

THE CHALLENGES OF MANUSCRIPT DESCRIPTION IN THE INFORMATION LANDSCAPE OF THE MODERN LIBRARY

CLAUDIA FABIAN

*Bavarian State Library, Manuscripts and Rare Books Department,
Munich, Germany*

ABSTRACT

The article begins with an overview of the broad traditions of manuscript description and a detailed account of the different types of manuscripts and their interlinking with general library collections. Today's cataloguing of manuscripts is integrated into the information landscape of the modern library: the article discusses the presentation of manuscripts on websites, specialized databases for manuscripts at local or national level and the integration of manuscript descriptions into general comprehensive online catalogue systems, illustrating the opportunities and problems of these options. The article also considers databases for particular features of manuscripts (bindings, watermarks, etc.), research surveys and discusses how they can be integrated into manuscript cataloguing. The use of authority files, the challenges of portals and the impact of digitization on manuscript cataloguing are discussed. The article considers the options for integrating existing material into modern internet architectures and presenting descriptions at different levels for different audiences.

KEYWORDS

cataloguing of manuscripts, integration of manuscript catalogues in the modern information landscape, databases for manuscripts

The tradition and legacy of cataloguing manuscripts

Historical manuscripts are the oldest holdings in libraries. It is not surprising then that they are not only witnesses to the history of text traditions, ownership and provenance, reading and studying (and thus knowledge and culture) but that they also have a long tradition of being assembled in collections and being described and catalogued. Manuscripts have been listed ever since they were first integrated into

collections and libraries, mainly in hand-written lists of holdings, inventories, lists of treasures, and then before long in catalogues. Usually these repertories comprised short entries but the knowledge they contain about the items and their relevance for today's tracing of the history of a manuscript should not be underestimated. From the beginning of the 19th century, libraries began to publish printed catalogues of manuscripts in order to make these unique items known to a growing international community of research and knowledge interested in these resources. Usually manuscripts have been the first part of a library's collection to be communicated to an audience wider than the local institutional community, and so they can be considered to be the pioneers of today's general availability of catalogues on the internet.

Printed catalogues have developed from original short entries into presenting more and more detailed descriptions of these unique manuscript items. The development of these descriptions has gone hand in hand with an increase in knowledge and intense research activities, mainly in codicology and palaeography but also through the interest of art history in illuminations.

Manuscripts are not only described in the context of library cataloguing. We also find detailed descriptions, sometimes exclusively, in exhibition catalogues. Whenever manuscripts are loaned for exhibition, they should be described to the highest standards by knowledgeable curators, and these descriptions should be published in the exhibition catalogue and made available on the Internet. As pioneers of this practice, early travel literature provides numerous descriptions of valuable items in library or museum collections. Detailed descriptions of manuscripts, mostly created by specialists, are given in auction catalogues. And finally, for the most outstanding manuscripts, there are commentaries in facsimile editions of prestigious mostly illuminated manuscripts. Critical text editions also provide detailed descriptions of manuscripts which are a primary or important source for establishing texts.

The modern library information landscape offers opportunities for presenting existing information thanks to the digitization of these resources and their appropriate linking to a machine-readable record of the manuscript.

The earliest examples of cataloguing from the 16th century show two different approaches to presenting manuscripts, both of which are still familiar to us today. One is 'holdings centred', with one catalogue for manuscripts and print; and the other is the creation of separate

catalogues for manuscripts and printed books, probably in order to reflect the uniqueness of manuscripts or to take account of the different methods and knowledge necessary for reading or consulting these items. Both approaches are also followed in the grouping of books into a broad subject approach: theology, law, medicine, literature, etc., reflecting the traditional systematic order of books in libraries. This subject approach is once again a requisite nowadays in order to create pathways into digital libraries.

It was not until the 19th and 20th centuries with the development of knowledge of palaeography and codicology for manuscripts, and the aim and increasing feasibility of international data sharing in the world of IFLA's universal bibliographic control for printed books, that differences in the cataloguing traditions and the needs of manuscripts and printed books began to be underlined and developed. In the modern information landscape of the late 20th and early 21st centuries, however, new technological approaches to cataloguing - with the use of authority files and standardized more flexible formats and character sets, the integration of different resources (maps, printed music, audiovisual material, electronic resources, etc.) into a single catalogue and the opportunities for digitization - allow us and, indeed, put pressure on us to create a common information retrieval environment for all forms of information resources. This encompasses manuscripts as well as printed books and reaches far beyond these two categories of traditional library material.

Differentiation of manuscripts

In order to understand the challenges of manuscript description in the information landscape of the modern library, it is worth considering what is meant by the term 'manuscripts' and the kind of differentiation which must be taken into account. In fact, the world of manuscripts is far from being homogenous, and this translates into cataloguing and description. The most common understanding is that manuscripts are the earliest records of written and surviving text culture, the 'books' existing prior to the invention of printing. The main materials are papyrus, parchment and paper but also textile materials (silk), etc. But there are even older records: cuneiform tablets, wax tablets and wood inscriptions, which pre-date manuscripts, but for descriptive purposes they are often considered as part of manuscript culture.

Manuscripts are appreciated as part of the text tradition and must be carefully described in this respect, but they are also unique cultural items as art and museum objects. They are therefore to be described in detail as material objects. This concerns their layout, for example, and in particular their illumination, bindings and provenance.

