

NEW CONCEPTUAL MODELS FOR INFORMATION ORGANISATION OF AUTHORITY DATA IN THE EUROPEAN CONTEXT

CLAUDIA FABIAN

Bayerische Staatsbibliothek, Munich, Germany

ABSTRACT

The paper focuses on the CERL Thesaurus as an example of building the European authority file relating to the early printed book production. It discusses Thesaurus contents and structure, methods of providing content, editing procedures, further developments, use of computer technology (assisted searching, access to local catalogues, provision of web-linking to other relevant resources of information), and the role of authority files for organising intellectual content.

KEYWORDS

authority data, early printed book production, CERL Thesaurus,
European context

Introduction – History of authority data

Cataloguing can be seen as an activity as old as gathering of books in order to establish collections and (later) libraries. It is a natural activity to draw up lists of books as – more generally – lists of goods and possessions. Compared to the archaic activity of cataloguing, authority data are in fact a new conceptual model for information organisation.

It is interesting to observe a close interaction between cataloguing rules (or concepts) and the physical medium as well as the size of a catalogue. The rules for a handwritten book catalogue, the one we are all familiar with from the 19th century, although much older catalogues exist (the oldest printed catalogue of holdings stored today in the Bavarian State Library listing Greek manuscripts of the Augsburg library dates back to 1575) are different from those used for a card catalogue which allows easier filing and additions, and those for ma-

chine-readable catalogues. Cataloguing rules are more complex and more explicit the bigger the quantity of objects to be listed. A smaller list, conceived by one person, can do with fewer rules than a union catalogue with thousands of contributing libraries.

Authority data – thus named and discussed – became a concept in cataloguing only with the considerable growth of catalogues, the development of union catalogues and the development of machine-readable cataloguing, the use of multi-file IT-systems and international cooperation.

Nevertheless, they have ancestors in cataloguing in traditional formats. In the old alphabetical card catalogue of the Bavarian State Library built up since 1841 and in use until 1981, for instance, special entries were made in separate sections called „PR“ („Personalrückweis“) or „KR“ (Körperschafts-Rückweis). These sections contain entry cards for persons or for corporate bodies – usually for „important names“. They give dates and provide information on books about this person or the entries for a corporate body.

Authority data also have predecessors outside the library and cataloguing world, especially in encyclopaedic works. Encyclopaedias describe all sorts of entities. Biographical reference works came up quite early in history, for example, in the form of lists of members of religious orders. Church or political registers were kept for centuries. In order to arrange the names in these kinds of books, rules or guidelines had to be established and followed. Information established and sorted to build up all kinds of registers or indexes to complex books (or catalogues) could also be called authority data. These ancestors of authority data keep playing an important role today in building up authority files as they name entities in a consistent and reflective way. Nowadays they are often used for the establishment of authority records.

International impact of authority data

The fundamental rule which led the way to authority records on the international level can be found in the Paris Principles of 1961. They prescribe the “Choice of uniform heading”: No. 7 – “The uniform heading should normally be the most frequently used name (or form of name), or title appearing in editions of the works catalogued or in references to them by accepted authorities”.¹ This implies that a person should always be entered under the same name, and references should be made from other forms of the name – a rule at that time already

well known and observed by local, regional and national cataloguing rules or codes, was addressed on an international level so as to give details about the choice of the heading in an international context (7.1 and 8.2). Catalogues always followed this rule, but the larger the catalogues became, the more explicitly this rule had to be spelt out, and the more difficult it proved to apply. Thus the need for authority data grows with the size of the catalogue, the number of contributors to a catalogue, particularly on an international level. With international cooperation in cataloguing, with the concept of Universal Bibliographic Control (UBC),² the term authority data came into being and the use of authority data gained importance. The earliest examples are IFLA's lists of uniform titles for classical anonymous works (published in 1978)³ – immediately as an international approach, and lists integrated into cataloguing codes, e.g., in the German cataloguing rules for uniform titles of biblical works or for printed music. The awareness of authority data and their importance thus can be traced back to its beginnings in the 1970s and 1980s.

Authority data are not only a modern concept in cataloguing, linked to the existence of machine-readable multi-file catalogues, they also show from the beginning a strong international impact or vocation or, to put it another way, they are immediately placed in the European (if not worldwide) context.

The idea of an international standard authority number (ISADN) not for a heading, but for an entity came up already in the 1980s, in connection with the establishment of *Guidelines for authority and reference entries* published as early as 1984.⁴ In 1991 IFLA published a study concerned with *Management and use of authority files*,⁵ and in 1993 about

- 1 Statement of principles adopted by the International Conference on Cataloguing Principles, Paris, October 1961. // Report / International Conference on Cataloguing Principles, Paris, 9th–18th October, 1961; edited by A. H. Chaplin and Dorothy Anderson. Reprint. London : IFLA International Office for UBC, 1981.
- 2 Anderson, Dorothy. Universal Bibliographic Control : a long term policy, a plan for action. Pullach/ München : Verlag Dokumentation, 1974.
- 3 Anonymous classics : a list of uniform headings for European literatures / compiled by the IFLA International Office for UBC ; edited by Rosemary C. Hewett. London : IFLA International Office for UBC, 1978.
- 4 Guidelines for authority and reference entries / recommended by the Working Group on an International Authority System, approved by the Standing Committees of the IFLA Section on Cataloguing and the IFLA Section on Information Technology. London : IFLA International Programme for UBC, 1984.
- 5 Beaudiquez, Marcelle; Françoise Bourdon. Management and use of name authority files (personal names, corporate bodies and uniform titles) : evaluation and prospects. München : K.G. Saur, 1991.

