Piloting a National Programme for the Digitization of Medieval Manuscripts in Germany

Claudia Fabian
Bayerische Staatsbibliothek, Munich
Claudia.Fabian@bsb-muenchen.de

Carolin Schreiber
Bayerische Staatsbibliothek, Munich
Carolin.Schreiber@bsb-muenchen.de

Abstract

An initiative to develop a concerted national programme for the digitization of medieval manuscripts in Germany was launched by five German manuscript centres in January 2013. This new proposal aims at the development of a master plan for the digitization of nearly all surviving medieval manuscripts in Germany and at the establishment of a new funding programme of the DFG (German Research Foundation), the largest research funding organization in Germany.

Besides this fundamental financial aspect, other objectives are the diffusion of technical standards and know-how to less experienced cultural heritage institutions and the creation of a nationwide network of competent partners, e.g. in the areas of digitization itself, long-term storage of digital data, the administration of persistent identifiers, and the internet presentation of digital collections. An important role will be played by the German manuscripts portal, Manuscripta Mediaevalia, which is intended as the central hub for both the presentation of digital collections of German manuscripts, but also of the relevant metadata, that is, digital and ideally searchable full-text versions of scholarly, in-depth manuscript descriptions. Of course, this
centralization and aggregation will also facilitate the exchange of data with other institutions.

The pilot phase will lead to the definition of a national digitization strategy with a clear prioritization of projects to be approached. It also aims at the creation of a sustainable technical infrastructure for the integration of both primary and secondary digital data. In this way, the current fragmentation of information in a plethora of local digitization projects is to be superseded by a unitary and complete virtual research environment for all those interested in medieval manuscripts from Germany.

Key Words: manuscripts; digitization; database; infrastructure

1. Background of the Initiative

Around 60,000 medieval manuscripts are preserved in German collections today, of which roughly 7.5 percent have been digitized so far. Time has come to take care of the 92.5 remaining percent and to do so in a systematic approach.

The DFG plays a key role both in financing in-depth scholarly cataloguing of manuscripts in various projects, in the collecting of this information in Manuscripta Mediaevalia, the national database for manuscripts, and in enabling and financing numerous digitization projects of cultural heritage material. A seminal, early example in the field of manuscript digitization was the Cologne project Codices Electronici Ecclesiae Coloniensis. It started in the year 2000; the library of the archdiocese of Cologne was the first library in the world to digitize all its medieval manuscripts.

To ensure that its investments in digitization are made in an efficient way, the DFG often stipulates the establishment of a national framework, a master plan or road map, in order to align different project initiatives and make them part of an integrated, research-oriented approach. A pilot phase of about two years, which comprises several coordinated pilot projects, usually precedes the main phase. The pilot phase is meant to establish procedures, set priorities and calculate costs.

In the area of early printed books, such DFG-funded programmes have financed the digitization of books printed in Germany in the sixteenth and
seventeenth centuries. Digitization is carried out in different libraries on the basis of and in order to widen the impact of the national bibliographic databases for early printed books.\(^4\) For the 18th century, the creation of the national bibliography and the digital library started in parallel,\(^5\) and is just between the pilot phase and the main phase. For incunabula, no such master plan has been established, but more than 7500 incunabula of the BSB (Bavarian State Library), which owns the largest collection worldwide, were digitized with financial support of the DFG. Links to the digitized copies are available in the specialized large incunabula bibliographies,\(^6\) and, for sure, the library’s own incunabula and general catalogue.\(^7\) The organizational framework ensures that – in principle – only one copy of an edition (manifestation) is digitized with DFG financing, and the Google digitization project at the BSB is fully taken into account. This provides a broad overview of German book production in digital form in a coordinated, straightforward, time effective approach, designed to make the best use of public money.

