatiale de l'espace, aux formes et aux couleurs qu'aux messages textuels. Cette propriété, largement utilisée dans les interfaces grand public, est en revanche largement ignorée dans les interfaces pour les chercheurs. Ainsi les informations de granularité physique et de méréologie sont très faciles à représenter graphiquement et très pénibles à gérer dans un affichage textuel. Pourtant c'est la solution majoritairement choisies par les bases que nous connaissons.

L'objectif est que le travail des chercheurs soit aussi simple que s'ils écrivaient avec un stylo et du papier et que la signification des données, problème implicite, reste visuellement implicite. C'est dans cette optique que nous avons piloté le projet *Formes à toucher*, qui a vu le développement d'un prototype d'application d'annotation des images sur écran tactile, aussi simple et rapide d'emploi que le stylo et le papier, en ajoutant des fonctions qui sont les avantages du numérique, par exemple la baguette magique. Pour l'analyse des variantes allographétiques, le chercheur, linguiste ou paléographe, rêve d'un système où sa lecture, son geste puisse être enregistré avec sa signification réelle. S'il trace des signes qui ressemblent à un x, un p un i et un tilde, il sait qu'il a tracé le mot *Christi*. A l'heure actuelle, cette interprétation pourrait être complètement automatisée et on aboutirait en un processus unique où le numérique demanderait confirmation seulement pour les cas ambigus à une transcription facsimilaire et à une édition normalisée, éventuellement pré-indexée par reconnaissance des entités nommées.

C'est ainsi qu'on libérera les forces de l'esprit pour de nouvelles questions.

3. Le développement de *Formes à toucher*

3.1. GIS Sourcem et financement

Le développement de Formes à toucher a été rendu possible par un financement du Groupement d'Intérêt scientifique Sourcem, portant sur les sources de la culture européenne et méditerranéenne, qui a lancé en mai 2012 un appel destiné à soutenir des projets portés par ses membres et dans le but de préparer des candidatures «de plus grande envergure en réponse à un appel d'offre (ANR, ERC, etc.)». L'enveloppe financière allouée étant nettement inférieure au coût réel des travaux à réaliser, même pour un prototype, il a fallu trouver un prestataire acceptant d'investir un nombre d'heures conséquent en recherche et développement. C'est finalement l'agence digitale Is&a bloom qui a été retenue pour sa démarche avant tout basée sur l'ergonomie d'usage de l'application, qui était un enjeu majeur du projet.

3.2. Fonctionnalités

Huit manuscrits ont été retenus pour la version 1 de l'application, quatre issus de Gallica et conservés au département des Manuscrits de la Bibliothèque nationale de France et quatre issus de la Bibliothèque virtuelle des Manuscrits Médiévaux et conservés à la Bibliothèque municipale de Toulouse. Ces huit documents permettent de refléter plusieurs usages potentiels d'annotation : ouvrages profanes et religieux, textuels et illustrés, en latin, français et hébreu. Les manuscrits de la BnF ont été numérisés à raison d'une page par image numérique alors que ceux de la BM de Toulouse l'ont été à raison de deux pages par image. Enfin, seule une sélection de feuillets sont présents dans chaque manuscrit dans la mesure où cette version 1 inclut toutes les images au sein même de l'application (la fonction de téléchargement devant être ajoutée dans une prochaine version).

La navigation se veut simple. La vue d'accueil présente les huit manuscrits, un toucher (ou «tap», dans le jargon des développeurs) sur l'un d'entre eux permettant de faire défiler l'ensemble de ses images. Une autre vue affiche les manuscrits qui comportent des annotations de l'utilisateur et une troisième n'affiche que les pages annotées (sans les autres pages de l'ouvrage). La quatrième vue intitulée «mes notes» sera complétée dans une version ultérieure avec la liste directe des annotations.

Une fois dans la page qui l'intéresse, l'utilisateur dispose d'une fonction de zoom fluide par pincement de l'écran (comme la plupart des opérations de ce type sur tablette). Un bouton unique en bas à droite permet d'accéder à l'ensemble des fonctions d'annotation.

Chaque note comprend trois types d'éléments : un titre, un corps de texte et une ou plusieurs formes, afin de pouvoir annoter aussi bien une lettre, une zone ou un groupe d'éléments distincts.

Les formes disponibles sont variées (rectangle, ellipse, sections de droites, tracé libre) et facilement modifiables ou déplaçables par double tap. Une fonction de baguette magique avec réglable de sensibilité a également été incluse. Elle reste cependant à perfectionner dans les cas assez fréquents où le contraste n'est pas très élevé.

Afin de ne pas alourdir la consultation, les annotations sont repérées sur la page par une étiquette numérotée (la suppression d'une note ne modifie pas les numéros, la rupture de séquence éventuelle étant ainsi assumée au dépend de la confusion que pourrait générer une renumérotation). On accède au tracé des annotations et à leur contenu par tap sur une étiquette ou par défilement de l'une à l'autre au moyen des flèches situées en bas de l'écran. Enfin, en haut à droite, un bouton «Notes» permet d'afficher directement le titre et le contenu textuel de toutes les annotations et de se positionner directement sur celle que l'on choisit.

Dans l'ensemble du processus de développement, un soin tout particulier a été apporté à la gestion du niveau de zoom ainsi qu'à la fluidité des déplacements. Il était en effet essentiel que l'utilisateur soit positionné au bon niveau dans l'image lorsqu'il sélectionne une annotation pour disposer à la fois d'un niveau de détail satisfaisant et du contexte entourant le secteur annoté. Un cadrage trop serré peut en effet dérouter l'utilisateur qui ne perçoit pas clairement l'endroit de la page où il se situe, faute de repères.

3.3. Études d'usage

En amont de la publication officielle de l'application sur l'AppStore, un certain nombre de personnes ont été sollicitées pour aider à définir les fonctionnalités et tester l'application en avant-première (bêta-testeurs). Le panel choisi est à la fois vaste et tente de refléter la diversité des utilisateurs de produits numériques autour des documents patrimoniaux : bibliothécaires spécialisés, développeurs informatiques travaillant sur les bibliothèques numériques, paléographes, historiens de l'art, diplomatistes. L'intérêt de chacun pour les nouvelles fonctionnalités apportées par l'application était divers.

Les personnes impliquées dans le développement des standards SharedCanvas et IIIF ont par exemple souligné l'intérêt de la génération d'un fichier XML incorporant les annotations et leur localisation car il permet de réexploiter ces données dans d'autres contextes applicatifs exploitant les mêmes normes. Cet export par e-mail devrait bien sûr dans l'idéal être exposé d'une manière plus conforme aux standards actuels d'échanges de données ouvertes mais la fonctionnalité actuellement proposée dans l'application permet d'en présenter le principe général.

D'autres testeurs au contraire ont fait de nombreux commentaires sur l'ergonomie de navigation et de saisie des formes graphiques, qui est un des points centraux de l'application. L'étude d'éléments précis de la mise en page ou de l'iconographie des

manuscrits nécessite par exemple le dessin de formes nombreuses et complexes qui se superposent les unes aux autres et peuvent parfois s'avérer délicats à visualiser dans leur ensemble ou à reprendre dans un deuxième temps pour les modifier ou les améliorer.

Une priorité a été de tenir compte des usages réels et de partir des besoins des utilisateurs tels qu'ils sont exprimés dans leur pratique quotidienne. Ainsi, par exemple, nous avons fourni à plusieurs chercheurs des tirages sur papier des images numériques. Cela peut sembler paradoxal, mais nous avons constaté que c'est réellement ainsi que s'effectue l'analyse d'un volume dans certains contextes. Ensuite, nous avons récupéré les feuilles annotées et identifié les différentes strates et les différentes significations des annotations : métadonnées générales concernant le manuscrit sur le premier feuillet (report de la cote, datation, localisation, identification du texte), avec expression de l'hésitation et du doute ; description d'initiales ornées, identifications de sujets iconographiques, mais aussi commentaires personnels liés intellectuellement, sans que les liens n'apparaissent. Dans l'un des cas, l'on constate trois annotations qui concernent l'origine du volume («France (Sud ?)» ; «prolongements / font assez italiens» ; «encre violette» qui est une caractéristique méridionale) ; sur les feuillets ultérieurs, les annotateurs se concentrent sur le relevé des points nouveaux et saillants. Dans un deuxième cas, le premier feuillet porte successivement les indications «Origine ? Date ?» puis une flèche vers «[astérisque] 1e/2 14e s.» ; «encre (pâle) claire» ; un peu plus loin «cf. initiales champies ; fond rouge + bleu = écartelé – et encoches dans aire => cf. 'Nord' ?» ; sur la page d'en face, on retrouve l'astérisque suivi de «v. 1333/1340» qui résout les questions posées à gauche, manifestement après vérifications intervenues après la consultation des autres feuillets ; pour le même manuscrit, on trouve de nombreux renvois vers d'autres feuillets («cf. f. 8v» etc.) ou commentaires («pas même enlumineur que f. 8v, cf. visages»). Un troisième chercheur n'a annoté que par des formes géométriques, encadrant des mots charnières dans le texte («Item»). Les pratiques de correction ou précision des informations que doit fournir la notice descriptive ont été identifiées, mais n'ont pas pu être réglées spécifiquement dans le cadre de cette application.

Au cours du développement, et en raison de l'importance accordée à l'ergonomie, nous avons observé des utilisateurs pendant la prise en main de l'outil. Confirmant l'exigence d'intuitivité, aucun des testeurs n'a pris la peine de lire les indications ajoutées pour expliquer comment activer certaines fonctionnalités. Ces tests effectués collectivement ou individuellement ont permis de lever en cours de développements des difficultés mineures, qui ne gênaient pas les développeurs et commanditaires, déjà trop experts. Les testeurs s'adressent selon les cas directement à l'un des responsables de l'application («comment tu reviens ?» ; «on ne peut pas avoir un stylet plus précis ?» ; «comment t'enlèves le chemin de fer ?»), soit mettent en lumière par leurs difficultés ou leurs agacements ce qui avait échappé («et comment je change

la couleur ?» révélant l'habitude d'annoter avec différents crayons ; «ah, il ne bouge pas ce crétin !» pour dénoncer l'absence d'accompagnement de mouvement si l'on va vers l'extérieur de l'image et pointant, inconsciemment la limite d'espace imposée par la tablette par rapport à une reproduction sur papier en grand format ou à l'échelle). Dans tous les cas, la conclusion était positive («c'est très bien, tout ça !» ; «c'est étonnant»).

Les différents types de commentaires correspondent aux différents comportements heuristiques. Le développement de l'application a cherché à répondre à tous les besoins et à contraindre le moins possible les chercheurs. Dans certains cas pourtant, il a été tenu compte des perspectives de collaboration et de la nécessité d'expliciter les habitudes dans un contexte interpersonnel ; nous avons en particulier renoncé à créer des formes de couleurs différentes : d'une part l'usage des couleurs n'est pas normalisé ; d'autre part car cet usage imposait un choix systématique de la couleur, donc une étape supplémentaire lors de l'annotation. En compensation, l'annotation comprend deux niveaux et peut accueillir une description de typologie.

3.4. Publication

L'ensemble du projet a été réalisé sous la licence de logiciel libre GPL v3.0 et le code source est accessible sur la plate-forme Github[1] où les développements peuvent être commentés et repris par n'importe quel utilisateur. Les applications libres ayant été réutilisées dans *Formes à toucher*, notamment pour la visualisation, la baguette magique et la transformation XML pour l'export sont également mentionnées. La diffusion de l'application sur l'AppStore d'Apple pour permettre son téléchargement par n'importe quel utilisateur a finalement été relativement rapide. La relecture du code par les services d'Apple donnant parfois lieu à des demandes de reprises importante, c'était un risque majeur pour la fin du projet, les ressources allouées pour cette étape alors très faibles. Les remarques du constructeur de l'iPad ont essentiellement porté sur des questions d'identification des utilisateurs et de l'utilisation de leur adresse mail. Il a finalement été décidé de n'utiliser que la fonction mail du client embarqué de l'iPad pour la fonction d'export.

4. L'avenir

4.1. SharedCanvas

L'application a été développée pour permettre un export des données d'annotation en utilisant la norme SharedCanvas, développée par un groupe d'experts internationaux pilotée par l'Université Stanford. Basée sur RDF et OpenAnnotation, cette norme

[1] <https://github.com/ivato/OAProto> (consulté le 6 mai 2014).

permet de documenter l'ensemble des informations disponibles sur un manuscrit numérisé, d'une simple liste d'images numériques jusqu'à une annotation complexe. Notre but n'est pas ici de rappeler dans les détails le fonctionnement de la norme[2], mais d'en expliquer l'application faite dans le domaine de Formes à toucher. Il faut également rappeler que nous sommes ici dans le cadre d'un prototype, vu qu'il n'existe pas encore de bibliothèque numérique utilisant la norme en production, mais les premières implémentations sont en cours dans plusieurs établissements et au sein de l'équipement d'excellence Biblissima[3].

SharedCanvas repose sur un principe simple mais essentiel : tout manuscrit numérisé est associé à un *manifeste*, un fichier librement accessible et reprenant de façon normalisée l'ensemble des informations dont dispose à son sujet : liste des images et de leur foliotation, identifiant de la ou des notices du manuscrit, liens vers les bases de données, les expositions virtuelles… Le manifeste de base est donc publié en même temps que l'ouvrage numérisé par l'institution de conservation qui en est responsable. Par la suite, toute personne qui souhaite réaliser un travail scientifique (édition électronique, annotation, transcription…) à partir du manuscrit numérisé a la possibilité de publier son propre manifeste donnant la description normalisée des travaux réalisé, et faisant référence au manifeste d'origine (celui de la bibliothèque numérique où se trouvent les images ayant servi de base au travail). Il est ainsi possible, par un jeu de renvois, de connaître la totalité des travaux entrepris sur un manuscrit, quelle que soit la structure qui en est porteuse.

Formes à toucher a été pensé dès le départ pour être compatible avec le protocole SharedCanvas. Comme aucun entrepôt à ce stade n'offrait de manifestes pour les manuscrits que nous avons sélectionnés dans la première version, nous avons implémenté à titre d'exemple une fonctionnalité permettant à l'utilisateur d'envoyer par mail (en utilisant le client mail standard de sa tablette) un fichier XML exprimé en RDF et comportant l'ensemble des formes tracées et les annotations associées. Les coordonnées vectorielles sont exprimées au format SVG, l'un des plus couramment utilisés, pour pouvoir être générées à nouveau dans un autre environnement à partir de la même image de manuscrit.

L'exemple ci-dessous détaille une annotation type générée par *Formes à toucher* en respectant ce principe :

---

[2] Matthieu Bonicel, «Hypertexte et manuscrits. Le défi de l'interopérabilité» in *Revue de la Bibliothèque nationale de France*, 2012, n° 42, p. 22-28.