*International cooperation in the field of authority data.*⁶ In 1991 the first edition of UNIMARC authorities showed the adaptation of this concept to the IFLA and European exchange format, followed by the second edition in 2001 and in 2009, the third edition, completely updated and enlarged version of the *UNIMARC Manual: Authorities Format* was published (with a change of terminology, especially “uniform heading” is replaced by “authorised access point”).⁷ Since 1998, the year of publication of the *Functional requirements for bibliographic records (FRBR)*⁸ IFLA established a group to extend the concepts of FRBR to authority data, the Working Group on FRANAR (Functional Requirements for Authorities and Numbering of Authority Records). In 2001 IFLA published the 2nd edition of the *Guidelines for authority records and references (GARR)*⁹ in which the idea of a uniform heading was abolished in favour of an “authorised” heading, authorised by a national agency. This is taken up by the *RDA: Resource Description and Access*¹⁰ in the term of “preferred” heading. The 2008 review on the feasibility of an ISADN decided to abandon this concept. In summer 2009 IFLA finally published the *Functional requirements for authority data*¹¹ as a result of the FRANAR Working Group. They aim at two user groups: authority data creators and librarians, and end users. Authority data help the user to “find” and “identify” – the same tasks FRBR defines for bibliographic records – and they allow the user to “contextualise” (create relationships) and to “justify”. These two qualities are particular for authority data.

- 6 Bourdon, Françoise. *International cooperation in the field of authority data: an analytical study with recommendations*. München : K.G. Saur, 1993.
- 7 UNIMARC manual : authorities format. 2nd revised and enlarged edition. München : K.G. Saur, 2001; UNIMARC manual : authorities format. 3rd ed. / edited by Mirna Willer. München : K.G. Saur, 2009.
- 8 *Functional requirements for bibliographic records : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records*. München : K.G. Saur, 1998. Available also at: <http://www.ifla.org/VII/s13/frbr/frbr.pdf> [cited 2010-02-08].
- 9 *Guidelines for authority records and references / revised by the Working Group on GARE Revision*. 2nd ed. München : K.G. Saur, 2001.
- 10 *RDA: Resource Description and Access* [cited 2010-02-08]. Available at: <http://www.rda-jsc.org/rda.html>; Full Draft 11/24/08 [cited 2010-02-08]. Available at: <http://www.rda-jsc.org/rdafulldraft.html>
- 11 *Functional requirements for authority data : a conceptual model / edited by Glenn E. Patton ; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Final report, December 2008 / approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, March 2009*. München : K.G. Saur, 2009.

Authority data for old books in the European context – The CERL Thesaurus

Theoretical reflection on and practical creation and application of authority data are parallel. On the European level, and as far as early printed books are concerned, the practicality of the European approach can be best illustrated by the CERL Thesaurus – an authority file built up and maintained by the Consortium of European Research Libraries since 1999. The Data Conversion Group in Göttingen is responsible for the development and maintenance of this tool which is hosted in Göttingen.¹² The CERL Thesaurus puts authority data in the European context insofar as it brings together into one file authority data created by different European libraries for different purposes, and makes them serve new purposes. It can thus be considered as a melting point of various European initiatives in the field of authority data serving the history of the book.

Another international project in this area is the Virtual International Authority File (VIAF) initiated by the Library of Congress, the Deutsche Nationalbibliothek, the Bibliothèque nationale de France and OCLC.¹³ VIAF concentrates on personal names and consists (as of August 2009) of 10 million names from 16 authority files of 13 participating national libraries. A user interface is available at <http://viaf.org/>. VIAF does not concentrate on early printed books but it will certainly have an impact on CERL's Thesaurus work.

European aspects

The CERL Thesaurus offers access to five files: imprint places, imprint names, personal names, corporate names and references. For the time being all files contain information which is of concern for the printed book production in Europe from ca. 1450 to 1830/50. Thus the CERL Thesaurus is closely related – both in content and in chronological limits – to the Heritage of the Printed Book Database (HPB) that has been built up by the same Consortium of European Research Libraries since 1990.¹⁴ It is interesting to see that the development of the Thesaurus comes about nearly ten years later than the concern for the bib-

12 CERL Thesaurus [cited 2010-02-08]. Available at: <http://cerl.sub.uni-goettingen.de/ct/>

13 VIAF: The Virtual International Authority File [cited 2010-02-08]. Available at: <http://www.oclc.org/research/activities/viaf/default.htm>

14 CERL: Consortium of European Research Libraries [cited 2010-02-08]. Available at: <http://www.cerl.org/>

liographic records: a typical situation for authority control. In autumn 2009 the CERL Thesaurus contained data coming from 54 authority files contributed by 12 European countries: Belgium, Croatia, Denmark, Germany, France, Italy, the Netherlands, Poland, Spain, Sweden, Switzerland, United Kingdom. An up-to-date list of contributions is maintained on the website of the CERL Thesaurus.