Usually a chronological division is made between medieval and modern manuscripts. The line is usually drawn around 1530/1550, at the point at which printed book culture was firmly established. But this should not allow us to forget that a flourishing manuscript culture continued into the 17th and 18th and even the 19th centuries, a subject that has not yet been sufficiently studied and appreciated both in its differences and complementarities to the printed book. A very important distinction is that between manuscripts which are books and codices and manuscripts which are (archival, literary) papers such as letters, autographs, notes and notebooks, a live tradition until the present when documents of this kind are slowly being abandoned and replaced by 'born digital' resources. These papers, which are sometimes (and misleadingly) called 'modern manuscripts' are also traditional library materials, shared with the world of archives.

Manuscripts have not only transmitted texts throughout the centuries; we also have manuscript maps, architectural plans and globes. Music manuscripts are among the most numerous and are still flourishing.

Inside a library's collection useful distinctions are made between West European, Oriental, Asian, Latin American, Australian and African manuscripts; languages or groups of languages may be distinguished; and a differentiation made between text manuscripts, illuminated manuscripts and music or maps manuscripts.

All these distinctions and categories have their own proper characteristics and demands and have an important influence on the description and cataloguing of these documents.

Last but not least, manuscripts do not stand alone in a library's collection but are interwoven in different ways with other parts of the collection. We become more and more aware of the unity of manuscripts and print in composite volumes, where a manuscript is bound with printed items. Manuscript fragments can be found on or in bindings. We find bound codices in archival papers; and very often a collection of letters, which has been bound as a codex in early modern times, is now part of the codex collection because of its physical aspect. Legacies of papers can often comprise 'libraries' mainly of printed books and other documents, for example, photographs and various collections, etc. These

variations have to be taken into account in description and cataloguing, and a way has to be found to deal with them. All this argues in favour of not isolating manuscripts from the rest of the collection but integrating them within general information tools in an appropriate way.

Modern cataloguing of historical manuscripts

Today's description and cataloguing of historical manuscripts is integrated into the information landscape of the modern library, which means that it is shaped by the opportunities and challenges of computerization and the Internet. Within this landscape three approaches to manuscript description can be distinguished: presentation on different websites; the establishment of specialized library databases for historical manuscripts (or archival papers) at local or national level; and the integration of manuscript descriptions into general comprehensive on-line catalogue systems.

All three approaches rely on the same technical infrastructure, and so they should not be considered as mutually exclusive but as converging: they may even be considered as a step-by-step approach aimed at the best description of the manuscripts item by item and the collection as a whole.

Websites

Since historical manuscripts are unique, they make a unique contribution to shaping and defining a library's special distinctive and different character. Libraries are therefore proud to describe their manuscript collections on their homepages and via the web.

Depending on the number of manuscripts and the state of cataloguing of the collection, libraries can offer either a short or a more complete overview. As well as general descriptive text, a website can make available lists of various types, for example, a list of manuscript shelfmarks, which are of particular importance, being the unique identifiers of these items. The lists can be in different order: chronological, linguistic, by subject and by provenance. This may be the opportunity to communicate or establish a structure within the collection, presenting collection level descriptions.

It goes without saying that digital copies of complete or parts of manuscripts can be linked to this type of information. Detailed manuscript descriptions created by a text processing system and established as a pdf file can also be linked to individual items in these lists.

Nowadays this information is easily retrievable by a simple Google search and so the manuscripts are widely disseminated without any major publishing effort, an enormous service to research worldwide. It is therefore important to give the name of the author of the description where it is of academic standard and to point to the 'in progress' state of the description along with a date.

The website also allows the presentation of a list of historical catalogues describing the manuscripts linked to a digital copy of the catalogues (without copyright restrictions). It is not important whether this catalogue or list is handwritten or printed, but it should contain some structured guidance for users in order to facilitate browsing and navigation in the digital images. Exhibition catalogues may be advertised or even made available here as well.

The website offers many opportunities for adding short-term established information and other features, for example, educational material, lists of bookmarks and library stamps, an overview of paleographical and codicological features or a detailed account of the restoration of an item.

Although presentation through a website is a necessary and important factor in making the collection known, this can never be a long-term investment even although websites are now harvested for long-term archiving. A living website is continuously reconstructed and recreated and links can break and features disappear, so it needs continuous care and migration. It is certainly not a final and secure place for information on manuscripts.

Specialized manuscript cataloguing databases

Cataloguing is more than a website presentation since its impact is far greater than merely making the collection known. It is essentially directed towards users, allowing them to find, identify, select and obtain access to items. Nowadays it also serves to ensure the visibility of an item in wider contexts (portals, on the Web), and it may allow direct access to a digital copy of the original. Thus today's cataloguing is closely linked to metadata provision and data migration and transfer, ensuring that data are not just kept in one context but can be integrated into other wider contexts. Cataloguing is governed by concepts, standards and rules. In an online environment formats are essential.

The academic cataloguing of historical manuscripts aims to describe them in extensive detail for researchers. Traditionally intended

for publication in printed catalogues and enjoying – before digitization – the rare opportunity of inspecting the unique manuscript thoroughly and from a scholarly perspective, manuscript cataloguing has tended to take quite an independent approach in the extension of the description and use of terminology. Nevertheless, in order to ensure a financially viable approach with comparable results, rules were established or emerged from following the practice of famous exemplary manuscript catalogues. In Germany, the main funding agency for academic manuscript cataloguing, Deutsche Forschungsgemeinschaft, published guidelines and rules since 1973¹ and set a target for the number of manuscripts to be catalogued every year, thus imposing a limit on the uncontrolled growth of information. Today experienced manuscript cataloguers realise that they are only guiding research in pointing to the item, indicating lines that should be followed for further research, and that enabling comparison and overview are important aspects for a consistent approach. On the other hand, the availability of a growing number of excellent information resources, not least digitized manuscripts, enables better identification and comparison, which also takes time for identification and evaluation. So manuscript cataloguing has gone from a very individual highly descriptive approach to greater standardization; and it will again be modified to some extent by the hitherto unimaginable visual availability of digital copies open to everyone on the Internet.