[3] Démarré pour une durée de sept ans en 2012, le projet Biblissima vise entre autres à mettre en place des infrastructures d'interopérabilité entre les bibliothèques numériques et les outils de la recherche dans le domaine du patrimoine écrit au Moyen Âge et à la Renaissance : <http://www.biblissima-condorcet.fr>. Un visualiseur de test exploitant les normes SharedCanvas et IIIF est accessible à l'adresse <http://demos.biblissima-condorcet.fr/mirador/> (consulté le 6 mai 2014).

```
<oa:Annotation rdf:about="url:uuid:UUID">
 Cible de l'annotation (image du manuscrit)
 <oa:hasTarget>
  <oa:SpecificRessource rdf:about="urn:uuid:UUID">
   <oa:hasSource>
    <sc:Canvas rdf:about="urn:uuid:5ab19cce-7e65-4618-8ae7-4b12e3f62c7d">
     Identification de l'image
     <rdfs:label>Toulouse,BM,315556101_MS0015_0028_N2.jpg</rdfs:label>
     Largeur de l'image
     <exif:width rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
      5440
     </exif:width>
     Hauteur de l'image
     <exif:height rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
      4080
     </exif:height>
    </sc:Canvas>
    A l'intérieur de l'image, forme tracée par l'utilisateur
    <oa:hasSelector>
     <oa:SvgSelector rdf:about="urn:uuid:f4ceb5d8-f7f1-43e7-9002-faeff479ab0e">
      <dc:format>image/svg+xml</dc:format>
      Standard utilisé (SVG), position, forme et taille du tracé.
      <cnt:chars>
       <![CDATA[<svg xmlns="http://www.w3.org/2000/svg" version="1.1">
        <ellipse cx="3190" cy="1842" rx="215" ry="233"
         style="fill:red;stroke:none;opacity:0.2" />
       </svg>]]>
      </cnt:chars>
     </oa:SvgSelector>
    </oa:hasSelector>
   </oa:hasSource>
  </oa:SpecificRessource>
 </oa:hasTarget>
 Titre général de l'annotation
 <rdfs:label>Salomon enseignant</rdfs:label>
 <oa:hasBody>
  Contenu textuel de l'annotation
  <cnt:ContentAsText rdfs:about="urn:uuid:UUID">
   <cnt:chars>
    XIVe èsicle (èpremire émoiti).
    Initiale O du livre de l'éecclsiastique.
   </cnt:chars>
  </cnt:ContentAsText>
 </oa:hasBody>
 Identification de l'auteur de l'annotation
 <oa:annotatedBy>
  <foaf:Person rdf:about="urn:uuid:UUID">
   <foaf:mbox rdf:ressource="mailto:matthieu.bonicel@bnf.fr" />
   <foaf:name>Matthieu Bonicel</foaf:name>
  </foaf:Person>
```

```
</oa:annotatedBy>
Horodatage de l'annotation
<oa:annotatedAt>2013—07—01T11:44:04Z</oa:annotatedAt>
</oa:Annotation>
```

Dans un second temps, il est évidemment prévu de pouvoir exposer les manifestes générés à partir de l'application *Formes à toucher* directement sur un serveur afin de mettre à jour automatiquement les manifestes d'origine des manuscrits numérisés. Mais cela nécessitera une mise en place généralisée de SharedCanvas, qui est en cours de réalisation notamment dans le cadre de l'equipex Biblissima en ce qui concerne les principaux entrepôts de manuscrits numérisés français.

Afin de pouvoir mettre en relation les manifestes, les images numériques et leurs contenus associés (annotations, transcriptions…), un ensemble applicatif a été développé en parallèle de SharedCanvas, nommé International Image Interoperability Framework, ou «IIIF». Disponible sous la forme d'une API à implémenter au sein des systèmes gérant les bibliothèques numériques, IIIF permet par le biais d'une syntaxe d'URL normalisée d'interroger depuis un visualiseur tiers des images hébergées dans un ou plusieurs entrepôts distants compatibles.

En s'appuyant à la fois sur SharedCanvas et IIIF, une version ultérieure de l'application *Formes à toucher* permettrait donc de visualiseur sur l'écran de la tablette des images provenant de n'importe quelle bibliothèque numérique disposant d'un serveur d'images interopérable, et d'exposer (soit un serveur propre à l'application, soit le serveur d'une institution partenaire) le résultat des annotations réalisées par l'utilisateur.

4.2. Collaboration : l'interopérabilité des chercheurs (en local)

La mise en place d'une version de l'application totalement connectée (contrairement au prototype actuel qui se contente d'exporter des données mais ne peut pas encore récupérer des images ou données venant de l'extérieur) permettrait aux différents utilisateurs de l'application de partager en temps réel les données qu'ils produisent à partir des images.

On pourrait imaginer de créer plusieurs couches d'affichage permettant d'afficher sur un feuillet de manuscrit, en plus de ses propres annotations, celles de son équipe de recherche, ou d'un collègue qui aurait souhaité les partager. Un perfectionnement du module d'identification actuel permettrait également de mettre en place une gestion des droits qui faciliterait le choix des utilisateurs avec lesquels on souhaite échanger des informations sur des travaux en cours qui n'ont pas nécessairement vocation à être connus de tous. *Formes à toucher* deviendrait ainsi une brique supplémentaire dans l'environnement numérique du chercheur qui travaille à partir des manuscrits et lui offrirait une solution ergonomique et portable pour saisir et partager des annotations.

Afin de favoriser le travail en équipe dans le cadre de projets spécifiques, il serait souhaitable également de pouvoir charger des images supplémentaires, qui ne soient pas obligatoirement issues des bibliothèques numériques, mais qui viennent enrichir le corpus. Cette fonctionnalité a été étudiée en cours de développement, mais est apparue difficile à maîtriser. En effet, sur iPad, l'import le plus immédiat est réalisé à partir de la pellicule. Or si une équipe de recherche veut charger un manuscrit entier, il faut décider comment affecter des métadonnées et une structure à l'ensemble des images : une par une ? en important un dossier ? comment renseigner les métadonnées.

4.3. Interopérabilité des bibliothèques numériques

Formes à toucher est un exemple parmi d'autre de l'usage qui sera à terme possible dans le monde des bibliothèques numériques patrimoniales. Si la masse et la qualité des interfaces ont considérablement progressé ces dernières années, concentrant la plupart des efforts fournis par les institutions de conservation, c'est dans le domaine de l'interopérabilité que les attentes sont aujourd'hui les plus fortes. En ce qui concerne le grand public par exemple, les fonctionnalités d'incorporation dans les réseaux sociaux se multiplient. Il est désormais possible d'intégrer un petit visualiseur d'un document de Gallica au sein même d'un autre site Web ou de sa page Facebook. On peut ainsi visionner le document sans changer d'interface.

Côté chercheurs, les attentes vont dans le même sens, mais avec un niveau d'exigence et d'ergonomie plus élevé. Le vieux rêve des spécialistes est en effet à portée de main : pouvoir comparer en un lieu unique des manuscrits conservés dans des bibliothèques différentes, parfois distantes de plusieurs milliers de kilomètres. Pour ce faire, il faut casser la logique des silos, qui contraint pour l'instant un utilisateur à ne pouvoir visualiser un document numérique que dans l'interface qui lui est proposée par l'institution conservant ce même document. L'enjeu de ce que l'on pourrait appeler une interopérabilité «profonde» c'est de déconnecter le stockage de la consultation. Quel que soit le serveur sur lequel le document numérique est conservé, l'utilisateur doit pouvoir le consulter dans l'interface de son choix, celle mise en place par son laboratoire ou son université, par exemple. Si ces derniers ont développés un visualiseur capable d'afficher dans le même écran deux images provenant de documents numériques différents, il est alors possible de comparer, voire d'annoter, deux ouvrages conservés et numérisés par deux bibliothèques différentes.

Au-delà d'une exposition normalisée des métadonnées nécessaires à cet opération, par le biais de SharedCanvas, dont nous avons déjà parlé plus haut, il reste encore à implémenter un protocole normalisé permettant l'échange d'information entre les applications de gestion des images, serveur d'un côté et client de l'autre. Une bonne partie de l'interopérabilité des documents numériques réside en effet dans la capacité d'un visualiseur à transmettre à un serveur qui ne le connait pas des instructions

permettant de récupérer la bonne image, dans le bon sens, au niveau de définition souhaité, et dans un délai correct. Des travaux sont actuellement en cours, notamment autour du standard IIIF[4].

Formes à toucher n'est rien d'autre qu'une version tactile d'un visualiseur exploitant ce processus et lu adjoignant un outil d'annotation, qui est le cœur de l'application. L'application a donc été développée en avance de phase par rapport aux travaux en cours dans les entrepôts numériques mais elle permet de proposer une preuve de concept sur les usages possibles des images interopérables dans le domaine de l'annotation de formes.

4.4. Développements supplémentaires et enjeux pédagogiques et scientifiques.

La version 1 de *Formes à toucher* a été mise en ligne début juillet 2013 et offre déjà un panorama d'usages possible du tactile dans le domaine de l'annotation de manuscrits médiévaux. Le retour des usagers devrait permettre d'envisager les développements futurs et d'identifier les besoins de fonctions complémentaires. Plusieurs d'entre elles sont déjà à l'étude, à commencer par l'intégration de l'application dans un environnement pleinement interopérable : récupération automatisée d'images de manuscrits dans les bibliothèques numériques, hébergement et mise à jour des manifestes produits par l'application sur un serveur dédié. La possibilité pour les chercheurs de charger dans l'application leurs propres images devra également être étudiée car il s'agit d'une demande forte, mais qui n'est pas sans poser un certain nombre de problèmes, notamment en ce qui concerne la normalisation de départ nécessaire à la production des fichiers d'annotation. Une solution pourrait être d'adosser *Formes à toucher* à un portail en ligne qui permettrait de créer des lots d'images numériques normalisés, de proposer une interface de visualisation collaborative des annotations produites par différents iPads et d'être lié à d'autres projets applicatifs, notamment dans le domaine des études paléographiques.

L'implémentation de ces diverses fonctionnalités nécessite avant tout la mise en place d'un serveur *Formes à toucher* capable d'interagir avec les différentes ressources qui seront mises à contribution pour les intégrer au sein de l'environnement propre aux tablettes. Cette solution nous semble un préalable nécessaire à la mise en place de services interopérables au sein de l'application mobile car la nature même de la technologie utilisée la rend peut robuste à l'échange direct de données ou d'images haute définition avec des services tiers. Le serveur jouerait donc le rôle d'une sur-

---

[4] International Image Interoperability Framework, <http://www.iiif.io> (consulté le 6 mai 2014). Un visualiseur exploitant ce standard est également disponible sur <http://dmstech.github.io/mirador/> (consulté le 6 mai 2014).

couche logicielle permettant de réaliser pour le compte des utilisateurs de tablettes les opérations nécessaires.

Concrètement, ce serveur, qui pourrait par exemple bénéficier des infrastructures mises en place par Biblissima, permettrait de rendre accessibles les images issues de bibliothèques numériques interopérables, de stocker les données d'annotations réalisées sur l'application *Formes à toucher* (formes graphiques et contenus textuels) et de les partager avec d'autres utilisateurs.

A ce stade d'avancée des projets de mise en interopérabilité on peut considérer qu'il sera possible dans un avenir proche de récupérer ainsi les images issues des deux principales bibliothèques numériques de manuscrits médiévaux en France, Gallica et la Bibliothèque virtuelle des manuscrits médiévaux (BVMM). Ce sont par ailleurs les deux entrepôts qui ont fournis les images utilisées pour la première version de l'application.

5. Conclusion

L'application *Formes à toucher*, développée avec les utilisateurs finaux et pour faciliter la recherche, n'a pas encore atteint son stade complet de développement. Outre des fonctionnalités de visualisation et de recherche dans les annotations, il lui est nécessaire que les grandes bibliothèques numériques se rendent elles-mêmes interopérables. Les deux institutions qui ont présidé au développement de *Formes à toucher* y travaillent et jouent un rôle pilote dans les développements actuels de l'interopérabilité (Biblissima, COST, moteur TRAME).

L'ergonomie doit rester le maître mot du développement puisque l'on désire que les chercheurs s'approprient l'outil pleinement et sas difficulté. Les études d'usage sont le meilleur moyen d'obtenir une application rapide, intuitive et fluide.

La mise en ligne de *Formes à toucher* sur l'AppStore a permis de faire la preuve de concept de l'utilité d'une application mobile et tactile pour annoter les documents manuscrits. Il est toutefois nécessaire de réaliser un certain nombre de développements pour améliorer sa mise en interopérabilité avec les entrepôts d'images existants pour la rendre pleinement opérante dans l'environnement de travail des utilisateurs. En effet, la masse de documents numérisés accessibles sur le Web s'agrandissant de plus en plus et de plus en plus rapidement, il devient fondamental de fournir aux chercheurs des outils permettant de les utiliser facilement, et sans fonctionner en vase clos. La mise en place de telles fonctionnalités va de pair avec une évolution constante des normes et des formats, qui nécessite une maintenance régulière des outils qui les exploitent. L'avenir de *Formes à toucher* est donc, comme dans de nombreux projets, soumise à la mise en œuvre de moyens réels pour permettre son maintien en conditions opérationnelles dans le contexte numérique actuel.

Plusieurs projets pourraient être envisagés pour assurer la maintenance et l'évolution de l'application. Il pourrait par exemple être utile d'explorer le développement d'une interface de visualisation collaborative des annotations réalisées par plusieurs utilisateurs et qui pourrait faciliter le travail en équipe, qui devient de plus en plus courant dans les projets de recherche.

Dans le domaine de la recherche, en particulier, l'apport de *Formes à toucher* étant la capacité à appréhender les formes, les données générées par les utilisateurs pourront être l'objet d'un programme de recherche sur le clustering et la reconnaissance des formes, et l'étude de leur variabilité, avec les principaux acteurs de l'analyse d'image et des formes appliquée aux objets patrimoniaux.

Bibliographie

L'application est téléchargeable sur l'iTunes Store à l'adresse <https://itunes.apple.com/fr/app/formes-a-toucher/id667357077>.

M. Terras, H. Cayless, et W. Noel, «Introducing Tile 1.0», in *TILE | Text-Image Linking Environment*, College Park : Maryland Institute for Technology in the Humanities (MITH), 2011.

M. Rehbein, P. Sahle, et T. Schaßan, Éd., *Kodikologie und Paläographie im digitalen Zeitalter – Codicology and Palaeography in the Digital Age*. Norderstedt : BoD, 2009.

F. Fischer, C. Fritze, et G. Vogeler, Éd., *Kodikologie und Paläographie im digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2*. Norderstedt : BoD, 2011.

M. Longcamp, «Etude comportementale et neurofonctionnelle des interactions perceptivo-motrices dans la perception visuelle de lettre. Notre manière d'écrire influence-t-elle notre manière de lire ?», Thèse de doctorat, Université de la Méditerranée - Aix-Marseille II, Aix-en-Provence, 2003.

Stanford University, Los Alamos National Laboratory, Open Annotation Collaboration, et Mellon Foundation, «SharedCanvas. A distributed canvas rendered from Linked Data Annotations», *SharedCanvas*, 2011. <http://www.shared-canvas.org/>.

WZIS – Wasserzeichen-Informationssystem: Verwaltung und Präsentation von Wasserzeichen und ihrer Metadaten

Erwin Frauenknecht, Maria Stieglecker

Abstract

The project *Wasserzeichen-Informationssystem (WZIS)* has been granted by the German Research Foundation for two stages. Upon conclusion of the first phase in 2012 the software *Wasserzeichen Studio* is available as »backend« for distributed data entry for watermarks and their metadata. The software allows for various ways of image manipulation. Common guidelines and an authoritative hierarchical classification guarantee the homogenous description. The online presentation at *www.wasserzeichen-online.de* offers broad possibilities to search the contents. *Piccard-Online* has been integrated as well as the digitised volumes of the printed Piccard finding aids. Thus, WZIS serves as a tool for watermark research which allows for interdisciplinary and cross-institutional communication on questions concerning the description and research about watermarks. The paper demonstrates the realisation of plans laid out by Christina Wolf in 2009.

Zusammenfassung

Das Projekt *Wasserzeichen-Informationssystem (WZIS)* wurde in zwei Phasen von der Deutschen Forschungsgemeinschaft unterstützt. Mit Abschluss der ersten Phase liegt seit 2012 die Software *Wasserzeichen Studio* als »Backend« für die dezentrale Eingabe von Wasserzeichen und ihrer Metadaten vor, die daneben auch verschiedene Möglichkeiten einer Bildbearbeitung bietet. Gemeinsame Richtlinien und eine verbindliche hierarchische Klassifikation gewährleisten eine homogene Erschließung. Die Online-Präsentation auf *www.wasserzeichen-online.de* bietet breit angelegte Möglichkeiten zur Abfrage der Inhalte. *Piccard-Online* sowie die digital vorliegenden Bände der gedruckten Piccard-Findbücher wurden in das Angebot eingebunden, so dass damit ein Werkzeug zur Erforschung von Wasserzeichen vorliegt, das einer fächer- und institutionenübergreifenden Verständigung über Fragen der Erschließung, Dokumentation und Auswertung von Wasserzeichen den Weg ebnen kann. Der vorliegende Beitrags erläutert die Umsetzung der 2009 von Christina Wolf vorgestellten Planungen.

