Kodikologie und Paläographie im digitalen Zeitalter 2

Codicology and Palaeography in the Digital Age 2

herausgegeben von | edited by

Franz Fischer, Christiane Fritze, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Malte Rehbein, Patrick Sahle

Introduction. Into the Wide – Into the Deep: Manuscript Research in the Digital Age

Franz Fischer, Patrick Sahle

Manuscript research is a wide field of scholarship which is integrated in core disciplines such as history, philology, or library science. Yet manuscript research is also crucial in other fields such as archaeology, history of arts, musicology or Egyptology, to name but a few. For all these disciplines, manuscripts are fundamental sources. There are different approaches to different types of manuscripts, but questions and perspectives, methodologies and tools are often quite similar. Innovations and new research strategies from one discipline can be transferred to and adopted by others. This introduction gives an overview of current aspects in the field of manuscript studies in both theory and practice by showing the relatedness of the contributions to the volume at hand as well as its predecessor, *Codicology and Palaeography in the Digital Age* (references given in parentheses). The texts are roughly assigned to five interrelated areas of manuscript research: (I) the photographic capturing of the manuscript surface, (II) the description of the manuscript for a catalogue, (III) the scientific examination of material aspects, (IV) the analysis of the script and (V) the deep encoding of the text itself.

I. Digital Reproduction

These days, the starting point for manuscript research projects is often digital reproduction. Digital facsimiles convey a great number of the original features and characteristics and can be easily provided and shared. Carried out on a large scale, digital reproduction is the cheapest way of making entire collections of manuscripts accessible. The criteria for selection vary according to research interests and institutions (cf. Kalning and Zimmermann in vol. 1).

The opening chapter of the first section, written by *Pádraig Ó Macháin*, is dedicated to one of the early digitisation projects, Irish Script on Screen (ISOS), initiated back in 1998 by the School of Celtic Studies at the Dublin Institute for Advanced Studies. This project's objective was, and still is, to digitise the entire Gaelic manuscript tradition—that is, all manuscripts in the Irish language—across all libraries and archives and to make the digital images freely available on the World Wide Web.

Research in art history can be greatly facilitated by means of manuscript facsimiles provided along with codicological data and descriptive texts. This can be more easily

achieved in the digital medium, as demonstrated by *Armand Tif,* whose chapter focuses on illuminated manuscripts from two particular monastic libraries in Austria. With the project described by *Alison Stones* and *Ken Sochats*, we leave the modern repository as the main organisational concept in favour of the manuscript tradition of one particular work, namely the popular Arthurian romance known as the Lancelot-Grail. Here, images and text are presented according to the narrative structure as well as the geographic dissemination of the manuscript witnesses. In a pre-analytical manner, this gives a promising starting point for comparative investigations beyond the flat surface of the reproduction.

An important caveat to manuscript research based on digital surrogates is articulated by *Melissa Terras*. Technical distortions can lead to the unintentional introduction of artefacts and errors into the digital representations of objects. These chapters show that digitisation is more than just a technical endeavour; it needs a methodology, and theoretical reflection upon the intersection of technical conditions and the requirements of critical scholarship.

II. Digital Catalogue and Semantics

On the other hand, any collection of digitised manuscripts would be of very limited use without inventories and catalogues indicating the content, material and provenance of each item in a particular collection. The new access to manuscript research by digital reproduction is still accompanied by a more traditional cataloguing approach. Codicology has always been the compilation and generation of knowledge about a manuscript, and cataloguing has been the dominant way of recording this knowledge. The creation of digital catalogues is increasingly common practice today, just as handwritten and print catalogues were common practice in archives and libraries before this (cf. Bernardi et al.; Cartelli et al.; Speer). Several software tools have emerged recently to facilitate this (for a description of just one such example, the M-Tool, see Uhlíř and Knoll). An adapted version of the word processing program TUSTEP, described by *Silke Schöttle* and *Ulrike Mehringer*, proved to be an appropriate tool for the creation of the online catalogue for the special collections of the Tübingen University Library.

Converting knowledge from analogue to digital is not just a technical issue of how to do this as quickly and effectively as possible. Rather, it sets its own methodological agenda. One of the changes occurring in codicology, in comparison to traditional print cataloguing, is the relationship between the codex (or its visual digital surrogate) and its description (cf. Stinson) and therefore the description itself.