In the late 1990s the German national database for manuscripts, *Manuscripta Mediaevalia*, was developed under the auspices of Deutsche Forschungsgemeinschaft.² In spite of various shortcomings, it is

- 1 Richtlinien Handschriftenkatalogisierung / Deutsche Forschungsgemeinschaft, Unterausschuß für Handschriftenkatalogisierung. 5. erw. Aufl. Bonn-Bad Godesberg : Dt. Forschungsgemeinschaft, 1992 [cited: 2012-05-10]. Available at: www.manuscripta-mediaevalia.de/hs/katalogseiten/HSKRICH.htm. 1973, a first separate edition was published, followed by a second in 1974 and a third in 1983. The latest 5th edition is currently under revision taking into account the new developments of cataloguing within a database.
- 2 Fabian, Claudia. Le catalogage des manuscrits médiévaux et de la Renaissance sur le réseau moderne informatique des bibliothèques : un aperçu des expériences allemandes. // *Linking the worlds of script and print : catalogues of European manuscripts and early printed books*. London : CERL, 2009. Pp. 1-15; Fabian, Claudia. Die Handschriftenerschließung im Wandel : modernste Methoden für ältestes Kulturgut. // *Information, Innovation, Inspiration : 450 Jahre Bayerische Staatsbibliothek* / hrsg. Von Rolf Griebel und Klaus Ceynowa ; Redaktion Klaus Haller. München : K. G. Saur, 2008. Pp. 167- 190; Positionspapier der Handschriftenzentren : Katalogisierung, Digitalisierung, Präsentation / Claudia Fabian, Robert Giel, Christian Heitzmann, Christoph Mackert, Eef Overgaauw, Alessandra Sorbello-Staub, Bernhard Tönnies, Bettina Wagner. // *Das Mittelalter* 14, 1(2009): Furor, zorn, irance. Pp. 140-148. [Berlin : Akademie-Verlag] [leicht aktualisierte Version]. // *Zeitschrift für Bibliothekswesen und Bibliographie* 56, 5(2009), 281-286.

an advantage for a library to catalogue, describe and retrieve this material in the widest possible context and not to build its own special catalogues using Word Access or individually defined XML structures. These local specialised databases tend to fall into oblivion, often not surviving their creators, since online catalogues need continuous care, migration and adaptation. If a specialised cataloguing system is needed because of the specific descriptive requirements of, say, historical manuscripts, the wider the context the better it is for their ongoing care and adaptation and for today's vital integration into the modern information landscape.

The German manuscripts database, *Manuscripta Mediaevalia*, is based on XML, TEI-P5. It thus clearly shows its origins in the web publishing community. The TEI –Text Encoding Initiative, P5 standard for manuscripts, was meant for Internet presentation. Another approach to historical manuscript descriptions is provided by EAD, the Encoded Archival Description, which is widely used in archives and for manuscript cataloguing by the Bibliothèque nationale de France, for instance.

Manuscripta Mediaevalia and similar manuscript databases provide highly differentiated structures and substructures, which allow extensive free text descriptions and handle layout issues. Their value lies in the flexibility of the format, which allows the hosting of rich scholarly detailed descriptions which cannot easily fit within the format restraints of a general library catalogue. They are structured within the P5 framework according to the needs of the material. It is possible to accommodate a number of descriptions for single items within *Manuscripta Mediaevalia*. The records are not only item specific but also author (or catalogue) oriented. A different record can be made available for the text and the illuminations and different records may be presented to translate different historical approaches. In order to regroup them for the user, the OPAC interface builds a 'shelfmark record' 'on the fly', which points to the different full records.

Like most manuscript databases, *Manuscripta Mediaevalia* serves two purposes: cataloguing and access. It therefore consists of two layers: a cataloguing layer offering an editor, with a database in MXML (*Manuscriptum XML*); and an access layer, providing the OPAC, the Internet-accessible user interface (<http://www.manuscripta-mediaevalia.de>). The OPAC is intended to serve the needs of a particular public, namely research in these areas and manuscript specialists, and it

is conceived as developing more and more into a portal approach. An English language user interface is under development. The OPAC offers differentiated search options for the 75,000 available manuscript descriptions via a drop down menu; searches can be limited to digital available copies or iconographic information. The portal gives a complete listing for Germany and a selective listing for some other countries where manuscripts are held ('Handschriftensammlungen') and of existing manuscript catalogues ('Handschriftenkataloge'), again according to place and language and material. The search facility could serve the research community very well if it had not to struggle with different degrees of completeness and different record structures which make satisfactory searches difficult. Some records comprise only the shelfmark and the index information from the printed catalogue with a link to the printed catalogue image; others consist of full text generated from OCR but do not give index information; and others make full use of the database with a full text description and index entries.

The database offers a number of linking options. URNs and URLs lead to the images of digitized manuscripts. From the very beginning, *Manuscripta Mediaevalia* was constructed to point to digitized images of printed catalogues, and it enables this after the necessary copyright issues have been cleared with printed catalogue publishers. The database allows the integration of authority files, although the inclusion of national authority files with regular updates has created foreseeable problems with maintenance. The database also allows the use of standard manuscript-specific vocabulary and terminology by enabling individual standardized vocabularies to be built, but again the maintenance of these files and vocabularies is another issue. Last but not least, OAI structures allow the distribution of information from this database to other relevant contexts. The integration of Web 2.0 features and semantic web options are on the agenda. A short glimpse at the content of a manuscript record shows the multiplicity of possible interlinking in a larger semantic web context, for example, to a bindings database like *Einbanddatenbank* (EBDB) (www.hist-einband.de) quoted by standard numbers;³ to a watermark database like *Piccard Online* (www.piccard-online.de), quoted by standard numbers; to the

3 *Einbanddatenbank* [cited: 2012-05-10]. Available at: <http://www.hist-einband.de/ziti-erweise.shtml>. Stempel (s) EBDB s000000, Rollen (r) EBDB r000000, Platten (p)EBDB p000000, Werkstätten/Buchbinder (w) EBDB w000000. With the prefix and the six digit number a clear link can be automatically generated.