A particular feature of the CERL Thesaurus and an issue of prime importance to authority data in the European context is multilingualism. As the records come from different sources, not only the language of cataloguing differs. Merging records from different European sources does not only mean merging different forms of names. It also implies merging comments and supplementary information in quite a number of different languages. The decision taken for the CERL Thesaurus was to leave these comments in whatever language they appear and to mark the language code in an \$8 subfield. The European Thesaurus is used by the Europeans – and they are supposed to understand (or at least to want to understand) the information in the record. In order to accommodate different scripts, particularly Cyrillic contributed through records of the National Library of Russia in St. Petersburg, the Thesaurus is fully UNICODE (UTF 8) compliant. Information in non-Latin script can be entered and retrieved making use of a virtual keyboard – made available to the user on the search screen of the CERL Thesaurus in the Internet.

The CERL Thesaurus is also firmly anchored in the European context because of its format. UNIMARC was chosen in order to use the European format for records exchange, but UNIMARC could not be fully applied as it allows for only one standard form in its authority format. The CERL Thesaurus however wishes and needs to retain all standard (authorised or preferred) forms defined by different contributing agencies without being selective or prescriptive for only one form. Therefore parallel headings had to be accommodated – and thus a vital principle of UNIMARC authorities had to be abandoned. UNIMARC is nevertheless the guiding structure to define the components of an authority record and the structure in fields and subfields. The format can therefore be qualified as “UNIMARC inspired”.

Contents of the CERL Thesaurus

With over 715,000 records the number of records in the CERL Thesaurus in autumn 2009 was impressive: 4,200 are for imprint places, 30,400

for imprint names, 670,000 for personal names, 8,300 for corporate names and 2,100 for references. Let us now consider in greater detail each of the components.

Place names

For the place name segment, the first contribution was an authority file for place names occurring in early printing created by the Bavarian State Library and published in 1991 as a book: “Druckorte des 16. bis 19. Jahrhunderts: Ansatzungs- und Verweisungsformen“ (Wiesbaden, Reichert). The file was then enhanced by standardised forms from the Swedish national bibliography for the 17th and 18th centuries and forms from the early book holdings from the National Library of Russia, St. Petersburg. In both cases these were not genuine authority records, but lists of names giving the standardised form and a variant. Today, all standardised place names available on files contributed for the Heritage of the Printed Book Database are also used to enhance the place name part of the CERL Thesaurus if they have standardised place names on the bibliographic records. Place names recorded in clearly defined fields of printers’ records are also integrated into this part of the Thesaurus. The place name file mirrors quite well the history of printing in Europe and shows – besides the standard forms defined by the various institutions or even projects – a huge number of variant forms in different vernacular languages, declinations or spelling differences and even mistakes (both ancient and modern) as found in the books themselves and/or in our catalogues. Fictitious place names present a particular issue. If they can be related to a real place in an individual edition of a book they are recorded in particular fields in the record for this real place name. Some of them cannot be located at all, and therefore have their own authority records. Geographic coordinates will be introduced in the foreseeable future and will allow CERL to build a map of printing places in Europe. The records for place names in the CERL Thesaurus have been edited in order to avoid duplicates among the different contributing files, and to make a single comprehensive record containing all relevant information, even references. This part of the Thesaurus not only helps to “find and identify” but also to “justify” and - by a special feature allowing the user to switch from the place name to the printers working at a particular place – also to “contextualise”. These are the tasks of an authority record as defined by IFLA’s FRANAR Working Group.

Authority control of place names has a long tradition before CERL's Thesaurus work if we remember the famous dictionary of Latin place names "Orbis Latinus" published by Graesse (Theodor Graesse: *Orbis Latinus oder Verzeichnis der wichtigsten lateinischen Orts- und Ländernamen*. 2nd edition. F. Benedict. Breslau 1909. New edition 1972). A lot of these forms are today part of the CERL Thesaurus and the heading of *Orbis Latinus* is always marked (*\$sOrbis*), but – most certainly – Graesse goes beyond printing places.

Imprint names

The imprint names file includes records of printers, publishers, booksellers, all those responsible for the physical production and distribution of a book. Authority files for imprint names have existed for a long time – one only needs to think of Josef Benzing's book on German printers of the 16th and 17th centuries (*Die Buchdrucker des 16. und 17. Jahrhunderts im Deutschen Sprachgebiet*, 2nd edition 1982). We also find quite a number of authority files established in the context of national retrospective bibliography. The authority files of the Short-Title Catalogue Netherlands, 1450–1800 (STCN) and the famous "Répertoire d'imprimeurs/libraires: vers 1500 – vers 1810" compiled by Jean-Dominique Mellot and Élisabeth Queval published as "Nouvelle édition mise à jour et augmentée" of the Bibliothèque nationale de France published in Paris in 2004 with its very extensive and well documented records are part of the CERL Thesaurus. The German file for printers used in the context of the GBV (Gemeinsamer Bibliotheksverbund) and for the VD17 was also integrated. As for VD16, authority records for printers and publishers were built up especially for the CERL Thesaurus containing the standardised name form (present in the bibliographic record) and reference forms (originally handled in separate records with a reference from an unauthorised other form of a name to the standardised form used in the VD16). A colleague in Stockholm is currently using the file to add records for early Swedish printers by working directly into the file.