These and other digitization projects allowed a number of institutions to gain experience in the field of digitization technology and workflows, which was also incorporated into the DFG’s Practical Guidelines on Digitization,\(^8\) a widely accepted national standard in Germany.\(^9\) The DFG also encouraged and financed the establishment of a central, standardized viewer to be used for all digital copies produced within these programmes.\(^10\) A central register of digital copies of printed books\(^11\) was built, even before Europeana or the German Digital Library (DDB) went online in 2008 and 2012, respectively.\(^12\)

2. The Digitization of Medieval Manuscripts in Germany

2.1 The Situation in the Past

It is obvious that the know-how acquired in all these projects must be used for the systematic digitization of medieval manuscripts. However, more experience has still to be gained in this particular field. Again, we have a national database, Manuscripta Mediaevalia, on which to build and where all information can be gathered. Nevertheless, the development of a master plan for the digitization of unique items, which should set priorities on material of importance for research, is a new issue and a challenge for the pilot phase.
Here, the issue of multiple copies of one edition does not arise, and the heterogeneity or even the lack of descriptions within Manu- scripta Mediaevalia must be taken into account. The experience gained in previous manuscript digitization projects serves as a starting point. In general, the digitization of manuscripts started quite early, sometimes with a pronounced background in research, very often with quite an individual approach, as “boutique digitization” (Altenhöner, Brantl, & Ceynowa, 2011). In the past, many libraries digitized in order to showcase their treasures, using their own resources. Digitization of manuscripts is currently also done on demand of and financed by customers. Digitization often accompanies exhibitions by virtual exhibits on the internet. It permits to present the manuscripts to a wider audience and to give virtual access to those parts of the collection where physical access is restricted or even excluded due to preservation aspects or the value of the items. There is quite a rich variety in funding, in quality, in completeness, in metadata. Among some of the largest projects is the Codices iconographici project of the BSB; other examples are Heidelberg University library’s Codices palatini, the virtual reconstruction of the monastic library at Lorsch, and – on a European scale – the Europeana Regia project, which involved the BSB and Wolfenbüttel.

2.2 The Design of the Pilot Phase

Time has thus come for a common approach to manuscript digitization. The DFG has entrusted this endeavour to the six German Manuscripts Cataloguing Centres, namely Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Universitätsbibliothek Johann Christian Senckenberg, Frankfurt a. M., Universitätsbibliothek Leipzig, Bayerische Staatsbibliothek Munich, Württembergische Landesbibliothek Stuttgart, and Herzog August Bibliothek Wolfenbüttel. They were established by the DFG and have enjoyed its financial support for more than 35 years, mainly for the detailed scholarly cataloguing of manuscripts, both from their own collections and from other libraries in their area. After preparing and submitting papers for the concerted digitization initiative since 2010, the proposals submitted for pilot Projects by five centres (Berlin, Leipzig, Munich, Stuttgart and Wolfenbüttel) and the institutions responsible for Manuscripta Mediaevalia (Bildarchiv Foto Marburg, Staatsbibliothek zu Berlin and Bayerische Staatsbibliothek) were accepted in June 2013. These institutions have a
broad expertise in the field of manuscript studies and conservation, an institutional collaboration with major digitization centres and centres for the preservation and conservation of manuscripts, a sound experience in standardization and good contacts with the academic world. All pilot projects are coordinated by the BSB.

The pilot phase aims at the development of a master plan for the digitization of all surviving medieval manuscripts in Germany and at the establishment of a new funding programme of the DFG. We are warned that DFG only sees its own financial involvement to cover the digitization of up to 50% of these manuscripts. The setting of priorities for collections to be digitized is therefore one of the key prerequisites to be defined in the pilot. This implies taking into account as wisely as possible all other sources for funding.

The pilot phase has to look carefully into those aspects of digitization which still remain to be defined in the special treatment of manuscripts and the definition of special standards whenever needed; another aspect is the identification of services to be offered by the digitization and/or manuscript cataloguing centres to those institutions that cannot digitize or catalogue their manuscripts on their own. A key factor is the coherent presentation of and access to digitized manuscripts to be ensured by *Manuscripta Mediaevalia*, which will be further developed to provide consistent information about the manuscripts and access to the digital copies.

Seven different collections of manuscripts comprising in total 900 manuscripts are part of the pilot. They represent a range of types of collections and cover different levels of documentation in *Manuscripta Mediaevalia*.