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 105–121.

1. Ausgangslage

Die Analyse von Wasserzeichen wird in verschiedensten Bereichen, von der Handschriftenbearbeitung bis zu Restaurierungsarbeiten genutzt, um zu Antworten nach zeitlicher und räumlicher Einordnung oder Fragen nach der Provenienz von Papierquellen zu gelangen. Moderne Methoden sowohl in der Visualisierung der Wasserzeichen wie deren Vermittlung über Online-Angebote wie »Piccard-Online«, »WZMA - Wasserzeichen des Mittelalters« oder »WILC - Watermarks in Incunabula printed in the Low Countries« erleichtern die Identifizierung und Zuordnung von Wasserzeichen wesentlich – auch wenn aufgrund der Heterogenität der Systeme und Sammlungen die Nutzung oft nicht einfach ist (Wolf 98-99). Das Portal »Bernstein – The Memory of Paper« bietet darüber hinaus eine Zusammenschau bestehender europäischer Datenbanken zu Wasserzeichen und Papiergeschichte. Hier findet sich auch seit 2012 die Online-Präsentation des von der Deutschen Forschungsgemeinschaft in zwei Projektphasen geförderten Projektes »Wasserzeichen-Informationssystem (WZIS)«. Als Partner fanden sich für die erste Projektphase (2010-2012) die Württembergische Landesbibliothek Stuttgart (WLB), das Landesarchiv Baden-Württemberg (LABW), die Bayerische Staatsbibliothek München (BSB), die Universitätsbibliothek Leipzig (UBL) und die Österreichische Akademie der Wissenschaften (ÖAW) zusammen, als weitere Partner traten für die zweite Förderphase (2012-2014) die Staatsbibliothek zu Berlin (SBB) und die Deutsche Nationalbibliothek Leipzig (DNB) hinzu. Ziel dieses Projektes war und ist der Aufbau eines gemeinsamen Informationssystems für Wasserzeichen in den DFG-Handschriftenzentren. Im Zuge der Katalogisierung mittelalterlicher Handschriftenbestände in den verschiedenen Bibliotheken wurden zum Teil sehr umfangreiche Sammlungen zu Wasserzeichen angelegt, die aber nur eingeschränkt und lokal nutzbar waren. Es handelt sich hierbei auch nicht um abgeschlossene Sammlungen, da sie durch neue Erschließungsprojekte laufend erweitert werden. Um sie einem breiten Publikum zugänglich zu machen wurde ein System zur Verwaltung von digitalisierten Abbildungen von Wasserzeichen sowie ihrer Metadaten entwickelt, wobei die Dateneingabe dezentral an den einzelnen Bibliotheken erfolgt.

Der Rückgriff auf bestehende und bewährte Systeme erleichterte die Umsetzung der Projektziele. Das Frontend wurde sehr ähnlich zu »Piccard-Online« gestaltet, der aufgrund ihres weltweit umfangreichsten Inhalts zu Wasserzeichen wohl am häufigsten genutzten Datenbank, und bietet so einen großen Wiedererkennungseffekt, hält aber auch wesentliche Erweiterungen bereit wie etwa eine Visualisierung geografischer und zeitlicher Verteilung bestimmter Wasserzeichen-Motive. Für die homogene Erschließung der Wasserzeichenmotive ist die Entwicklung einer dynamischen Klassifikationsstruktur Voraussetzung, wofür als Basis die im Rahmen des Projektes »Bernstein« entwickelten Hilfsmittel dienen (Frauenknecht u.a. *Watermark Terms*; Frauenknecht u.a. *Bernstein Systematics*). Schließlich wurde als grundlegende

Software das Programm »Watermark Processing and Database Management Toolkit (Watermark Toolkit)« verwendet (s. Stieglecker), das an der Österreichischen Akademie der Wissenschaften für das System »WZMA« entwickelt und nun für die Bedürfnisse einer dezentralen und erweiterten Dateneingabe adaptiert wurde.

2. Inhalte

Die in WZIS einzubindenden Wasserzeichensammlungen entstanden aus verschiedenen Projekten heraus und spiegeln hinsichtlich abgedeckter Region, Chronologie oder Inhalt unterschiedliche Intentionen wider.

Den Grundstock bildet die Wasserzeichensammlung Piccard, die im Landesarchiv Baden-Württemberg verwahrt wird und die mit etwa 92.000 auf Karteikarten verzeichneten Einzelbelegen als weltweit größte ihrer Art gilt. In digitalisierter Form steht sie seit 2006 komplett als »Piccard-Online« zur Verfügung und ergänzt die bis dahin gedruckt vorliegenden Findbücher wesentlich, da mit diesen nur eine Auswahl an Motiven und Einzelbelegen der umfangreichen Sammlung zur Verfügung steht. Für eine sinnvolle Integration der Sammlung in WZIS ist eine sukzessive Überführung in die neue Klassifikation und Systematik, die im Rahmen des Projektes entwickelt wird, unumgänglich. Da Untersuchungen erwiesen, dass zum einen nicht mehr alle in den gedruckten Findbüchern abgebildeten Wasserzeichen als Karteikarte im Archiv vorliegen und zum anderen nicht alle Karteikarten in dem dem Motiv entsprechenden Band aufscheinen, wurde die mühselige Aufgabe in Angriff genommen, einen Abgleich »Piccard-Online« mit den gedruckten Findbüchern durchzuführen, um so eine einfache Nutzung der gesamten Piccard-Sammlung über nur eine Benutzeroberfläche zu ermöglichen.

Eine Ergänzung erfährt diese Belegsammlung von Gerhard Piccard durch seine sich an der Württembergischen Landesbibliothek befindlichen Expertisen zu Wasserzeichen, deren Zeichnungen nun digitalisiert und mit ihren Metadaten über WZIS zugänglich gemacht werden. Weiters werden an der WLB Durchzeichnungen von Papiermarken, die im Zuge der Katalogisierung der mittelalterlichen lateinischen Handschriften sowie neuzeitlicher Manuskripte der Universität Tübingen erstellt wurden, in das System eingepflegt sowie Wasserzeichen aus neuzeitlichen Musikhandschriften hessischer Archive.

Ebenfalls aus Katalogisierungsprojekten zu lateinischen, griechischen, deutschsprachigen und illuminierten Handschriften aus eigenen Beständen sowie anderer bayerischer Bibliotheken wie der Universitätsbibliotheken Augsburg und Eichstätt oder des Germanischen Nationalmuseums Nürnberg entstandene Belegsammlungen zu Wasserzeichen werden von der Bayerischen Staatsbibliothek München in WZIS integriert. Neben der bereits erfolgten Einarbeitung der Papiermarken von 90

Blockbüchern in bayerischen Sammlungen werden auch solche aus primär datierten Inkunabeln aus Druckorten im süddeutschen Raum erfasst.

Einen in filigranologischer Hinsicht bisher stiefmütterlich behandelten Raum deckt die Universitätsbibliothek Leipzig zunehmend ab. Seit 2004 wird hier kontinuierlich eine Wasserzeichenkartei zu in Ostdeutschland verwendeten Papieren aufgebaut, da nicht zuletzt aufgrund der politischen Verhältnisse der zweiten Hälfte des 20. Jahrhunderts entsprechende Marken in bestehenden Repertorien nur unzureichend dokumentiert sind. Neben dieser Kartei finden Wasserzeichen aller datierten Papierhandschriften aus dem Bestand der UBL sowie von datierten Leipziger Frühdrucken Eingang in WZIS; bereits eingespeist sind Marken aus deutschsprachigen mittelalterlichen Handschriften der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden.

Mit der an der Staatsbibliothek zu Berlin beheimateten Sammlung von Eva Ziesche findet der Bestand einer der hervorragenden Kennerinnen der Filigranologie Aufnahme in WZIS und bietet so umfangreiche Ergänzungen an Belegen zu Wasserzeichen aus mittelalterlichen und jüngeren Handschriften sowie Inkunabeln. Aus am Handschriftenzentrum der SBB angesiedelten Erschließungsprojekten kommen Wasserzeichenreproduktionen weiterer Einrichtungen hinzu. Neben Greifswalder und Rostocker Handschriften werden auch Bestände aus rheinländischen Sammlungen wie der Universitäts- und Landesbibliothek Bonn, des Landeshauptarchivs Koblenz oder der Stadtbibliothek Trier auf ihre Wasserzeichen hin untersucht.

Von einer ganz anderen Seite her bringt das Deutsche Buch- und Schriftmuseum der Deutschen Nationalbibliothek in Leipzig seine Wasserzeichensammlungen ein, die mit ihren etwa 400.000 Einzelbelegen als die weltweit größte Zusammenstellung neuzeitlicher Papiermarken gelten. Für die Integration in WZIS bot sich der Zugang über die nach Papiermühlen geordneten Sammlungen an, wobei die ›Papiermühlen in Thüringen‹ den Anfang machen, da sich dieser Bestand durch sehr differenzierte Erschließungsdaten auszeichnet. Mit den Sammlungen der DNB werden und sollen auch über Projektende hinaus Wasserzeichen der Neuzeit - die bisher weder in gedruckter noch digitaler Form ausreichend publiziert wurden - einem größeren Interessentenkreis leicht zugänglich gemacht werden.

3. Erfassungsmodul

Hauptkomponente der ersten Projektphase war die Entwicklung eines Erfassungsmoduls zur dezentralen Dateneingabe und Verwaltung von Wasserzeichenbildern und ihrer zugehörigen Metadaten. Die Bereithaltung sämtlicher Daten in einer zentralen MySQL-Datenbank, die am Landesarchiv Baden-Württemberg in Stuttgart lokalisiert ist, bietet neben der damit einhergehenden homogenen Eingabe den Vor-

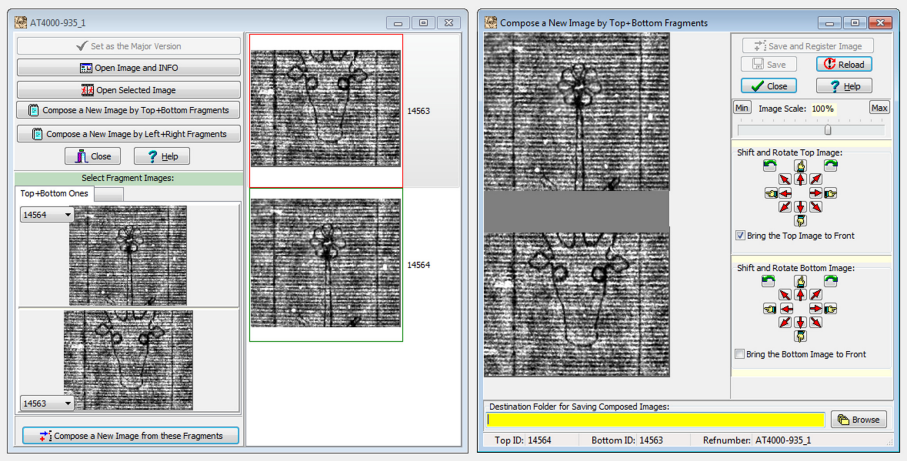

Abbildung 1. Wasserzeichen-Studio: Zusammensetzung eines Bildes aus zwei Bildteilen.

teil der schnellen Abfrage sowie der differenzierten Recherchen mit ausgewogenen, strukturierten Ergebnissen. Die Integration sowohl bestehender Sammlungen wie »Piccard-Online« sowie in Zukunft entstehender wird über das Tool »Wasserzeichen Studio« ermöglicht.

Dieses Tool basiert auf dem in Zusammenarbeit der Österreichischen Akademie der Wissenschaften mit der Russischen Akademie der Wissenschaften von Victor Karnaukhov entwickelten Programm »Watermark Processing and Database Management Toolkit (Watermark Toolkit)«.

Verschiedene Bildbearbeitungsfunktionen ermöglichen eine möglichst qualitätsvolle Darstellung der digitalisierten Wasserzeichenbilder der sehr heterogenen Vorlagen an Reproduktionen in Form von Durchzeichnungen, Abreibungen, Durchlichtbildern, Radiographien oder Thermographieaufnahmen: Stufenloses Rotieren, Spiegeln oder Invertieren des Bildes ist ebenso möglich wie eine Kontrastverbesserung. Die Darstellung von Wasserzeichen aus Folio-Bänden erweist sich – je nach Qualität der Digitalisate – meist problemlos, schwieriger stellt sich die Lage bei Quart- oder Oktavhandschriften dar. Hier befinden sich die Wasserzeichen im Falz, sind sozusagen zwei- oder viergeteilt, es müssen also Bildteile zu einem Bild zusammengesetzt werden. Die Funktion »Compose a New Image by Fragments« bietet hier eine komfortable Lösung, bei der Komponenten eines Bildes in einem gemeinsamen Rahmen in die jeweils gewünschte Position gebracht und als neues Bild abgespeichert werden können (Abb. 1). Die genannten unterschiedlichen Reproduktionsverfahren machten ein weiteres, nun implementiertes Feature wünschenswert, die Ablagemöglichkeit mehrerer

Bilddateien pro Wasserzeichen. Immer wieder liegen für das Wasserzeichen ein und desselben Blattes Papier mehrere Belege des Wasserzeichens vor, für die verschiedene Abnahmeverfahren gewählt wurden, etwa ein Abreibung, eine Durchzeichnung und eine Thermographieaufnahme. Zeigen diese nun verschiedene Einzelheiten der Papiermarke in einmal besserer, einmal schlechterer Qualität, kann es durchaus von Vorteil sein, mehrere Reproduktionen nebeneinander vergleichen zu können. Das Wasserzeichen Studio erlaubt nun, dass mehrere Bilddateien – mit systeminternen IDs versehen – unter einer Referenznummer verwaltet werden (Abb. 2). Diese Referenznummer, die auch der Zitierung des Wasserzeichens dient, ist eindeutig und setzt sich zusammen aus einem Ländercode, einem Institutionscode, der Signatur der Quelle und der Blattnummer, von dem das Wasserzeichen abgenommen wurde.

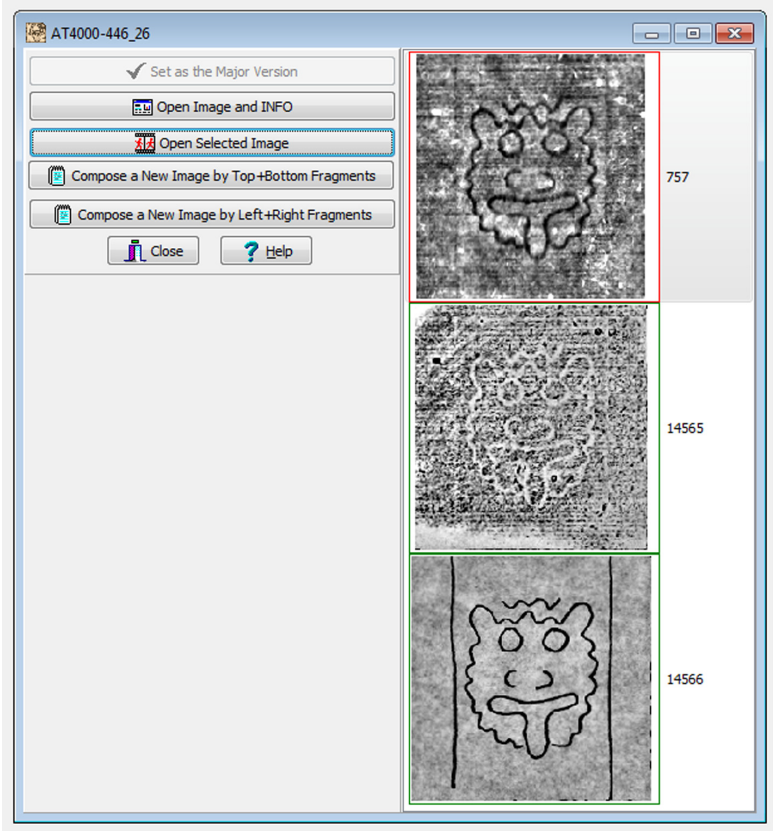

Abbildung 2. Wasserzeichen-Studio: Verwaltung mehrerer Bilddateien unter einer Referenznummer.