Moreover, there is a tendency towards open forms of collaboration in providing information and access to the manuscript heritage. The MaGI project presented by *Marilena Maniaci* and *Paolo Eleuteri* is an example of cataloguing facilitated by flexible digital tools and software (cf. Bernardi et al.; Cartelli et al.): across institutional boundaries, this project aims at cataloguing and selectively digitising all Greek codices held in Italian libraries.

The mere existence of digital reproductions and online catalogues prompts us to consider connecting catalogues by bringing all the available documentation together in comprehensive portals (cf. Uhlíř and Knoll) and Virtual Research Environments (cf. Deckers et al.). Assessing the current state of digitisation, *Ezio Ornato* draws some radical conclusions. Based on the conviction that manuscripts are indeed written for the reader, his chapter reads like a codicological manifesto: researchers and cataloguers are called upon to unite as a community and to express their particular research needs, and databases must be created systematically and structured in order to realise the vision of a "Bibliotheca universalis librorum Medii Aevi", freely accessible by means of a "Catalogue grand ouvert". The condition for both of these would be a radical change and liberalisation in digitisation and publication policies of most of the manuscript libraries. This, in return, would have an impact on scholars who rely on a code of research ethics rigorously banning plagiarism, appropriation, forgery and obliteration.

Tendencies towards ever more comprehensive portals offer new opportunities for comparative studies and a global perspective on our cultural heritage. The next 'evolutionary' step in integrated manuscript descriptions is triggered by ideas from the 'Semantic Web'. Here, the prevailing approach is to enrich already available data. This includes making semantically explicit what has been previously implicit in mere strings of characters. This means in turn that concrete objects as well as abstract concepts need to be identified in manuscript descriptions in order to connect these to entities from authority files and to bind them together via taxonomies and multilingual vocabularies. Yet the comprehensive usage of catalogue information across language borders and cultural practices in describing codices is only one side effect of 'semantisation'. The chapters written by *Toby Burrows* and *Robert Kummer* both, independently from each other, sketch the basic concepts and current state of technical solutions for a "semantic codicology". Both chapters reveal the enormous potential of semantic data and open a wide horizon for future research, where completely new questions may arise that scholars could not have previously imagined.

III. Manuscripts and the Sciences

In an ideal digital world, all knowledge of the handwritten tradition would be collected, connected, enriched and accessible from a single point of entrance. The sheer quantity

of all that easily accessible information might itself allow for qualitative progress in manuscript research. In addition, new approaches are emerging from the fields of information technology and the sciences. These too enable codicological research to gain new insights into the material aspects of cultural artefacts that have been already subjects of study for centuries.

Once manuscripts are available as digital facsimiles, we have the grounds for systematic analysis based on computational methods. Similarities and distinctions in scripts and individual hands can be measured and calculated (cf. Aussems and Brink). This sheds new light on the conditions and processes of manuscript production and on the number of scribes involved (cf. Stokes). Besides that, computational methods can also be applied to solve the problem of identifying fragments of documents that have been scattered and should be joined again. *Nachum Dershowitz*, *Yaacov Choueka*, *Roni Shweka* and *Lior Wolf* show how automated image analysis can produce significant new information and lead to well-founded suggestions about which fragments originate from a single document.

There is always more to a manuscript than meets the eye. An example of how hyperspectral imaging can be used to aid in text recovery is given in the first volume of this series (Shiel, Rehbein and Keating). The problem of faded and illegible writing is now addressed again in the chapter by *Daniel Deckers* and *Leif Glaser*. This contribution demonstrates how high-flux storage ring x-ray radiation can be applied to make script that has been erased visible again. Another 'deep' insight beyond multiple layers of written text and into the history of the supporting material can be gained by looking at the genetic makeup of the animal skin that is now parchment. The processed skin bears all the DNA information of the individual goat, sheep or calf from which it was taken. Systematic DNA sets of a large number of folia would fundamentally change the traditional way of dating and localising the creation of writing support, and the study of accidental characteristics would be completed by a scientifically grounded and possibly more reliable methodology. While the option of obtaining and analyzing such DNA strings has already been described in principle elsewhere, *Timothy Stinson* now ties the scientific approach back to the knowledge and evidence from the humanities again. Information on the DNA of parchments will contribute to codicological research and our understanding of parchment production as well as the history of animal husbandry. However, it always needs to be understood within the context of additional historical and archaeological evidence.

What the DNA is to parchment, the watermark is to paper. Watermarks give important indications regarding date and place of production of the writing support. *Peter Meinlschmidt*, *Carmen Kämmerer* and *Volker Märger* introduce thermography as a non-invasive method that yields clear pictures of watermarks. These images can be processed and integrated into comprehensive databases (cf. Wolf) in order to be compared and searched using pattern recognition techniques.

