



OLIVER DUNTZE TORSTEN SCHASSAN GEORG VOGELER (HRSG. | EDS.)

KODIKOLOGIE & PALÄOGRAPHIE IM DIGITALEN ZEITALTER 3  
CODICOLOGY & PALAEOGRAPHY IN THE DIGITAL AGE 3

SCHRIFTEN DES INSTITUTS FÜR DOKUMENTOLOGIE UND EDITORIK 10

# Schriften des Instituts für Dokumentologie und Editorik

---

herausgegeben von:

Bernhard Assmann	Alexander Czmiel
Oliver Duntze	Franz Fischer
Christiane Fritze	Ulrike Henny
Malte Rehbein	Patrick Sahle
Torsten Schaßan	Markus Schnöpf
Martina Semlak	Philipp Steinkrüger
Georg Vogeler	

Band 10

# Inhaltsverzeichnis

Vorwort

Preface

Oliver Duntze  
Einleitung

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

Computation and Palaeography: Potentials and Limits

## **Digitale Reproduktion als paläographisches Werkzeug**

### **Digital imaging as a palaeographic tool**

Fabian Hollaus, Melanie Gau, Robert Sablatnig, William  
A. Christens-Barry, Heinz Miklas  
Readability Enhancement and Palimpsest Decipherment  
of Historical Manuscripts

Christine Voth

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

## **Verwaltung von Erschließungsdaten Organizing descriptive information**

Rombert Stapel

The development of a medieval scribe

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

Erwin Frauenknecht, Maria Stieglecker  
WZIS – Wasserzeichen-Informationssystem: Verwaltung  
und Präsentation von Wasserzeichen und ihrer  
Metadaten

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

## **Appendices**

Kurzbiographien – Biographical Notes

KPDZ 1 – CPDA 1

KPDZ 2 – CPDA 2

# 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 Herausbergremium und einem anonymisierten Peer-reviewing externer Fachgutachter unterzogen.

Verweise zu Webseiten und Online-Ressourcen 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 Feinheiten 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).<sup>1</sup> 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.

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 Multispektalfotografie 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-Einleitung XI tische Auswertung verschiedener paläographischer Merkmale in der Transkription der Handschriften Henrik van Vianens ermöglicht es, einen skripturalen »Fingerabdruck« des Schreibers zu erstellen.<sup>2</sup> 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<sup>3</sup> 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* 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 paläogra-phischer, 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.

---

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

Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. IX–XII.

- 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>

# **Computation and Palaeography: Potentials and Limits<sup>4</sup>**

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 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:

- A clear articulation of standards for digital image acquisition followed by all digital imaging projects when possible.

- 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.
- 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.
- 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 method-ologies. 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?*