Vom Papierproduktionsprozess her bedingt stehen verschiedene Wasserzeichen untereinander in Verbindung. Zum einen wurde beim Papierschöpfen an der Bütte arbeitsteilig mit zwei Sieben gearbeitet. Auf beiden Sieben wurde je eine Drahtfigur zur Herstellung des Wasserzeichens befestigt, die im Motiv übereinstimmen, aber natürlich nicht deckungsgleich sind. Aufgrund dieser Sieb-Paare gehören auch immer zwei Papiermarken als Wasserzeichenpaar oder Zwillingsmarken zusammen. Zum anderen sind modernere Schöpfsiebe oft mit zwei oder mehr Drahtfiguren pro Sieb versehen, die dann auf dem geschöpften Papierbogen entsprechend viele Wasserzeichen zeigen, die als Gegenmarken bezeichnet werden. Die Eingabemaske der Datenbank sieht die Möglichkeit vor, sowohl Zwillingsmarken wie Gegenmarken miteinander zu verknüpfen und entsprechend darzustellen.

Rückschlüsse auf Aufbau und Zusammensetzung einer Handschrift, eventuell verwendetes Restpapier, Verbindungen zwischen Handwechsel und Papiersorten kann über ein Lagenschema graphisch übersichtlich dargestellt werden. Im Laufe der Zeit haben sich verschiedene Vorgehensweisen und Vorlieben entwickelt hinsichtlich eines handschriftlichen Überblicks oder einer Erfassung in Tabellenform. Hier verbindliche Richtlinien aufzustellen und diese nachträglich umzusetzen, um die Lagenzusammensetzung und Verteilung der Wasserzeichen der eingebundenen Handschriften in homogener tabellarischer Form in WZIS abzubilden, hätte den Rahmen des Projektes bei weitem gesprengt. Um die Informationen dennoch zu erfassen wurde eine Upload-Möglichkeit für Dateien im PDF-Format in die Programmstruktur integriert.

Mit der zweiten Projektphase werden verstärkt Wasserzeichen aus neuzeitlichen Papieren erfasst, für die – im Gegensatz zu mittelalterlichen – zum Teil sehr umfangreiche Informationen zu ihrer Herkunft bekannt sind. Die Zuordnung der Papiere zu bestimmten Papiermühlen und Papiermachern gibt hinsichtlich Verbreitung, Handelswegen oder Lokalgeschichte wichtige Aufschlüsse für verschiedenste Fachdisziplinen. Im »Wasserzeichen Studio« können deshalb die Wasserzeichen über erweiterbare Listen mit Ortsangaben zu Papiermühlen und Namensnennungen von Papiermachern verknüpft werden.

Um ein Überfrachten des Systems mit zusätzlich vorhandenen Informationen zu vermeiden bzw. um einer Inhomogenität des Systems vorzubeugen, besteht die Möglichkeit, auf verschiedenen Ebenen mittels Link auf externe Datenbanken zu referenzieren. Von einem Papiermacher kann etwa auf die »Gemeinsame Normdatei« verlinkt werden, von Handschriften auf Handschriftendatenbanken wie »Manuscripta Mediaevalia«, »Handschriftencensus« oder »Manuscripta.at«, von einzelnen Wasserzeichen auf weitere Wasserzeichendatenbanken wie »WZMA« oder »WILC«.

Um verschiedene Inhalte der Datenbank im Frontend in ihren geographischen Zusammenhängen sichtbar machen zu können, erfolgen im Erfassungsmodul entsprechende Eintragungen zu Aufbewahrungs- und Beschreiborten der Papiere oder Papiermühlen, soweit diesbezügliche Informationen vorliegen (Abb. 3).

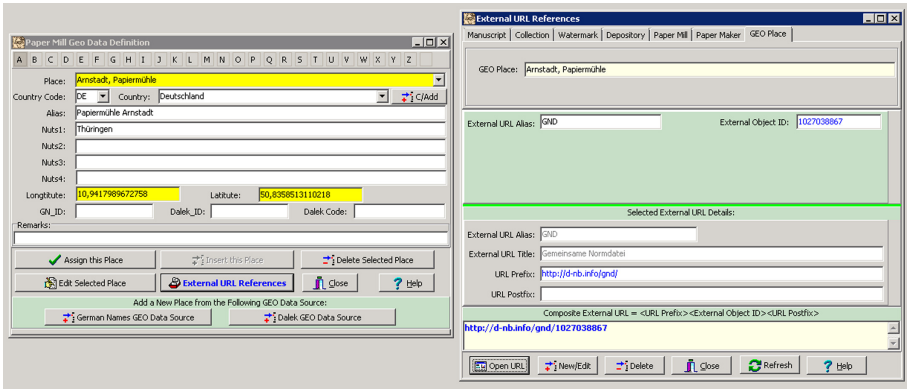

Abbildung 3. Wasserzeichen-Studio: Georeferenzierung von Papiermühlen.

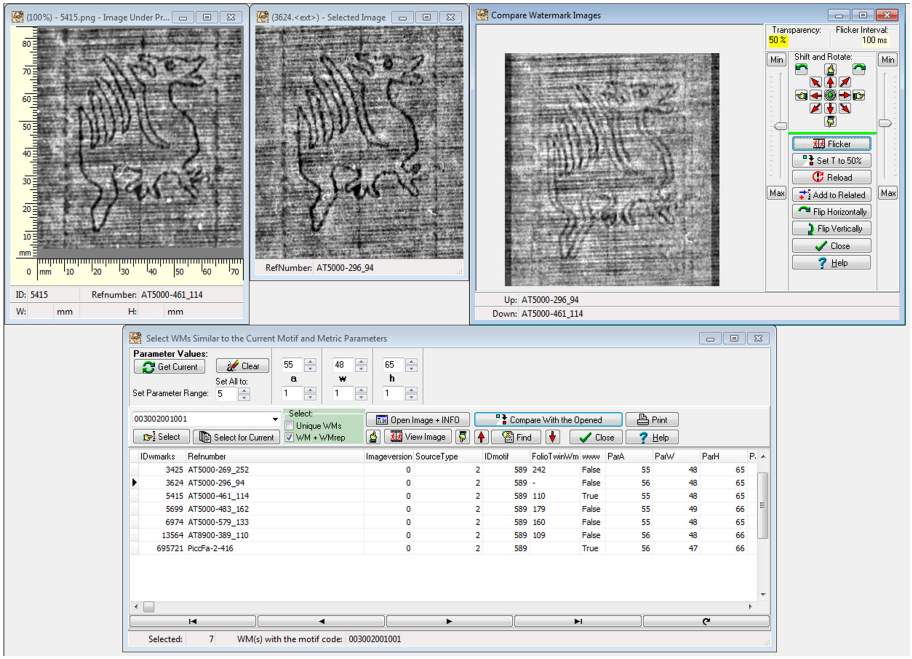

Abbildung 4. Wasserzeichen-Studio: Überprüfung von Identität mittels Layer-Technologie.

Die Einbindung der einzelnen Wasserzeichen in die Datenbank erfolgt nach vorgegebenen Richtlinien, um eine homogene Erschließung und in der Folge differenzierte Suchfunktionen mit ausgewogenen Ergebnissen im Frontend zu gewährleisten. An vorderster Stelle steht die Zuordnung des Wasserzeichenbildes zu einem Motiv über ein Baum-Menü, das - ausgehend von zwölf Motivgruppen - hierarchisch auf bis zu zehn Ebenen angeordnet ist. Diese hierarchisch strukturierte Klassifikation ist dynamisch angelegt, sodass auf neu auftretende Motive reagiert werden kann. Ein Tool zur semiautomatischen Vermessung des Zeichens sorgt für eine möglichst genaue Verzeichnung der absoluten Höhe wie Breite des Wasserzeichens sowie des Abstandes zwischen den nächstgelegenen Stegdrähten. Mithilfe dieser Vorgaben kann nun nach bereits in die Datenbank integrierten identischen oder in Varietät vorliegenden Zeichen gesucht werden, die in Motiv und Abmessungen mit dem Neuzugang übereinstimmen. Mittels Layer-Technologie können Bilder transparent übereinandergelegt und so unkompliziert miteinander verglichen und auf mögliche Identität hin überprüft (Abb. 4) und entsprechend miteinander verknüpft werden.

Zu jedem Wasserzeichenbild bietet schließlich eine »Watermark Description Form« einen Überblick über die getätigten Eingaben (Abb. 5). Von Angaben zum Bild (wie Imageversionen, Gegenmarken oder Zwillingsmarken) über Informationen zur Quelle (wie bewahrende Institution, Signatur oder Datierung und Lokalisierung) bis zu verwandten Wasserzeichen (identische Zeichen, Varianten und Angaben zum Typ).

4. Klassifikation

Ein zentrales Element der Erschließungsrichtlinien bildet ein hierarchisch strukturiertes Klassifikationsmodell mit verschiedenen Motivebenen. Die Grundstruktur dieser Klassifikation wurde im Rahmen des Bernstein-Projektes aus dem Abgleich bestehender Wasserzeichensammlungen wie »Piccard-Online« (PO), »Wasserzeichen des Mittelalters« (WZMA) und »Watermarks in Incunabula printed in the Low Countries« (WILC) gewonnen. Auch das klassische Wasserzeicheninventar von »Briquet« sowie der »IPH-Standard« fanden Berücksichtigung.

Auf der Ebene der Hauptmotive gelang so im Vergleich mit diesen Modellen eine Reduktion auf zwölf Hauptmotive, während der »IPH-Standard« hier 25 Motive und »Piccard-Online« sogar 38 Motive vorsehen (Abb. 6) Für die Benutzer bedeutet dies, dass der Einstieg in die Klassifikation leichter fällt, weil weniger Gruppen zu berücksichtigen sind. Folgende Hauptmotive stehen zur Verfügung: Anthropomorphe Figuren, Fauna, Fabelwesen, Flora, Berge/Himmelskörper, Realien, Symbole/Herrschaftszeichen, Geometrische Figuren, Wappen, Marken, Buchstaben/Ziffern, unbestimmte Zeichen. Diese Motive sind weiter hierarchisch untergliedert und zwar bis zu

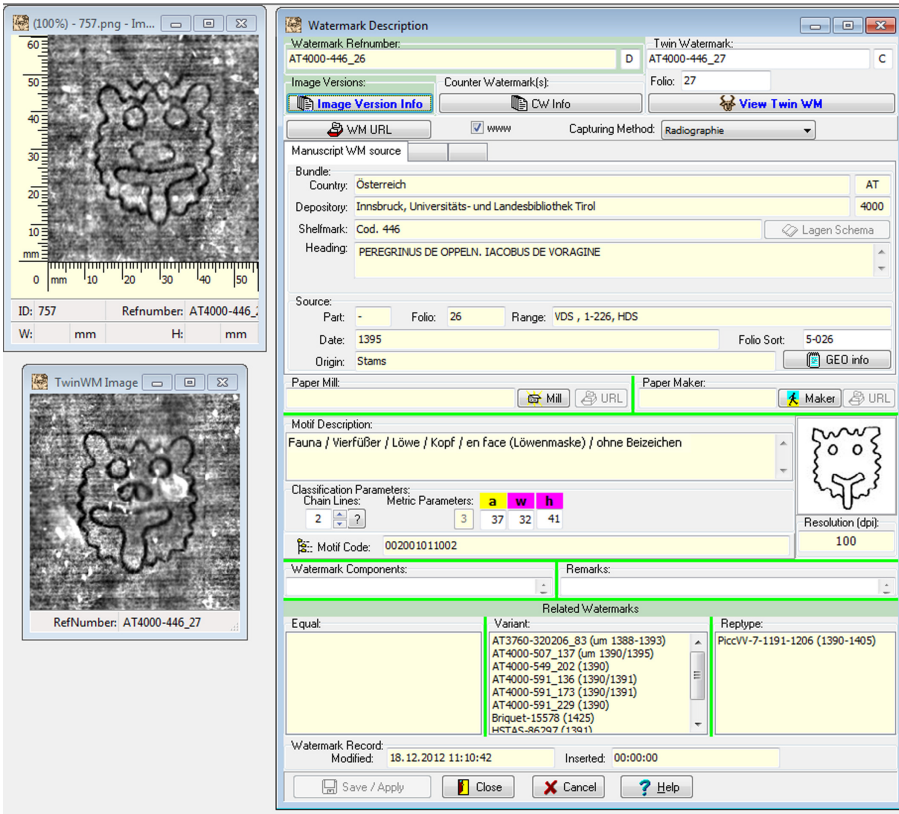

Abbildung 5. Wasserzeichen-Studio: »Watermark Description Form« – Überblick über die getätigten Eingaben.

einer Erschließungstiefe von zehn Ebenen. Damit ist eine Tiefenerschließung möglich, die auch Raum lässt, um auf neu auftretende Motive zu reagieren.

Im Folgenden soll ein Detailbeispiel zeigen, welche Schwierigkeiten bei der Erstellung der hierarchischen Klassifikation zu bewältigen waren. Weit verbreitet ist das Wasserzeichenmotiv ›Zwei Türme (mit Tor oder anderen Beizeichen)‹, sehr häufig ein Wasserzeichen, das auf Papiermühlen aus Ravensburg hinweist. Nach dem »IPH-Standard« wäre dieses Motiv unter K7/4 ›Türme (zwei) mit Tor‹ einzuordnen, eine weitere Differenzierung ist hier nicht vorgesehen. Das würde konkret bedeuten, dass in diese Gruppe rund 4.800 Zeichen einzuordnen wären, die jedoch noch zahlreiche weitere Unterscheidungen aufweisen.

Abbildung 6. Vergleich der Hauptmotive zwischen dem IPH-Standard (links, 25 Motivgruppen), Piccard-Online (Mitte, 38 Motivgruppen) und der WZIS-Klassifikation (rechts, 12 Motivgruppen).

Diese Differenzierung wurde in »Piccard-Online« getroffen, dort ist das Motiv in insgesamt elf Untergruppen unterteilt. Allerdings sind die Bezeichnungen in einigen Fällen etwas unklar. So ist zum Beispiel nicht sofort ersichtlich, worin der Unterschied zwischen den Motivgruppen ›zwei Türme, dazwischen freistehender einkonturiger Torbogen‹ und ›zwei Türme, dazwischen einkonturiger Torbogen‹ besteht. In der verbalen Beschreibung mag das deutlich werden, in der visuellen Darstellung mit dem entsprechenden Vorschaltbild ist das weniger deutlich zu erkennen.

In »WZIS« wurde versucht, eine einfachere Struktur zu wählen, die gleichwohl die verschiedenen Motive berücksichtigt (Abb. 7). Sie beschränkt sich zunächst auf die Unterscheidung, ob die zwei Türme Beizeichen aufweisen oder nicht. Aus diesem Grund ist das Motiv ›zwei Türme‹ in »WZIS« aufgeteilt in ›zwei Türme, ohne Beizeichen‹ und ›zwei Türme, mit Beizeichen‹, erst danach erfolgt eine weitere Differenzierung einzelner Merkmale wie Fenster im Turm oder Beizeichen wie Buchstaben, Kreuz oder Horn (Abb. 8)

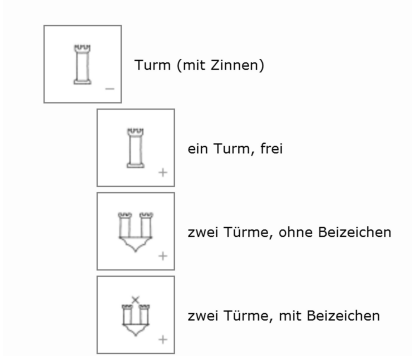

Abbildung 7. Ausschnitt aus dem WZIS-Strukturbaum: Die Motivgruppe ›Turm« ist in drei weitere Untergruppen differenziert: ›ein Turm, frei« – ›zwei Türme, ohne Beizeichen« – ›zwei Türme, mit Beizeichen«.