In the first case group are collections currently being catalogued at a scholarly level with funding from the DFG, such as the manuscripts from the Benedictine monastery of St Emmeram in Ratisbon, preserved at the BSB, and a project of the Leipzig university library, which concerns scattered holdings in very small institutions in Saxony. Detailed online descriptions of a manuscript and its digital images are provided at the same time for *Manuscripta Mediaevalia*. The digitization will be useful for the manuscript cataloguer as well as for the researcher; it will be supplementing the catalogue descriptions. In this case group, the impact of digitization on scholarly cataloguing has to be studied carefully. The question is if and where expensive cataloguing can be reduced because of the availability of a digital copy.
The second group comprises well-documented, but not recently catalogued collections. An example is the Munich project for the digitization of vernacular manuscripts from the Munich Court library (bearing the shelfmarks Cgm 1-200), for which a detailed catalogue from the year 1920 is available in searchable form via Manuscripta Mediaevalia (Petzet, 1920). Another project concerns the manuscripts of the Lüneburg council library, to be digitized by Wolfenbüttel. Here, the manuscript descriptions, dating from printed catalogues from the 1970s and 1980s (Fischer, 1972; Stähli, 1981), need to be converted into full text and fed into Manuscripta Mediaevalia. Similarly, Stuttgart will digitize about seventy Codices biblici, for which satisfactory descriptions exist, but have not been published. In these cases there are good quality descriptions available, although they vary in formats (mostly printed catalogues) and are represented in various ways in Manuscripta Mediaevalia: not yet present, full text description without standardized access points, image scans of the printed catalogue with some access points. These different forms of information have to be harmonized in a cost-effective way within the database in those areas vital for access – no new cataloguing is intended.

A third group consists of collections which have not been adequately described, such as a fund at the university library of Leipzig or part of the large Berlin manuscript collection, for which only short descriptions of the 19th century exist. Here digitization will be accompanied by the creation of short records to be put into Manuscripta Mediaevalia, thus giving access to the digital copy. Models for such short records have been developed in the context of DFG cataloguing projects and for the manuscripts of the BSB within the online library catalogue. This may be a pioneering procedure also for the numerous manuscripts digitized so far without any link to Manuscripta Mediaevalia.

A fourth group of projects – proposed but unfortunately not financed in the pilot phase – is that induced by research interests. Here not only the integration of information into Manuscripta Mediaevalia has to be assured, but also access to this information has to be provided for research purposes – scholarly descriptions of manuscripts and digital copies have been considered primary research data for the humanities for a long time.

These different corpora should help to define clearly the relationship between different levels of cataloguing (all to be stored within Manuscripta Mediaevalia) and digitization. They will also allow the definition of different
ways of access to the digital copy – kept within the framework of *Manuscripta Mediaevalia* or on local servers.

### 2.3 A Master Plan for the Digitization of Medieval Manuscripts in Germany

For the master plan, this choice of pilots implies addressing the following questions concerned with prioritization – to be discussed with the academic community; this discussion is planned at the occasion of two conferences in September 2014 and in spring 2015, but also throughout the pilot phase.

The first question in this context concerns the amount of scholarly interest a collection has attracted. Is it more important to digitize otherwise unknown (uncatalogued) collections – and thus to make them accessible and available for research – or is it preferable to focus on well-known collections for which research interest and thus user demand are already manifest? Is it mainly research projects or teaching purposes that should initiate digitization?

The second aspect must deal with the size and accessibility of collections intended for digitization. Should manuscripts scattered in small collections be preferred? Since they are often comparatively unknown and difficult to access, digitization would make them available for research and be a safeguard against oblivion and loss. These collections are often privately owned or ecclesiastic, their holding institutions cannot apply for DFG funding – in spite of their relevance for research. Middle-size collections in German libraries are often much better catalogued and known than parts of very large collections, which are often only listed in catalogues compiled in the 19th century. Should their digitization be preferred or postponed? What about very large collections, which are normally easier to access, but of which frequently only parts have been re-catalogued at a scholarly level, whereas the greater part is described only in quite concise and often outdated historical catalogues?