Imprint names have special problems because the definitions of the standardised forms differ very much. Many bibliographic files do not use authority control for printers' names. According to the different cataloguing rules for early printed books the name as found on the book may be found in the publishers' statement – without any differentiation of homonyms or standardisation of variants like differ-

ent spellings or name forms of both first and family names. In some cases the whole imprint statement as found on the book is shown in the record. The German cataloguing rules restrict the indication of the name to the family name. When authority records exist, printers are sometimes considered as personal names, sometimes as corporate bodies and this is reflected in the differing structures of the names. Another problem is the definition of an entity: when is a new record needed for a printer or publisher – when the name does not change? Some tend to have a long history – family history and printer's shop history. All these rules are now well-defined in a paper accessible on CERL's website. Nevertheless, editing of the records present in the CERL Thesaurus according to these rules is still outstanding and even merging of the contributed records has not yet been undertaken. So the number of the records in the file is higher than the number of entities. Merging will soon start at least for those records which have identical standardised forms as headings.

In the records of the CERL Thesaurus we find again standard forms and variants, even if the variants are less frequent than for place names and will need a lot of editing, especially adding the non-standardised or differently standardised forms present in the bibliographic records. Editing should only start once the purpose these variants may serve is better known – especially for assisted searching. Again we see contextualisation taking place in linking the names to predecessors or successors and by the wealth of information contributed by some files. Imprint names are not only linked to place names in the internal structure of the CERL Thesaurus but also to an ever increasing number of printers' devices coming either from original files which already show a link to this kind of image information or from digitisation projects undertaken by CERL in order to enhance this part of the Thesaurus based on printed resources. Not only records contributed by the Italian Edit16 database allow direct linking to digitised printers' devices, images are also accessible from Ronald B. McKerrow's *Printers' & Publishers' Devices in England & Scotland 1485-1640* (London, 1913), Francisco Vindel's *Escudas y marcas de impressores y libreros en Espana durante los siglos XV a XIX (1485-1850)* (Madrid, 1935), and Philippe Renouard's *Les marques typographiques parisiennes des 15e et 16e siècles* (Paris, 1926) thanks to the digitisation of the printed volumes. This work is ongoing.

Personal names

In the personal names file – the most important part of the CERL Thesaurus – all persons contributing to the intellectual and artistic creation of a printed book are taken into account. The first input in this part of the CERL Thesaurus came again from the national retrospective bibliography, STCN, created by the Royal Library in The Hague. 52,000 authority records of the ESTC were integrated at an early stage. The biggest number of records is contributed by the German Personennamendatei (PND). This national authority file is highly specialised in names of early printing and they are marked by a special code which allows their automated selection for inclusion into the CERL Thesaurus. The PND also contains all the classical and mediaeval authors published in printed editions (Personennamen der Antike, PAN, Wiesbaden, Reichert, 1993 and Personennamen des Mittelalters, PMA first published in 1989, and in the 2nd enlarged edition in 2000) and since then maintained within the PND. The CERL Thesaurus receives a weekly update from the PND. Again the records contain standard and variant forms, sometimes pseudonyms, mostly biographical data, often sources. A particular strength of the list are names related to the provenance of the books, mainly former owners, marked as a special kind of record type – thus allowing particular features in retrieval. This is described in more detail by David Shaw in his book. Internal links within the Thesaurus may lead to associated corporate names or related personal names. Manual editing – described later – and inclusion of other authority files for personal names, especially those which control bibliographic data recorded in the Heritage of the Book Database are ongoing.

Corporate names

The corporate names included into the CERL Thesaurus have been less developed until now as names of corporate bodies in early book cataloguing are extremely difficult and heterogeneous among files. Most of the records have been contributed in the context of provenance information. They may also contain links to successors or predecessors, and show the same particular feature for provenance as the personal names file.

References

This part of the CERL Thesaurus is a sort of a help file. It gives full bibliographic information for all those titles which are quoted in different authority records, mostly in abbreviated form, and relates this

to the abbreviations which might vary considerably between different agencies contributing the files. The reference helps to create the mouse-over presentation in the Internet version of the Thesaurus for the abbreviated sources.

Providing content – editing the CERL Thesaurus

The CERL Thesaurus brings together authority records created for different authority files. That is very much in the tradition of CERL: merging in one file originally separate files coming from different libraries and/or cataloguing projects. That is exactly what CERL does in building up the Heritage of the Printed Book file hosted by OCLC. For internationally shared authority work this methodology is interesting, as there are experiments elsewhere with search engines, cross-file searching and record linking (especially in the VIAF project) whereas the CERL Thesaurus is applied for physical merging of files.

The CERL Thesaurus is not prescriptive of any standard form, but by quoting the agencies using different standard forms it may help cataloguers in choosing an appropriate standard form, if it is used for this purpose. As the records come from different files, it is not selective in content but aims at accommodating in a common structure all kinds of records and all information coming in. The value of the Thesaurus is to list all name forms in one record. As for all authority records the name forms thus become equally eligible as access points, the entity described in an authority record is clearly defined and named more by its number than by any heading. Therefore keeping already assigned authority record numbers of the source files is another important feature of the CERL Thesaurus. This is not only vital for automated updating but may also allow one day for directly linking to the bibliographical records in the Heritage of the Book Database or within the CERL Portal in which these authority numbers appear.

One of the interesting features of the CERL Thesaurus is that it does not use – or better, reuse – only genuine authority data. It also allows the use of standardised forms contained in a bibliographic record in order to construct an authority record out of information gathered in the bibliographic record. This is especially interesting for recording place names – but also for making best use of reference forms for personal names contained in a bibliographic record and for further developing the imprint names section.

The integration of all these records into a common file structure allows the creating of added-value features, e.g. links to places, successors, predecessors, provenance marks, etc. Thus the CERL Thesaurus also creates new records and enables the enrichment of the existing records by adding new information.

