IMPRINTS AND OWNERS

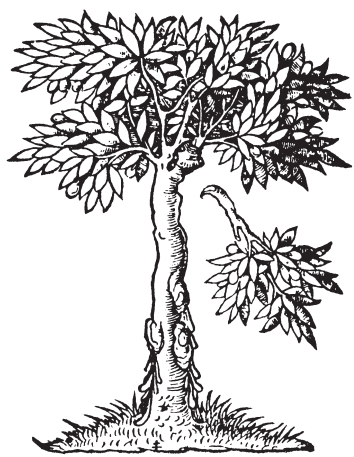
CERL PAPERS · VII

Imprints and owners:

Recording the cultural geography of Europe

Papers presented on 10 November 2006
at the CERL Seminar hosted by the
National Széchényi Library, Budapest

EDITED BY
David J. Shaw



London 2007
Consortium *of* European Research Libraries

Published in 2007 by
The Consortium of European Research Libraries
The Finsbury Business Centre, 40 Bowling Green Lane,
Clerkenwell, London ECIR ONE
Telephone 020 7415 7134 Fax 020 7970 5643
www.cerl.org

ISBN 0–9541535–6–1

Designed by James Mosley, Justin Howes and Derek Brown, and typeset
by Derek Brown in Matthew Carter's Galliard CC

Produced by Oblong Creative Ltd, Wetherby, West Yorkshire

Contents

Character encoding in the MOKKA-R

ÁDÁM HEGYI AND RÓBERT ZAWIASA

A Magyar Országos Közös Katalógus – Régi Nyomtatványok (MOKKA-R) során többször találkozunk a karakterkódolás problémájával. Karakterek szempontjából az adatbevitelt, a tárolást és a megjelenítést külön-külön tárgyaljuk. Az általunk működtetett adatbázisba alapvetően két módon kerülnek be adatok: offline konverzióval és online szerkesztéssel. Konverzió esetén rendszerint nem UNICODE-ot, hanem DOS 852-es vagy Windows 1250-es kódolást használtak a készítők. Ilyen esetben legtöbbször automatizálható a karakterek UNICODE-ra való átkonvertálása. Online szerkesztés esetén a karakterkódolás szempontjából a kliens-szerver viszony okozhat gondot, mivel a böngészőprogram fontkészlete az adott gép operációs rendszerétől függ.

The Hungarian National Shared Catalogue – Early Prints (MOKKA-R) digital project aims to carry out a complex system of tasks. The basic aim of the project is to create a national finding list of early printed books, and also to promote the more efficient and speedier processing of such documents. It is very important for the history of books that in addition to the bibliographic description of books printed before 1850, the individual characteristics (e.g. possessor, binding, marginalia, etc.) should also be recorded. Books published before 1850 found in the Carpathian Basin constitute an entity from the viewpoint of cultural history. It is therefore justified to extend the project to beyond national borders. As far as possible, access must be created to the holdings of all collections in the Carpathian Basin which are linked to the history of books in the period between 1450 and 1850. This can make the study of the collections – which are already related historically – far more effective.

The project covers all printed documents (books, journals, pamphlets, maps, etc.) produced between 1450 and 1850. Two considerations together determine the scope of collecting: one is the period, that is, documents printed between 1450 and 1850, the other is the geographical approach,

that is, all printed materials to be found in the present territory of Hungary and in the Carpathian Basin.

The MOKKA-R will be operated not by an integrated library system but with the use of XML which is much easier to modify. It comprises two main parts: the OPAC module and the cataloguing module. The system is accessible on-line, meaning that both the OPAC and the cataloguing module can be reached through the TCP/IP network. Independently of that, there is also a possibility of cataloguing off-line. The search interface is open to everyone, while the editing part is password protected.

Presently there are more than 100,000 records in this catalogue, and a certain part of it is already available in the Hand Press Book Database. In my presentation I am going to speak about the problems of character encoding which we had to face during the past years. These problems are handled and solved by our programmer, Róbert Zawiasa.

During the processing of old books we have been regularly confronted with the difficulty that it is not standard founts that we have to record. Several different character encoding systems are used in parallel worldwide. The MOKKA-R intends to adopt open standards and so we chose UNICODE because nowadays this is the standard that is able to handle the ordinary and specific characters of most of the languages.

When working on the MOKKA-R there are several work stages where the problem of character encoding arises. The majority of the problems emerge when the texts prepared with different code tables have to be loaded during data entry. In Hungary many catalogues still use the DOS code table in which the fonts are stored in a way which is very different from the present-day usage. The demand for displaying non-standard fonts during retrieval also may arise. Unfortunately, loss of data may also occur in the case when the MOKKA-R transfers records to other databases. Let us examine this in detail!

From the point of view of characters, data entry, storing and displaying have to be discussed separately. Basically, our database is filled with data in two ways: by off-line conversion and by on-line editing. Very often we have to convert data that were prepared with DOS 852 or Windows 1250 encoding, and not with UNICODE. In such cases, generally, the conversion of characters to UNICODE can be automated. Naturally, there are always exceptions. If errors are left and noticed after the conversion the correction is done manually by editing the text. Very rarely there are cases when even this manual method cannot be used for correction because there are letters which are not known even by UNICODE. Usually, these

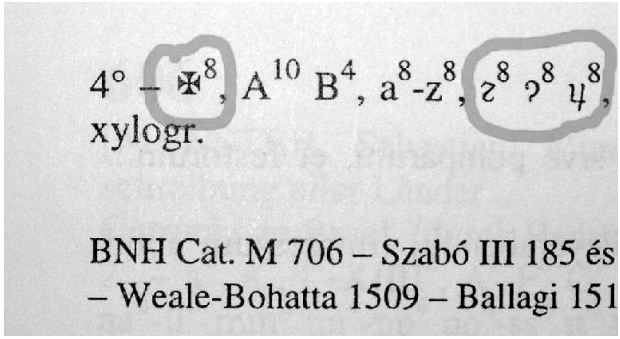


Figure 1

are specific characters used in signatures [Fig. 1]. For example, that is why the following figure was given the 'weird bird' name [Fig. 2]. In the source code it is solved by XML tags: <*char name="weird bird"*/>. Unfortunately, in such a case we violate our own principle that we store everything in UNICODE, but because this is marked specifically, later we can replace it with an image.

New records can also be entered to the database of the Mokka-r via a form on the Internet. In this case, from the point of view of character encoding the client-server architecture may cause problems because the font set of the browser depends on the operating system of the respective computer. For example, if we try to edit a text in Hungarian in a German Windows, we will not find Hungarian characters among the keys. An error will also appear when our system offers a value list of the available fonts because the character seen in the browser depends on the character table of the client and not on that of the server. A good example of this is the HTML file that displays full letters of the UNICODE: the Hungarian version of Windows which I use displayed many Asian letters incorrectly [Fig. 3]. No needs for improving this have been indicated yet and so we have not elaborated a solution for the problem of client-side character encoding.

Now some words about storing. The general principle is that we store every character in UNICODE. However, this depends, to a great extent, on the data entry. The examples mentioned above cause problems in storing, too, since the characters entered incorrectly will be stored defectively, and the 'weird bird' not recognized by

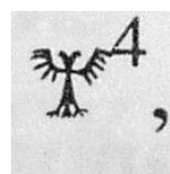
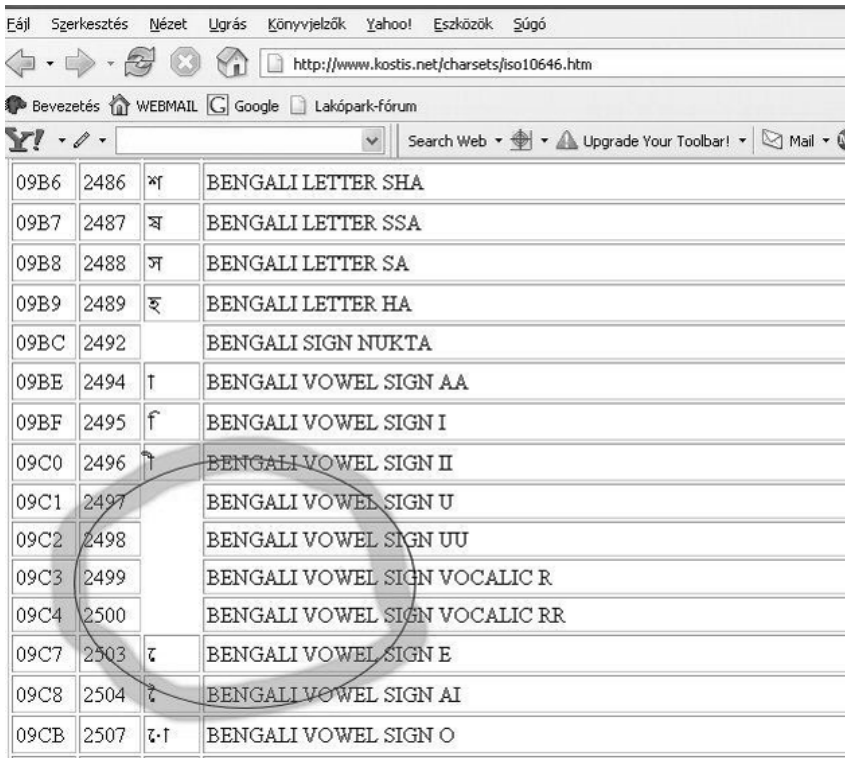


Figure 2



09B6	2486	শ	BENGALI LETTER SHA
09B7	2487	ষ	BENGALI LETTER SSA
09B8	2488	স	BENGALI LETTER SA
09B9	2489	হ	BENGALI LETTER HA
09BC	2492		BENGALI SIGN NUKTA
09BE	2494	া	BENGALI VOWEL SIGN AA
09BF	2495	ি	BENGALI VOWEL SIGN I
09C0	2496	ী	BENGALI VOWEL SIGN II
09C1	2497	ু	BENGALI VOWEL SIGN U
09C2	2498	ূ	BENGALI VOWEL SIGN UU
09C3	2499	ৃ	BENGALI VOWEL SIGN VOCALIC R
09C4	2500	ৄ	BENGALI VOWEL SIGN VOCALIC RR
09C7	2503	ে	BENGALI VOWEL SIGN E
09C8	2504	ৈ	BENGALI VOWEL SIGN AI
09CB	2507	ো	BENGALI VOWEL SIGN O

FIGURE 3

UNICODE cannot be handled properly. Still, it was important for us to choose this solution because the platform independence of UNICODE makes possible providing long-range services and simple communication with different databases.

In every case, searching in the MOKKA-R is executed with normalization. This means that the accented letters of the Latin alphabet are considered equivalent with each other: if an 'i acute' ('í') is typed in, it will automatically correspond to an 'i' ('i'). A special emphasis was put on the Hungarian alphabet, and so the letters 'a acute' ('á') and 'e acute' ('é') are not identical with the 'a' and 'e', similarly the letters 'o with double acute' ('ő') and 'u with double acute' ('ű') are not identical with the 'o' and 'u'. Presently this search principle applies only to the Latin alphabet; in the

course of searching we can normalize neither Greek, nor Cyrillic letters. Naturally, if demands appear, there is a technical solution for this.

At the beginning of our presentation we referred to the relationship between the MOKKA-R and other databases. Unfortunately, problems may occur in this area, too. The MOKKA-R also provides records to other systems. However, there is no use in our storing fonts in UNICODE if the receiving database does not recognise them. The MOKKA-R is able to export different data exchange formats, for example MARC, and others. However, MARC often uses ANSEL character encoding which does not know for example the dagger or the paragraph symbols. Therefore letters stored correctly in UNICODE are converted improperly to the databases using MARC.

What conclusions can be drawn from all this? It is worth using UNICODE because with this our texts preserve their value for a long time. Its usage in case of the MOKKA-R enables the system to handle the specific characters appearing frequently in old printed materials. All of the databases loaded into the MOKKA-R can be converted into UNICODE unless the UNICODE itself does not know a sign or letter, either. In the course of data entry and display we cannot influence the character table and font set of the client and this can result in errors, too.

As you may see, still many tasks are before us in connection with this in the future.

The Union Catalogue of Rare Book Collections in the Carpathian Basin: about the Expert System on Book History in Hungary

ISTVÁN MONOK

A mai Magyarország könyvtáraiban közel 1 millió 1850 előtti könyvet őriznek. Könyvtáraink a régi könyvek feldolgozását kezdettől fogva a CERL elveinek megfelelően végzik; a leírás kezdettől fogva kiterjedt a könyvek proveniencia vizsgálatára, fizikai állapotának rögzítésére, és az egyes kötetekben megőrzött rejtett művek feltárására is.

A magyar olvasmánytörténeti, ERUDITIO adatbázis (www.eruditio.hu) ma négy nagyobb egységből áll:

- *a korabeli könyvjegyzékek adatbázisa (1500–1750),*
- *az ezeken szereplő egyes könyvtételek adatbázisa,*
- *a könyvtárak tulajdonosairól készült adatbázis és*
- *a könyvtárak történetére vonatkozó szakirodalmi adatbázis.*

Ettől a négy adatbázistól függetlenül a nemzeti könyvtárban elkészült a Magyarországon 1850 előtt megjelent könyvek nemzeti bibliográfiája, az 1700-ig bárhol megjelent magyar vonatkozású könyvek bibliográfiája. Külön épül a magyarországi nyomdászok adatbázisa, és a magyar könyvtörténeti szakirodalmi adatbázis.

Mivel Magyarországon az állam azokat a projekteket támogatja, melyekben a résztvevő tagkönyvtárak száma több tucat, vagy néhány száz, javasoljuk a filozófiaváltást a CERL számára is: sokkal szélesebb körűvé kellene tennie a tagkönyvtárak hozzáférését a professzionális könyvtörténeti adatbázisokhoz.

The system of Hungarian libraries has been partially transformed since 1990 and certain units of the library network have been assigned different functions. In the new distribution of tasks, the role of the National Library has increased. At the same time, as there were no immediate state measures taken, the spread of information technology in libraries offered

excellent opportunities to IT enterprises to introduce into the processing activity a number of systems incompatible with each other. Fortunately, these systems were based on international standards both in the sense of library work and in that of informatics. As a result, establishing a shared national catalogue is not impossible, and so Hungary's 52 major libraries are building a shared catalogue, an activity which in coming years will be extended to ecclesiastical collections.

In financing library work, the most important role is given to the state, to local governments and to churches, all three of which are struggling with serious financial difficulties. Consequently, libraries cannot look forward to very positive developments in the near future. In this dire situation, the shared catalogue is of even greater significance: if we are lucky, maintaining one might be a way of saving costs.