Throughout all projects, the demands for conservation and preservation and the special requirements for the digitization of particularly valuable, often illuminated parchment manuscripts have to be evaluated (5th case group). It is evident that manuscript digitization must be carefully monitored by preservation experts, since it puts a high stress on the objects. On the other
hand, once digitized – and well digitized, making sure that all information is transmitted even when there is individual scribbling in the margins, fleuronnée, tight bindings – the digital availability of the manuscripts will considerably reduce the use of the original, and often allow a thorough investigation of the manuscripts for the first time.

Although preservation is an important cost factor in manuscript digitization, it is not covered by the DFG’s financial support. It is considered a basic service for the maintenance of collections, to be provided at the libraries’ own expense – quite a challenge for these digitization projects. Nevertheless, preservation issues have an impact on workflows and on the average number of pages to be digitized per day – and thus on the overall cost of the projects. The assumption is that “normal” manuscripts can be digitized at a ratio of 250 images per day, illuminated, precious, parchment manuscripts with gold, purple, very tight bindings etc. at only 100 images per day. The pilot projects must permit the quantification of the financial commitment necessary for the digitization of the German manuscripts heritage. An average price per digitized page including work on metadata, presentation etc. is expected as one of the outcomes of the pilot – as it was defined for prints of the VD 18 (55 cents per image) and VD 17 bibliographies (77 cents per image).

This raises more questions of prioritization, namely those of accessibility of the manuscripts and their nature. Shall priority be given to those precious manuscripts that are not normally available for consultation – thus opening up for research material that is otherwise inaccessible, but particularly valuable, which can be done probably at a high price and low speed? Even if some of these treasures are not illuminated, many of them will be. Illuminated manuscripts in general might be a priority, since art history relies heavily on images. Or is it the special duty of the DFG to finance the digitization of those “dull” text manuscripts which would otherwise never be transformed into a digital format, thus providing access to masses of hidden texts? Can we imagine entrusting the digitization of normal paper codices to Google one day? Or are profit-oriented companies more interested in the precious material, but would make it only available for a fee? 22

The master plan for the digitization of manuscripts must also define responsibilities and name competent partners. Services, which the digitization centres or the manuscript centres could offer to smaller institutions, need to be defined. Although the long-term archiving of the digital copies has not been
kept on the agenda of the pilot phase – it is covered by other programmes – it must be ensured that all digital copies of manuscripts have a persistent identifier (a URN) and are integrated into a long-term archiving system. The awareness of metadata standards, digitization techniques, the administration of persistent identifiers, and the internet presentation is an important part of the project. The creation and definition of a sustainable technical infrastructure for the integration of both digital metadata and digital images is therefore an important aim of this pilot phase.

And again this may imply a priority question, this time concerned with the institutional framework of projects: Is the existence of a well-working digitization centre an argument for prioritizing a collection? Is transport to a digitization centre an obstacle, since it causes costs for shipping and insurance?

Last but not least, the partners of the pilot should learn from each other in defining efficient workflows from the preparation of a manuscript for digitization to the presentation of the digital copy. Here several approaches will be tested, taking into account the various local scenarios (catalogue, workflow tools for digitization, different viewers), the different scanning devices (cameras, scanners), and the different data structures within Manuscripta Mediaevalia. Various ways lead to the result that the DFG wants to see: a useful record within Manuscripta Mediaevalia connected to a digital copy with meaningful metadata displayed in the DFG viewer.

Another important topic to discuss within this pilot phase is the impact of digitization on the current practice of scholarly cataloguing of medieval manuscripts in Germany, for which guidelines have been defined by the DFG. The accepted principle that every digitization must be based on a catalogue record will be maintained. Every digitization must be complemented by at least a basic, copy-right free description of the manuscripts within Manuscripta Mediaevalia. The manuscript cataloguing centres consider digitization a first step towards the full scholarly description of manuscripts, but not as a substitute for it. However, the rules for the cataloguing of certain aspects traditionally described in detail may need to be reformulated because of the accompanying visual information now available. Last but not least, a new relationship must be found between the full scholarly information kept within Manuscripta Mediaevalia and the metadata related to the digital copy. A concise set of data has to be defined, which will be accessible in Manuscripta Mediaevalia, but also passed on to a number of different portals,
e.g. Europeana,27 The European Library,28 or the Deutsche Digitale Bibliothek.29 It may be generated in Manuscripta Mediaevalia, but also transmitted to it from local catalogues, digital libraries or local presentations etc.