ISTC or Gesamtkatalog der Wiegendrucke, quoted by standard numbers; to persons and places accessible through national authority files or the Virtual Authority File (VIAF). There already are and will continue to be many more options, too.

Manuscripta Mediaevalia is not the only national database for manuscripts, although it is one of the earliest and it is the one that attempts to make the best use of legacy cataloguing. There are other national initiatives as, for example, Calames in France and Manus in Italy. The Swiss digitization project 'e-codices' also offers a centrally fed and administered cataloguing database and a very good OPAC interface. This project shows very clearly how existing manuscript descriptions can be utilized by recataloguing or by intellectual control over entering information in a database.

Nowadays local presentations of manuscripts should either be subsets of larger databases or they should be a first temporary approach, in which case Word or Excel files are just as good as more sophisticated XML structures. As automatic transfer into a larger database is a long and time-consuming process, a new manual input from a machine-readable document using simple cut and paste seems to be the safer and quickest way. This also allows for the necessary standardisation of search terms which might differ or go beyond index terms in a printed catalogue.

The OAI structuring of the manuscript database will also allow the harvesting of this information into modern central indexes for data retrieval. This means that although it is being kept in a different information resource, the data can be made accessible in an overall information landscape.

Records for manuscripts in such a specialized environment present a number of particular challenges. There is the issue of legacy data from printed catalogues. This can and should be made available in the manuscript database. The first step may be to add images of the printed catalogue to a short record for the manuscript. The shortest imaginable form is a note of the owning library and the shelfmark. The approach taken by Manuscripta Mediaevalia has been to integrate index information into the database in order to allow users to make the type of searches for which the index of a printed catalogue was previously consulted. What the user will not find in this record though is the most basic and concise information about authors and title, content, year and place and collation, all of which is information which could easily be keyboarded from the headings of a printed catalogue.

Today it is easy to OCR a digitized version of a printed catalogue and to tag it in an XML structure making use of the layout principles. In this way a more structured full-text search may be offered to the user. This was the second approach in *Manuscripta Mediaevalia*. The format today (MXML, *Manuscripta XML*) allows for this information, making a distinction between descriptive full-text paragraphs of what is called the 'print version' and access points which are index entries.

Germany's manuscript cataloguers now have different ways of inputting their descriptions into *Manuscripta Mediaevalia*. Some use the MXML editor offered by the database and catalogue in an online system which hides incomplete or not yet authorized 'in progress' records from users. The editor also helps in the processing of a 'printable' catalogue. Others, like those in the Bavarian State Library, prefer to continue working in a simple Word environment. As soon as a description is finished, it can be published as a pdf file on the library's website and OPAC environment. The pdf file is labelled 'work in progress', with a date and the name of the cataloguer. When all the descriptions to be published in a printed catalogue have been completed, the cataloguers re-edit their work, establish the necessary index entries, update the reference literature information, and make comparisons and draw conclusions from this broader overview. All this implies a reworking of the Word document with the manuscript description and it is only when it is finalized for print that this record, which is then considered stable, is also keyboarded by a librarian into the *Manuscripta Mediaevalia* database. The task of the librarian is not only to cut and paste the correct parts of the Word document into the appropriate full-text paragraphs of the 'print version', but also to edit the standardized entries and establish links to the full text as well as to authority files inside and outside the context of *Manuscripta Mediaevalia*.

The digital images from the printed catalogue are added to the database at a much later stage, in line with licence agreements with publishers. They are then also accessible from each individual description. The added value of this exercise is that it gives full access to the printed catalogue, and to those parts such as the introduction or, in the case of catalogues of illuminated manuscripts, comprehensive stylistic chapters which for the time being cannot yet be reproduced in the database. This shows that *Manuscripta Mediaevalia* is essentially a cataloguing database and not a publishing database.

This procedure highlights a major concern for database maintenance. How should the editing of existing records be handled? How

should new information be added when considering a longer time span than the actual life of the database? For the time being, *Manuscripta Mediaevalia* is much more of a static tool, intrinsically related to printed catalogues and trying to provide online access, than it is a living database in its own right and development. This issue also has to do with copyright issues – the author of an academic manuscript description cannot be considered a normal cataloguer but rather the creator of a work for which the author has copyright until 70 years after death, even if the exploitation of this work lies with the library or funding institution. How can one keep the description made by its creator and allow for updating, integration of new information, modification of outdated or even incorrect information or simply adding more detailed information about, say, illuminations in a manuscript which has not been described by an art specialist first but through a text-oriented approach? There are options for creating different versions which might help to find a solution to this question. For the time being, very few libraries add information with a date and a name into existing descriptions and most create a new record, especially for specific new descriptions for texts or illuminations. The updating of reference literature is also a serious question. The description has been created on the basis of consulting reference works quoted within the record. Can new references be added without reworking the description? Again, there is no common vision yet on how to deal with this issue. For larger collections the solution lies in a reference database outside *Manuscripta Mediaevalia* but accessible from each record.⁴