The experience of processing older printed books using IT tools varies widely depending on the type of library. What I mean is that the attitude to them is different in the National Library, in specialist scientific libraries, in the general public and in church libraries. Accessibility through the Internet is not very welcome in ecclesiastical libraries. One reason is that it will increase the demand for their books, which they fear will disturb the intimacy of the religious community. Another explanation is that churches still remember the 1948–1952 period when their books were secularised and confiscated. What they managed to save at that time was what nobody knew existed. Therefore, serious powers of persuasion have been required so that churches would agree to having their pre-1850 stocks processed in a union catalogue. It needs to be added that in Hungary nearly half of all old books are still to be found in church libraries.

In Hungarian libraries, dealing with old books, as well as historical scholarship in general is almost a daily task for librarians. It is a tradition which did not disappear in the second half of the twentieth century. On the contrary, it was rather strengthened when a number of distinguished historians and philologists who, for political reasons, were not allowed to lecture at universities or to be active in the research teams of the Academy were 'exiled' to libraries, archives and museums. Thanks to these phenomena, scholarly work in libraries has managed to maintain its outstandingly high standards.

For the reasons outlined earlier, identifying and cataloguing old books has not always followed strictly the guidelines librarians would apply. From the very beginning, it has been much more thorough, including recording the provenance of books, the description of their physical state

and the listing of works hidden and preserved in each volume (introducing content). The main criteria of processing have been those of cultural and book history.

In 1979, for example, a fundamental research project into literary history conducted at the University of Szeged set the aim of mapping out the intellectual field of force within which old Hungarian literature emerged (in the 16th and 17th centuries). In brief, what the team wanted to document was what the various professional, religious and social groups in the various geographical areas were reading in the Hungary of the time. This was the starting point of two research projects: firstly, uncovering the archival sources (book lists) of the history of reading, and secondly, the examination of manuscript notes in printed books. Naturally, at the same time retrospective bibliographic work continued in the National Library. That is how three complementary sets of databases have been made (1) of the books that were published in Hungary, (2) of the ones preserved in Hungary and its successor countries, and (3) of those that, according to archival documents, were available in contemporary Hungary.

Today the so-called ERUDITIO database is composed of four major units, maintained jointly by the Hungarian National Library and the Library of Szeged University: the database of contemporary book lists, (1500–1850), the database of individual book items in them, the database of owners of libraries, and finally, the secondary literature database of the history of libraries.

Independently of these four databases, the National Library has produced and regularly updates the national bibliography of pre-1850 books published in Hungary and the bibliography of Hungary-related books published anywhere in the world. In this context, a book is Hungary-related if its author, language or theme is Hungarian. The Hungarian printers' database (*Clavis typographorum regionis Carpaticae*), as well as the secondary literature database of Hungarian book history have been developed separately. By the end of the year 2006, the authority control file of authors publishing in Hungary before 1800 will be completed, and we are also elaborating the authority file for contemporary geographical names. We would like CERL to accept these as the valid authority data for the Carpathian Basin, as this material also includes the alternative versions in the various national languages (i.e. Latin, Slovak, Romanian, Croatian, Serbian, German and Yiddish). The form of classification is the way the author used his name in his day (be it a German, Hungarian or any other national language form). For geographical names, we recommend that

CERL should use the official names of the given time, which in the Carpathian Basin is always the Latin version, as up to 1844 the official language of Hungary was Latin.

Another, recently launched project is the union catalogue database of the pre-1850 manuscripts preserved in the Carpathian Basin. This database is intended to include the manuscripts that we know of but have been destroyed. For building it, in the same way as before, cultural historical criteria are applied.

Work on the National Shared Catalogue started as early as 1996, but it only began its genuine operation when the National Library as an institution undertook the tasks of organisation and professional co-ordination (2002). Today there are already 52 libraries acting together on extending this catalogue. Initially, there was no plan for a separate catalogue of old books, but it soon became clear that the standardised forms for describing modern books are not suitable for describing old books. Thus, many libraries started their processing work based on the ISBD/A standard. By 2006 we have reached the stage whereby we have created a common bibliographical worksheet and OPAC; also the new system is capable of receiving records from other software used in Hungary.

We are obtaining continually from the relevant institutions records of the 97,000 old books which feature in the shared catalogue of modern books, adding to them the National Library's catalogue of sixteenth-century imprints and are starting to prepare church libraries for joining. It is estimated that nearly one million pre-1850 books are kept in the libraries of present-day Hungary. The cataloguing practices of libraries vary, mainly as regards the depth they go into: reading manuscript notes causes a lot of difficulty in many places, as does giving a professional description of bindings.

The shared catalogue is already working interactively with the ERUDITIO databases. Our goal is that within a single unified system it should be possible to identify book items that archival sources describe in no more than a few words. This will be enabled by recording provenance notes and will contribute to the full value of a hit: For example, if an owner's inventory list features a 'Vergilius', but this is an item which it is impossible to identify and we know no more about the given book, and if today there is a work by Vergil in which the given owner's note is available, then the system will automatically identify the item. Naturally, this falls into the category of the 'ideal case'. We hope, however, that on the basis of provenance notes it will become possible to reconstruct several lost and dispersed libraries.

The Expert System on Book History in Hungary

One of the high-priority tasks of the Book History Research team set up within the National Library this year, will be to create the 'Libraries of the Carpathian Basin 1000–2000' database. Thus, existing databases will be complemented by documenting libraries known only from descriptions and secondary sources.

I wish to talk separately about the co-operation in describing books that has been going on between the major libraries in the countries of the Carpathian Basin since 2000. In an ideal world, the union catalogue of old Hungarian books would receive the records of Slovak, Romanian, and also partly Croatian, Austrian and Serbian libraries, because pieces from various collections in the former Hungary have ended up in other countries' libraries. We have reached a theoretical agreement on this, and hope that the technical framework will allow us to realise this joint goal.

In our co-operation with the CERL Hand Press Book Database, it would be ideal if it did not have to receive records separately from each Hungarian library, as records are kept in many different systems, and the Hungarian state has already spent considerable sums on conversion. It is vitally important that the CERL community should understand that in Hungary, and probably in the whole of the Central European region, today the state still takes a much more considerable role in such undertakings and in such types of international co-operation than a number of Western European states do, which in many is a laudable thing. The union catalogue of old books is not a business venture, as it cannot be marketed. However, it is in the interest of culture and scholarly policy in general to create it.

At present, there are six Hungarian member libraries in the CERL organisation, but the high membership fees pose a serious difficulty for them. Referring to the fact that compared to the cost, very few people can use the CERL database, the state is torn between subscribing for services that provide information to many and the CERL subscription, which may be seen as a luxury. State sponsorship of cataloguing church libraries, familiarity with which is a priority to CERL as well, is a major investment in itself, while the profit will be shared by the CERL community. For this reason, we recommend that CERL should modify its philosophy: for the same subscription fee we could gain far more records if access were extended. According to our estimates, if we had access to 500 IP addresses, the state would be willing to contribute more funding to the union catalogue of old books, which in turn, would mean more records and more associate libraries for CERL. For church libraries, CERL is of special importance, as they rarely have paper-based reference books, which makes

cataloguing even more difficult for them. Thus, they have to rely on freely accessible large catalogues with large numbers of old books records (i.e. the Austrian national library, ICCU, GBV and the BNF catalogue, as well as the South German shared catalogue, etc.). I think that the view is shared both inside and outside Hungary that the realisation of the wonderful mission of CERL is often aggravated by a highly materialistic attitude.

Thus, from the very beginning, Hungarian libraries have applied the CERL principles in processing old books, applying the criteria of cultural history as their main guiding principle. Their processing serves more than just the purpose of information provision; they do more than trying to ensure free access to information: In fact, they are creating a professional system of book history that is suitable for introducing the book history of the whole region of the Carpathian Basin.

APPENDIX

Bibliotheca Eruditionis
A Database of Old Hungarian Written and Printed Materials (1500–1800/1850)

This database system is of fundamental importance for studying sixteenth- and eighteenth-century cultural history in Hungary, processing the results of basic research, and using quantitative methods, which is quite rare in our discipline. As a unique supplement, a large part of the corpus will be presented in multimedia images as well (i.e. books, bindings and illustrations).

In practice, this means that the databases will answer the questions of who read what and when in the Carpathian Basin of the 16th and 18th centuries. In the system, all this is linked to access to the visual image of the main archival sources of book history and to the illustrative material of the most characteristic volumes. In addition to being of use to scholarship, the interdisciplinary presentation gives a personal experience of Hungarian scholarly life in the 250 years following the protestant Reformation, as well as of the full reception of European intellectual movements and reading material. All this can meet many different users' needs, starting from researchers' primary needs to those of secondary students.

PARTS OF THE PROJECT

Book history databases

The actual goal of the project is to reconstruct the books available (the potential reading material) in the 16th and 18th centuries in the Carpathian Basin, which is done in the *Virtual Library of Old Hungarian Prints*. Digitization of

the 16th- to 18th-century sections of the Hungarian national bibliography. This database is comprised of three parts:

- bibliographical descriptions supplemented by the full list of existing copies
- the publishing history of individual texts (only up to 1700)
- title-page photos, dedications and forewords to individual editions (in the second stage of development, we wish to present them in their totality).

The shared catalogue recording pre-1850 books in the Carpathian Basin (MOKKA R) will serve as background database for the book history databank. Its model has been completed, today containing some 97,000 records. The catalogue of the National Library's sixteenth-century books is also made up of two parts:

- Bibliographical descriptions, of which it is an internationally unique feature that in every case the contents of volumes are recorded (*contenta*)
- The title-page photos of the collection, its dedications, forewords, the most frequent and the least frequent *in extenso* (at full length)

In addition to a detailed description of individual copies, it is our intention to supplement the full catalogue with visuals.

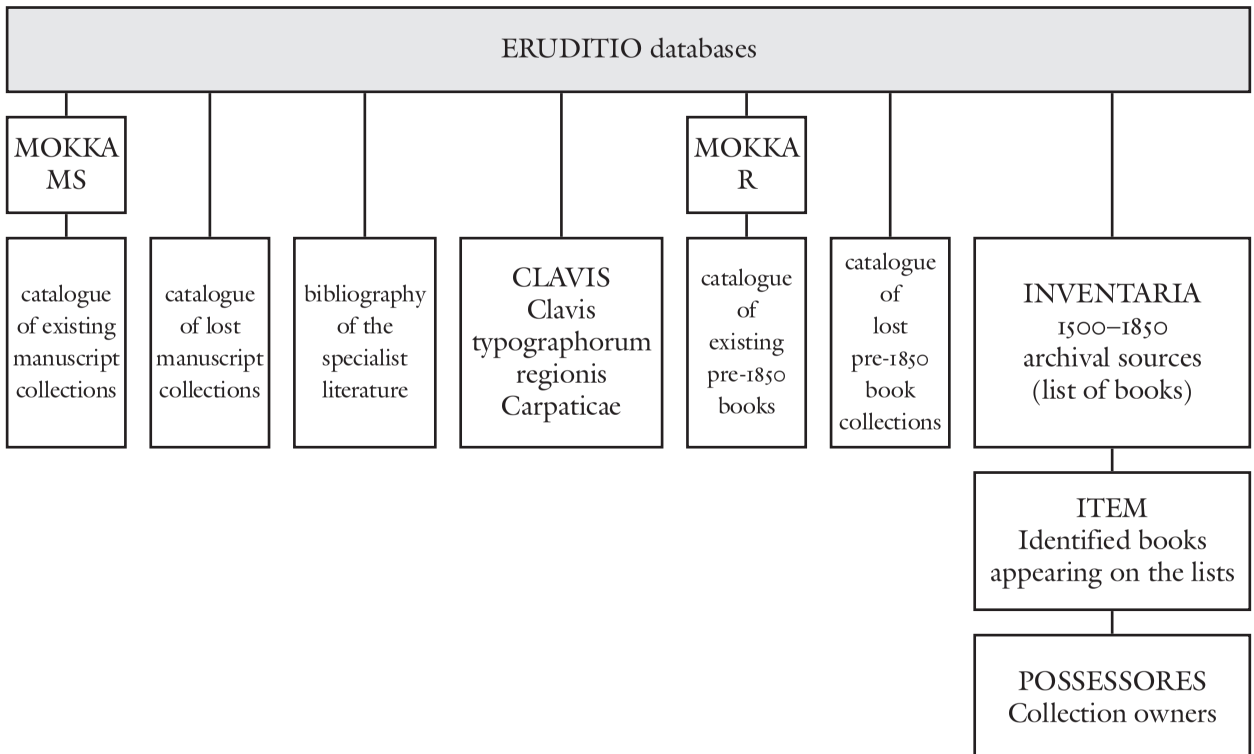
Reading history databases

Based on the two decades of foundation work mentioned before, an SQL database comprising Hungarian reading culture in the entire period from 1526 to 1750/1800 has been completed. This databank contains the following:

- Between 1750 and 1800 archival sources, which in the given period say whether certain people owned books in the Carpathian Basin, and how they described them.
- In this database, some 150–200,000 books are mentioned, which means about 60,000 definitely identifiable Hungarian and foreign items.
- The other database contains the full texts of documents, as well as facsimile copies of the main ones.
- The third database contains the titles of actual works identified with the help of considerable professional tools and on the basis of contemporary descriptions. Initially, this was uploaded according to the frequency of occurrence, thus the most frequently featuring books were the first to be included at the beginning of the project.
- The fourth database contains the secondary literature related to sources, to the owners of the books for which data is available, to collections and to individual books.

The four databases are searchable at the SQL level; and on the hypertext interface they are interlinked, thus fully integrated.

Experts'system of Hungarian book history
1500–1850



Cataloguing rare books in the Library of the Hungarian Academy of Sciences: problems of incunabula and copy-specific data

MARIANNE ROZSONDAI AND
BÉLA ROZSONDAI

Az MTA Könyvtárát 1826-ban Teleki József alapította azzal, hogy családi könyvtárát az Akadémiának ajándékozta. Innen és Ráth György (1905), majd Vigyázó Ferenc (1928) adományából ered a régi könyvek gyűjteményének és benne a mintegy 1200 ősnyomtatványnak túlnyomó része. Az ősnyomtatványok jórészt német és itáliai nyomdák termékei, főleg egykori német tulajdonosoktól. Jelentős a magyar vonatkozású anyag (Hungarica).