Editing the CERL Thesaurus is done in three ways:

(a) The preferred way is an automatic process. New records are integrated by machine process, records are created and prepared by automatic routines, duplicates are detected and merged by machine procedures. Further information is supplied by machine processes. Links to digitised images or to other authority records in the file are detected and added automatically into the records.

(b) Not all editing and merging of duplicates can be done by machine, as the records do not contain all the information needed for successful (i.e. correct) merging without intellectual input. Nevertheless, in view of the quantity of records, it is helpful to apply machine routines for detecting possible duplicates. Machine-supported manual editing is used to work in the files. The likely duplicates are listed in a file for editorial checking. The editorial control may approve a machine-detected duplicate, in which case a particular software routine merges the records, that means maintains all information given in the record marked for deletion which is not present in the record to be kept. The decision about merging has to be taken by a human eye and brain, comparing the two records and pressing the yes or no button. A good number of de-duplications can thus be achieved per hour. Sophisticated machine power – for the merging process – enables rather quick human editing with reliable results. Once the editor decides that there is no duplicate, the two records remain in the file and they are marked to be withdrawn from the duplicates list or to be maintained for further checking. For the time being, editorial control remains at the level of the authority records. According to the content of the two records, de-duplication takes place or the record is discarded. Quite a number of duplicates remain in the file which can only be detected by closer bibliographical and biographical studies.

(c) Manual editing also allows an editor to create new records, to enhance existing records or to merge identical records. New records can be created online in the CERL Thesaurus. This is done for Swedish printers names and was used for adding some place names in the past. This also needs to be done to edit user comments inserted via the an-

notation tool which allows the user to contribute to the development of the CERL Thesaurus. Manual editing is also necessary if a complete revision of the content of the records is envisaged.

All manual editing is done via the WinADH cataloguing client, maintained by the Data Conversion Group in Göttingen. A special login is required which can be obtained through CERL's Secretariat.

Use of the CERL Thesaurus

The CERL Thesaurus can be accessed and used in two ways. It is accessible free of charge for everybody via the Internet and WWW (<http://cerl.sub.uni-goettingen.de/ct/>). Unlike some other CERL products, especially the Heritage of the Book Database, the use is not limited to the CERL membership. A second access is through SRU – this allows the Thesaurus to be integrated into other applications, for the time being especially into the Heritage of the Printed Book Database and into the CERL Portal, but in principle and upon agreement of CERL, the Thesaurus could be integrated into other Internet applications and databases dealing with early imprints and needing to overcome the problem of different forms of names for one and the same entity. The SRU Interface is SRU 1.1 compatible and provides the CT information in a MARCXML syntax or in a proprietary format for using variant names in assisted searching. The indexes are the same as in the web interface. The WEB interface for the general user and access is supposed to be self-explaining. The search box on the first screen allows the user to choose a section of the Thesaurus with the help of a pull-down menu, to select a method of searching (truncated or exact name) and – at a very prominent place – to limit the search to provenance information only. “Browse” and “search” are the two possible access options. I would always recommend browsing in order to get a feeling of the content of the files. The use of a virtual keyboard for searching in non-Roman script is offered. “Browse” leads to an alphabetically sorted list of names, preceded by the number of hits. If the name is a reference form within the authority record the (first) standardised heading is also indicated by “=”. If one or more forms from this browse list – which also allows viewing previous or next page – are chosen, the system automatically ticks all other names contained in the same authority record and presented on this screen. From there the user is taken to a short display – in case of two or more hits – or to the full record in case of a single hit. Icons with mouse-over text display allow further navigation once the record

has been identified. The display of the internal format is very helpful for librarians who thus understand the history and creation of this record, especially if it was merged from several different and multilingual files. Other options are the display of related records, the display of standard forms, showing the chosen headings according to different files and/or cataloguing rules. If the icon for provenance information shows up on a record of a person or a corporate body, this allows access to different remote databases from which provenance information has been included into the CERL Thesaurus. The CERL Thesaurus is thus – to my knowledge – the first tool that allows provenance searching in different local catalogues through one access point: the record in the Thesaurus. In printers' records a link may be present which allows the user to access digitised printers' devices hosted on different remote systems.

One of the very important and future oriented WEB2 features of the CERL Thesaurus is to allow for annotations. They can be linked to a particular record by clicking on the annotation icon, which also allows reading the existing comments or annotations. By this means everybody – researcher or librarian – can contribute to the content and quality of the CERL Thesaurus.

Assisted searching

When CERL started to conceive and subsequently to build up the Thesaurus in 1999, this was governed by two complementary reasons: firstly, that great number of files arrived for the inclusion into HPB which in their original context were based on authority files, mostly for names of persons, but sometimes also for printers and publishers. This original link to an authority file was maintained in the HPB, recording the respective ID Number of the authority record, but was no longer of any help in searching as this number lost its linking value. The information contained in the authority records referred to was lost for HPB, which in most cases only retained the standardised form (there is nevertheless the option to include all name forms in the bibliographical record). The second and principle reason was HPB itself. This is a highly specialised information tool for early printed books from 1450 to 1850. The information contained in the records is carefully maintained, mapped and thus retrievable in a way that is most adequate for early printing, much better than what our OPACs can do. Still, this sophisticated retrieval cannot by itself overcome different traditions of cataloguing transported in the records. So the CERL