Another challenge for the database lies in the language of cataloguing. The language is not only German but a highly specialist language, full of abbreviations and requiring a good level of understanding of this type of material. On the Internet this information is available to everyone – specialists and the ‘interested general public’. This is desirable when the digitized manuscripts themselves are visible on the Internet and require further explanation. Technical facilities (such as mouse-over, linking to general information tools, eventually even using ‘text mining’ methods) may cater for better explanations but nevertheless

4 E.g.: *Forschungsdokumentation Handschriften* [cited: 2010-05-10]. Available at: <http://www.bsb-muenchen.de/Forschungsdokumentation.172.0.html>; see: Gullath, Brigitte; Wolfgang-Valentin Ikas. Zum Nutzen von E-Zeitschriften, Datenbanken und Internet-Publikationen. Neue Wege im Ausbau der BSB-Forschungsdokumentation zu Handschriften und Seltenen Drucken. // *Zeitschrift für Bibliothekswesen und Bibliographie* 58 (2011), 72-76.

most of the issues again involve manual editing of information. There are two approaches which would help in making our manuscripts more accessible and comprehensible to the general public, and which need to be included in our cataloguing and our databases in order to spread from there into wider contexts: a (broad) subject approach which can then be of guidance particularly in conjunction with a well prepared chronological and geographical approach. We see examples today in the World Digital Library and Europeana. Temporal and spatial information is already included in the records but it needs to be modified or developed in order to allow meaningful searching, for example, just think about statements like 'late 15th century' or 'Southern Germany'. A subject approach needs to be defined in a very broad sphere taking account of the ancient and medieval structures of knowledge. A list of generic terms relevant for this kind of material could be defined and made available in a multilingual approach through the CERL Thesaurus.

The second requirement is for a generally readable and intelligible description or presentation of a manuscript, a kind of abstract of all the knowledge detailed within the record, giving succinct information on what is special about the manuscript to non-specialist users. This sort of description is used in the World Digital Library (where the text is even read out); it is proposed for Europeana; and it is what makes the Swiss e-codices appear user friendly. The format definition of manuscript databases should allow for the insertion, presentation and conveying of this type of information. It could then be translated into English or even presented in a multilingual approach.

The language of cataloguing is also a concern as manuscripts from one country may be dispersed to several holding institutions in today's world or described by specialists worldwide in their own languages, and so there are many reasons for favouring a multilingual approach to manuscript cataloguing. In the context of *Manuscripta Mediaevalia*, this is a concern for the full text structure and for the index terms which since they use a controlled standard terminology, ideally an authority file, could also allow for a multilingual approach.

Another issue is the scope of the database. Calling the database *Manuscripta Mediaevalia* makes one wonder immediately how descriptions of post-medieval manuscripts, the *recentiora*, are handled. Many are already in *Manuscripta Mediaevalia* although because of its title they would not be expected there. Too narrow a definition of a specialized database implies the distribution of descriptions for items

in one collection into various specialised systems, in different cataloguing tools according to the material. It is no surprise that we have a papyrus database, an oriental manuscripts database, a music manuscripts database (RISM online) and a database for archival papers, autographs and letters (Kalliope). It is up to the library to decide which tools to use to catalogue its collection. Each of these databases is optimized for its particular material and that indeed forms its strength and value, but each demands different rules, standards and formats and thus presents a challenge for curators.

The time may have come for a European or international consolidated approach to material of this kind, but there does not seem to be a European or international tool for manuscript cataloguing, although both Manuscriptorium and Manuscripta Mediaevalia offer their databases and services to a number of countries. There are, however, a growing number of portals ensuring presentation and access as, for instance, The European Library, the CERL portal and Europeana. It remains both a challenge and an ongoing task for specialized databases to present their data as meaningfully and exhaustively as possible to these and other portals.

Integration of manuscript records into general catalogues

Nowadays the integration of manuscript records into general library catalogues has reappeared as an alternative or parallel approach in manuscript cataloguing. This approach depends on general library policy. It will only be taken by those libraries whose aim is to integrate all their material into one cataloguing system. This approach has been enabled by the development of the cataloguing rules and database formats which allow the recording of such diverse materials as printed books, maps, music, photographs, audiovisual material and born digital material in one system. It is eased by the application of UNICODE, which allows the use of a wide character set, and by a more and more flexible definition of fields and field lengths. Although a MARC format for manuscript encoding has existed since the late 1990s, developed in the US in parallel to the definition of the MASTER DTD in Europe, it is still evident that a detailed description for a manuscript – as in the example of Manuscripta Mediaevalia - cannot be fully integrated into a general system. It is difficult to describe the different parts of a manuscript in sufficient detail, especially with incipit and explicit information, to give detailed listings of illustrations and to describe the

binding and the page layout, sometimes even the provenance. All this would require the building of a particular and sophisticated application profile. Nevertheless, even quite a standard MARC record allows the inclusion of enough concise information about a manuscript to make it retrievable in a global context. Such a concise MARC record must be carefully defined and structured but it should make the best reuse or reinterpretation of existing fields. The definition may be guided more by access points than by the descriptive elements. It is of primary importance to document this format well and to ensure consistent coding (stating that it is a manuscript or even an illuminated manuscript) in order to treat the record in a particular way when it is technically possible to do so. Special attention should be given to the unique identifier, which is usually the shelfmark. Manuscript specialists tend to say that this is the best and most reliable way to identify and cite a manuscript, much better than author and title. In the Bavarian State Library, we therefore decided to put the shelfmark in the title field of each manuscript record. As manuscript shelfmarks are rather similar, the name of the institution precedes the shelfmark. If a title exists or has been established, the location and shelfmark are added after the title preceded by a blank dash blank. Even the uninitiated user of a general OPAC can understand that this record denotes this particular manuscript.