Az ősnyomtatványokat az ALEPH könyvtári rendszerben katalogizáljuk, a más régi könyvekkel és a kéziratokkal egységes szerkezetben, de külön-külön albázisban. A 15. századi nyomtatványokról rövid bibliográfiai leírást (short-title catalogue) készítünk, eltérve az ISBD(A) rendelkezéseitől, és hivatkozunk az ismert katalógusokra. Viszont közöljük a példány jellemzőit: az esetleges kiadás- vagy nyomásváltozatot, a tulajdonosi jegyeket, bejegyzéseket, a kötést, belső díszítést, a könyvben talált fragmentumokat, a kolligátumokat, a példány állapotát, a hiányokat. Az adatbázisban keresni lehet a nyomdahely, nyomdász és kiadó, a példány tulajdonosa és a könyvkötő szerint. A jelen cikkben bemutatunk néhány érdekes példányt vagy tulajdonost.

Több kérdés még kielégítő megoldásra vár: a kiadvány címének megválasztása, egységes címek, személy- és földrajzi nevek (történelmi vagy modern), a különböző forrásokból vett adatok és a szögletes zárójel [] használata, a katalógus nyelve (pápák, uralkodók neve, a megjegyzések szövege), az azonosítatlan tulajdonosok szerinti keresés (ismeretlen vagy töredékes név, monogram, címer, pecsét), speciális jelek reprodukálása betű szerinti közléskor; a 16. század eleji nyomtatványok ISBD(A) szerinti vagy rövid leírása, stb. A rekordhoz képet csatolva majd bemutathatjuk a példány egyes jellemzőit.

THE INCUNABULA COLLECTION

The Library of the Hungarian Academy of Sciences holds the second largest collection of incunabula in Hungary. Approximately 1200 books and single sheets printed in the fifteenth century are preserved here. The bulk of the collection was formed by gifts from aristocrats, other generous book-collectors, and some members of the Academy.[1]

The first major donation came from the Teleki family. In 1826, half a year after the Academy was founded in November 1825, its first president, Count József Teleki (1790–1855), offered his family library of thirty thousand volumes to the Learned Society. Thus he laid the foundations of the Library of the Academy. He had the augmentation of the Library at his heart, and enriched it by purchasing several invaluable volumes for it until his death. In 1834, he bought and donated to the Academy the book collection of Ferenc Kresznerics (1766–1832), an honorary member of the Academy, with 87 incunabula among them. However, it is difficult to establish which books belonged to Count László Teleki (1764–1821), father of József, or to his own private library, presented to the Academy later in 1850, since all these books bear the stamp: 'G. Telekiek' Alapítványa' (Foundation of the Count Telekis). It rarely occurs that an invoice is stuck in a book, like that from Matthäus Kuppitsch, Vienna, 1834, in Schedel's *Liber chronicarum*, Augsburg, 1497 (Inc. 156: CIH 3040, HCR 14509), although we know that József Teleki employed agents abroad to buy him rare books. Through Teleki, 409 incunabula (in 356 volumes) passed into the possession of the Academy.

Following some minor and sporadic donations, the collection was considerably increased by György Ráth (1828–1905), a lawyer and a man of broad learning.[2] He was the first director-general of the Museum of Applied Arts, Budapest, a member of the Upper House of the Parliament, and author of a number of books on legal matters and the arts. At the end of 1895, he concluded a contract of donation, according to which he would bequeath his books to the Hungarian Academy of Sciences, i.e., to the Academy's Library. After his death in 1905, the Academy received György Ráth's 2300-volume library of Hungarian-related works (Hungarica), including 145 incunables. The word 'Ráth' in the shelfmarks distinguishes them within the Hungarica and the incunabula collections.

Count Ferenc Vigyázó (1874–1928) was a politician and law graduate, and an expert book collector. He left his entire wealth, including his highly valuable 17,000-volume library, to the Hungarian Academy of Sciences in

1928. With his library, a further 431 incunabula (in 424 volumes) passed into the possession of the Academy Library. At the beginning of each volume, Vigyázó noted in pencil the date, price and name of the antiquarian bookseller from whom he bought the book. He consciously endeavoured to acquire as many Hungarica as possible. The Ráth and Vigyázó books are well preserved and mostly conserved copies.

Other smaller but still significant augmentations were made by several Hungarian bibliophiles. Mention should be made of Ferenc Pulszky (1814–97), who enriched our Library with some especially fine early printed books. One of them is Justinus' *Epitome*, [Roma, *c*. 1470] (Inc. 415: CIH 1987, HC 9646) with beautiful, multicoloured Florentine foliated scrolls on the first leaf. The entire book is emended by a Hungarian humanist Paulus Schaider.[3] This is therefore an example of a printed book with signs of early Hungarian use. The emendations mark the corruption of the text in comparison with, e.g., that of the edition produced in Venice in 1470 by Nicolas Jenson (Inc. 682: CIH 1986, HR 9647), which must have been based on an uncorrupted manuscript. Count Vigyázó bought the copy of the latter from T. de Marinis et Co. in Florence in 1904.

After World War II, the number of incunabula increased by 64 until 1970, the date of publication of the national catalogue (CIH),[4] and by further 31 since then.[5] Fifty-six of them were bought or received as a present; 7 editions came from confiscated monastic libraries after 1948. The rest, 32 items, arose from two internal sources: some undated books formerly kept in the main holdings of the Library could be identified as incunabula (the opposite change also occurred); besides, some prints or fragments were discovered in bindings during conservation work. (Almanacs and broadsides or other publications which became worthless for contemporaries were often used for making book covers or paste-downs.)

The earliest printed item, a single leaf (tom. 2, fol. 85) from Gutenberg's first great undertaking, the 42-line Bible from 1454/55 (Inc. 1000: CIH 610, GW 4201), came to the Academy Library as a gift from Gabriel Wells in 1922. The Library owns eight incunables which are unique. One of them is the Esztergom-rite funeral ceremonial *Obsequiale Strigoniense*, Nürnberg, 1496 (Ráth 1049: CIH 2461), which Ráth purchased in Germany. The short, satirical epic — written in the form of macaronic verses in Latin and Italian — known as the *Boazana*, [Bologna, *c*. 1495] (Inc. 889: CIH 693a, H 3664) tells about the life of university students of the time in Bologna. The *Almanach pro anno 1483*, [Augsburg, 1482] (Inc. 1012: CIH

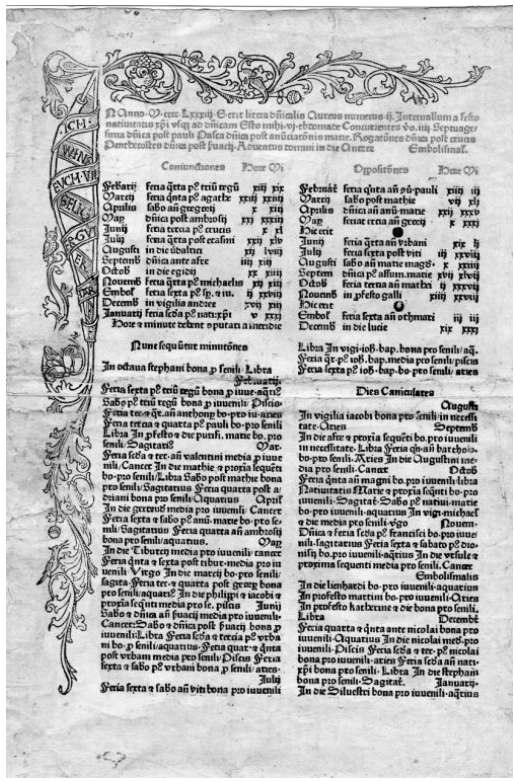


FIGURE 1 *Almanach pro anno 1483*. [Augsburg, 1482] (Inc. 1012). (Reproduced by permission of the Library of the Hungarian Academy of Sciences)

144) (Fig. 1) emerged from the binding of *Sermones de sanctis* by Jacobus de Voragine, also printed in Augsburg in 1484 (Inc. 514: CIH 1867, C 6550). Its Gothic leather binding was made in Augsburg, too. King Matthias's statute-book, *Constitutiones incliti regni Hungariae*, Leipzig, 1488 (Inc. 353: CIH 1074, C 1758) is not unique, but it is the most complete copy known today.

The Library has a number of incunabula in more than one copy: there are 112 incunabula in two, 13 in three copies (e.g. *Schedel's Liber chronicarum*, Nürnberg, 1493: CIH 3039, HC 14508*, owned by Teleki, Ráth and Vigyázó, too) and 4 in four copies (for instance *Chronica Hungar-*

orum by Thuróczy, Augsburg, 1488, var. C: CIH 3324, HC 15518* = H 15516). It is also due to the purposeful collecting activity of György Ráth and Count Vigyázó that the rare book collection has become rich in Hungarica. About one-fourth of the incunabula collection consists of such books: works by Pelbartus de Themeswar, Osualdus de Laskó, Michael de Hungaria, Georgius de Hungaria, Johannes Thuróczy, works dedicated to King Matthias Corvinus and to Hungarian humanists, works on the struggle against the Ottoman Empire, Hungarian ritual books and publications of printers of Hungarian origin.

As we have seen, the greater part of the Library's rare book collection came from donations by private persons, who were able to obtain such old prints from second-hand dealers abroad, partly from Northern Italy, but mainly from the German-speaking parts of Europe. This fact is reflected in the composition of the incunabula collection. The prevalent language is of course Latin, represented by 1,118 copies, followed by German 49, Italian 11, and some Greek, Dutch and Hebrew. The distribution of countries as places of publication is, in descending order:

> Germany 541, Italy 524, Switzerland 82, France 18, Netherland 9, Bohemia 4;

and the order of the towns:

> Venice 334, Strasbourg 144, Nuremberg 117, Augsburg 73, Köln 69, Basel 63, Rome 52, Bologna 41, Hagenau 31, Leipzig 25, etc.;

while the representation of printers/publishers in our collection:

> Anton Koberger (Nuremberg) 84, Heinrich Quentell (Köln) 33, Martin Flach (Strasbourg) 26, Johann Prüss (Strasbourg) 25, Bonetus Locatellus (Venice) 24, Benedictus Hectoris (Bologna) 23, etc.

CATALOGUING INCUNABULA

Old printed books are valuable sources of texts, pictures or music, but the book as a physical object is perhaps more often studied nowadays than its contents. Attention is directed either to the history of publishing, printing, etc., or to the story of the individual copy.

Modern cataloguing of rare books means the recording of (1) bibliographic data, (2) headings, i.e., main/added entries, and (3) copy-specific data. Bibliographic data serve the identification of the edition, and therefore, according to ISBD(A) rules, certain data should be given exactly as they are in the book in the prescribed source of information. Headings, on the other hand, provide access points to the bibliographic units, giving the

possibility of finding them in the catalogue, i.e., retrieving and sorting records from the database. Thus, they should be given in standardized form, as uniform titles, personal and geographical names, etc. Copy-specific data include possessors' notes, illumination, binding, and the physical condition of the book, so they inform on the reader, the history of book collections, etc., and record or help book preservation.

The incunabula of the Academy Library were first described in a scholarly way by Árpád Hellebrant.[6] He quoted the *incipit*, colophon, etc. – it is not his fault that the abbreviation characters were expanded – and gave collation, initials, woodcuts, and notes on the copy. The catalogue was published in 1886, with an introduction and notes on the individual entries in Latin, so that its international use might be facilitated. The 466 entries cover 493 copies including duplicates. The items are arranged and numbered in chronological order (and in alphabetical order of towns within a year), and there are six types of indexes. About a quarter of the editions lack a colophon. The 'books without the year' of printing are listed in the second part of the catalogue, beginning with entry 355, as '*libri nota anni carentes*'.

The chronological arrangement soon became unsuitable because of new accessions, to say nothing of composite ('bound-with') volumes containing more than one publication. Unfortunately, no other such printed catalogue has been published since then, nor was the card catalogue of the Vigyázó collection completed by the time of World War II. These copies were thus not recorded in GW. The Collection of Rare Books was set up within the Department of Manuscripts in 1954. The books were arranged in a new system: the whole collection was identified and catalogued anew, and each item received a *numerus currens*. The incunabula were treated and are kept separately.

In 1970, a catalogue of all incunabula kept in public libraries of Hungary was published; this is known in abbreviated form as CIH. It is also referred to as 'Sajó-Soltész', even though the title page – in smaller type – indicates that it had been prepared in cooperation with Csaba Csapodi and Miklós Vértesy. It was Csapodi who identified and described the incunabula of the Academy Library, while Vértesy accomplished the same task for the holdings of the University Library, Budapest. In the Academy Library, CIH registers 994 fifteenth-century editions, which makes a total of 1,159 copies with duplicates included. Since then, the number of editions has increased by 31, as mentioned above. There was a need for a modern and up-to-date catalogue of our rare book collection.

The software for the computer catalogue was already given: ALEPH, an integrated library system, which is used for the modern books in the Academy Library, and in many libraries in Hungary, in Europe, and overseas. The record structure is based on HUNMARC. By little modifications that do not conflict with existing records of modern books, we have devised a uniform data structure: fields, indexes, display formats, for incunabula and other rare books, as well as for manuscripts and bequests. In the OPAC of the Academy Library, the bibliographic records are contained in two databases. The incunabula records are a subset (INC) of the special collections (dissertations, bequests, etc.), and the other rare book categories form a subset (RBK) within the database of modern books and periodicals. Cataloguing of incunabula and of sixteenth-century printed books is in progress; the data records are not yet available to the public.

Incunabula editions have been identified and described in detail, and all the extant copies, so far as possible, have been recorded. Thus, the first requirement of a catalogue, to identify an edition, could be fulfilled by reference to the standard incunabula catalogues. We prefer the following sources: Hain, Copinger and/or Reichling (H, C, R and combinations), GW, BMC, BSB-Ink, CIH, and ISTC. For practical purposes, however, some kind of bibliographic description is necessary. So we decided to build a short-title catalogue.

Our practice of cataloguing and some problems of incunabula will be demonstrated by the following examples.

Incunabula, like manuscripts, often lack a title page, and bibliographic information has to be found elsewhere in the book. The 1490 Venice edition of the works of Robertus Caracciolus (Inc. 228: CIH 901, GW 6042) has a title page with three words altogether; leaf 3 begins with an *incipit*, and the last page before the Register ends with the explicit and the colophon (abbreviations are not reflected):

> *fol. 1a*: sermones fratris roberti
>
> *fol. 3a*: Sacre theologie magistri … // … fratris Roberti epis- // copi … opus // quadragesimale … incipit.
>
> *fol. 337b, col. 2:* … fratris Roberti episcopi Aqui // nensis … sermones qua // dragesimales: de aduentu: et de timore iudiciorum // dei … expliciunt. // Impressi in ciuitate Uenetiarum per Ioannem // de Forliuio et Gregorium fratres Anno domini. // M.CCCC.LXXXX. die. 15. Martij.

followed by the printers' device (Z G = Zuane [Johannes] and Gregorius).