Thesaurus was meant to directly help the user of HPB to find his way through the multiplicity of cataloguing traditions and to give him the best and the most complete results for a particular search without asking him to perform himself the whole intellectual work (like knowing all variants of a place name). When searching for a place name or a personal name the user does not only get the hits for his search term, he is also invited to find more results by using assisted searching (clicking on the link “find variations of this ... name”). If he chooses this option, the search term is transferred to the CERL Thesaurus (a separate and distant file) and the hits within the CERL Thesaurus are displayed. According to the choice of the user, a record identified on the Thesaurus can be used to enhance the search on HPB (pressing the “search” button displayed on each retrieved record) by taking into account all name forms present in the CERL Thesaurus record, translated into “or” searched within HPB. This feature is available and works very well for place names (called: “publisher location” in the HPB) and personal names. For imprint names (“publisher” in the HPB) it is offered but has still to be more developed and better adapted. The Heritage of the Book Database hosted by OCLC thus makes use of an external file in order to help the user to find his way in the heterogeneous content of the records and to get a better retrieval result. This feature is undoubtedly of immense help in retrieval, but it cannot overcome all search problems in such a composite file. In some cases only a firm linking structure could do this. Three examples taken from the place names may illustrate this. Homonymous names (like Frankfurt am Main, Frankfurt an der Oder, sometimes only called “Frankfurt” as a variant) create huge problems in assisted searching and falsify the retrieval results. Homonymous name variants create the same kind of confusion. For example, Lugdunum is a variant for Lyon and for Leiden (Lugdunum Batavorum). Fictitious place names (especially coded) cannot be used for assisted search facility. A fictitious name like “Utopia” or “Eleutheropolis” can be used for several real places or to describe a separate (unidentified) entity. Fictitious names of the type “Rom i.e. Schwabach” would give wrong search results in assisted searching as all imprints published in Rome would be included, and by a special marker (410 11) they can be excluded from assisted searching. In order to obtain acceptable, albeit not perfect results careful mapping between the search queries of the questioning system and the indexing of the Thesaurus needs to be done and de-

signed in the query routines. Imagine the number of false hits you get when your querying system translates a name like “Victorius, Petrus” (\$aVictorius\$bPetrus) into a search “Victorius or Petrus” instead of “Victorius and Petrus”. Even worse things can happen when parts of the name are not taken into account, particularly surnames, in the CERL Thesaurus in the subfield \$r, e.g., “\$aThomas\$r Aquinas”, as a query was not answered because the CERL Thesaurus listed Aquinas in \$e and not \$b. The integration of the CERL Thesaurus into another Search application is in no way an easy and automated procedure but needs careful intellectual defining and monitoring.

The CERL Thesaurus is also used for assisted searching within the CERL Portal. On the advanced search screen the search for “persons” and “places” can be extended by clicking the “add variant forms” link next to the search field. Here the application is a bit different as the user himself may choose name forms which he feels are helpful to get a better recall of records. This is a useful feature as some records – for places and persons – contain a huge number of variants, taken from all kinds of sources, and which are not necessarily present on any bibliographic file. Automated searching for all these many forms is machine time consuming and easily leads to time-out results (which do not help at all) especially addressing remote indexes via Z 39.50, or brings up false results if only part of the files provided show hits. The use of the CERL Thesaurus in the context of the CERL portal also brings up the question of relevant content. As the CERL portal also gives access to manuscripts – both mediaeval and modern – the entities of the CERL Thesaurus are no longer sufficient.

Further development of the CERL Thesaurus – Semantic Web

The value of the CERL Thesaurus is also found in itself – in a tool that gives us the possibility to bring together and to edit authority information for early printing in a single and authoritative environment which can be moved closer and closer to research context. The Thesaurus allows not only editing of existing records – even by using the annotation facility – but creation of new records for information not yet available in the file by all sorts of procedures, to link to all kinds of external resources in order to enhance the impact of the information, but it also provides the structure to bring in new and different entities relevant for early printing.

Enhancing existing content

There are lots of wishes and examples: new records could be created and added by checking the names present in the HPB and the CERL Portal which are not yet included in the Thesaurus. The records can be enhanced, e.g., by geographic coordinates from the Bernstein project to be added to the place names. With these coordinates we can create an electronic map which would free us from the historically doubtful use of country codes and maybe give a better result to research questions like “French imprints of the 16th century”. Geographic historical information is also provided by other projects not limited to the history of printing. Bavaria, for instance, creates an electronic tool for its regional history, Bayerische Landesbibliothek online, Ortsdatenbank. This information can be linked to records in the CERL Thesaurus so that you can switch from one instrument to the other. The imprint name records allow for further connections. All started by linking to the digitised images of printers’ devices of the Italian project in the context of EDIT16. Since then a number of reference works for printers’ devices (other examples of predecessors of authority files, by the way) have been added and much more can be done in this area. The links to digital images available on the internet through full text digitisation may not only be interesting for access from the bibliographic record to the full text but also for printers’ devices and examples of printing characters to be related to a particular printer. The records for persons can be linked to portraits of persons, more and more accessible in different files, or digitised copies of people’s handwritings, important both for provenance and autograph studies. And there are many more ideas and possibilities...