Example:

Augustinus, Aurelius
De magistro – BSB Clm 6322
Frankreich, 2. Hälfte 9. Jh.
I + 26 Bl. - Pergament

The integration of manuscript records into the general library catalogue has a lot of advantages. It ensures that they have the same long-term survival prospects as all the library's other records. Today the online catalogue is the library's only memory of holdings. It will thus be maintained, archived, saved, migrated and cared for to the best of the institution's capacity. This implies continuously improved user access to the information. The records do not only remain in the context of the library but are disseminated through various channels in line with different agreements. Accordingly, knowledge about the most valuable and unique items in the collection is distributed in the same way as other material, and, thanks to OAI structures, the easy transfer of data is a

reality. The data can thus go into various portals, for example, the CERL portal, Karlsruher Virtueller Katalog, WorldCat and Manuscriptorium.

In today's information landscape the various linking structures provided by online catalogues are a particular asset. The records are linked to authority files, at least for personal and institutional names, and they are accessible by OPEN URL or SFX linking methods or they can make use of these links to navigate to other information resources. In the Bavarian State Library we integrate the manuscript shelfmark preceded by a unique text string into a particular field. If this is present, the system knows from the introductory string to which other systems to interrelate, for example, to the database for research documentation. Along with the uniform titles from the printed catalogue in yet another field, the system can automatically offer a link to a digitized catalogue image (in *Manuscripta Mediaevalia*, on a local server, in the digital library) or to a full catalogue description of the manuscript in *Manuscripta Mediaevalia*.

Nowadays there are even more options for access to fuller descriptions of manuscripts. There is the possibility of catalogue enrichment, adding a pdf file to a record and a file which is full text indexed and searchable. This can be done by the same methods which we use to add a table of contents for modern books. For this enrichment we can even think about using entries, say, in exhibition catalogues when they give interesting descriptions of manuscripts. In the near future it will be possible to unite the full information in the machine-readable *Manuscripta Mediaevalia* description to a common index. The manuscripts thus profit in a particular way from the technologically most advanced IT structures. Even the fear of some users that manuscripts may be lost in the overall general catalogue is outdated nowadays. It is possible to create individual 'views' – for example, manuscripts only – and faceted options can guide users to select or deselect manuscripts identified by a search. The full development of the OPAC, with suggest functions or implementations of Web 2.0 facilities and services, is also available for manuscripts. In a more traditional way, once manuscript records are in the general catalogue, the circulation or acquisition tools of an integrated library system can be made fully beneficial for manuscripts administration.

Even for those who cannot imagine cataloguing manuscripts in a general catalogue, the question of defining metadata for manuscripts must be addressed when it comes to digitization of the originals. If the

digital objects are catalogued within the general catalogue, the day will come when digitized manuscripts must be entered in the system, and then the question will arise why OPAC access is being given to a surrogate without ensuring this for the original.

The general catalogue provides the broadest context. It shows the cultural tradition translated in the library collection as a whole. The integration of manuscripts allows users to make true and unexpected discoveries, keeps the awareness of these testimonies alive, and serves not least as one of these necessary propedeutics and incentives to manuscript studies.

Today's challenge is to prepare manuscript information for the semantic web. There are good examples of successful datamining in XML texts, a procedure which could be applied to detailed descriptions more easily than to short MARC records. The RDF approach, which allows the provision of linked open data, will be applied very soon to general catalogues where manuscript information will be of particular value.

It is only fair to mention that the integration of manuscript records in the general catalogue also has its shortcomings and challenges. It needs particular forms of adaptation and strict observation of rules and format prescriptions; and it does not yet allow the full description of the item. There are real issues of data integrity. Authorship is not an issue with these short records and so data can be changed quite easily by anyone; and the records are not sufficient to trace the history of knowledge about a manuscript. There is also an effect on data retrieval: a lot of information is left to full text search and does not thus allow the building of object specific indexes. The use of particular authority files or standardized vocabularies may be more difficult and will be a matter for the cataloguer without much technical assistance.

The sophisticated pathways from the short record to the full information (on another server) or to the research documentation via SFX or OPEN URL may only be intelligible to an informed public. They are not easily - and certainly not automatically - transferable into secondary portals and information environments.

Authority files

Authority files are a particular concern both in specialised and general catalogues. Whereas in a general catalogue the information to be controlled or linked to authority files is pre-defined, a specialised environment may allow for much more control and linking. Consequently, the

entities in manuscript cataloguing which can be linked to an authority file need to be carefully defined. These certainly include names of persons and institutions for which existing general authority files can be used. New records required for manuscripts should be created in these files, e.g. for places, where a number of regions may not yet be authority file controlled, and for names of scribes and former owners. Besides this, a number of particular vocabularies in codicology, palaeography, illumination and titles are needed for a manuscript description. It is a real challenge to establish these in the context of an authority file or a thesaurus file. In some areas authoritative data are already available. They can be linked to the description by using standard identifiers, e.g. to special databases for bookbindings and watermarks.⁵ A standard file for titles, also giving incipit and explicit information, and a standard file for reference works would be of great help in manuscript cataloguing. As these tasks are large and shared within the international community, the ideal place to gather this kind of information is at international level. The CERL Thesaurus built up by the Consortium of European Research Libraries already brings together a lot of this information and allows the maintenance and development of new resources in a multilingual environment making the best use of existing resources.⁶