In the bibliographic record, the short title proper is entered in the 'Title and Statement of Responsibility' field. The author's name appears in standard form in the 'Main Entry – Personal Name' field. Both fields are indexed. Usually we prefer the form specified in PMA[7] and the DDB website,[8] but sometimes we retain the name given in CIH. Alternative names can be added and linked in the authority file if necessary. References in the repeated 'Other Standard Identifier' fields are also indexed (two of them are shown here):

Other Standard Identifier	GW 6042
Other Standard Identifier	CIH 901
Main Entry – Personal Name	Caracciolus, Robertus
Title and Statement of Responsibility	Opera varia

This edition is actually a collection. The BSB-Ink catalogue lists the titles of all *Sermones*, dedications, etc. We do not describe the contents in such a detail, at least if it is given elsewhere: we usually do not mention a dedication, privilege, letter or a poem. The titles in the collection and statements of responsibility are listed in the 'Contents Note' field with prescribed punctuation, while the uniform titles and responsible persons' names of further works are indexed in the respective 'Added Entry' fields. In this case we add only the work by Bonanus:

Other Standard Identifier	BSB-Ink C-106
...	...
Contents Note	Opera varia / Robertus Caracciolus. De conceptione virginis Mariae / Dominicus Bonanus
Added Entry – Personal Name	Bonanus, Dominicus
Added Entry – Title	De conceptione virginis Mariae

When we read in an early printed book (Inc. 43: CIH 1726, H 8801*) at the beginning of a work: *fol. 1a, lin. 1*: [T]res modi ...; *lin. 3*: vnus quo asseritur quod soli *predestinati* // saluentur. ...; *lin. 5*: Tertius quoque dicitur quod // solum *liberum arbitrium* salutem operetur: – it is difficult even to find out a title (Fig. 2). As the author is not named but has been ascertained from an external source, we repeat it in the prescribed brackets in the 'Statement of Responsibility' ($c) subfield:[9]

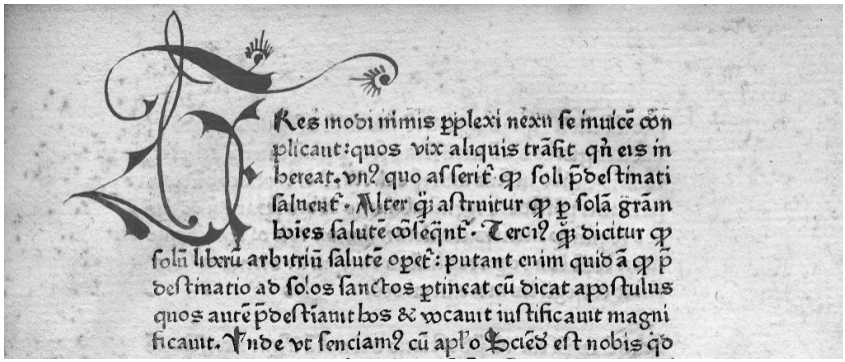


FIGURE 2 [Honorius Augustodunensis]: *De praedestinatione et arbitrio libero*. [Strassburg, *c.* 1471], fol. 1a (detail) (Inc. 43: Library of the HAS)

Main Entry – Personal Name	Honorius Augustodunensis
Title and Statement of Responsibility	$a De praedestinatione et arbitrio libero $c [Honorius Augustodunensis]

or rather, we put it in the 'Contents Note' field – adding here in this case the title of the second, anonymous work in the collection:

Contents Note	De praedestinatione et arbitrio libero / [Honorius Augustodunensis]. Cognitio vitae
Added Entry – Title	Cognitio vitae

For works without an author, the 'Main Entry' will be the title. Parts or books of the Bible are always filed under the uniform title, e.g., *Biblia* [Lat.]. A typical case of indecision is the *Codex Iustinianus*, which is classified differently in the incunabula catalogues. The 'Main Entry' is either Justinianus imperator, or the uniform common title: *Corpus iuris civilis*, of which it is a part. Two remarks are appropriate here, concerning not only incunabula. (1) In fact, in a record, where persons' names or titles concerned are all in indexed fields, it does not matter too much, whether they are termed 'main entries' or 'added entries', while this distinction used to be important in a traditional card catalogue. (2) In naming a pope, a sovereign or a prince, we are facing the dilemma: should we use the Hungarian or the Latin expressions?[10]

In the area 'Publication, printing or distribution, etc.' we voted for short descriptions as well. In the respective subfields ($a, $b, $c), we put the place name – with a few exceptions – in its nominative present-day national form, the name(s) of the publisher(s), etc. in its (their) short Latin form, and only the year of the date. This field is not (and should not be) indexed, therefore we created 'Added Entry' (A. E.) fields for indexing 'Special Personal', 'Corporate' and 'Geographical Names' in the 'Local fields' area, where, in general, printers' names are set in their modern (national?) forms, i.e., last name, comma, first name(s). The year for an index entry is extracted from the MARC control field '008'. The relevant fields for the Caracciolus edition (GW 6042):

008/07-10	1490
Publication, Printing, etc.	$a Venezia $b Johannes et Gregorius de Gregoriis $c 1490
A. E. – Personal Name, Special	$a Gregoriis, $j Johannes de $4 typogr.
A. E. – Personal Name, Special	$a Gregoriis, $j Gregorius de $4 typogr.
A. E. – Geographical Name, Special	Venezia (loc. impr.)

The Honorius Augustodunensis edition (H 8801*) has no colophon, and the imprint data come from sources outside the publication. Thus, they are enclosed in brackets, and alternative assignments are given in a 'Note' field – and here is again the question of language in the notes:

Publication, Printing, etc.	$a [Strassburg] $b [C. W.] $c [c. 1471]
Publication, Printing, etc. Note	Assigned to C. W. (= Clas Wencker? Conrad Wolfach?); see Needham, Paul, in British Library Journal 6 (1980) 130–143. – Also recorded as: [Nürnberg : Anton Koberger, c. 1471] or [c. 1475]

In the case of *Sermones de tempore* ... by Bernardus Claravallensis (Inc. 271: CIH 571, GW 3942), the stated time limits require a more complete specification of the date:

[Speyer : Peter Drach, post 31. Aug. 1481, non post 1482]

Strabo's *Geographia* was printed twice in a year by Johannes Rubeus Vercellensis in Venice, therefore we record the exact day given in the imprint:

24. Apr. 1494 (Inc. 445: CIH 3149, HC 15090*)
and
28. Jan. 1494/95 (Inc. 149/koll. 1: CIH 3150, C 5661).

As to the latter edition, January at that time belonged to 'last year', i.e. 1494, but it would be 1495 in today's calendar.

The 'Physical Description' area is also treated differently in ISBD(A) and in incunabula catalogues. We specify the 'Extent' as the number of leaves, blank leaves included, irrespective of the kind of the numbering of leaves or pages, without brackets. E.g., in Strabo's edition of 28. Jan. 1494/95 we write: 166 fol., but it would be: [16], CL fol., according to ISBD(A) rules. Both format *and* dimension(s) should be given. Collation, signatures are inserted in a note field only if they are not found elsewhere, or if they help understand other notes in the record.

COPY-SPECIFIC DATA

For the copy-specific data we use four MARC fields in the 'Note' area:

Provenance Note: owners' notes, stamps, etc., marginal notes
Copy Identification Note: binding, illumination, rubrication
Physical Description Note: collation, wanting leaves, misbound parts
Action Note: preservation actions done or required

Owners and bookbinders, other persons, and the corresponding geographical names are recorded in the 'Special Added Entry' fields already mentioned. If we have more copies of an edition (or more volumes of a work with different characteristics), all copies or volumes are described in the same bibliographic record. The above fields are then repeated for each copy. The shelfmark, being a piece of copy-specific information, is also given here, of course. The 'Catalogue card' display format has been modified to achieve an appropriate appearance on the screen.

Although in a short-title catalogue no literal transcriptions are expected, we realize the importance of recording the variant (issue, state) of the copy in hand, representing an edition. The ratio and distribution of variants may contribute to our knowledge of early printing. BMC and GW often indicate text variants. If we identify our copy with one of the variants, we describe it in the 'General Note' field. Here we face two difficulties. (1)

Sometimes we observe a small discrepancy between the text (or title) in our copy and in one of the standard catalogues. Can we regard it as a variant, or is it a printer's error in the catalogue? Therefore, though bibliographic description represents the 'ideal' copy, it should always be stated, which copy was used for it. (2) Reflecting special characters and abbreviations in the database is practically impossible. We try to find a substitute, as â or ô implying a nasal, 9 for 'us', or, in other cases, we enclose the letter(s) in braces {}, like {con}, or just {p} for 'per', 'prae' or 'pro'. ISTC, ALEPH and other systems allow the attachment of images to the record. In a printed catalogue, the special characters must be reproduced anyway.

Copy-specific information is also the enumeration of items in a composite ('bound-with') volume. The Latin word *colligatum* is used in Hungarian (kolligátum), both for the volume and its components. You would never retrieve the editions bound together in a composite volume from an alphabetic card catalogue, although the information is there. In our database, as it must be, each unit is described in a separate bibliographic record. The first unit in the volume is considered a node, which is linked to each of the rest. Besides, each record contains the short description of its linked partners in 'Contents Note' fields. This kind of information should always be given if the catalogue includes copy-specific data.

Four incunables (Inc. 356 – Inc. 359) and a manuscript (K 39) had formed a 'bound-with' volume, before they came to the Academy in 1851 as a present from the Franciscan Monastery in Gyöngyös. The volume used to belong probably to a Benedictine or Premonstratensian nunnery. After adventurous events the parts were separated and rebound. The manuscript is a prayer-book, 35 leaves, a language record in Hungarian and Latin from the early sixteenth century. It reads on *fol. 4a*: 'Kezdetyk Jesusrol walo tyzen // ewth yeles ymadsaag kyt zent brigida . . . mond wala . . .' (Begin fifteen notable prayers about Jesus that St Birgitta had said). The other parts – all are works needed in a convent – were printed in the 1490s (CIH 1873, CIH 1760, CIH 3397, CIH 3163).

Our copy of the *Quaestiones* . . . by Albertus Magnus, [Basel, non post 1475] (Inc. 121: CIH 69, GW 679) contains a manuscript of 69 leaves (and 3 blank) at the end, noted by Hellebrant, but not studied so far: 'incipit discipulus de sanctis per annum', that is, sermons by Johannes Herolt. Our immediate task was to note its existence in the catalogue, while the textual analysis awaits a specialist in the field.

In earlier centuries, printed books were in general not bound in the place where they were printed, but rather where they were purchased, read

and used. Binding, therefore, reveals the owner/reader of the book and the place where it was used. This is very important from the point of view of a book's provenance. Nearly one quarter of the Library's incunabula still preserve their original, historical bindings. There are 183 original Gothic bindings and 53 Renaissance bindings. In addition to these, bindings in the Baroque, Rococo and even Historicist style are also regarded as historical bindings, although they are, as a rule, the second or even third binding of an early printed book.

An interesting example is the blind-tooled leather binding in Mudéjar style, which covers Thomas de Aquino: *Summa contra gentiles*, Venezia, 1480 (Inc. 509: CIH 3258, HC 1389*). On fol. 1a, there is a Spanish note addressed to the bookbinder. A counterpart of this bookbinding was shown at the exhibition '*Europalia 85 España*' in Bibliotheca Wittockiana, Brussels. The latter book belongs to the Cathedral Library in Segovia. The two books were printed in the same year, and must have been bound in the same workshop one after the other.[11]

Fortunately, many of the incunables bear some sign of ownership. About one tenth of our incunables used to be in Austrian – private or monastic – possession, according to owners' notes in 103 volumes.[12] Forty-three incunabula of the former Kresznerics collection used to belong to the mediaeval library of the Dominican Friars in Vienna; it is not only the monastery's entries but also their Gothic bookbindings – and the parchment fragments found in some of them – that indicate their Viennese origin.[13] For instance, our *Biblia*, cum postillis Nicolai de Lyra, Pars 2, Venezia, 1482–1483 (Inc. 320: CIH 665, GW 4287) has saved, as paste-downs, a fragment of two leaves (shelfmark: T 328) of *Tractatus de arte magica* from the 14th c. probably from Vienna.[14] The number of books of German provenance is also significant.

One of the best-known owners is the humanist Konrad Celtis, whose note in ink is seen on the title page of *Scriptores rei rusticae*, Bologna, 1494 (Inc. 898: CIH 3057, HC 14568*). The book-plate of Hartmann Schedel, physician, compiler of *Liber chronicarum*, is stuck inside the upper cover of Guido de Cauliaco: *Chirugia*, Venezia, 1498 (Inc. 750: CIH 1531, GW 11696). He registered the place and time of purchase, also the fact that it was unbound when he bought it. Later on, this *Chirurgia* went to the possession of a Fugger, and Vigyázó bought it from Jacques Rosenthal for 250 Mark in 1904.

The title page of Caracciolus (Inc. 228) bears an owner's note: 'Ex Bibliotheca Mandelliana 1788', and also a black ink stamp 'F G', which is

seen in several pieces of our rare book collection. David Mandelli (Pozsony, *c.* 1780 – Paris, 1836), the 'Hungarian Diogenes', an odd scholar, was a child in the year given. Could the owner be his father? – The 'F G' stamp on fol. 1a of Honorius Augustodunensis (Inc. 43) is accompanied in this book by a note in German on the fly-leaf signed by F. G., but this doesn't help, either, to disclose the person behind the initials.

Our copy of Lucanus' *Pharsalia*, Brescia, 1486 (Inc. 279: CIH 2096, BMC VII 975) has two small peculiarities. Quire 'u' is indicated in the *Registrum* to have three sheets, and quire 'n' – two sheets. Our copy, and one of the two copies in the British Library were gathered accordingly. The third sheet assigned to 'u', however, is in fact the middle sheet of quire 'n', with signature 'n iii' and catchword 'ueluti'. In the other copy of the British Library, the Register was emended with pen-and-ink, and the sheets were collated correctly. The Venice, 1492 edition (CIH 2099, BMC V 431) inherited the displacement of the text, since it was printed from the Brescia edition.The other interesting feature occurs in the *Epigramma* on fol. 4a, in the last line but one: the word *bardos* has been struck, and replaced in ink by *foetus*, in all three copies mentioned! (It would be interesting to know, whether it was done by the same hand.) In the Venice 1498 edition (Inc. 159/koll. 2: CIH 2102, BMC V 523) this line is correctly printed: 'O utinam tales producas brixia foetus. // Vnde habeas nomen perpetuumq3 decus.' (O, Brixia, may you bring forth offspring, who would win you fame and eternal glory!). Why did the printer in Brescia set blockheads (bardos) instead of offspring (foetus)? Was it a bad joke, or a revenge upon Brescians?