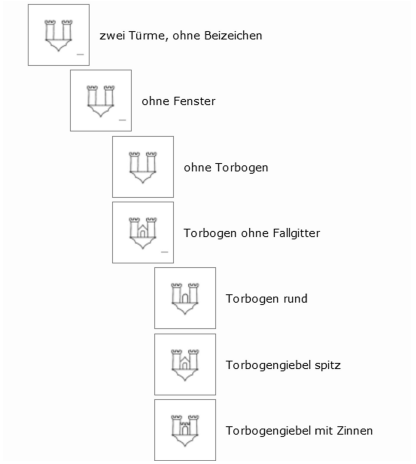

Abbildung 8. Beispiel für die weitere Klassifizierung der WZIS-Motivgruppe ›Zwei Türme, ohne Beizeichen«.

Gleichzeitig wurden bei diesem Motiv ›zwei Türme‹ einige Gruppen aus »Piccard-Online« genauer differenziert, wie das folgende Beispiel zeigt: In »Piccard-Online« enthält die Gruppe ›Zwei Türme – dazwischen zweikonturiger Torbogen ohne Zinnen – Türme mit Fenster – zwei Buchstaben im Sockel‹ insgesamt 326 Wasserzeichen, die im Sockel ganz unterschiedliche Buchstabenkombinationen aufweisen, wie CM, ES, DO oder HL. Gerade diese Siglen liefern wichtige Informationen zum Papiermacher, weil sich hinter den Siglen sehr häufig die Anfangsbuchstaben des Namens verbergen. In vielen Fällen ist damit eine individuelle Zuweisung auf einen einzelnen Papiermacher aus Ravensburg möglich. CM sind beispielsweise die Initialen von Christoph Mieser, der von 1636 bis 1644 als Beständer der äußeren Papiermühle in der Schornreute bei Ravensburg nachweisbar ist. ES weist auf Eustachius Sauter, der um 1602 in Ravensburg nachweisbar ist, während DO die Initialen für Daniel Dorn darstellen. Unter diesem Namen sind allerdings drei gleichnamige Personen im 17. Jahrhundert auf der Ravensburger Papiermühle in der Schornreute belegt.

Um diese Differenzierung nach den Namenssiglen sichtbar zu machen, mussten diese 326 Wasserzeichen in »WZIS« aufgeteilt und nach ihren Siglen geordnet werden: ›Turm – zwei Türme mit Beizeichen – zwei Buchstaben – zwei Buchstaben im Sockel – CM‹ bzw. ›ES‹ oder ›DO‹. Das neue Programmfeature der Papiermühlen und Papiermacher erlaubt es zudem, über die hierarchische Klassifikation hinaus diese namentlichen Identifizierungen im System zu erfassen (siehe Abb. 3).

Bei bestimmten Motiven kommt die hierarchische Klassifizierung an ihre Grenzen. Gerade im Bereich der Wappenmotive stellt sich die Klassifikation im Moment als work in progress dar. Eine differenzierte Erschließung steht hier noch aus, und eine ganze Reihe von Wasserzeichen bereiten hier Probleme bei der Einordnung. Das sind vor allem komplexe Wappenmotive, deren verbale Beschreibung nur über spezifische, heraldische Fachtermini möglich ist. Erfasser und Benutzer, die mit dieser heraldischen Fachsprache nicht vertraut sind, stehen damit vor einer ganz großen Herausforderung. Als zusätzliches Erschließungsmittel bei komplexen Zeichen kann im Wasserzeichen-Studio eine Beschreibung von einzelnen Bestandteilen des Wasserzeichens vorgenommen werden. Durch die Zusatzbeschreibung des Thesaurus im Feld ›watermark components‹ kann das Motiv in seinen Einzelteilen bestimmt werden, ohne den exakten Platz in der hierarchischen Gliederung zu kennen (Abb. 9). Die Begriffe aus dem Thesaurus können über die Suchfunktion erfasst werden. Eine spätere Redaktion kann dann solche komplexen Wasserzeichen in die jeweilige exakte hierarchische Motivgruppe verschieben.

5. Präsentationsmodul

Die vorläufigen Ergebnisse des Projektes WZIS stehen bereits als frei zugängliches Online-Präsentationsmodell unter www.wasserzeichen-online.de zur Verfügung.

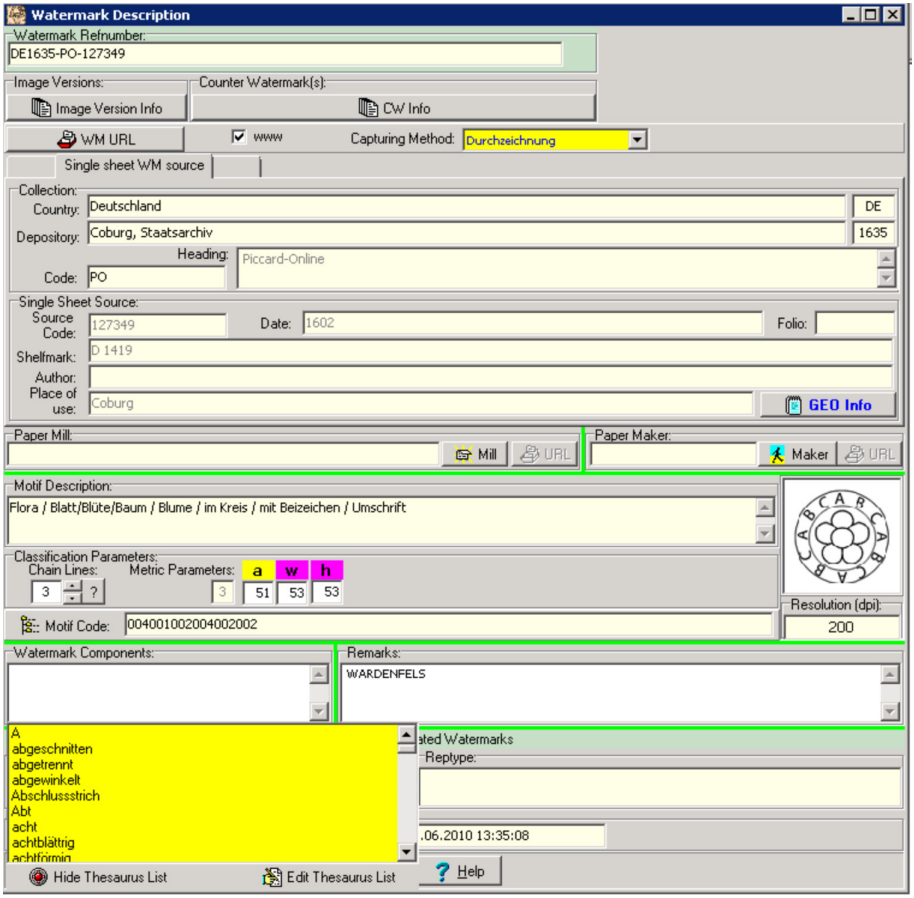

Abbildung 9. Einzelne Bestandteile komplexer Motive können über einen Thesaurus ergänzt werden.

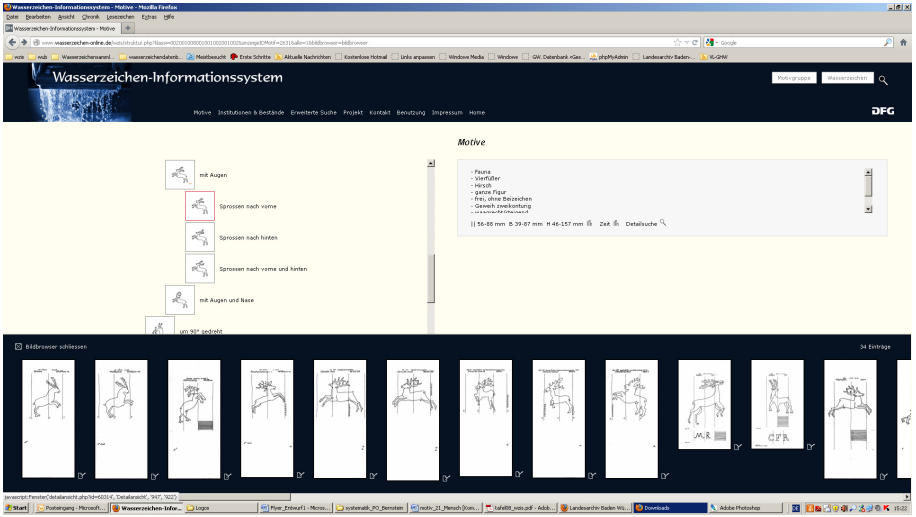

Abbildung 10. Ein Bildbrowser ermöglicht in WZIS eine schnelle Durchsicht auch großer Treffermengen.

Über die Suchmaske (›erweiterte Suche‹) können Wasserzeichen nach den unterschiedlichsten Kriterien gesucht werden: Motiv, Größenparameter, Beschreibort, besitzende Institution, Verwendungszeitraum oder Referenznummer. Die Suchabfrage kann variabel kombiniert werden. Auch auf Motivebene ist eine Suche nach bestimmten Klassifikationsgruppen möglich. Ein Bildbrowser (Abb. 10) ermöglicht die schnelle Durchsicht auch größerer Treffermengen. Zu jeder Motivgruppe kann der Verwendungszeitraum graphisch dargestellt werden, außerdem ist eine kartographische Visualisierung möglich, die räumliche Zusammenhänge leichter erkennbar macht. Auch die externen Verlinkungen auf andere Datenbanken (wie etwa »Manuscripta mediaevalia« oder »Gesamtkatalog der Wiegendrucke«) werden im Frontend unter der Rubrik ›Bezüge‹ dargestellt (Abb. 11).

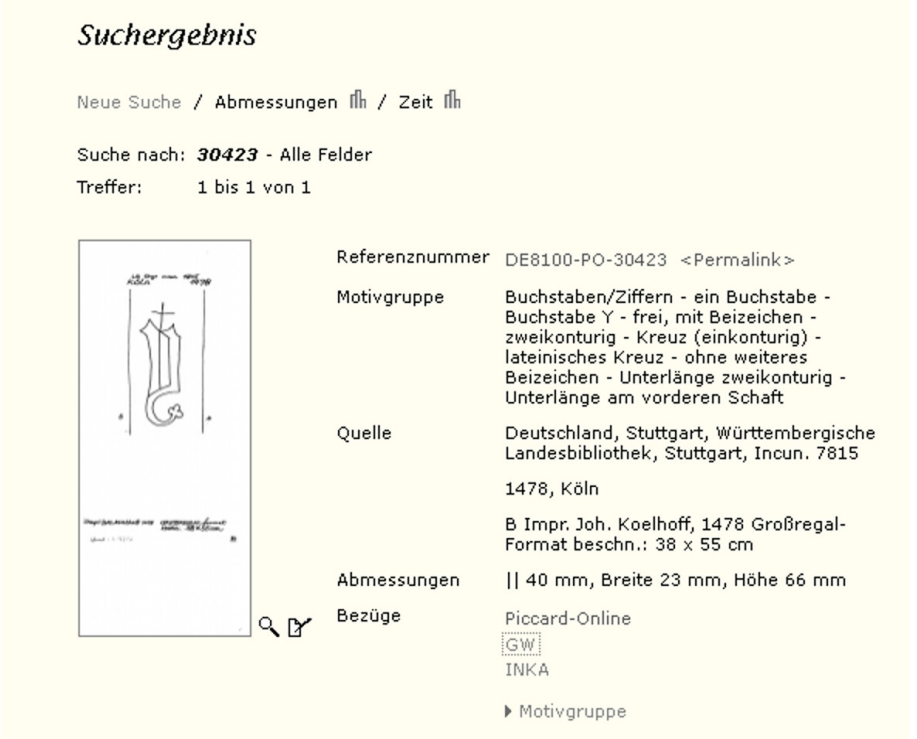

Abbildung 11. Externe Verlinkungen erschließen weitere Informationsquellen zu den einzelnen Wasserzeichen. Das abgebildete Wasserzeichen (DE8100-PO-30423) ist einem Inkunabeldruck von 1478 entnommen. Der entsprechende Verweis auf GW, den Gesamtkatalog der Wiegendrucke, liefert wichtige Informationen zu diesem Druck.

Bibliographie

Bernstein: The Memory of Paper. Wien: Österreichische Akademie der Wissenschaften, 2009. <http://www.memoryofpaper.eu:8080/BernsteinPortal/appl_start.disp>.

Frauenknecht, Erwin, Carmen Kämmerer, Peter Rückert und Maria Stieglecker. *Watermark Terms. Vocabulary for watermark description.* Version 9.1h: November 10, 2012. <http://www.memoryofpaper.eu/products/watermark_terms_ge.pdf>

Frauenknecht, Erwin, Peter Rückert und Maria Stieglecker. *Bernstein Systematics.* Version 1.6.1a: 13 February 2012. <http://www.memoryofpaper.eu/products/Bernstein_systematics.pdf>

Frauenknecht, Erwin und Maria Stieglecker. »Das Projekt Wasserzeichen-Informationssystem (WZIS). Innovative Wege bei der Erfassung und Präsentation von Wasserzeichen«. *2012 IPH Congress Book (im Druck).*

Meinlschmidt, Peter, Carmen Kämmerer und Volker Märgner. »Thermographie – ein neuartiges Verfahren zur exakten Abnahme, Identifizierung und digitalen Archivierung von Wasserzeichen in mittelalterlichen und frühneuzeitlichen Papierhandschriften, -zeichnungen und -drucken.« *Kodikologie und Paläographie im Digitalen Zeitalter 2*. Hrsg. Franz Fischer, Christiane Fritze und Georg Vogeler. Norderstedt: BoD, 2010, 209–226. <http://kups.ub.uni-koeln.de/4351>.

»Ochsenkopf und Meerjungfrau. Papiergeschichte und Wasserzeichen vom Mittelalter bis zur Neuzeit«. *Begleitbuch und Katalog zur Ausstellung des Landesarchivs Baden-Württemberg, Hauptstaatsarchiv Stuttgart und der Österreichischen Akademie der Wissenschaften, Kommission für Schrift- und Buchwesen des Mittelalters, Wien*. Stuttgart/Wien^{3} 2009.

Piccard, Gerhard. »Die Wasserzeichenkartei Piccard im Hauptstaatsarchiv Stuttgart. Findbuch I–XVII.« *Veröffentlichungen der Staatlichen Archivverwaltung Baden-Württemberg, Sonderreihe. 25 Bände*. Stuttgart: Kohlhammer 1961–1997.

Piccard-Online. Stuttgart: Landesarchiv Baden-Württemberg. <http://www.piccard-online.de/>.

Rückert, Peter, Jeannette Godau und Gerald Maier, Hrsg. *Piccard-Online. Digitale Präsentationen von Wasserzeichen und ihre Nutzung*. Stuttgart: Kohlhammer 2007.

Stieglecker, Maria. »Zur Methode der Wasserzeichenerfassung für die Sammlung *WZMA*«. *Piccard-Online. Digitale Präsentationen von Wasserzeichen und ihre Nutzung*. Hrsg. Peter Rückert, Jeannette Godau und Gerald Maier. Stuttgart: Kohlhammer 2007, 55–63.

Wolf, Christina. »Aufbau eines Informationssystems für Wasserzeichen in den DFG-Handschriftenzentren.« *Kodikologie und Paläographie im Digitalen Zeitalter*. Hrsg. Malte Rehbein, Patrick Sahle und Torten Schaßan. Norderstedt: BoD, 2009, 97–107. <http://kups.ub.uni-koeln.de/volltexte/2009/2963>.

WILC: Watermarks in Incunabula printed in the Low Countries. Koninklijke Bibliotheek – National library of the Netherlands, 2000–2007. <http://watermark.kb.nl/page>.

WZIS: Wasserzeichen-Informationssystem. Stuttgart: Hauptstaatsarchiv, 2012. <http://www.wasserzeichen-online.de/wzis>.

WZMA: Wasserzeichen des Mittelalters. Wien: Österreichische Akademie der Wissenschaften, 2007. < http://www.ksbm.oeaw.ac.at/wz/wzma.php>.