Distribution of records

In our information landscape, records do not remain within the system in which they are created. The implementation of OAI structures allows for the harvesting of data and the definition of RDF triples for reuse within the Open Data environment, including Internet-based research projects. It is in the best interests of a data provider to see his data distributed as widely as possible. Here again general catalogues have an easier time than specialized databases. It is also noticeable that most portals provide simple broad access to data based on the central Dublin Core fields. A lot of portals like Europeana and the World Digital Library and OCLC's OAISTER focus only on digitized material. Manuscripts can play a major role in cultural portals, and many por-

5 Einbanddatenbank. Ibid.

6 Consortium of European research Libraries. CERL Thesaurus [cited: 2012-05-10]. Available at: http://www.cerl.org/web/en/resources/cerl_thesaurus/main; see: Fabian, Claudia. New conceptual models for information organization of authority data in the European context. // Summer School in the Study of Old Books : proceedings / edited by Mirna Willer and Marijana Tomic. Zadar : Sveuciliste, 2010. Pp. 237- 258.

tals such as the very early Digital Scriptorium, the online manuscript library of the University of California and the Swiss e-codices project specialize in manuscripts. A number of subject portals or projects offer different research approaches in themes, texts, reconstruction of libraries: examples include the Roman de la Rose Project, the New Testament Project and the Avicenna Project.

Other portals such as the CERL Portal and The European Library wish to integrate information about manuscripts whether they are digitized or not, and the inclusion of existing specialised databases is of great interest. The CERL portal already gives access to Calames, Manuscriptorium and MANUS among others, but it is an ongoing effort to integrate all of these specialised databases and to make sure that careful mapping and indexing of the very heterogeneously structured information in these databases permits overarching, meaningful and material specific retrieval. The way in which the CERL Thesaurus is integrated into this search is an interesting example of using external authority files to enhance the quality of retrieval in a heterogenous and multilingual environment.

The number of these portals shows the importance for the owning library of not only providing digital copies of its manuscripts but also of presenting full and detailed descriptions, which can then be enhanced by more specialized information in other portals or projects. The EC-funded project *Europeana Regia* shows how important it is for manuscripts not only to be added to the general gateway to culture, *Europeana*, but also to be maintained in their home environment (which may be a local catalogue as well as a national database) and define standards for metadata, digitization of manuscripts and quality control in this area.⁷ The portals offer secondary multi-aspect faceted approaches to

7 *Europeana Regia* [cited: 2012-05-10]. Available at: <http://www.europeanaregia.eu/en>. *Europeana Regia* funded by the European Commission under the Information and Communication Technology Policy Support Programme (ICT PSP), from January 2010 to June 2012 digitised more than 874 rare and precious manuscripts, with the collaboration of five major libraries (Bibliothèque nationale de France (BnF), project leader, Bayerische Staatsbibliothek, Munich (BSB), the Universitat de València Biblioteca Històrica (BHUV), the Herzog August Bibliothek, Wolfenbüttel (HAB) and the Koninklijke Bibliotheek van België – Bibliothèque royale de Belgique (KBR). It comprises three collections that are currently dispersed and which represent European cultural activity at three distinct periods in history: the Biblioteca Carolina (8th and 9th centuries), the Library of Charles V and Family (14th century) and the Library of the Aragonese Kings of Naples (15th and 16th centuries). The reports of the different workpackages (among them metadata management for manuscripts, best practice for digitization of precious manuscripts, quality control aspects) will become publicly available at the end of the project.

a document, which is their particular value, but they cannot replace or supersede the responsibility of the institution for describing as exhaustively as possible all the parts of its collection and maintaining this information in a living, harvestable and re-usable form for the long term.

Digitization of manuscripts

With the increasing digitization of manuscripts new challenges arise for manuscript description. The metadata for the header of the digital object can be derived from a concise catalogue record and translated into the XML structure. New issues arise when it comes to creating a table of contents or a page by page analysis of the digital copy, indicating so-called structural metadata directly linked to the page of the digital object. Not only must the structure of information (METS, TEI) be defined but also the content provided in this context and the use of a consistent and intelligible vocabulary. This kind of information will have a great influence on manuscript cataloguing. Whereas for the time being it is part of a full academic description, usually in a comprehensive synoptic analysis from where it is taken and translated into an analytical form to help and guide the user browsing the digitized document, the question arises as to how cataloguing may be modified taking into account the scope of this analytical information structure. It is conceivable that an incipit will no longer be listed in a detailed catalogue description but only as structural metadata pointing immediately to the image. The same is true for information on illuminations. The success and implementation also depends on the possibility of making these structural metadata searchable and indexed in a differentiated way, which is likely to happen. Full-text retrieval based on transcriptions or editions of a given text is the next step. We see that this has already been achieved in Google Book Search based on OCR reading of printed texts; the route for manuscripts is much longer but it is open.

Research surveys

For all manuscript collections it is advisable to keep an up-to-date research survey, listing published (and even, as far as available, unpublished) information about individual manuscripts. This information is not only helpful to manuscript cataloguers but also to all academic users. Nowadays, a machine-readable research survey can be made accessible from a manuscript description. Whereas a description always relies on the knowledge about a manuscript at a given time, the re-

search survey will also help the user to be aware of ongoing research on a manuscript – and that is what the cataloguing of the manuscripts is meant to encourage. A lot of these surveys already exist in machine readable form and can be addressed through manuscript-oriented portals. A more recent approach is to disclose restoration documents to users as well, as it has been done in the Codex Sinaiticus project (<http://codexsinaiticus.org>). This is particularly helpful if digital images of the original are available. The restoration surveys help users to understand the appearance of individual pages of the manuscript, to know about missing parts, and sometimes to read or decode restoration information given within a codex.