We have a copy of the Mainz, 1485 edition of *Ein Gart der Gesundheit* (Inc. 403: CIH 1743, H 8948*), with hand-coloured woodcuts. The names of herbs, or some medicaments of mineral or animal origin, are printed in German, Latin, Greek and mostly in Arabic. Rather often, the Hebrew name is written in ink beside the picture. There are many words, some longer marginal notes in Hebrew and in German, some of them by the same hand. We find an owner's note in red ink by C.H.V. from the year 1542 at almond-tree, Cap. xxxv, fol. 39a, and some notes on misbound quires. This book could be a valuable language record of the period. Regrettably, it was trimmed, when rebound, and the notes were more or less truncated. Again a warning that cannot be repeated often enough: librarian and conservator must co-operate!

Finally, let us recall the problems that can arise in preparing a short-title catalogue of incunabula, e.g.: treating external data, using brackets [] in

the 'Title and Statement of Responsibility', the 'Edition', the 'Publishing/Printing' area, etc.; the choice of titles, uniform titles, personal names (authors: ancient, medieval; printers: modern?), place names, assignment to a country (historical, up-to-date?); the language of the catalogue: the names of sovereigns, popes et al., and the language of the 'Notes' (Hungarian, Latin, English?); headings for unidentified possessors (unknown or fragmentary names, initials, coats of arms, stamps); recording special characters, symbols in the record; whether to apply ISBD(A) rules or short-title cataloguing for early sixteenth-century books? Some of these will be solved when images are attached to the bibliographic records.

NOTES

1. Csaba Csapodi, 'Die Inkunabelsammlung der Bibliothek der Ungarischen Akademie der Wissenschaften', *Gutenberg-Jahrbuch* (1966), 74–79.
2. Marianne Rozsondai, 'The Ráth collection of the Library of the Hungarian Academy of Sciences', in *A Hungarian citoyen György Ráth and his lifework*, ed. Hilda Horváth (Budapest, 2006), pp. 31–47.
3. János Véber, 'A Magyar Tudományos Akadémia Könyvtára egyik Justinus-ősnyomtatványának bejegyzései', *Magyar Könyvszemle*, 122 (2006), 1–15.
4. CIH = *Catalogus incunabulorum quae in bibliothecis publicis Hungariae asservantur*, ed. Géza Sajó et Erzsébet Soltész, in colligenda materia et identificandis incunabulis socii fuerunt Csaba Csapodi et Miklós Vértesy, vol. 1–2 (Budapest, 1970).
5. Marianne Rozsondai, 'Az Akadémia Könyvtára ősnyomtatvány-gyűjteménye 2000-ben / The incunabula collection in the Library of the HAS in 2000', in *Örökségünk, élő múltunk: Gyűjtemények a Magyar Tudományos Akadémia Könyvtárában* (Budapest, 2001), pp. 34–62.
6. Árpád Hellebrant, *Catalogus librorum saeculo XV^o^ impressorum quotquot in Bibliotheca Academiae Litterarum Hungaricae asservantur* (Budapest, 1886).
7. *Personennamen des Mittelalters*: PMA, Teil 1–2, Regeln für die alphabetische Katalogisierung: RAK, 6 (Wiesbaden, 1989).
8. *Deutsche Nationalbibliografie online: Katalogdatenbank ILTIS*, Deutsche Nationalbibliothek <http://z3950gw.dbf.ddb.de>
9. Subfields are indicated in some cases only to separate data elements, but have no significance in this paper.
10. Mauro Guerrini, 'The language of the catalogue and the form of heading: vernacular, original, conventional', in *Many into one: Problems and opportunities in creating shared catalogues of older books*, CERL Papers, 6 (London, 2006), pp. 11–20.
11. Marianne Rozsondai, 'Ein spanischer Einband im Mudéjar-Stil, ca. 1480', *Philobiblon*, 35, No. 3 (Sept. 1991), 237–39.

12. Marianne Rozsondai, 'Österreichische Possessorvermerke in Inkunabeln der Bibliothek der Ungarischen Akademie der Wissenschaften', in *Ex libris et manuscriptis: Quellen, Editionen, Untersuchungen zur österreichischen und ungarischen Geistesgeschichte*, hrsg. István Németh und András Vizkelety (Budapest; Wien, 1994), pp. 53–77.
13. Marianne Rozsondai, 'Wiener Dominikanereinbände in der Bibliothek der Ungarischen Akademie der Wissenschaften', *Gutenberg-Jahrbuch* (1981), 234–44.
14. Thanks are due to Dr Edit Madas for identifying the fragments.

Provenance and the Itinerary of the Book: recording provenance data in on-line catalogues

TONY CURWEN AND GUNILLA JONSSON

A provenancia-adatok többé-kevésbé rejtve maradnak a könyvtárak gyűjteményeiben, annak ellenére, hogy az elsdleges történeti forrásanyagok jelentős részét alkotják. Manapság a könyvtárosok már egyre jobban felismerik ennek az információnak a fontosságát, de a rögzítés technikájában jelentős hiányosságok tapasztalhatók. A formátumok különbözősége csak töredéke a problémának, a legrosszabb talán az, hogy nincsenek egységes katalogizálási alapelvek és egységes thesaurusok. A provenancia adatokat általában rendezetlen formában jegyzik fel, nem szisztematikusan és megszerkesztetten, ami a tudósok munkáját segítené.

A provenancia adatok használható módon való rögzítése időigényes és speciális tudást igényel több különböző területen. A könyvtárosoknak azon kell dolgozniuk, hogy megalkossák a szerkezetet, de össze is kell fogniuk a kutatókkal és könnyen használható eszközökkel kell őket ellátniuk annak érdekében, hogy a kutatók tudását is hasznosíthassák a szakterületen. Ily módon ez a fontos történelmi forrásanyag lassan elérhetővé válik a kutatás számára.

> I keep six honest serving men
> (They taught me all I knew);
> Their names are What and Why and When
> And How and Where and Who.[1]

These lines, familiar to two or three generations of English speakers, contain the seed of many an investigation, and certainly provide a framework for a consideration of aspects of provenance. It is our thesis that while Who – the ownership and custodianship of books – has long been at the heart of provenance studies, Where and When have been relatively neglected, as has What, thus making finding the answers to How and Why far

more difficult, largely for lack of full and systematically structured data in our catalogues. The core of provenance is ownership, but maybe other aspects of provenance information have been neglected in comparison.

Before continuing, we need to establish a common understanding of what we mean by provenance information. We propose that any piece of information that may bear witness to the itinerary of a book is provenance information. Those data may cover the path from the origin of a book to the current owner and the stops the book has made between those two stations. A brief list would comprise names, coats of arms, dedications, initials, dates, prices, numbers or codes, booksellers' signatures, margin notes, bindings, etc.

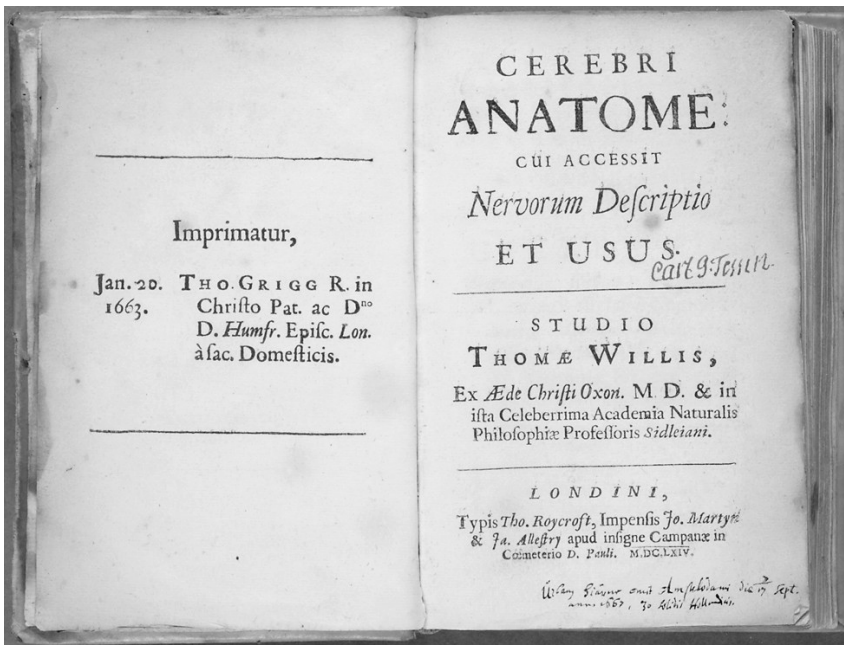
When you are dealing with hand-press publications in libraries you are inevitably dealing with books and collections that have individual stories and itineraries. The large, historical library collections consist of donations, purchases, inherited collections, mergers of collections, war booties. Provenance information of many different kinds is present in almost every old book on a library shelf. These data are, however, seldom added to the catalogue records, and if they are, it is usually not done in a way as to make them systematically retrievable. Provenance information is more or less hidden in library collections, although they constitute a substantial fund of primary, historical source material. If it could be exploited it would add considerably to historic research in a wide sense.[2] To do so, however, it must be recorded more often than it is today, but primarily, it must be recorded in a more structured way.

Before you can start to record any information in a useful way, there must be an appropriate framework for the data, and we will demonstrate shortly that this framework is more or less lacking today. When CERL does focus on provenance information, it is the framework for catalogue records, the format structure, thesauri and cataloguing guidelines that is in focus.

Properly organised provenance data should systematically help the researcher to discover *Where* items were and *When*, and *Why* they were collected or sold/dispersed, and *How* they illustrate the history and society of their times. Format extensions are one step in providing a framework for this. CERL is actively trying to influence format developments in this area and has achieved certain extensions to UNIMARC for provenance data already. We believe this to be the primary task for librarians.

Some examples will demonstrate the current shortcomings.

EXAMPLE 1 *From the collections of the National library of Sweden*



Imprimatur,

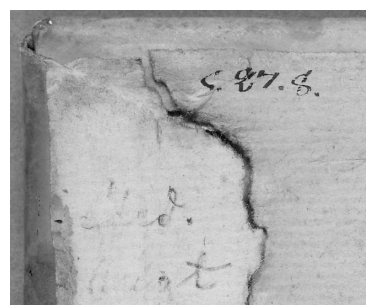
Jan. 20. 1663. THO. GRIGG R. in Chriſto Pat. ac Dⁿᵒ D. *Humfr.* Epiſc. *Lon.* à ſac. Domeſticis.

CEREBRI
ANATOME:
CUI ACCESSIT
Nervorum Deſcriptio
ET USUS.
Carl G. Tessin

STUDIO
THOMÆ WILLIS,
Ex Æde *Chriſti Oxon.* M. D. & in iſta Celeberrima Academia Naturalis Philoſophiæ Profeſſoris *Sidleiani.*

LONDINI,
Typis *Tho. Roycroft*, Impenſis *Jo. Martyn* & *Ja. Allestry* apud inſigne Campanæ in Cœmeterio D. *Pauli.* M.DC.LXIV.

EXAMPLE 1A and 1B Imprimatur and title page, and part of inner front cover with shelfmarks



In example 1, we are dealing with well-known owners, Carl Gustaf Tessin and Urban Hiärne, who were also active as authors; there are already authority records for them in the library's database. MARC21 has provisions for recording the information in a structured way in the bibliographic record for the manifestation. There are even standardized codes or terms provided for the kind of role – ownership in this case – of

a person. A record with a provenance note and as much structured data as possible could look like this, in part (*note that MARC examples throughout this paper have been spaced out slightly for ease of reading*):

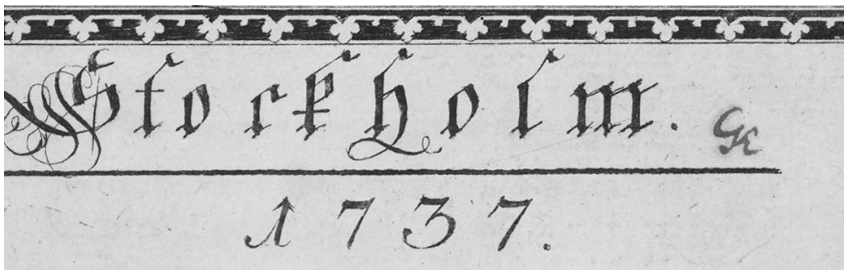
561 ## **$a**Urban Hiärne's note of purchase on the title page and a natural observation note by him on inside of back cover. Date of purchase: 1667-09-17. Price: 30 Dutch Soldi. Place of purchase: Amsterdam. Carl Gustaf Tessin's signature on title page. The Royal library's shelfmark from the middle of the eighteenth century on upper left corner of front cover inside (Anders Wilde's catalogue U.126:41). **$5**SwSKB
655 #7 **$a**Annotations **$2**rbprov **$5**SwSKB
700 1# **$a**Hiärne, Urban, **$d**1641-1724. **$4**fmo **$5**SwSKB
700 1# **$a**Tessin, Carl Gustaf, **$d**1695-1770. **$4**fmo **$5**SwSKB

Field 655 provides a way of indicating provenance annotations. The term used here is taken from the thesaurus of the Rare Books and Manuscripts Section of the Association of College and Research Libraries, a division of the American Library Association. In the 700 fields, subfield 4 provides standard codes, for owner 'own' or former owner 'fmo'. Subfield 5 holds the code for the institution to which the field applies – always necessary to add when dealing with copy-specific information.

There is, obviously, a skeleton of a framework for structured provenance data, but it covers far from everything we would like to do. There is no way of coding date of purchase or price or other dates of former ownership, and it is not possible to code Amsterdam as the place of purchase or connect it with provenance in any structured way. Historical shelfmarks, as the one which proves the national library as owner from the middle of the eighteenth century in this case, can only be recorded as unstructured note text. However, the main problem is that it is not normal cataloguing praxis to record ownership in a structured way. If it is recorded at all, it is usually only recorded in a note field, from which no systematic retrieval is possible.