The relationships between a full scholarly description in Manuscripta Mediaevalia and the so-called structural metadata, linked to single digital images, which allow the user to navigate and serve as “tables of contents”, need to be defined. Structural metadata should at least comprise a concordance of image files to page or folio numbers in the manuscript, but they should also point to interesting details, such as bindings, text divisions, important illuminations, marks of ownership etc.

2.4 Technical infrastructure

An important role is accorded to Manuscripta Mediaevalia, which is intended as the central hub not only for the metadata but also for the presentation of digital collections of German manuscripts.30 The database will contain – as today – descriptions of a manuscript (at least a short record, but preferably a full scholarly description, sometimes several per manuscript) and the scanned pages of the corresponding printed manuscript catalogue. It must provide access to further documentation about the manuscript and to a digital copy, if available. Of course, the portal itself must be adapted to this new central role. A lot of internal format aligning has to be done. Integration of records from various sources and the easy transfer of concise and consistent manuscript information to other services must be established. Manuscripta Mediaevalia must develop into an aggregator for information on manuscripts and provide an interface to other portals, e.g. national or international heritage portals. In order to provide access to the DFG viewer, the metadata must be provided in the METS and TEI formats. In this way, the current fragmentation of information in a plethora of local digitization projects is to be superseded by a unitary and complete virtual research environment for all those interested in medieval manuscripts in German collections and beyond.

An important task is standardization. The scholarly metadata should provide searchable full-text versions of the manuscript descriptions. The normalization of the specialized vocabulary used must be tackled in order to create various authorized access points for sharing within the internet.
While persons’ names have long been subject to and integrated into national and international authority files, work on the standardization of other entities has only just begun. Thesauri or classified and ideally hierarchically grouped lists of terms should be developed for the most important concepts or subjects in the disciplines of palaeography, codicology and art history. New developments in cataloguing have to be taken into account, especially the so-called “primary relationships” (work – expression – manifestation – item) defined by the new RDA rules. They need to be adapted to accommodate the peculiarities of unique medieval manuscripts. A very important entity in the framework of RDA is undoubtedly that of uniform titles for (literary) works, which is especially complex in the case of medieval manuscripts. Their notoriously unstable text form (with omissions, local additions, revisions etc.) is an enormous challenge. Within the context of the pilot phase, Bildarchiv Foto Marburg will attempt to exploit the internal authorities file created in *Manuscripta Mediaevalia* for text titles and transfer them to the national *Gemeinsame Normdatei* created by the Deutsche Nationalbibliothek and its partners. Another prerequisite will be to establish manuscripts themselves, that is, unique artefacts, as clearly identifiable entities, namely individual works of art by assigning them a unique identifier or standard number useful in the semantic web environment.

### 3. Conclusion

Many new features will have to be considered. The integration of user feedback and contribution is of primary importance – made more attractive and realistic when displaying a digital copy, and possible with the aid of Web 2.0 technologies. The integration of or links to specialized databases such as the existing ones for watermarks or blind-tooled book bindings or provenance portals will widen the impact of the database for research. A new challenge – outside the pilot phase – will be the administration of transcriptions of texts connected with the manuscripts and online editions.

Although they are essential for the good use of the digitized manuscripts, aspects of the presentation of the digital copy beyond metadata and DFG viewer issues are not on the agenda of the master plan. The DFG is quite right in considering this as a central task of the digitizing institution and/or
dedicated research projects. We already see a number of approaches to presentation emerging – zoom and turning the pages are already yesterday’s technology, today superseded by virtual exhibitions, apps with innovative features, 3-D-presentation, virtual desktops allowing for comparison, etc.\textsuperscript{34} How useful cutting-edge technology can be for digitized manuscripts is shown by the image search implemented in the Bavarian cultural heritage site, \textit{Bavarikon}.\textsuperscript{35}