Conclusion

The cataloguing and description of manuscripts has been on the agenda of libraries ever since they first came into collections. Thanks to the technological advances of the modern information landscape, the impact of manuscript cataloguing keeps growing in importance and in audiences. It is fascinating to see how the most modern of technological developments can be used to serve the oldest testimonies of culture in our libraries.

English language editor Dr Ann Matheson

References

- Consortium of European research Libraries. CERL Thesaurus [cited: 2012-05-10]. Available at: http://www.cerl.org/web/en/resources/cerl_thesaurus/main
- Einbanddatenbank [cited: 2012-05-10]. Available at: <http://www.hist-einband.de/zi-tierweise.shtml>
- Europeana Regia [cited: 2012-05-10]. Available at: <http://www.europeanaregia.eu/en>
- Fabian, Claudia. Le catalogage des manuscrits médiévaux et de la Renaissance sur le réseau moderne informatique des bibliothèques : un aperçu des expériences allemandes. // *Linking the worlds of script and print : catalogues of European manuscripts and early printed books*. London : CERL, 2009. Pp. 1-15.
- Fabian, Claudia. Die Handschriftenerschließung im Wandel : modernste Methoden für ältestes Kulturgut. // *Information, Innovation, Inspiration : 450 Jahre Bayerische Staatsbibliothek* / hrsg. Von Rolf Griebel und Klaus Ceynowa ; Redaktion Klaus Haller. München : K. G. Saur, 2008. Pp. 167- 190.
- Fabian, Claudia. New conceptual models for information organization of authority data in the European context. // *Summer School in the Study of Old Books : proceedings* / edited by Mirna Willer and Marijana Tomic. Zadar : Sveučilište, 2010. Pp. 237- 258.
- Forschungsdokumentation Handschriften [cited: 2010-05-10]. Available at: <http://www.bsb-muenchen.de/Forschungsdokumentation.172.0.html>

- Gullath, Brigitte; Wolfgang-Valentin Ikas. Zum Nutzen von E-Zeitschriften, Datenbanken und Internet-Publikationen. Neue Wege im Ausbau der BSB-Forschungsdokumentation zu Handschriften und Seltenen Drucken. // Zeitschrift für Bibliothekswesen und Bibliographie 58 (2011), 72-76.
- Positionspapier der Handschriftenzentren : Katalogisierung, Digitalisierung, Präsentation / Claudia Fabian, Robert Giel, Christian Heitzmann, Christoph Mackert, Eef Overgaauw, Alessandra Sorbello-Staub, Bernhard Tönnies, Bettina Wagner. // Das Mittelalter 14, 1(2009): Furor, zorn, irance. Pp. 140-148. [Berlin : Akademie-Verlag] [leicht aktualisierte Version]. // Zeitschrift für Bibliothekswesen und Bibliographie 56, 5(2009), 281-286.
- Richtlinien Handschriftenkatalogisierung / Deutsche Forschungsgemeinschaft, Unterausschuß für Handschriftenkatalogisierung. 5. erw. Aufl. Bonn-Bad Godesberg : Dt. Forschungsgemeinschaft, 1992 [cited: 2012-05-10]. Available at: www.manuscripta-mediaevalia.de/hs/katalogseiten/HSKRICH.htm.

Biographical sketch

Dr. Claudia Fabian is Head of Manuscripts and Rare Books Department, Bayerische Staatsbibliothek, München since 2004, she was Head of Department of Users Services (2000-2004), and Head of Alphabetical Cataloguing within the Cataloguing Department (1986-2000). She holds Dr. phil., Classical Philology, has professional training as academic librarian and as teacher for Latin and French. She studied Latin and French at the universities of Munich and Paris, Sorbonne.

She is member of the Consortium of European Research Libraries (Advisory Task Group and Executive Committee) since 1990, IFLA's Rare Book and Manuscripts Section since 2009, and LIBER's Steering Committee for Preservation and Heritage Collections and member of the LIBER board since 2010. She is chair or member of various national and regional bodies for manuscripts and early printed books. She is involved in a number of national or regional projects in the area of manuscripts and early printed books, co-responsible for the mass digitization projects funded by Deutsche Forschungsgemeinschaft for manuscripts and early printed books, and responsible for the EU funded project *Europeana Regia* under the leadership of Bibliothèque nationale de France since January 2010.

IZAZOVI OPISA RUKOPISA U INFORMACIJSKOM KRAJOLIKU MODERNE KNJIŽNICE

Sažetak

Članak započinje pregledom opsežnih tradicija opisa rukopisa i detaljnim prikazom različitih vrsta rukopisa i njihovog povezivanja s općim knjižničnim zbirkama. Danas je katalogizacija rukopisa integrirana u informacijski krajolik moderne knjižnice: u članku se raspravlja o prezentaciji rukopisa na mrežnim mjestima, o specijaliziranim bazama podataka za rukopise na lokalnoj i nacionalnoj razini i integraciji opisa rukopisa u opće, sveobuhvatne online kataložne sustave te se ilustriraju mogućnosti i problemi ovih opcija. U članku se raspravlja i o bazama podataka za posebne značajke rukopisa (uvezi, vodeni žigovi itd.) i pregledima istraživanja te se razmatraju mogućnosti njihova uključivanja u katalogizaciju rukopisa. Raspravlja se i o korištenju normativnih datoteka, problemima portala i utjecaju digitalizacije na katalogizaciju rukopisa. U članku se razmatraju mogućnosti integracije postojeće građe u modernu internetsku arhitekturu i prikaza opisa na različitim razinama i za različite vrste korisnika.

Ključne riječi: katalogizacija rukopisa, integracija kataloga rukopisa u suvremen informacijski krajolik, baze podataka za rukopise