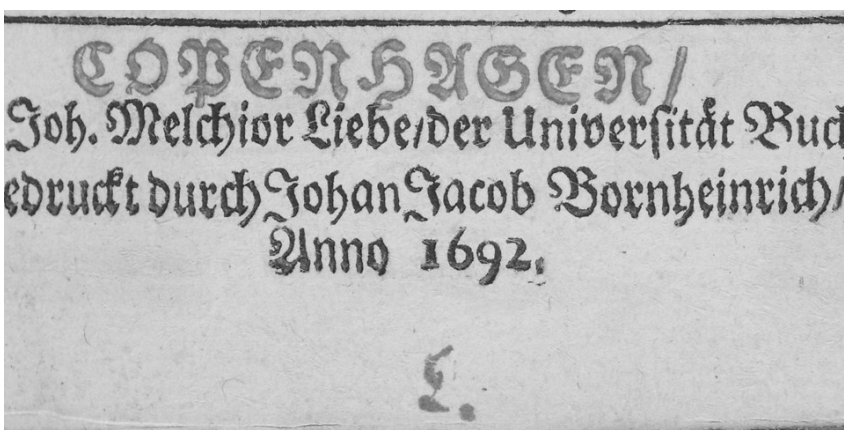
This was an example with well-known historical figures as owners. What about obscure names or unverifiable initials? Booksellers' initials, for example, that are often found on the title pages of old books, are of great interest and would be worth recording systematically.

Example 2, a and b *From the collections of the National Library of Sweden*



Example 2a

It is almost certain that the little signature in example 2a is for the publisher and bookseller Gottfried Kiesewetter, German of birth but active in Sweden in the middle of the eighteenth century. This particular signature occurs in other books, and it can be identified from known examples of his handwriting. Inconspicuous signatures, like his, are not uncommon on old books. It may be assumed that they are usually booksellers' signatures.



Example 2b

These signatures are not always easy to decipher – is the signature in example 2b 'CL' or 'GL'? Still, they should be recorded in some way. A scanned image linked to the record would be desirable, obviously, but it would still not provide an access point for systematic searches.

In December 2004 the Folger Library proposed on the ExLibris rare-book discussion list that it might be a good idea to use MARC21 field 720 to record unverified names from provenance information. This started a discussion within CERL, which ended in the conclusion that MARC21 720 was not a good solution.[3] But what is? Both UNIMARC and MARC21 offer certain specific fields for provenance data, but their functionality is limited.

MARC21 records for these cases could look like this:

2a:

561 ## **$a**Possibly published or imported for sale by Gottfried Kiesewetter; copy A in the National library has his signature on the title page. **$5**SwSKB
700 1# **$a**Kiesewetter, Gottfried **$d**1700-talet. **$4**pbl
700 1# **$a**Kiesewetter, Gottfried **$d**1700-talet. **$4**bsl **$5**SwSKB
700 0# **$a**GK **$4**bsl **$5**SwSKB

2b:

700 0# **$a**GL (?) **$4**bsl **$5**SwSKB
700 0# **$a**CL (?) **$4**bsl **$5**SwSKB

When the identity is known, the structured entry for initials is supposed to be recorded as a variant form in a 400 field in the authority record, not in a 700 field in the bibliographic record as well. When we need to link the data to a specific copy, however, it is simpler and more logical to have the information in the bibliographic record.

It is quite possible to create MARC21 authority records for unverified names, or to add structured access points in the bibliographic record for these unverified names, but it is not cataloguing praxis to do it. Likewise, there is nothing to prevent the creation of an authority record for initials – but the cataloguing praxis is not to do it. There are, in fact, no cataloguing guidelines at all for cases like these.

Binding information, not limited to owners' stamps or coats of arms on the covers, is another area of relevance in this context. Useful fields for genre/form terms (MARC21 655; UNIMARC 608) give controlled access for bindings, armorials, etc. – but there is no generally accepted thesaurus. A comprehensive list of provenance terms is offered by the German *Thezaurus der Provenienzbegriffe* compiled by the Anna Amalia Bibliothek in Weimar.[4] Another source that has already been mentioned is the Rare Books and Manuscript Section (RBMS),[5] which offers thesauri for

binding and provenance terms on their website; their thesauri are also defined in the MARC Standards code list. In UNIMARC, there are a number of codes available in field 141 for copy-specific attributes.

A MARC21 example from a union catalogue, with the 563 note fields for binding information of copies in different collections,[6] could look like this:

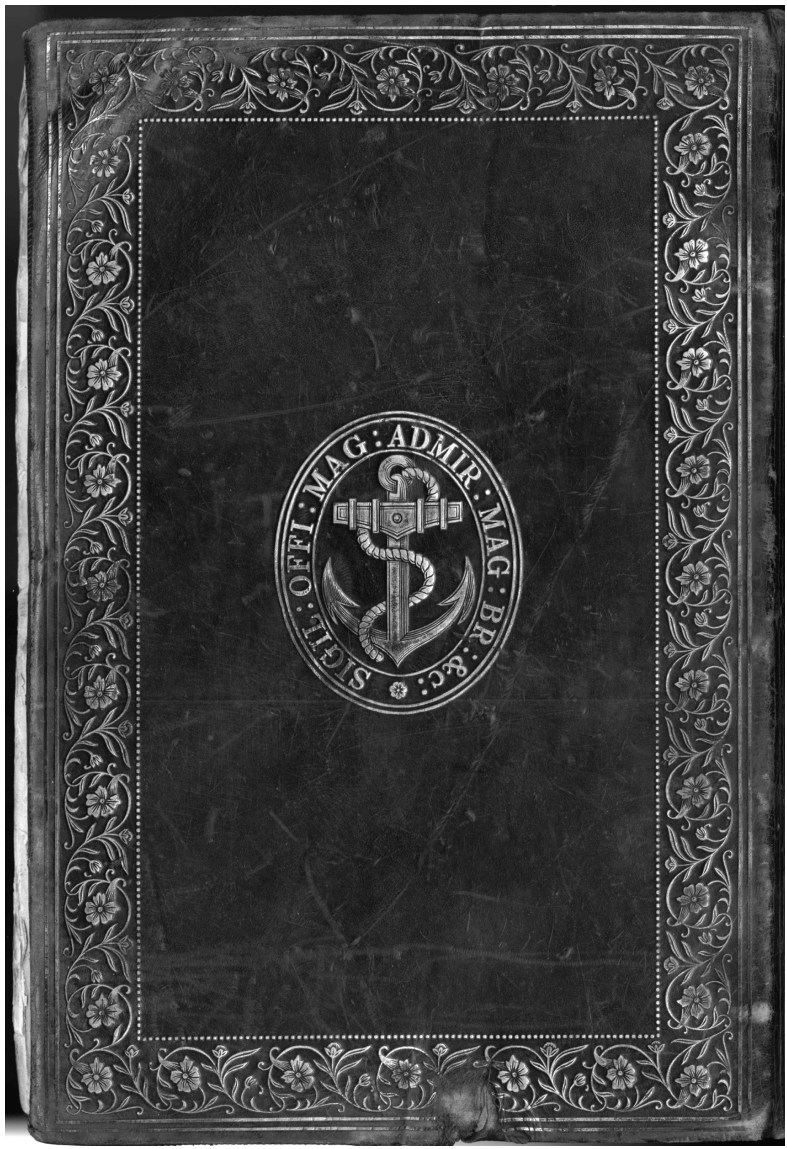
563 ## **$a**Full calf skin binding with monogram, 'V E', of Ulrika Eleonora the younger on front and back covers. **$5**SwSKB
563 ## **$a**Cordovan binding with blind-tooling on spine and boards. Wooden boards, clasps. **$5**Sko
655 #7 **$a**Calf bindings **$2**rbbin **$5**SwSKB
655 #7 **$a**Monogrammed bindings **$2**rbbin **$5**SwSKB
655 #7 **$a**Goatskin bindings **$2**rbbin **$5**Sko
655 #7 **$a**Wooden boards **$2**rbbin **$5**Sko
655 #7 **$a**Clasps **$2**rbbin **$5**Sko

'Cordovan' is not in the RBMS thesaurus, we have to resort to the broader term 'goatskin bindings'. Subfield 2 holds the MARC standards code for the thesaurus that has been used.

Provenance data for which we find no suitable field or subfield in MARC21 are purchase prices, historical shelfmarks, dates that apply to ownership, either of an exlibris or in an ownership note, and places connected with ownership, as has been pointed out above. At least in the case of female owners, it would also be of interest to record the gender of owners. It is possible to do that in a MARC21 authority record, and in the access forms in the bibliographic record, but not in a way that would make it systematically retrievable.

The next two examples show some other aspects of what might be called 'extended provenance'. We make no apology for the fact that the first one is quite modern, long after the hand-press period. Provenance is still being created today. In both cases the examples were kindly brought to the attention of the author by the librarians from their own knowledge of their collections.

EXAMPLE 3 *From the collections of the Imperial War Museum, London*



EXAMPLE 3

Example 3 at first sight appears to be a very uninteresting Bible, in rather poor condition. It is an edition of the 1611 Authorized or King James version. Thousands of copies of this particular edition must have been printed, probably about 1910. The stamp of the Imperial War Museum appears in the bottom right-hand corner of the title page.

More interesting is the back cover, shown here, embellished with gilt decoration around the edges. In the centre is a device consisting of an anchor with a rope twisted around it, a 'fouled anchor', enclosed in a highly abbreviated Latin phrase, SIGIL: OFFI: MAG: ADMIR: MAG: BR: &C: – 'The seal of the Lord High Admiral of Great Britain, &c.' *This* is the sign of provenance in the sense of ownership: the book was government property, issued by the British Admiralty to a unit of the Royal Navy.

The illustration may look poor, but in fact it is the book which is in poor condition, explained by a note on the flyleaf in the handwriting of the ship's chaplain. It is rather faded and does not reproduce well. A line-by-line transcription reads:

HMS Warspite / Commissioned April 1915 / (quam Deus Protegat) / The injuries to this book were / caused by a fragment of shell / in the Battle of Jutland May 31 / 1916. / A shell penetrated the deck / immediately outside the Church, / blowing in the armoured door on / the port side of the Church & killing / several men; and a fragment / injured this Bible. /

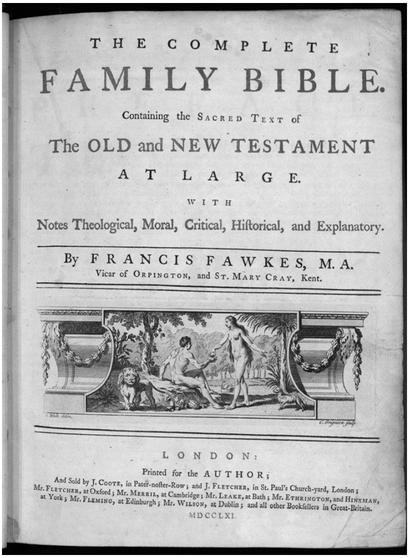
Signed / Walter Julius Carey / Chaplain / H.M.S. Warspite

The provenance is the same: there has been no change in ownership, but the note records a very significant event in the itinerary of this book. It ought to be possible to record such information in a structured way. This is addressed in a new UNIMARC field for Place and date of provenance, which should be published in mid-2007. Both the field and this specific example will be mentioned and discussed in more detail later:

$m North Sea
$e Warspite (battleship)
$f 19160531
$h Battle of Jutland
$5 UK-ImpWarMus : 02/1122

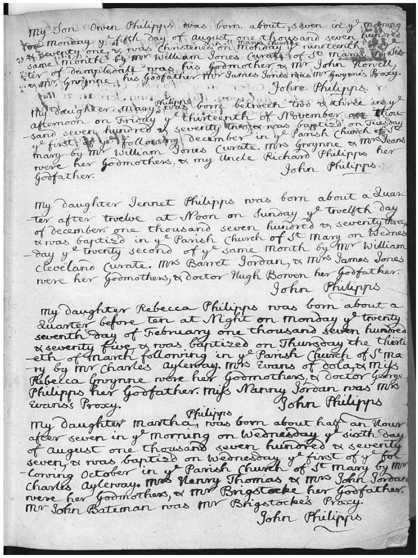
The structure and potential for indexing here are perfectly clear and self-explanatory, but there can still be problems: even translated literally, 'Battle of Jutland' is not what it is called in German – *Skagerrakschlacht*. Can thesauri help?[7]

EXAMPLE 4 *From the collections of the National Library of Wales, Aberystwyth*



THE COMPLETE
FAMILY BIBLE.
Containing the SACRED TEXT of
The OLD and NEW TESTAMENT
AT LARGE.
WITH
Notes Theological, Moral, Critical, Hiſtorical, and Explanatory.
By FRANCIS FAWKES, M.A.
Vicar of ORPINGTON, and ST. MARY CRAY, Kent.

LONDON:
Printed for the AUTHOR;
And Sold by J. COOTE, in Pater-noſter-Row; and J. FLETCHER, in St. Paul's Church-yard, London; Mr. FLETCHER, at Oxford; Mr. MERRIL, at Cambridge; Mr. LEAKE, at Bath; Mr. ETHRINGTON, and HINXMAN, at York; Mr. FLEMING, at Edinburgh; Mr. WILSON, at Dublin; and all other Bookſellers in Great-Britain.
MDCCLXI.



My Son Owen Philipps was born about seven in ye morning
on Monday ye fifth day of August one thousand seven hundred
& seventy one, & was Christened in St Mary's church on Monday ye nineteenth of ye
same month by Mr William Jones Curate of St Mary's. My Sis-
-ter of Dampleoale was his Godmother, & Mr John Nowell,
& Mr Gwynne, his Godfather. Mr James Jones was Mr Gwynne's Proxy.
John Philipps.

My daughter Mary Philipps was born between two & three in ye
afternoon on Friday ye thirteenth of November one thou-
-sand seven hundred & seventy two, & was baptizd on Tuesday
ye first of ye following December in ye Parish Church of St
Mary by Mr William Jones Curate. Mrs Gwynne & Mrs Jeans
were her Godmothers, & my Uncle Richard Philipps her
Godfather.
John Philipps.