Particular care will be given to develop the CERL Thesaurus to better accommodate information on manuscripts, to better serve the CERL Portal which includes manuscript information. The Thesaurus already has important assets as the names of mediaeval and ancient authors from the German authority files PAN and PMA are included. The place name section would have to be developed for regional information about production places for manuscripts. The personal names must be expanded to include scribes or manuscript illuminators and the corporate names file to include manuscript holding institutions, both modern and historical, scriptoria, monasteries, etc. A large number of these standardised name forms are available in the records of the manuscript files.

More entities can be added in both the place names and the imprint name files taking into account information on bindings – again providing places, workshops and digital images, e.g., through the German file for bindings, Einbanddatenbank – or on paper production in the context of the Bernstein project, again providing places, workshops and digital images. In order to maintain the relevance, special codes for different kinds of places will be introduced for printing places.

The CERL Thesaurus is not a completed authority file but a growing and inspiring tool for information on early printing. Therefore nowadays it is very much at the heart of CERL's activities.

Files for new entities

Further development goes even beyond enhancing the files which are already there. The CERL Thesaurus may need to become hospitable to other kinds of authority records. For some time already we have been discussing about a file for uniform titles – so very important in a FRBR environment, and vital for finding a way through manuscripts and early printing traditions of the core works of the European cultural heritage.

And the need already arises – especially through the Portal development – to have multilingual subject approaches in all kinds of areas, to mention only book illustration, book binding vocabularies, provenance information, manuscript cataloguing terminology, historic events, etc. I am not saying that the Thesaurus is the instrument to be developed into a multilingual dictionary for all information of this kind, but it already provides structures which are useful for this kind of development.

Inclusion into the CERL Research Portal

The CERL Thesaurus is also the core element in the development of what we call the CERL Research Portal giving an even better access to the various and rich information gathered by CERL for the users. It might allow them to enter a search term and receive information about persons, when the name is found in the Thesaurus personal name file, about organisations, if found in the corporate name file, about titles using data coming from the HPB. Information can be further differentiated into books showing titles from HPB, book ownership from the Thesaurus provenance data, related publications from the subject headings in the bibliographic record, libraries which own copies from the holding information in the HPB, etc.

Semantic Web

In February 2009 I was invited to demonstrate the CERL Thesaurus to a community of researchers – mediaevalists – interested in Semantic Web applications. They all did quite different research projects, creating databases and web pages for their particular concerns. Our concerns – as librarians – and their concerns wanting an ongoing impact of their findings may be joined in a common care for the CERL Thesaurus. In fact the value of the Thesaurus and its future lies in its role of providing on the European level authority data to organise intellectual content.

The CERL Thesaurus provides an object identifier (and maintains all object identifiers of contributing libraries). It gathers all kinds of name forms (thus having a dictionary or vocabulary function), it provides the necessary (basic) information about an entity in a multilingual environment. It allows the establishment of relationships between entities, both controlled: e.g., predecessors, successors, pseudonyms, fictitious name forms, and machine processed: related place names for printers, etc. The CERL Thesaurus has a pathfinder function and works as a finding aid. It points to the relevant material and allows for “post-coordination” in a context where pre-coordination – the usual library and cataloguing task – is no longer sufficient, this means in the WWW. It thus helps to discover more. This requires a sophisticated implementation so that the useful functions of the CERL Thesaurus help the user to navigate the interlinked cosmos.

In the European context the CERL Thesaurus invites both reuse and cooperation in the area of early printing and – one day – in the manuscript tradition. Authority records are expensive to build and to maintain. It is necessary to join forces for better and fuller information, serving not only one’s own particular purpose but multiple interests. Therefore it is better to standardise – using existing tools – than to reinvent. Every name form is admitted in the Thesaurus – but it must be integrated into one entity, into one name record. The common denominator is the object identifier, not any name. Individual projects may use any free text, standardised entries – provided usually by libraries – to ensure long-term retrieval and access. Tools – like the CERL Thesaurus – allow searching of any structured information with a wide vocabulary. The researcher describes, the librarians standardise in order to provide wider access. So everybody is invited to enrich the CERL Thesaurus as a backbone of Semantic Web application, thus caring for its growing relevance in content and ensuring its long-term sustainability and effectiveness.

References

- ANDERSON, Dorothy. *Universal Bibliographic Control : a long term policy, a plan for action*. Pullach/ München : Verlag Dokumentation, 1974.
- Anonymous classics : a list of uniform headings for European literatures / compiled by the IFLA International Office for UBC ; edited by Rosemary C. Hewett. London : IFLA International Office for UBC, 1978.
- BEAUDIQUEZ, Marcelle; Françoise Bourdon. *Management and use of name authority files (personal names, corporate bodies and uniform titles) : evaluation and prospects*. München : K.G. Saur, 1991.
- BOURDON, Françoise. *International cooperation in the field of authority data : an analytical study with recommendations*. München : K.G. Saur, 1993. See also French version: Bourdon, Françoise. *International cooperation in the field of authority data = Coopération internationale en matière de données d'autorité auteurs*. // BBF 4(1994), 142-144 [online]. Available at: <http://bbf.enssib.fr/> [cited 2010-02-08].
- Functional requirements for authority data : a conceptual model / edited by Glenn E. Patton ; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Final report, December 2008 / approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, March 2009. München : K.G. Saur, 2009.
- Functional requirements for bibliographic records : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München : K.G. Saur, 1998. Available also at: <http://www.ifla.org/VII/s13/frbr/frbr.pdf> [cited 2010-02-08].
- Guidelines for authority and reference entries / recommended by the Working Group on an International Authority System, approved by the Standing Committees of the IFLA Section on Cataloguing and the IFLA Section on Information Technology. London : IFLA International Programme for UBC, 1984.
- Guidelines for authority records and references / revised by the Working Group on GARE Revision, 2nd ed. München : K.G. Saur, 2001.
- RDA: Resource Description and Access [cited 2010-02-08]. Available at: <http://www.rda-jsc.org/rda.html>; Full Draft 11/24/08 [cited 2010-02-08]. Available at: <http://www.rda-jsc.org/rdafulldraft.html>
- Statement of principles adopted by the International Conference on Cataloguing Principles, Paris, October 1961. // Report / International Conference on Cataloguing Principles, Paris, 9th–18th October, 1961 ; edited by A. H. Chaplin and Dorothy Anderson. Reprint. London : IFLA International Office for UBC, 1981.
- UNIMARC manual : authorities format. 2nd revised and enlarged edition. München : K.G. Saur, 2001.
- UNIMARC manual : authorities format. 3rd ed. / edited by Mirna Willer. München : K.G. Saur, 2009.