Un corpus di iscrizioni medievali della provincia di Viterbo: Metodologia d'analisi e alcune riflessioni sulla sua informatizzazione

Elisa Pallottini

Riassunto

Il contributo propone alcune riflessioni sulla metodologia con cui è stato condotto lo studio di un corpus di iscrizioni dell'alto Lazio (provincia di Viterbo, Italia) databili tra il VI e il XII secolo, che ha costituito l'oggetto della mia tesi di dottorato di ricerca in Paleografia Greca e Latina. La ricerca si muove sulla scia e nella suggestione del progetto editoriale e culturale, lanciato nel 2001 dal Centro italiano di studi sull'alto medioevo di Spoleto, di costituzione del Corpus delle IMAI (Inscriptiones Medii Aevi Italiae, sec. VI-XII) e mira a contribuire alla conoscenza del patrimonio storico e paleografico pertinente al territorio e all'arco temporale presi in considerazione. La ricerca, inoltre, ha voluto offrire una concreta riflessione personale sulle problematiche teoriche e metodologiche sottese all'informatizzazione di questa tipologia di documenti; a tale scopo, è stato realizzato un database sperimentale in cui sono stati riversati i dati desunti dall'analisi delle circa 100 epigrafi che compongono il corpus, in ciò tenendo conto dei database epigrafici disponibili on-line al momento in cui fu avviata la ricerca.

Esporrò dunque i criteri che hanno guidato la selezione e l'analisi del corpus e l'organizzazione del materiale sia in un catalogo cartaceo sia nel database. In questa sede sarà dedicata particolare attenzione alla descrizione del database, concepito come strumento di archiviazione e di ricerca dei dati. La maschera di data-entry consente di inserire, per ciascuna epigrafe, immagini e record contenenti i dati necessari alla descrizione di un'iscrizione in tutti i suoi aspetti costitutivi, testuali e materiali, secondo regole e criteri di classificazione prestabiliti. In questa sede, ci si soffermerà in modo particolare sulle problematicità emerse nella descrizione formale della paleografia delle iscrizioni del corpus, prive nel complesso di canoni grafici autonomi ed omogenei e per le quali, pertanto, risulta difficile individuare criteri oggettivi di classificazione basati sul rilevamento di costanti nello spazio e nel tempo.

Zusammenfassung

Der Beitrag stellt einige methologische Überlegungen vor, die sich aus der Erstellung einer Datenbank zum Corpus der Inschriften des oberen Latium (Provinz Viterbo)

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 123–134.

aus dem 6.-12. Jahrhundert ergeben. Das Projekt ist Teil der Forschungs- und Editionsarbeiten des Centro italiano di studi sull'alto medioevo in Spoleto und der Arbeit am Corpus der Inscriptiones Medii Aevi Latini. Bei der Erstellung einer experimentellen Datenbank zur Analyse der etwa 100 Inschriften des Corpus ist ein Onlinedatenbank der Inschriften entstanden, die zu theoretischen und methodologischen Überlegungen anregten.
Der Beitrag stellt die Kriterien vor, die bei Auswahl und Analyse des Corpus und der Organisation des Material sowohl als gedruckter Katalog wie als Datenbank angewendet wurden. Die Datenbank ist als Archivierungs- und Recherecheinstrument erstellt worden. Die Datenbank erlaubt es, zu jeder Inschrift ein Bild und zentrale Beschreibungsdaten zu erfassen. Die Beschreibungsdaten zu Text und Material folgen einem etablierten Schema. Diese formale Beschreibung ist besonders problematisch für die Paläographie der Inschriften im Corpus, die keine unabhängigen und homogenen Modelle besitzt, so daß es schwierig wird über Zeit und Raum stabile Klassifikationsmerkmale zu entwickeln.

Abstract

In this paper I aim to present some methodological observations on the study of a corpus of Latin Medieval inscriptions found in the Northern Latium (Viterbo's province, Italy) dating between the 6th and the 12th century, that I carried out during my PhD research in Greek and Latin Palaeography. This work is inspired by the cultural and editorial initiative of constituting the IMAI Corpus (Inscriptiones Medii Aevi Italiae), launched in 2001 by the Centro italiano di studi sull'alto medioevo di Spoleto, and it aims at achieving a better comprehension of the inscription of the historic and paleographical framework concerning this area and period. Furthermore, this research offers a personal and solid consideration about some methodological questions involved in the formalization of this type of documents. For this purpose, an experimental database has been created to collect all corpus' data, more than 100 items, keeping an eye to the epigraphic on-line database accessible to public when the research started.
I will explain the criteria through which the corpus was selected, analyzed and arranged at the same time in a repertoire of paper-form and in the database. The paper for this anthology focuses on the database, conceived as a tool for datastorage and research. The input mask allows to add, for each item, images as well as description records of an inscription in all its constitutive aspects, textual and material, according to predetermined rules and classification criteria. Particular attention will be given to the difficulties encountered for the formal description of the Palaeography of the corpus' inscriptions, since the lack of homogeneous and specific graphical

canons make it difficult to provide objective criteria for classification based on the identification of constants in time and space.

1. Introduzione

Da poco più di un decennio, il censimento e lo studio delle iscrizioni medievali d'Italia è divenuto oggetto di una ricerca programmatica che ha dato corpo alla collana delle *Inscriptiones Medii Aevii Italiae (saecc. VI–XII)*, istituita su iniziativa del Centro italiano di studi sull'Alto medioevo di Spoleto per rispondere alla sentita esigenza di un *corpus* epigrafico italiano che si affiancasse ad analoghi *corpora* di altri paesi europei. Di questo *corpus*, organizzato su base territoriale moderna, esistono oggi tre volumi al primo dei quali, edito nel 2002 e dedicato alle iscrizioni del settore sud-orientale della provincia di Viterbo (*IMAI* 1), sono seguiti quelli relativi alle epigrafi della provincia di Terni (*IMAI* 2), per l'Umbria, e di Belluno, Treviso, Vicenza (*IMAI* 3), per il Veneto, pubblicati, rispettivamente, nel 2009 e nel 2012. Si muove sulla scia e nella suggestione di questa iniziativa l'analisi di un *corpus* di iscrizioni condotta per l'elaborazione della tesi di Dottorato in Paleografia Latina (Pallottini 2009–2012). All'epoca in cui fu avviata la ricerca, erano in corso di ridefinizione, da parte del Comitato scientifico di *IMAI*, i criteri generali stabiliti per il primo volume della collana secondo le direttive impartite a quel tempo dai membri del Comitato (*IMAI* 1; *IMAI* 2). Ciò ha offerto l'occasione per elaborare una riflessione personale sui criteri metodologici che meglio rispondessero all'obiettivo che il repertorio in oggetto si era proposto: contribuire alla conoscenza qualitativa e quantitativa del patrimonio epigrafico medievale dell'alto Lazio, attraverso il censimento, l'analisi e l'informatizzazione di un *corpus* di epigrafi della provincia di Viterbo. Accanto alla costituzione di un repertorio cartaceo, infatti, è stata approntata una banca dati informatica sperimentale che permette di archiviare i dati ricavati dall'analisi delle singole testimonianze e di eseguire ricerche basate su molteplici parametri.

La sistemazione del repertorio in chiave informatica è motivata fondamentalmente dall'interesse di avviare una riflessione concreta sui possibili metodi – concettuali prima ancora che tecnici – relativi alla formalizzazione dei dati concernenti le iscrizioni medievali. Per le testimonianze epigrafiche di questo periodo, infatti, non sono ancora disponibili su larga scala strumenti tecnologici e digitali analoghi a quelli implementati per la gestione e la fruizione dei dati concernenti le iscrizioni greche e latine di datazione anteriore al VII sec. d.C. Tale documentazione epigrafica è liberamente consultabile in rete attraverso banche dati appartenenti alla Federazione internazionale *Electronic Archive of Greek and Latin Epigraphy* (EAGLE), costituitasi nel 2003 con l'obiettivo di costituire un archivio generale dell'epigrafia greca e latina di età classica mediante la creazione di un portale condiviso che riunisse i database

epigrafici allora esistenti; fanno parte attualmente della Federazione textitEAGLE l'*Epigraphic Database Roma* (EDR), l'*Epigraphic Database Bari* (EDB), la *Epigraphische Datenbank Heidelberg* (EDH) e, dall'anno 2009, *Hispania Epigraphica* (HE). Tali banche dati epigrafiche sono strutturate tutte secondo lo standard *EpiDoc*, basato sul linguaggio di archiviazione XML e convenzionalmente adottato per la codifica digitale e per lo scambio dei testi in ambito epigrafico e papirologico (EpiDoc).

Nel campo dell'epigrafia medievale si segnalano singoli progetti, diversi tra loro per tematiche e per stadio attuale di sviluppo, concernenti il trattamento informatico e digitale di determinati *corpora* di iscrizioni: oltre alla digitalizzazione del *corpus* di epigrafi medievali della città di Padova dell'Università di Padova, si ricorda la piattaforma informatica, già attiva in rete, nata dal progetto *Da Canusium tardoantica a Bari medievale: le radici della Puglia moderna* dell'Università di Bari e basata sullo standard *EpiDoc* per quanto riguarda l'edizione dei testi epigrafici.

Alla luce, dunque, della recente implementazione di progetti e della moltiplicazione delle occasioni di dibattito sull'applicazione di risorse tecnologiche e digitali al trattamento dei dati relativi alle iscrizioni medievali, si è ritenuto interessante elaborare un database informatico sperimentale che permettesse di archiviare, di ricercare e di visualizzare le informazioni del *corpus* di epigrafi in esame e, soprattutto, di focalizzare alcune problematiche sottese alla formalizzazione dei dati delle iscrizioni, con particolare attenzione all'aspetto paleografico.

2. Metodologia del censimento e organizzazione dei dati

Nella prima fase della ricerca sono stati stabiliti i parametri fondamentali per la selezione del *corpus*; si è provveduto, cioè, alla delimitazione dei confini geografici e cronologici e alla definizione delle caratteristiche del materiale da analizzare.

L'area presa in considerazione dalla ricerca fa riferimento alla partizione territoriale moderna e coincide con il settore della provincia di Viterbo non considerato all'interno del primo volume delle *IMAI*; dalla sezione del bacino provinciale oggetto dell'indagine è stata esclusa anche la città di Viterbo.

In osservanza ai criteri che guidano la catalogazione nella collana *IMAI*, si è scelto di tenere conto delle testimonianze epigrafiche di datazione compresa tra gli anni 500 e 1199, scritte in qualunque lingua e su qualunque supporto ad eccezione delle monete e dei sigilli (*IMAI* 1, 1, pp. VII–VIII).

Il repertorio è stato concepito come uno strumento per rappresentare la realtà epigrafica attuale del territorio indagato e, pertanto, accoglie le iscrizioni conservate oggi nell'area considerata o attestate qui per l'ultima volta[1]. A ciascuna epigrafe è stata dedicata una scheda di analisi in forma sia cartacea che digitale. Le iscrizioni

[1] Si fa riferimento sia alle iscrizioni andate distrutte sia a quelle perdute o comunque non rintracciabili, queste ultime potenzialmente conservate ancora oggi sul territorio e che bisogna, dunque, continuare a cercare.

provenienti con sicurezza dal territorio in questione ma conservate attualmente fuori da esso sono state raccolte in un *Appendice* articolata in brevi paragrafi contenenti per ciascun pezzo le informazioni essenziali e i principali riferimenti bibliografici. Tale scelta riflette la modalità con cui è stato condotto il censimento delle epigrafi, attraverso una duplice ricognizione bibliografica e sul campo, e, pertanto, è parsa obbligata per conferire alla ricerca carattere il più possibile oggettivo e sistematico: l'unico dato certo, rispondente cioè a criteri di valutazione più oggettivi, è legato all'evidenza materiale di ciò che si conserva sul territorio, motivo per cui come principio operativo e organizzativo della schedatura è stato adottato quello relativo al luogo di conservazione attuale delle iscrizioni o, nel caso di pezzi non rintracciabili, del luogo presso cui furono visti per l'ultima volta. Il censimento, almeno su un piano teorico, deve considerarsi comunque incompleto, stante l'evidente possibilità che una parte della produzione epigrafica medievale superstite sfugga tuttora alla conoscenza, diretta o indiretta che sia, degli studiosi per l'inevitabile fattore di casualità nei rinvenimenti e dei meccanismi stessi della conservazione che condizionano inevitabilmente l'esito di ogni raccolta di materiale condotta su base territoriale. La presenza di materiale epigrafico in centri di cui non si è avuta notizia, dunque, è quantomeno presumibile, come del resto si può anche supporre che alcune iscrizioni considerate perdute siano, in realtà, conservate altrove, prive di un'adeguata documentazione che permetta di rintracciarne il luogo di origine o di rinvenimento; queste, infatti, sono informazioni indispensabili, almeno nella maggior parte dei casi, per ricondurre a un certo ambito territoriale una testimonianza che giace, decontestualizzata, in un luogo diverso rispetto al contesto in cui essa è stata prodotta e utilizzata.

Si è tenuto conto, invece, dell'insieme delle testimonianze provenienti sicuramente dal territorio nella parte di sintesi che precede il catalogo e che offre un quadro sufficientemente rappresentativo, delineato in un'ottica diacronica, delle caratteristiche della produzione epigrafica locale dal punto di vista della consistenza, della distribuzione geografica, della tipologia e del contenuto dei testi e della paleografia, aspetto, quest'ultimo, su cui si tornerà più avanti data la notevole rilevanza che esso riveste tra le problematiche sottese alla formalizzazione dei dati delle iscrizioni.

I criteri utilizzati per la strutturazione delle schede epigrafiche cartacee e per il loro ordinamento all'interno del repertorio cartaceo si ispirano a quelli che guidano la catalogazione nel *corpus* delle *IMAI* (*IMAI* 1, p. VII–VIII; *IMAI* 2, p. IX): le iscrizioni, numerate progressivamente, sono state ripartite secondo i comuni di pertinenza che procedono in ordine alfabetico, e, all'interno delle singole località, in base ai contesti di attestazione, raggruppati per tipologie funzionali (*IMAI* 2, p. IX), e alla cronologia a partire dall'epigrafe con la datazione più alta[2]. Le schede contengono tutte

[2] Rispetto al *corpus* delle IMAI, i raggruppamenti considerati in questo studio sono disposti secondo un ordine diverso e non includono i contesti funzionali dell'edilizia privata e delle strutture assistenziali

le informazioni necessarie all'analisi completa di un'iscrizione; la loro numerazione nel catalogo librario è corrispondente a quella con cui esse sono identificate nella banca dati.

Nel repertorio cartaceo, la scheda di analisi è introdotta da una stringa in cui sono indicati, oltre al numero identificativo dell'epigrafe, il luogo di conservazione o di ultima attestazione, la funzione del testo e la cronologia, espressa in secoli e seguita, quando possibile, da indicazioni utili a una sua determinazione più puntuale. Nella prima parte della scheda trovano posto tutte le informazioni sulla tradizione del pezzo (ubicazione e funzione originaria, luogo di provenienza, eventuali spostamenti), indicando eventualmente l'assenza di notizie al riguardo o il loro carattere particolarmente ipotetico; seguono i dati ricavati dall'analisi del supporto, quelli tecnici relativi alla modalità esecutiva della scrittura, la descrizione paleografica, l'indicazione dell'anno dell'ultima ricognizione e la bibliografia completa sul pezzo. E' fornita, poi, la riproduzione fotografica digitale dell'epigrafe effettuata possibilmente dall'originale e corredata dal rilievo, quando eseguito; seguono la trascrizione e l'edizione del testo con l'apparato critico.

La scheda informatica di *data entry*, a uso del compilatore, si presenta come un foglio-dati che permette di inserire le informazioni necessarie alla descrizione pressoché completa dell'epigrafe, in ciò evidenziando la scelta di rappresentare, attraverso il sistema informatico, il documento epigrafico in tutti i suoi aspetti costitutivi.