This pilot phase will give us an opportunity to learn and to reflect a lot on how to deal best with our manuscript heritage in the two years to come. We will share our progress on a European and international level – within the Consortium of European Research Libraries (CERL) and the newly established Ligue des Bibliothèques Européennes de Recherche (LIBER) forum for digital cultural heritage and other national initiatives like France’s \textit{Bibliissima}.\textsuperscript{36}

\textbf{References}


Piloting a National Programme for the Digitization of Medieval Manuscripts in Germany


Notes

1 On the more than 130 manuscript catalogues financed by the DFG and published by the German Manuscripts Cataloguing Centres by 2009, see B. Wagner, ed. (2009), at pp. 17–28.

2 http://www.manuscripta-mediaevalia.de.

3 http://www.ceec.uni-koeln.de.


6 Incunabula Short Title Catalogue (ISTC), see http://www.bl.uk/catalogues/istc/ and Gesamtkatalog der Wiegenderucke (GW), see http://www.gesamtkatalogderwiegenderucke.de/.

7 See http://www.bsb-muenchen.de/Inkunabelkatalog-BSB-Ink.181.0.html and http://www.bsb-muenchen.de/OPACplus.92.0.html.

8 http://www.dfg.de/formulare/12_151/index.jsp.

9 An earlier important guideline is IFLA’s Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives (http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf); aspects of the digitization of medieval manuscripts are discussed in detail in Bacher, Fabian, Ikas & Schreiber (2008).

10 For this DFG viewer, see http://dfg-viewer.de/en/regarding-the-project/.

11 Zentralverzeichnis Deutscher Drucke (ZVDD), see http://www.zvdd.de/startseite/.


13 http://www.bsb-muenchen.de/Codices-Iconographici-CodIcon-online.174.0.html.


15 Bibliotheca Laureshamensis, see http://www.bibliotheca-laureshamensis-digital.de/de/index.html.


17 On these Manuscripts Cataloguing Centres, see http://www.manuscripta-mediaevalia.de/info/partner.html and Wagner, ed. (2009), pp. 8–16.

18 On catalogues published by these centres, see above, footnote 1.

19 There is a substantial number of recommendations and current practices in the context of digitization activities, but most of these are related to text/printed documents. A convenient summary is found in Parkoła (2014), a paper published in the EU’s Succeed project on digitisation standards, see http://www.succeed-project.eu/. On metadata standards for manuscripts and their interoperability in the semantic web, see the EU-funded DM2E-Project (http://dm2e.eu).

20 For the DFG Guidelines for short descriptions of manuscripts, see Mackert (2011); on the short entries for manuscripts in the BSB’s online catalogue, see Fabian (2008), especially at pp. 183–190.

21 A survey of German manuscript collections and the catalogues available for them can be found at http://www.manuscripta-mediaevalia.de/area/2/Handschriftensammlungen.html.
An example is the Parker Library on the Web project, which is available only for subscribed users (http://parkerweb.stanford.edu/parker/actions/page.do?forward=home).

The German competence network for digital preservation is called NESTOR; for its activities see http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home and Schrimpf & Keitel (2014). For a national strategy for the long-term archiving of library materials, a call for proposals has been issued by the DFG in July 2013: http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_ueberregionale_informationsservices_131115.pdf.

Technical standards are discussed in Bacher, Fabian, Ikas, & Schreiber (2008) and Bacher, Fabian, Ikas, & Schreiber (2011).


http://www.europeana.eu.


On the German authorities file GND, see http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html; its international counterpart, VIAF, is available at http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html.

See above, note 31.

The German projects Wasserzeichen-Informationssystem (WZIS, see http://www.wasserzeichen-online.de/wzis/index.php) and Einbanddatenbank (http://www.hist-einband.de/).

Three-dimensional presentations are included, for example, in Bavarian (http://www.bavarikon.de/de/find/result/1), virtual exhibitions are offered, for example, by the British Library (http://www.bl.uk/catalogues/illuminatedmanuscripts/tours.asp), The European Library and Europeana Regia (http://www.theeuropeanlibrary.org/tel4/virtual/regia).

For the image-based search mode in Bavarian, see http://bildsuche.bavarikon.de/.