My daughter Jennet Philipps was born about a Quar-
-ter after twelve at Noon on Sunday ye twelfth day
of December one thousand seven hundred & seventy three,
& was baptizd in ye Parish Church of St Mary on Wednes-
-day ye twenty second of ye same month by Mr William
Cleveland Curate. Mrs Barret Jordan, & Mrs James Jones
were her Godmothers, & doctor Hugh Bowen her Godfather.
John Philipps

My daughter Rebecca Philipps was born about a
Quarter before ten at Night on Monday ye twenty
seventh day of February one thousand seven hundred
& seventy five & was baptized on Thursday the thirti-
-eth of March following in ye Parish Church of St Ma-
-ry by Mr Charles Aylewav. Mrs Evans of Dola, & Miss
Rebecca Gwynne were her Godmothers, & doctor George
Philipps her Godfather. Miss Nanny Jordan was Mrs
Evans's Proxy.
John Philipps

My daughter Martha Philipps, was born about half an Hour
after seven in ye Morning on Wednesday ye sixth day
of August one thousand seven hundred & seventy
seven, & was baptizd on Wednesday ye first of ye fol-
-lowing October in ye Parish Church of St Mary by Mr
Charles Aylewav. Mrs Nanny Thomas & Mrs John Jordan
were her Godmothers, & Mr Brigstocke her Godfather.
Mr John Bateman was Mr Brigstocke's Proxy.
John Philipps

EXAMPLE 4

Example 4 is another English Bible, with notes, in two volumes, published in 1761–62. There do not seem to be very many copies to be found, but that is not the reason for its inclusion here. The bookplate inside the front cover states that the Bible was purchased in January 1921 by the National Library of Wales from the library of Sir Owen Henry Philipps Scourfield of Williamston, Pembrokeshire, Wales.

A little research reveals that Sir Owen's father was John Henry Philipps, who took the name Scourfield when he succeeded his maternal uncle.

What is remarkable about these two volumes is the extremely detailed record of births and baptisms, maintained by John Henry Philipps's grandfather for the years 1771–78, 1806 and 1808, the last entry being that for John Henry Philipps himself (not shown here).

For example, the last entry on this page reads:

My daughter Martha Philipps was born about half an hour after seven in ye Morning on Wednesday ye sixth day of August one thousand seven hundred & seventy seven, & was baptizd on Wednesday ye first of ye following October in

ye Parish Church of St Mary by Mr Charles Ayleway. Mrs Henry Thomas & Mrs John Jordan were her Godmothers, & Mr Brigstocke her Godfather. Mr John Bateman was Mr Brigstocke's Proxy. [*signed*] John Philipps.

This is priceless information for family and local historians. It also records provenance, for it shows that the book was in the possession of the family for a number of years. Charles Parry at the National Library of Wales remarked that there are several Bibles there containing information like this which does not appear in the general catalogue. It is pertinent to wonder how many Bibles (and other works) collected for their significance *as editions of the Bible* similarly contain valuable unindexed, and therefore untraceable, information in libraries around the world.

In this situation the computer format is not as helpful as it might be. Entries for former owners can be added to a MARC21 record easily:

700 1# **$a**Scourfield, Owen Henry Philipps, **$c**Sir **$4**fmo

and so on. However, although it is possible to make a subject entry for a published history of the Philipps family just as easily:

600 34 **$a**Philipps family **$z**Pembrokeshire (Wales)

one cannot attach a subfield $5 for the location and shelfmark of specific copy of a book which is *not* primarily about the family, such as a Bible, but nevertheless contains valuable added subject information. (That currently applies to MARC21; it should be possible in UNIMARC from mid-2007 following a change to the format).

PROVENANCE INFORMATION IN THE CATALOGUE RECORD

Provenance information is to be found in free-text notes fields, often largely unstructured, although there may be conventions governing their style and sequence. It is relatively easy to find provenance information about a known book: one has only to find the record for the book and look for the notes.

Working the other way round is another matter entirely. In order to find books which might have useful provenance information for one's topic it is necessary to decide on terms which are likely to appear in the context of that topic (plus equivalents in various languages?) and search the notes fields in the catalogue for them. This is a jump into the dark unknown: has one chosen the right terms? And has the provenance information been recorded in the first place?

The obvious places to search would seem to be the designated provenance notes fields :

561 (MARC21)
563 (MARC21; for information about bindings)
317 (UNIMARC)

and these will certainly retrieve relevant information, but in fact a lot more information can be found in General Notes (500 in MARC21; 300 in UNIMARC), with or without introductory text (e.g., *Prov:*, *Proveniens:*, etc), or by using an 'all notes' search. Most search systems have a fairly broad option probably favoured by most people most of the time ('Title words', 'Author', 'Subject words', 'Notes' and so on) plus an advanced searching option for more experienced users who wish to search more precisely, making use of the structure of the online bibliographic record. The results from the latter can prove surprising.

The online catalogue of the Library of Congress, for example, gives the user the option to look for terms in specific fields, by prefixing the letter K to the MARC21 tags. A search for 'Ex libris' produced the following hits:

all notes fields	(KNOT)	502
General Note	(K500)	324
Provenance Note	(K561)	97
Library of Congress copy	(K051)	53

The 'All notes' search finds *all* the occurrences of the phrase 'Ex libris' in notes fields, but many of these records may turn out to be totally irrelevant. The Provenance Notes search retrieves far fewer records, all of them relevant, but misses many others. (Field 051 is peculiar to the Library of Congress, and is used for information about their 2nd, 3rd ... copies of a work. This field should not appear in other libraries' catalogues.) Simple arithmetic shows that there are still 29 records with other notes containing 'Ex libris' hiding somewhere else.

Why are there so few hits for the Provenance Note field? The answer lies in the history of the format, coupled with library policy and economics. The USMARC format, including the General Note field (500), was more or less established as standard by 1973 – but the designated Provenance Note field (561) was added to the format in 1983. So for the first decade LC and countless other libraries were creating machine-readable catalogues and using 500 because there was nowhere else for provenance information.

Many libraries may have continued to do this, rather than change to 561; others may have started to use 561 but decided not to alter older records (including the Library of Congress, apparently); a few may have been able to make the complete change and bring their usage of the format up to date. One suspects that the reality is very few indeed.

This is a continuing problem: we argue for improvements in the formats, but whether improved and new fields, however desirable, are adopted, and to what extent, remains in the hands of the individual chief librarians, their cataloguers and their accountants.

STRUCTURED SEARCHING: ACCESS POINTS IN THE CATALOGUE

Relator terms or codes for owners, auctioneers, donors, binders, etc., can be added to the controlled headings for persons and institutions. In this way, the headings are distinguished from those for authors or subjects. For example,

100 1#	**$a**Molnár, Ferenc	works by Molnár
600 14	**$a**Molnár, Ferenc	works about Molnár
700 1#	**$a**Molnár, Ferenc, **$e***former owner*	specific copies once owned by Molnár
700 1#	**$a**Molnár, Ferenc. **$4***fmo*	[the same, in coded form]

Some questions remain. Is this done, and done consistently, or is the provenance information left solely in unstructured notes fields? If it is done, are these qualified headings visible not only in the library's own catalogue, but also in any union catalogue to which the records may have been contributed? And is it possible to *search* for these qualified headings to the exclusion of the others?

TOWARDS MORE STRUCTURED SEARCHING

There is a need for more and better structured fields to aid systematic searching for provenance information in online catalogues. Some work has been done recently on the UNIMARC format, building on developments in both UNIMARC and MARC21.

Both formats had almost identical fields for Hierarchical Place Access, used very largely, although not exclusively, for places of publication:

	UNIMARC 620	**MARC21 752**
$a	*Country*	*Country*
$b	*State, Province &c.*	*State, province, territory*
$c	*County*	*County, region, islands*
$d	*City*	*City*

In UNIMARC, field 620 was expanded for in two ways. First, new indicator values for *type* of publication to provide not only for regular book-trade publications but also for sound recordings (perhaps recorded on one date, issued on another, and reissued on a third one), production and issue of films, theatre programmes valid for a season, and so on. Second, new subfields were added for *circumstances* of publication, to provide for venue, date(s), season and occasion or event.

For example:

$a	*Country*	Hungary
$b	*State, etc.*	
$c	*County*	
$d	*City*	Budapest
$e	*Venue, specific location*	Szent István bazilika
$f	*Date*	18480423
$g	*Season*	
$h	*Occasion, event*	Easter Day
$i	*Final date*	

At almost the same time, MARBI was discussing the expansion of MARC21 field 752, but with a totally different objective, namely a more detailed specification of *places* (to the level of thesauri, e.g. *Getty Thesaurus of geographic names*) including Continents, City sections, Oceans, Space …

For example:

$a World
$a Europe
$a United Kingdom
$b England
$c Greater London
$d City of Westminster
$f Westminster

At last Caxton's output can be correctly located in Westminster, rather than London, while at the other extremes of place and time, MARC21 is ready for the first book to be published on the moon.

However, neither of these UNIMARC and MARC21 fields deals with the history of specific items after their publication: the implications for provenance do not seem to have been considered at all. With help and encouragement from CERL, a proposal for a new field 621 for Place and Date of Provenance was put to, and accepted by the Permanent UNIMARC Committee in 2006. This drew on both UNIMARC 620 for

dates, seasons, occasions and venues, and MARC21 652 for a wider range of places. In this way, the answers to the Provenance questions *Where* and *When* can be recorded in a systematically searchable way.

It should be made clear that it is not necessary to use *all* the detail every time – $dBudapest by itself is legitimate, if it is not normal practice to record country and all the intermediate levels of jurisdiction as well. Moreover, the field can record incomplete data:

$e The Old Mill	*a location/venue, if that is the only geographic information given*
$f 1762	*a year, without month or day; or*
$f uuuu07	*'July' with no year, etc.*

Provenance information is often incomplete, but that is no reason for not recording the fragments that we have. By matching fragments, books related by provenance may be brought together.

Here, then, is the *HMS Warspite* example once again:

317 ## **$a**Inscription on flyleaf: HMS Warspite Commissioned April 1915 [*etc.; inscription recorded in full*] … Signed Walter Julius Carey Chaplain H.M.S. Warspite **$5**UK-ImpWarMus : 02/1122
500 11 **$a**Bible **$m**English **$q**Authorized
621 ## **$m**North Sea **$e**Warspite (battleship) **$f**19160531 **$h**Battle of Jutland **$5**UK-ImpWarMus : 02/1122
700 1# **$a**Carey **$b**Walter **$4**??? **$5**UK-ImpWarMus : 02/1122

This is part of a UNIMARC record using the new field 621 for Place and date of provenance. Subfield **$m** holds a geographic area which is not a jurisdiction like a state or city, $e a building, vehicle or other venue, in this instance a ship, **$f** the precise date, **$h** the name of the event, and **$5** the present location of the item. In field 700, is there a good code (**$4**) which would describe the Walter Carey's contribution? '060' for Associated Name seems a rather feeble and unsatisfactory solution; perhaps '020' for Annotator is better, although that might suggest a person who has made notes throughout a text. He was not the former owner of that Bible.

CONCLUSIONS

We do see today a growing understanding among librarians that provenance information is important, but the framework for recording it has serious gaps. Format functionality is piecemeal, but maybe still worse is that common cataloguing guidelines and common thesauri are lacking. It is not too uncommon to record provenance information in unstructured

notes, but that does not offer the systematic access that would support scholars.

There is also the issue of bibliographic records versus holdings records, which is complicated by inconsistent thinking and practice about what is 'public' and what is 'private' information in the catalogue. Far too many libraries appear to have adopted a policy of treating all copy-specific information as local and 'private'. Very often the information (or much of it) is visible in a library's own OPAC, but vanishes when records are loaded into a union catalogue. One way in which this is done is the widespread use of 59X fields (39X in UNIMARC). 59X (and other '9' fields) are left free in the format for local definition, and one library's 590 is not necessarily the same as another's – so they are all suppressed in union catalogues or universal, bibliographic databases. $5 tends to go the same way. Then there is the $4/$e problem with names which have relator codes or terms attached: can one see them, and can one search them? As the reader has seen, our examples and arguments assume that provenance information belongs in the bibliographic record. We believe this to be the proper place for it, both on logical and practical grounds.

The question of resources is also pertinent in this context, resources in budget terms and resources in terms of competence. There is today a large number of catalogue records for hand-press publications, not least thanks to CERL's efforts and achievements. Most of those records, however, have very little or only unstructured provenance information. It is not likely that libraries will be able to afford to extend the cataloguing task to a comprehensive recording of copy-specific information, and we do not think cutting back radically on traditional cataloguing data is a good solution either, although we have seen it proposed recently by a librarian and researcher. Anyway, it will not be possible to go back and add this information to records produced in completed retrospective projects, and in the regular operations of libraries we have to face the fact that librarians are usually not trained for this kind of task.

Is it necessarily the librarian's job to put in provenance data? In many or even most cases, it would be more efficient for the researchers to record provenance information. They are the experts in reading hands from different times, in binding techniques and styles, or in the book trade of earlier times. In specific cataloguing projects, it is usually possible to achieve co-operation between librarians and researchers.[8] For single books, other solutions must be sought, and in our vision for the future, we see a web template offered to the researcher by the library's system and

a paper slip handed out with every old book, urging the researcher to add whatever data about the book he/she would want to add to the catalogue record. The so called 'scholar's notepad' which CERL is developing is maybe a first step towards that future.

So, what should CERL do and what should libraries do? Our answer is that we should continue to extend format functionality, but also develop cataloguing guidelines and thesauri, and not least, develop easy-to-use tools for researchers and establish co-operation with them.

NOTES

1. Rudyard Kipling (1865–1936), from 'The Elephant's Child' in *Just so Stories for Little Children*, 1902.
2. Monique Hulvey, in a presentation in a CERL seminar in Zagreb 2005, made a full account of the research potential for provenance data and a very comprehensive account of projects. (Her presentation is available on CERL's web site.)
3. See G. Jonsson and M. Willer, *Provenance information and authority control – a discussion paper*, in the provenance section of CERL's web site.
4. See http://www.klassik-stiftung.de/einrichtungen/herzogin-anna-amalia-bibliothek/projekte/provenienzportal/informationsmittel/t-pro.html
5. See http://library.osu.edu/sites/users/russell.363/RBMS%20Thesauri/index.htm
6. 'Sko' is not an approved code, but is still used in the Libris database.
7. It should be noted that, at the time of writing, the library of the Imperial War Museum was not using a MARC-based system for its cataloguing and OPAC.
8. A current example is APIS 'Advanced Papyrological Information System', which records provenance information: http://dpg.lib.berkeley.edu/webdb/apis/apis2?invno=&apisid=1450&item=1.