Readings and website references

- BOURDON, Françoise. *Modelling authority data for libraries, archives and museums : a project in progress at AFNOR*. // *Cataloging & classification quarterly* 38, 3/4(2004), 505-516, and in: *Authority control in organizing and accessing information : definition and international experience* / Arlene G. Taylor; Barbara

- B. Tillett, editors. New York [etc.] : The Howarth Information Press, 2004. Pp. 505-516. Available also at: http://www.sba.unifi.it/ac/relazioni/bourdon_eng.pdf [cited 2010-02-08].
- CERL: Consortium of European Research Libraries [cited 2010-02-08]. Available at: <http://www.cerl.org/>
- CERL Thesaurus (Homepage information) [cited 2010-02-08]. Available at: <http://cerl.sub.uni-goettingen.de/ct/>; http://www.cerl.org/web/en/resources/cerl_thesaurus/main
- FABIAN, Claudia. The CERL Thesaurus File. // Cataloging & classification quarterly 39,1/2 (2004), 413-420, and in: Authority control in organizing and accessing information : definition and international experience / Arlene G. Taylor; Barbara B. Tillett, editors. New York [etc.] : The Howarth Information Press, 2004. Pp. 413-420. Available also at: Authority control: definizione ed esperienze internazionali / Atti del convegno internazionale, Firenze, 10 - 12 febbraio 2003. Firenze, 2003. Pp. 327-333 [cited 2010-02-08]. Available at: http://www.sba.unifi.it/ac/relazioni/fabian_eng.pdf
- Library of Congress authorities [cited 2010-02-08]. Available at: <http://authorities.loc.gov/>
- MARC 21 format for authority data [cited 2010-02-08]. Available at: www.loc.gov/marc/authority
- PATTON, Glenn E. FRAR : extending FRBR concepts to authority data. // World Library and Information Congress: 71st IFLA General Conference and Council, 14-18 August 2005, Oslo, Norway [cited 2010-02-08]. Available at: <http://archive.ifla.org/IV/ifla71/papers/014e-Patton.pdf>
- TILLET, Barbara B.; Corey Harper. Library of Congress controlled vocabularies, the Virtual International Authority File, and their application to the Semantic Web. // World Library and Information Congress: 73rd IFLA General Conference and Council, 19-23 August 2007, Durban, South Africa [cited 2010-02-08]. Available at: http://archive.ifla.org/IV/ifla73/papers/147-Tillet_Harper-en.pdf; German translation: Kontrollierte Vokabulare der Library of Congress, der Virtual International Authority File und ihre Anwendung im Semantic Web [cited 2010-02-08]. Available at: http://archive.ifla.org/IV/ifla73/papers/147-Tillet_Harper-trans-de.pdf
- VIAF: The Virtual International Authority File [cited 2010-02-08]. Available at: <http://www.oclc.org/research/projects/viaf/>; Beta search engine [cited 2010-02-08]. Available at: <http://viaf.org/>

Biographical sketch

Dr Claudia Fabian has worked in the Bavarian State Library, Munich, Germany since 1984, after receiving the MA and teaching degree in Classical Philology and French. After being responsible for cataloguing and authority control for years, she was Head of Users Services, and since 2004 she has been the Head of the Department of Manuscripts and Early and Rare Printed Books. She is a member of several groups in national library committees and represents the Bavarian State Library in CERL's (Consortium of European Research Libraries) committees.

NOVI KONCEPTUALNI MODELI ZA ORGANIZACIJU INFORMACIJA AUTORIZIRANIH PODATAKA U EUROPSKOM KONTEKSTU

Sažetak

Na opisu Tezaurusa Konzorcija europskih znanstvenih knjižnica (CERL Thesaurus) rad prikazuje izgradnju europske datoteke autoriziranih podataka za stare tiskane knjige. Opisuje se sadržaj i struktura Tezaurusa, načini na koji se prikuplja njegov sadržaj, postupci uređivanja datoteke, daljnji razvoj, uporaba računalne tehnologije (potpomognuto pretraživanje, pristup lokalnim katalozima, mehanizam mrežnih veza s drugim relevantnim izvorima informacija), te uloga datoteka autoriziranih podataka za organizaciju intelektualnog sadržaja.

Ključne riječi: autorizirani podaci, stara tiskana knjiga, CERL-ov Tezaurus, europski kontekst