La struttura del foglio-dati è articolata in una maschera principale e in una sottomaschera. La prima contiene tutte le informazioni univoche per l'epigrafe: numero identificativo, tipologia contenutistica, cronologia, ubicazione e funzione originaria, stato di conservazione del supporto (integro, integralmente ricomponibile, frammentario – frammento isolato, frammentario – frammenti contigui / parzialmente contigui / solidali), dati tecnico-grafici, bibliografia, trascrizione ed edizione del testo, varianti di lettura e riproduzione fotografica. La sottomaschera, invece, consente l'immissione di dati che possono ammettere valori multipli per la stessa iscrizione. Ciò si è ritenuto indispensabile, infatti, per la schedatura informatica di alcune iscrizioni frammentarie che richiedono la registrazione di dati diversi per ciascun elemento in cui è diviso il supporto: differenti, infatti, oltre alle misure, possono essere lo stato di reperibilità o di irreperibilità del singolo elemento, il luogo di conservazione o di ultima attestazione e l'anno dell'ultima ricognizione[3]. I dati immessi in questi campi, invece, sono univoci se l'iscrizione è integra oppure ricomposta da frammenti contigui conservati nello stesso luogo o, infine, quando è costituita da un frammento isolato.

Per questioni di praticità, considerato il numero sostanzialmente limitato di iscri-

poiché non attestati tra i casi esaminati.

[3] E' esemplificativo a tale proposito il caso di un'epigrafe proveniente da Orte composta da due frammenti non contigui conservati in luoghi diversi.

zioni che compongono il *corpus* in questione, si è scelto di inserire le immagini delle epigrafi in forma di allegato.

Oltre alla riproduzione fotografica dell'iscrizione è fornita anche la scansione digitale del rilievo, quando disponibile; esso costituisce la prima tappa del processo di analisi di un testo epigrafico da parte dell'editore che, attraverso il disegno, rende esplicita la personale lettura e interpretazione dei segni grafici che lo compongono.

Un campo della maschera principale è riservato all'inserimento della trascrizione, seconda tappa fondamentale del processo di codifica del testo iscritto attraverso l'uso di segni convenzionali prestabiliti; la sua funzione è quella di rendere pienamente intellegibile la lettura dell'epigrafe fornita dall'editore e di chiarire, quindi, le scelte operate nella fase successiva di edizione dell'epigrafe.

Prima di procedere oltre, è opportuno soffermarsi brevemente sulle norme di restituzione delle epigrafi adottate nel presente lavoro; esse compendiano in un unico sistema quelle utilizzate nel primo volume delle *IMAI* e il sistema *Krummrey-Panciera*, sviluppato per l'edizione delle iscrizioni di età classica ed estese oggi, con i necessari adattamenti, anche a quella delle epigrafi medievali di cui, secondo quanto previsto dalle stesse norme, è stata abolita la trascrizione (*IMAI* 2, p. X; Panciera 1991, pp. 9–21). Essa, invece, è stata preservata per la restituzione dei testi delle iscrizioni del *corpus* in oggetto, in ciò tenendo conto della sua utilità anche in rapporto all'applicazione informatica che permette di ricercare parole nella forma che esse presentano nel documento originario, preventivamente trascritto. I testi sono trascritti in carattere maiuscolo utilizzando il font Cardo (*IMAI* 2, p. X); sono trascritte in maiuscolo anche le lettere che originariamente hanno forma minuscola, e così anche le lettere sovrascritte, incluse o inserite in interlinea, scritte di seguito all'interno del binario di scrittura. Tutte le singole parole sono separate da uno spazio bianco anche in presenza di *scriptio continua*. E' rispettata l'impaginazione originaria del testo, a meno che le lettere o le parole non siano disposte in colonna; in tal caso, il testo è trascritto di seguito indicando i cambi di riga con la barra verticale singola; la doppia barra segnala il cambio di pagina o di lato di un monumento. Le abbreviazioni e i segni di interpunzione, descritti preventivamente nel commento paleografico della scheda epigrafica cartacea, sono segnalati, rispettivamente, con il tratto soprascritto e con un punto medio, a prescindere dalla particolare forma che essi presentano nel documento epigrafico; a questo proposito, è quantomeno opportuno sottolineare che la trascrizione costituisce uno strumento interpretativo del testo epigrafico e, dunque, non ha alcuna pretesa di riprodurre in forma mimetica le caratteristiche grafiche del documento originario. Le lacune sono segnalate mediante il segno diacritico convenzionale della parentesi quadra; per la specificazione della sua tipologia sono stati adottati, con alcune semplificazioni, i criteri introdotti con il secondo tomo delle *IMAI*.

L'edizione costituisce l'ultima tappa del processo di codifica dell'iscrizione. Il testo è digitato in carattere minuscolo tondo del font Cardo; viene ripristinato l'uso delle

maiuscole e della punteggiatura secondo i criteri moderni; i rimandi a capo riga sono segnalati dal consueto trattino. Vengono emendate le inversioni di lettere ma non le trascrizioni fonetiche e tutti i possibili esiti grammaticali, sintattici e fonetici degli usi linguistici dovuti alle consuetudini linguistiche o al livello di educazione grafica dello scrivente. Lo scioglimento delle abbreviazioni è reso tra parentesi tonde quando segnalate dal segno, tra parentesi uncinate se prive del segno.

Nel campo «edizione normalizzata», necessario per la ricerca basata sul parametro testuale e non visibile nelle schede digitali di visualizzazione dei risultati di una ricerca, è inserita l'edizione dell'epigrafe priva dei segni diacritici eventualmente presenti.

3. La formalizzazione dei dati: metodo e problematiche

Per l'efficace gestione dei dati da parte del programma informatico, ogni campo richiede l'immissione di una singola, precisa informazione sull'iscrizione. Nonostante il carattere sperimentale del database, ad accesso riservato e compilabile da un solo estensore, la formalizzazione dei dati è parsa comunque assolutamente necessaria, in particolare per quelle informazioni che costituiscono i parametri per eseguire una ricerca all'interno della banca dati, cioè:

- luogo di provenienza;
- luogo di conservazione;
- tipologia contenutistica;
- datazione;
- tipologia grafica;
- testo.

La classificazione degli aspetti descrittivi del documento epigrafico non ha sollevato particolari problemi per quelle informazioni che, anche nelle schede cartacee, fanno riferimento a classificazioni o a terminologie prestabilite o comunque uniformi quali la tipologia contenutistica, la cronologia, la materia del supporto, il suo stato di conservazione, la tecnica di esecuzione dell'epigrafe e la tipologia del solco, nel caso in cui l'iscrizione sia incisa; la compilazione di questi campi è eseguita di volta in volta attraverso la selezione di una voce da liste predefinite, informatizzate come menù a tendina e redatte *ad hoc* sulla base delle caratteristiche della documentazione epigrafica analizzata. In altri campi, invece, l'immissione del testo è libera ma avviene secondo regole prestabilite, per conferire anche alla compilazione di queste voci carattere il più possibile uniforme e rigoroso.

Per consentire l'esecuzione di una ricerca basata anche sul parametro paleografico, è stato necessario associare un elenco di valori al campo descrittivo della tipologia grafica. Si tratta di un aspetto estremamente problematico da formalizzare in ragione dell'attestazione, nell'epigrafia di questo periodo, di scritture non inquadrabili il

più delle volte in una specifica tipologia scrittoria e, spesso, misti di lettere desunte da modelli grafici diversi tra loro. La sistematizzazione delle scritture epigrafiche, necessaria al trattamento informatico del dato paleografico, è risultata possibile ed efficace soltanto per quelle che fanno riferimento a determinate tipologie scrittorie descrivibili mediante la terminologia della nomenclatura grafica tradizionale. Nella maggior parte dei casi, però, il vocabolario paleografico disponibile è inadeguato alle esigenze descrittive delle scritture epigrafiche che, come già si è detto, si caratterizzano spesso per restituire sistemi grafici complessi, valutabili singolarmente in base alle specificità che essi di volta in volta presentano. La documentazione epigrafica del viterbese analizzata nell'ambito di questa ricerca si caratterizza per restituire scritture graficamente molto variegate, databili, talvolta, in modo alquanto approssimativo sulla base esclusivamente dell'analisi paleografica e dei confronti istituibili con altre iscrizioni affini. Tale difficoltà si pone soprattutto in relazione alle testimonianze iscritte di età altomedievale, frutto per lo più di sperimentalismi condotti all'interno di botteghe artigianali o in ambienti esterni a queste con esiti grafici complessivamente dissimili tra loro poiché direttamente connessi all'abilità tecnica e alla cultura grafica del lapicida e, pertanto, non riconducibili nella maggior parte dei casi ad una determinata tipologia scrittoria; dall'analisi tecnico-grafica di alcune iscrizioni emerge l'impressione di una conoscenza, da parte del lapicida, degli elementi stilistici e formali di un determinato modello grafico e, allo stesso tempo, l'assenza delle capacità e degli strumenti tecnici necessari per realizzarlo.

In linea generale, l'unica costante che sembra potersi individuare sotto l'aspetto paleografico nella produzione epigrafica di livello qualitativo più o meno modesto di questo periodo è proprio l'assenza di regole grafiche precise nella realizzazione delle scritture epigrafiche, prive per lo più di canoni formali autonomi. A rendere assai problematica l'individuazione di categorie generali attraverso le quali classificare, in modo sufficientemente rappresentativo e non equivoco, tutte le differenti, possibili espressioni grafiche restituite dalle iscrizioni ascrivibili all'arco temporale considerato in questo studio, concorrono: la variabilità di forme e di disegni diversi di una stessa lettera all'interno di un'unica epigrafe, l'attestazione della medesima forma di una lettera in iscrizioni che restituiscono scritture complessivamente anche molto dissimili tra loro e variamente databili, nonché la mancanza di riferimenti cronologici interni o esterni – facile a riscontrarsi anche a causa delle cattive condizioni di conservazione dei pezzi – che obbliga spesso a ricavare la datazione di un'epigrafe soltanto dall'analisi paleografica e dai confronti istituibili con altre testimonianze affini per tecnica esecutiva e per contesto culturale e geografico. Delle circa 50 iscrizioni che compongono il *corpus*, soltanto alcune possono essere ricondotte a una determinata tipologia grafica; più spesso, come si è accennato, si rilevano scritture ibride di lettere desunte da sistemi grafici diversi o per le quali il riferimento a un determinato modello può essere colto soltanto in termini di generiche assonanze. Nella banca dati

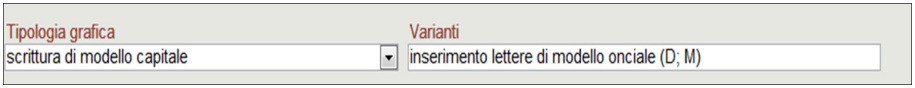

Figura 1. Campi per l'immissione dei dati paleografici. (Palaeographic data input fields).

approntata per le epigrafi di questo *corpus*, la sistematizzazione dei dati concernenti la paleografia delle iscrizioni ha condotto alla realizzazione di un elenco, in forma di menù a tendina, dal quale selezionare la voce che meglio descrive la scrittura nel complesso e di un campo a testo libero, denominato «varianti», che permette di indicare le eventuali intrusioni di lettere di forma o di modello diversi rispetto a quello prevalente. Le definizioni inserite nella lista fanno riferimento quando a determinate tipologie grafiche quando alle categorie generali di «maiuscola» e di «minuscola» eventualmente associate a un modello grafico di riferimento, per cui è possibile scegliere tra:

- scrittura di modello capitale;
- scrittura di modello onciale;
- scrittura mista di lettere capitali e onciali, da utilizzarsi nel caso in cui essa presenti una mescolanza di forme desunte dai due modelli scrittori;
- scrittura maiuscola, da utilizzarsi nel caso in cui essa presenti forme grafiche comprese nello schema bilineare e non riconducibili ai canoni formali di un modello;
- scrittura minuscola;
- scrittura mista di forme maiuscole e minuscole.

Un esempio che mostra l'immissione dei dati nei campi descrittivi del sistema grafico è offerto nella figura sottostante:

Per il campo «varianti», occorre ripeterlo, è utilizzato un lessico descrittivo non formalizzato e, pertanto, esso non è stato incluso tra i parametri sui quali basare la ricerca; per quanto riguarda i dati paleografici, la maschera di ricerca permette soltanto di selezionare una voce dal campo «tipologia grafica», mentre le «varianti» grafiche eventualmente presenti sono segnalate nella scheda digitale di visualizzazione dei risultati della ricerca (fig. 2). E' chiaro che una ricerca impostata soltanto su definizioni generali di scritture spesso altamente complesse, in nessun modo può considerarsi effettivamente indicativa del dato paleografico, pur tenendo conto della segnalazione delle «varianti» concepita come necessario completamento della descrizione generale di una scrittura; a tale proposito, è doveroso evidenziare l'assenza nella banca dati di un sistema di archiviazione degli altri aspetti parimenti importanti per la descrizione di una scrittura, quali quelli relativi all'apparato dei segni interpuntivi e abbreviativi o agli elementi riservati al trattamento stilistico delle lettere. Del resto, non soddisfa

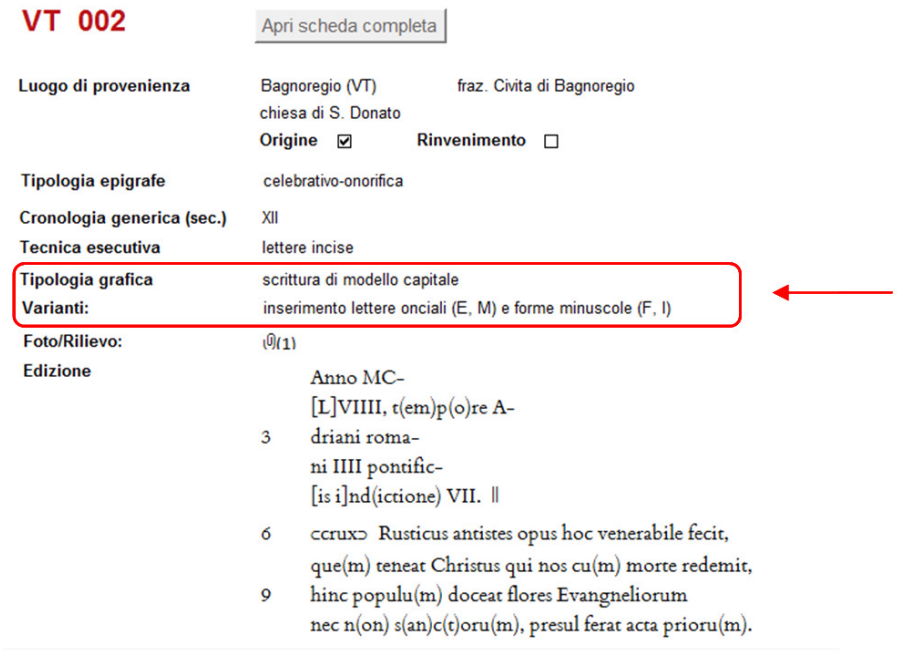

Figura 2. Scheda sintetica di visualizzazione dei risultati di una ricerca; evidenziati in rosso, i dati paleografici restituiti dai campi, complementari, «tipologia grafica» e «varianti». (Summarized display form of research results; in red, the palaeographic data from the complementary fields «graphical typology» and «variables»).

appieno neppure il sistema di categorie generali elaborato per la definizione generale delle scritture poiché soggetto a un margine di interpretazione troppo ampio: manca, ad esempio, un criterio proporzionale che permetta di distinguere una scrittura di impianto prevalentemente capitale con intrusioni di lettere onciali da un'altra che potrebbe essere considerata ibrida di lettere desunte dai due modelli grafici, come anche puramente discrezionale può essere la distinzione tra maiuscole e minuscole, considerata, ad esempio, la frequente attestazione di lettere di forma minuscola comprese nello schema bilineare e, dunque, utilizzate come maiuscole.