IV. Digital Palaeography

The growing mass of palaeographic information available online has changed the conditions for both research (cf. Ciula) and teaching (cf. Kamp) of what can now be called "digital palaeography". Practical experiences in teaching are reflected upon in the chapter by *Peter Stokes*. How has the teaching of traditional skills changed, and to what extent should digital content be explicitly introduced into the curriculum for the study of medieval manuscripts? The author claims that a deep integration of both the digital and the traditional approaches has to be a fundamental principle, and that technical aspects are not a mere addition or something arbitrary. As such, palaeography should be taught in the wider context of Digital Humanities but at the same time "digital palaeography" should not be treated separately from palaeography in general.

In the tradition of a "quantitative palaeography" *Dominique Stutzmann,* by analysing medieval charters from Burgundy, demonstrates that the encoding of variant letter forms is an appropriate way to examine allographetic characteristics (cf. Hofmeister et al.) and to draw conclusions about provenance and dating.

From the very beginning of palaeography as a discipline in the late 17th century, scholars have always been trying to classify scripts. This endeavour, again, has very much changed under the new terms and conditions of digital information and software tools (cf. Stansbury; Stokes; Aussems and Brink). Palaeography demands ever more detailed data and research (cf. Hofmeister et al.; Gurrado). The availability of information, and a transfer of methodologies, creates new possibilities for the study of handwriting as a cultural phenomenon across time and space. In recent years the interest in manuscripts and writing has increased beyond the occidental focus. This may comprise medieval oriental codices (contributions on this may be included in a planned volume III) as well as the study of texts from ancient Egypt under a palaeographic paradigm as introduced in the chapter by *Stephen Quirke*. Dating from about 1850–1750 BC, the several thousand fragments from Lahun form a promising collection of material for research that is based on computer-aided palaeography but that also aims at new insights into literacy and power in the ancient world.

Improving legibility is another major task in palaeography. Palaeographers, working mostly with digital surrogates of manuscripts, should make extensive use of advanced image processing techniques to improve the legibility of script for machine processing and human reading (cf. Fusi; Tomasi and Tomasi). *Markus Diem*, *Robert Sablatnig*, *Melanie Gau* and *Heinz Miklas* present their work on the pre-processing of images from Slavonic manuscripts in Glagolitic script, taking a new approach in applying OCR (optical character recognition) software to handwritten documents. Similarly, *Julia M. Craig-McFeely* presents techniques and tools for restoration and recognition of faded and degraded script in sources of special interest for musicologists. This contribution

goes even further, however, discussing digital tools for the next step in musicological research: the transcription of music notation and the creation of fluid scholarly editions.

V. Transcription and Text Encoding

In a similar manner, an attempt to bridge the gap between learning to read manuscripts (cf. Kamp; Cartelli and Palma) and learning to transcribe and edit digitally handwritten text is undertaken by the "Ad fontes" project as presented in the chapter by *Isabelle Schürch* and *Martin Rüesch*. But what is used here as a label for a project with pedagogical purposes in palaeography also describes a strong tendency in approaching our cultural heritage in general, and manuscripts in particular, under the new conditions of the digital age: *ad fontes*—to the sources! Manuscripts have become far more visible. And while the visibility and accessibility of manuscripts improve, their perception as documents in their own right—which are inseparably interconnected with their content—becomes more commonly accepted among scholars. As a matter of course, digital facsimiles are becoming an integral and natural part of scholarly editions. Two examples of these are the electronic edition of one of Montesquieu's notebooks, presented in the chapter by *Carole Dornier* and *Pierre-Yves Buard*, and the digital archive of Foucault's notes which he wrote when preparing "Les mots et les choses" (English: "The Order of Things"), as presented by *Philippe Artières*, *Jean-François Bert* and *Samantha Saïdi*. More than just illustrative examples, these projects also prove how documents can make visible the intellectual evolution of the thinking of these famous authors. With this in mind, and supported by a wide range of examples well worth looking at, *Elena Pierazzo* and *Peter A. Stokes* start to work out a revision of the prevalent perception of a manuscript text. In the guidelines of the Text Encoding Initiative (TEI), which may be regarded as the 'de facto' standard for the description and digital encoding of texts, the focus is primarily the text, a linguistic object, rather than the document, a physical object. In contradiction to this traditional attitude, the concluding chapter takes a codicological approach and argues for the establishment of a more document-centred markup standard for the transcription of manuscripts, that is, *putting the text back into context*.