4. Conlusioni

La ricerca presentata in questa sede è stata concepita e realizzata con l'intento di contribuire al progetto di schedatura delle iscrizioni medievali d'Italia e, per quanto

riguarda l'aspetto relativo all'approntamento del database, per riflettere concretamente sulle problematiche sottese alla formalizzazione dei dati, soprattutto di quelli paleografici, nella prospettiva di avviare un confronto con i singoli, diversi progetti già esistenti inerenti l'informatizzazione di determinati *corpora* di epigrafi di datazione post-classica. Proprio la formalizzazione dei dati sulla paleografia delle iscrizioni è, si ritiene, l'aspetto per il quale si rivela di importanza fondamentale la discussione tra gli studiosi delle diverse discipline che fanno della scrittura e del suo supporto l'oggetto dei propri studi. Tale confronto, infatti, costituisce il presupposto necessario affinché le soluzioni di volta in volta proposte, secondo differenti approcci teorici e in funzione di determinate problematiche, per il trattamento informatico di una specifica documentazione epigrafica possano tradursi in piste di ricerca per lo sviluppo di strumenti tecnologici e digitali che affianchino e potenzino la metodologia tradizionale di analisi e di fruizione delle testimonianze epigrafiche di epoca medievale.

Bibliografia

Corpus delle epigrafi medievali di Padova. Padova: Università di Padova. <http://www.cem.dissgea.unipd.it>.

Da Canusium tardoantica a Bari medievale: le radici della Puglia moderna. Bari: Università di Bari. <http://www.imaap.uniba.it>.

EAGLE: *Electronic Archive of Greek and Latin inscriptions*. <http://www.eagle-eagle.it>.

EDB: *Epigraphic Database Bari*. <http://www.edb.uniba.it>

EDH: *Epigraphische Datenbank Heidelberg*. <http://www.uni-heidelberg.de>.

EDR: *Epigraphic Database Roma*. <http://www.edr-edr.it>.

EpiDoc: *Epigraphic Documents*. <http://www.epidoc.sourceforge.net>.

HE: *Hispania Epigraphica*. <http://www.eda-bea.es>.

IMAI 1: *Inscriptiones Medii Aevi Italiae*, 1, Viterbo – 1, a cura di L. Cimarra, E. Condello, L. Miglio, M. Signorini, P. Supino, C. Tedeschi, Spoleto: Centro Italiano di Studi sull'Alto Medioevo, 2002.

IMAI 2: Guerrini, Paola, *Inscriptiones Medii Aevi Italiae, 2, Umbria, Terni*, Spoleto: Centro Italiano di Studi sull'Alto Medioevo, 2009.

Felle, Antonio Enrico, «Esperienze diverse e complementari nel trattamento digitale delle fonti epigrafiche: il caso di EAGLE e di EpiDoc», *Diritto romano e scienze antichistiche nell'era digitale. Convegno di studio (Firenze, 12–13 settembre 2011)*, Torino 2012, pp. 117–130.

Pallottini, Elisa, *Iscrizioni medievali della provincia di Viterbo (secc. VI-XII): censimento, analisi, informatizzazione*, tesi di Dottorato, Università di Roma «La Sapienza», a.a. 2009–2012.

Panciera, Silvio, «Struttura dei supplementi e segni diacritici. Dieci anni dopo», *Supplementa italica* 8 (1991), pp. 9–22.

Appendices

Kurzbiographien – Biographical Notes

Bernhard Assmann studierte an der Universität zu Köln Informationsverarbeitung, Mittlere und Neuere Geschichte und Historische Hilfswissenschaften. Danach betreute er das Digitalisierungsprojekt «Die Werke Friedrichs des Großen» an der Universitätsbibliothek Trier. Gegenwärtig ist er beim Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen in Köln beschäftigt.

Matthieu Bonicel graduated from the École des Chartes and obtained a PhD at the Université Paris 1 Panthéon Sorbonne focused on a digital approach of performing arts and performing people in Avignon in the late Middle Ages. Since 2007 he is a curator at the Manuscripts department of the Bibliothèque nationale de France, where he is in charge of digitization and information technologies. He took part in several international projects such as the Roman de la Rose digital library and Europeana Regia and he is a member of the International Image Interoperable Framework community, a group of programmers and digital librarians committed to opening access to digital libraries and tools. He is also the coordinator of Biblissima, a digital Observatory for the Written Heritage of the Middle Ages and the Renaissance.

William A. Christens-Barry is the founder of Equipoise Imaging, LLC. He specializes in optical measurement, imaging, and analytical techniques. His dissertation in physics adapted spectroscopic and mathematical techniques to investigate nucleic acid secondary structure and sequences. Dr. Christens-Barry has participated in numerous imaging, optical analysis, and spectral imaging projects, with an emphasis on the special requirements of manuscript and antiquities spectral imaging. He originated the narrowband illumination approach to spectral imaging of manuscripts and textual materials. He has served as a core scientific member of the Archimedes Palimpsest Project and the ongoing St. Catherine's Palimpsest and Dead Sea Scrolls imaging teams, with key responsibilities in the development and use of optical imaging and computational processing techniques.

Oliver Duntze hat an der Universität zu Köln Germanistik, Philosophie und Malaiologie studiert. 2005 promovierte er im Fach Buchwissenschaft mit einer Arbeit über den Straßburger Druckerverleger Matthias Hupfuff zum Dr. phil. 2007 bis 2010 war er im Projekt «Deutsches Textarchiv» an der Berlin-Brandenburgischen Akademie der Wissenschaften tätig. Seit April 2010 ist er wissenschaftlicher Mitarbeiter beim Gesamtkatalog der Wiegendrucke an der Staatsbibliothek zu Berlin

– Preußischer Kulturbesitz. Seine Forschungsschwerpunkte liegen im Bereich der Druck- und Buchhandelsgeschichte des 15. und 16. Jahrhunderts.

Erwin Frauenknecht ist Historiker und hat 1995 an der Universität Regensburg promoviert. Er war Post-doc-Stipendiat an der Universität Gießen, 1998-2004 als wissenschaftlicher Assistent an der Universität Tübingen tätig und hat Lehraufträge an den Universitäten Regensburg, Tübingen und Stuttgart wahrgenommen. Seit 2008 ist er Projektmitarbeiter am Landesarchiv Baden-Württemberg im Bereich der Wasserzeichenforschung (Bernstein-Projekt, Wasserzeichen-Informationssystem WZIS).

Tal Hassner received the M.Sc. and Ph.D. degrees in applied mathematics and computer science from the Weizmann Institute of Science in 2002 and 2006, respectively. He later completed a postdoctoral fellowship, also at the Weizmann institute. In 2006, he joined the faculty of the Department of Mathematics and Computer Science, The Open University of Israel, where he currently holds a Senior Lecturer position (Assistant Professor). Tal Hassner was awarded a full graduate student scholarship, by the Feinberg Graduate School, the best student paper award at the IEEE International Conference on Shape Modeling and Applications, and his work on approximate nearest subspace search was selected as the editor's choice for a spotlight paper in IEEE Transactions on Pattern Recognition and Machine Intelligence (TPAMI). Tal Hassner has co-organized the Schloss Dagstuhl - Leibniz Center for Informatics, Perspective Workshop on «Computation and Palaeography: Potentials and Limits», in 2012, and is co-organizer of the Schloss Dagstuhl - Leibniz Center for Informatics, Seminar on «Digital Palaeography: New Machines and Old Texts», 2014.

Fabian Hollaus has studied Visual Computing at the Vienna University of Technology. He is working since 2011 in the «The Enigma of the Sinaitic Glagolitic Tradition» project at the Computer Vision Lab, Institute of Computer Aided Automation, Vienna University of Technology. His research areas are document analysis and cultural heritage applications.

Melanie Gau has specialised in Computational Linguistics and Palaeoslavic Studies. Currently she is taking part in the Austrian Science Fund project «Critical Edition of the New Sinaitic Glagolitic Euchology (Sacramentary) Fragments with the Aid of Modern Technologies» and is writing her PhD-thesis on the «Psalterium Demetrii Sinaitici (Sin. slav. 3/N)».

Heinz Miklas is former professor at the Institute of Slavic Studies at the University of Vienna. He is mainly engaged in the comparative history of Slavic writing systems, Slavonic codicology and palaeography and the edition of Glagolitic and Cyrillic

texts. He was head of the Austrian Science Foundation project «The Enigma of the Sinaitic Glagolitic Heritage», is founder of the *Vienna Archaeographic Forum* (WAF), and one of the founders of the *Centre of Image and Material Analysis in Cultural Heritage* (CIMA).

Elisa Pallottini has a PhD in Greek and Latin Palaeography from *La Sapienza* University of Rome and her main research area lies in Latin Medieval Epigraphy, with particular regard to the Palaeography of Latin inscriptions. She took part in the research project *Titulus_base des données épigraphique* (*Centre d'Etudes Supérieures de Civilization Médiévale, University of Poiters*), within the investigation methodology phase aimed at the formalization of data from the Corpus of Medieval French inscriptions. She is currently employed at *La Sapienza* University of Rome as Post-Doctoral researcher in Latin Palaeography.

Malte Rehbein is Professor and Chair of Digital Humanities at the University of Passau. He has published on manuscript studies, digital editions, text encoding, information visualisation, and pedagogy. He is editor-in-chief of the Digital Medievalist Journal and is currently member of the Executive Board of the German-speaking Digital Humanities association DHd.

Robert Sablatnig is an associate professor of computer vision heading the Computer Vision Lab (which was part of the Pattern Recognition and Image Processing Group), and is head of the Institute of Computer Aided Automation, engaged in research, project leading, and teaching. His research interests are 3D Computer Vision, Automatic Visual Inspection, Hierarchical Pattern Recognition, Video data analysis, Automated Document Analysis, Multispectral Imaging, Virtual- and Augmented Reality, and Applications in Industry and Cultural Heritage Preservation. He is Vice President of the Austrian Association for Pattern Recognition (AAPR/OAGM), the Austrian branch of IAPR.

Torsten Schaßan ist Mitarbeiter an der Abteilung Handschriften und Sondersammlungen der Herzog August Bibliothek Wolfenbüttel. Er betreut dort die digitale Erschließung der historischen Bestände, insbesondere die Handschriftenkatalogisierung und digitale Editionen. Er studierte in Köln Mittlere und Neuere Geschichte, Germanistik und Philosophie.

Rombert Stapel is a researcher at the International Institute of Social History in Amsterdam, working at the «Global Collaboratory for the History of Labour Relations, 1500-2000». He is currently finishing his PhD thesis at Leiden University (in a partnership with the Fryske Akademy (KNAW)) which concerns the Croniken van der Duytscher Oirden or Jüngere Hochmeisterchronik. He has published on a stylometric analysis of this chronicle and on the history of the

military orders in general, and is preparing a joint publication on collaborative authorship in the medieval Low Countries.

Maria Stieglecker promovierte an der Universität Wien in Geschichte und absolvierte den Ausbildungslehrgang am Institut für Österreichische Geschichtsforschung. Sie ist wissenschaftliche Mitarbeiterin an der Österreichischen Akademie der Wissenschaften, Institut für Mittelalterforschung, Abteilung Schrift und Buchwesen und betreut den Forschungsschwerpunkt «Wasserzeichen des Mittelalters». Weiters arbeitet sie an verschiedenen nationalen und internationalen Projekten zur Wasserzeichenforschung und Handschriftenkatalogisierung mit.

Peter A. Stokes is Senior Lecturer at the Department of Digital Humanities, King's College London, and Director and Principal Investigator of the ERC-funded DigiPal, «Digital Resource and Database of Palaeography, Manuscripts and Diplomatic». As well as palaeographical method and the application of computing to manuscript studies, his other primary interests include the vernacular English scripts of the late-tenth through twelfth centuries. He has also published on computing in lexicography, Anglo-Saxon charters and bounds, and early-modern book collectors; has developed software for digital humanities; and is currently Director of Digital Medievalist.

Dominique Stutzmann est docteur en histoire, chargé de recherche au Centre National de la Recherche Scientifique (UPR 841 – Institut de Recherche et d'Histoire des Textes, section de paléographie) et chargé de conférences en paléographie médiévale à l'École Pratique des Hautes Études. Il a étudié les lettres classiques, l'allemand et l'histoire à la Sorbonne (universités Paris 1 et Paris 4) et obtenu le diplôme d'archiviste paléographe de l'École nationale des Chartes. Il a été conservateur à la Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (département des Manuscrits) et à la Bibliothèque nationale de France (département de l'Information bibliographique et numérique).

Georg Vogeler ist Assistenzprofessor am Zentrum für Informationsmodellierung in den Geisteswissenschaften der Karl-Franzens-Universität Graz. Er war wissenschaflicher Mitarbeiter am Lehrstuhl für Geschichtliche Hilfswissenschaften in München, wo er bei Prof. Walter Jaroschka, Prof. Walter Koch und Prof. Joachim Wild zum Thema „Spätmittelalterliche Steuerbücher deutscher Territorien" promovierte. Seit 2003 forscht er zur „Digitalen Diplomatik", zum Gebrauch der Urkunden Kaiser Friedrichs II. und zur Digitalen Edition von mittelalterlichem und frühneuzeitlichem Rechnungsschriftgut. Er ist Gründungsmitglied des Instituts für Dokumentologie und Editorik, technischer Direktor des Monasterium-Konsortiums bzw. des ICARus und wirkt in verschiedenen internationalen Projekten mit.

Christine Voth has recently completed her PhD in the Department of Anglo-Saxon, Norse and Celtic at the University of Cambridge. Her thesis, «An Analysis of the Anglo-Saxon Manuscript London, British Library, Royal 12. D. xvii, and its Context in the Re-emergence of Manuscript Culture in Tenth-Century England», is a combined palaeographic, codicological and historical analysis of the earliest surviving medical manuscript from Anglo-Saxon England. Her research interests include the material culture of the early medieval world, particularly the production, use and transmission of medieval manuscripts.

Lior Wolf is a faculty member at the Computer Science Department at Tel-Aviv University. Previously, he was a post-doctoral associate in Prof. Poggio's lab at MIT. He graduated from the Hebrew University, Jerusalem, where he worked under the supervision of Prof. Shashua. Lior Wolf was awarded the 2008 Sackler Career Development Chair, the Colton Excellence Fellowship for new faculty (2006-2008), the Max Shlumiuk award for 2004, and the Rothchild fellowship for 2004. His joint work with Prof. Shashua in ECCV 2000 received the best paper award, and their work in ICCV 2001 received the Marr prize honorable mention. He was also awarded the best paper award at the post ICCV 2009 workshop on eHeritage.

KPDZ 1 – CPDA 1

Kodikologie und Paläographie im Digitalen Zeitalter / Codicology and Palaeography in the Digital Age, hg. v. Malte Rehbein, Patrick Sahle und Torsten Schaßan unter Mitarbeit von Bernhard Assmann, Franz Fischer und Christiane Fritze. Schriften des Instituts für Dokumentologie und Editorik 2. Norderstedt: Books on Demand, 2009.
ISBN 978-3-8370-9842-6
Online: <http://kups.ub.uni-koeln.de/volltexte/2009/2939/>

Der gedruckte Band kann zum Preis von € 49,- über den Buchhandel, über amazon.de und über die Webseite des Verlages bezogen werden:
<http://www.bod.de/index.php?id=1132&objk_id=217805>.

You can order the printed version at the price of € 49,- from your local bookstore, via amazon.de or via the website of the publishing house:
<http://www.bod.de/index.php?id=1132&objk_id=217805>.

Beiträge – Contributions

KPDZ 2 – CPDA 2

Kodikologie und Paläographie im Digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2, hg. v. Franz Fischer, Christiane Fritze und Georg Vogeler unter Mitarbeit von Bernhard Assmann, Malte Rehbein und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 3. Norderstedt: Books on Demand, 2010.
ISBN 978-3-8423-5032-8
Online: <http://kups.ub.uni-koeln.de/4337/>

Der gedruckte Band kann zum Preis von € 59,- über den Buchhandel, über amazon.de und über die Webseite des Verlages bezogen werden:
<http://www.bod.de/index.php?id=296&objk_id=477356>.

You can order the printed version at the price of € 59,- from your local bookstore, via amazon.de or via the website of the publishing house:
<http://www.bod.de/index.php?id=296&objk_id=477356>.

Beiträge – Contributions