   See also the project list in Monique Hulvey's presentation referred to above, and the database of Bibliothèque municipale de Lyon, http://sged.bm-lyon.fr/Edip.BML/(j3d1gmjtbqqsne45qpq43zrd)/Pages/Redirector.aspx?Page=MainFrame

Accessing the record of European printed heritage: The CERL Thesaurus as an international repository of names from the hand-press era

ALEXANDER JAHNKE

A CERL Thesaurus személyek, intézmények, nyomdászok és helyek neveit tartalmazza a kézisajtó korából (1450-kb. 1830). A CERL tagkönyvtárak által rendelkezésre bocsátott file-okból épült, és elsősorban az egységesített névalakok hiányát pótolja a CERL HPB adatbázisában. Túllépi az authority file-oknak tulajdonított határokat, központi csomópontot adva a nyomtatástörténetben való eligazodáshoz. A CERL Thesaurus átfogó hivatkozásgyűjteményeként használható ezen a területen, de elérhetővé válnak olyan kapcsolódó web-es források is, mint a nyomdász- és kiadói jelvények digitális képei, vagy a könyvtárak katalógusaiban (OPAC) rejlő provenancia-információk. Továbbá párbeszédet is biztosít azáltal, hogy a szakmának lehetősége van megjegyzéseket fűzni az egyes rekordokhoz.

INTRODUCTION

The CERL Thesaurus is a database containing the names of printers, authors, places and corporate bodies who were in some way involved in the production of books from the times of Gutenberg until the emergence of rotary printing in the first half of the nineteenth century. The word *thesaurus* should here be understood in the original Greek sense of 'treasure' rather than in the way this term is used in information science as defined in the international standard on Guidelines for the establishment and development of monolingual thesauri (ISO 2788). According to the official ISO definition, a thesaurus is a representation of concepts and the relationships among them.[1] It controls the naming of those concepts by stating preferred terms and indicates concepts being super- or subordinated or somehow related to others. Since the entities represented in the CERL

Thesaurus by nature defy being structured hierarchically, it does not meet these requirements and therefore cannot be called a thesaurus in the technical sense of the word.

Calling the CERL Thesaurus, on the other hand, an *authority file* would not give a proper description of this database either. There are of course certain elements of an authority file present in the CERL Thesaurus, such as see-reference tracings and see-also references. And further, the primary sources for the CERL Thesaurus are authority files used in various libraries throughout Europe. But as soon as it comes to the nitty-gritty of authority work, which is establishing *one* controlled heading that is used as an access point throughout a library catalogue, the CERL Thesaurus will quickly disappoint its user, for one of the basic concepts applied in the CERL Thesaurus is to abandon the indication of one preferred entry form.

While failing in respect of an authority file's main purpose, the CERL Thesaurus on the other hand exceeds the scope of an authority file by far. In addition to providing variant name forms, the CERL Thesaurus offers additional information on a printer's, a person's or a corporate body's background, such as biographical information, the described person's occupation or the corporate body's activities and more.

It further represents relationships between the described entities. For example, a personal author being a member of a certain corporate body, or a printer taking over the workshop from his predecessor etc. By this representation of relationships it might probably be justified to call this database a thesaurus, though, as we have said above, the type of relationships that the CERL Thesaurus records is by nature non-hierarchical. However, by providing this kind of information the CERL Thesaurus functions as a source of reference for the field of historic printing. This aspect is not unknown to authority files, and most of the larger (national) authority files show the tendency to be used as reference works as well. But the CERL Thesaurus proceeds towards transcending the boundaries of common authority files by providing not only access to information recorded in the database, but also to information that is recorded in external sources. These external information resources are for example digitised reference works on printers and their devices such as Ronald B. McKerrow's *Printers' & Publishers' Devices in England & Scotland 1485–1640*, or databases like the *Censimento nazionale delle edizioni italiane del XVI secolo* also known as *EDIT16*. The latter provides detailed information on Italian printers, images of the devices used by them and bibliographic descriptions of the books printed by them.

Another type of external information objects which the CERL Thesaurus provides access to, is the recording of provenance information in local library catalogues. Collections of books owned by a certain person or institution that are now scattered among several libraries may be virtually reconstructed by using the CERL Thesaurus as an access point to the bibliographic records in question.

Finally the CERL Thesaurus is a platform of scholarly communication. By the integration of a feature that allows users of the CERL Thesaurus web interface to leave comments on a record 'right next to it', anybody can contribute to the CERL Thesaurus. This communication platform allows the scholarly community to share information on the described entities that are beyond the actual scope of the database and help to improve the Thesaurus's quality and usability.

Taking the last three points into account, the CERL Thesaurus appears to adopt elements of a portal application. Its appearance as a single database contradicts this denotation, though. Speaking about the CERL Thesaurus as a 'repository of names', as we did in the title of this paper, implies that there is more to this database than being an authority file. On one hand it does in fact serve in several ways the purposes of a portal. But it is still a stand-alone database application and therefore naturally lacks a number of elements that are on the other hand typical to library portals.

In order to better understand the nature of this crossbreed, we will now approach it from the perspective of the five aspects outlined above and investigate how the CERL Thesaurus functions as an authority file, a reference source, an access point to digitised material and provenance information and as a means of scholarly communication.

THE CERL THESAURUS AS AN AUTHORITY FILE

The CERL Thesaurus has been developed by the Data Conversion Group at the Niedersächsische Staats- und Universitätsbibliothek Göttingen since 1999. From the beginning this development was closely related to the evolvement of the Hand Press Book Database (HPB). The HPB, CERL's main field of activity, is a database that assembles bibliographic files describing printed material from the hand-press era. These files are contributed by a large number of European and North-American libraries. The Hand Press Book Database aims to give a comprehensive view on the entirety of Europe's printed heritage. To ensure this comprehensiveness at acceptable expense, the approach has been taken to build up this database from existing bibliographic files provided by CERL's member libraries. As

these bibliographic records are produced by different cataloguers in different countries, following different cataloguing rules and cataloguing traditions, they inevitably lack coherence to a certain extent, when integrated into one data source. So far as bibliographic description is concerned, this incoherence is hardly noticed. But with regard to the provision of access points, the use of different cataloguing rules to set up entry forms within one catalogue is a violation of the principle first enunciated by Charles A. Cutter and later reiterated in the Paris Principles, stating that publications by one author should be accessible under the very same heading.[2]

Regardless of the fact that the current system which HPB is hosted on does not support the employment of authority records, it would be an impractical endeavour to apply authority control retrospectively on now 2.8 million records. As far as authority control is abandoned in HPB, the collocation of works written by the same author or printed by the same printer etc. has to be left to the user. When searching in HPB it is his task to expand the search on all possibly occurring name forms.

This is the point where the CERL Thesaurus comes in: similar to the approach of HPB, the CERL Thesaurus brings together in one database authority records from the libraries contributing to HPB. There are records from the German *Personennamendatei*, the authority files for printers and authors from the STCN, the authority files for personal and corporate authors from the *English Short Title Catalogue*, authority files on imprint places and printers from the National Library of Croatia, files from Poland, Italy, Sweden, France and many more[3]. This aggregation of authority records particularly – but not only – from libraries that contribute to the HPB, should ensure that a heading form occurring in the HPB is also recorded in the CERL Thesaurus.

Unlike the CERL Thesaurus, the HPB is a database where records are not edited manually once they have been uploaded. It is obvious that in terms of manual editing the CERL Thesaurus as a resource of collected authority records is easier to handle than the large HPB. The main task is to ensure that authority records referring to the same entity are joined together, which is achieved by merging parallel records into one. This merger of records is supported by an automated process that marks records containing the same name form as possible duplicates, or if evidence can be given from matching of biographical dates, the merging is also conducted automatically.

As a result, intellectually revised records end up containing a number of controlled headings established by different libraries according to their

particular cataloguing practice. These headings exist equally in parallel, with none of them being considered a preferred name form, since they coexist in the same manner as controlled catalogue entries within the HPB.

The data format used for the CERL Thesaurus is mainly based on but not exactly compliant with UNIMARC Authorities.[4] Due to the approach of abandoning the use of one preferred heading, the entry forms are repeated, in deviation from the format.[5] In the same way as the different headings come together in one record, all see-reference tracings from the originating records come together as well and now point to the group of controlled headings and no longer to one single preferred form.

A user of the Hand Press Book Database can use these controlled headings and the variant name forms to expand a query. The variant name forms (or see-reference tracings) might comprise entry forms from files within HPB that do not apply authority control at all. By including all possible entry forms into the query, an effect similar to the employment of direct authority control in a library catalogue is achieved. Though admittedly this procedure is not as precise as a direct linking of bibliographic records to authority records would be. The main purpose of authority control is to improve the precision of a system's response to a query by reducing the effects of synonymy as well as of homonymy. While synonymy can be handled with this technique of query expansion quite well, the homonymy problem still remains unsolved. There is as yet no way of dealing with homonymy other than linking directly to distinct authority records and performing a two-step search, first on authority records and than retrieving the related bibliographic records.

In order not to impose the work of typing all relevant search terms into the database's search interface on the end-user, a function has been added to RLG's Eureka system that supports the formulation of expanded queries.[6] This feature, called *Assisted Searching*, retrieves controlled headings and variant name forms from the CERL Thesaurus based on the search term entered by the user. From the search result all name forms are extracted (standard entry forms as well as see-reference tracings) and concatenated by a Boolean OR-operator, thus creating an expanded query. Based on this expanded query a new search is automatically triggered in the HPB.

This technique is demonstrated in Figure 1 showing the result screen produced by a search for *Budapest* as an imprint place: there are 33 bibliographic records in the database that describe books with the place name

Budapest appearing in the imprint as place of printing or publishing. This is *prima facie* an astonishing small result,[7] but one has to be aware of the fact that the search matches only those records where exactly the word *Budapest* is present in the imprint place field. Neither does this search find any grammatical deviations, nor name forms in Latin or other languages, nor publications that hold either *Buda* or *Pest* or any variant of these in the imprint.

When the user clicks on the link labelled *Variant Place Names*, the Eureka system executes a remote search for the word *Budapest* in the CERL Thesaurus's place names index. The CERL Thesaurus responds by sending back the controlled headings and the variant name forms.

The user may now start a search in HPB using all the name forms provided, by just a mouse-click. Figure 2 shows the search result on a search for *Budapest* as imprint name after the employment of Assisted Searching. The recall has been increased from 33 to 1485 hits, though with the constraint that some of the variant name forms might also be used for other printing places and thus a certain redundancy might deteriorate the search's precision.

Through Assisted Searching the CERL Thesaurus functions to a certain extent as a substitute for an authority file in the Hand Press Book Database. Since this mechanism is not based on authority control, but on query

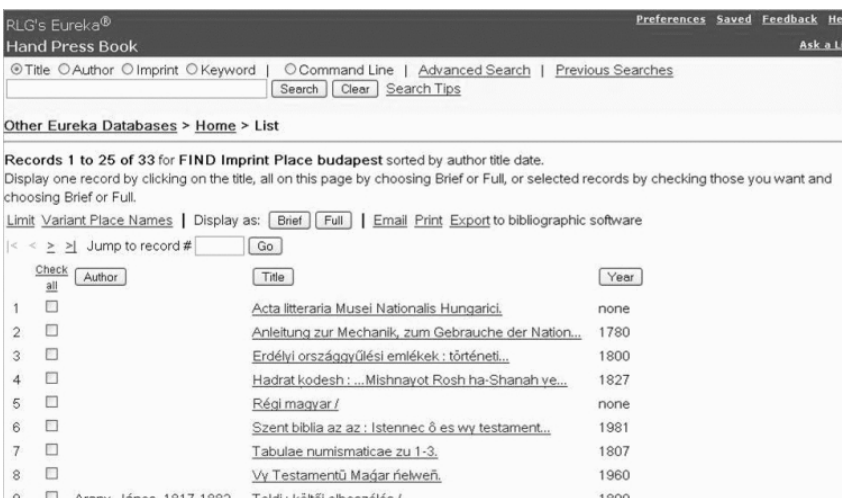
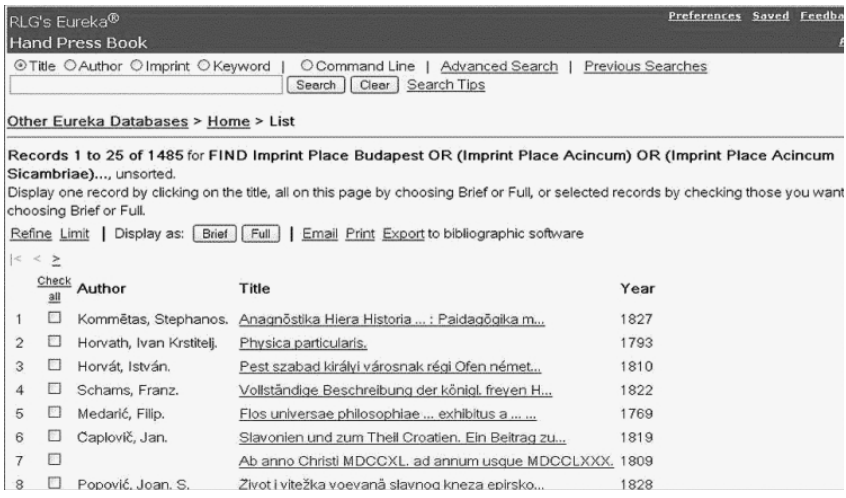


FIGURE 1 Search result for *Budapest* as imprint place in HPB


RLG's Eureka®
Hand Press Book
Preferences Saved Feedba
Title Author Imprint Keyword | Command Line | Advanced Search | Previous Searches
Search Clear Search Tips
Other Eureka Databases > Home > List
Records 1 to 25 of 1485 for FIND Imprint Place Budapest OR (Imprint Place Acincum) OR (Imprint Place Acincum Sicambriae)..., unsorted.
Display one record by clicking on the title, all on this page by choosing Brief or Full, or selected records by checking those you want
choosing Brief or Full.
Refine Limit | Display as: Brief Full | Email Print Export to bibliographic software
Check all

		Author	Title	Year
1	☐	Kommētas, Stephanos.	Anagnōstika Hiera Historia ... : Paidagōgika m...	1827
2	☐	Horvath, Ivan Krstitelj.	Physica particularis.	1793
3	☐	Horvát, István.	Pest szabad királyi városnak régi Ofen német...	1810
4	☐	Schams, Franz.	Vollständige Beschreibung der königl. freyen H...	1822
5	☐	Medarić, Filip.	Flos universae philosophiae ... exhibitus a	1769
6	☐	Čaplovič, Jan.	Slavonien und zum Theil Croatien. Ein Beitrag zu...	1819
7	☐		Ab anno Christi MDCCXL. ad annum usque MDCCLXXX.	1809
8	☐	Popović, Joan. S.	Život i vitežka voevanâ slavnog kneza epirsko...	1828

FIGURE 2 HPB search result after query expansion

expansion, it helps to handle the phenomenon of synonymy but does not tackle the issue of homonymy. To distinguish homonymous forms in a search would only be possible if direct linking to authority records had been applied in the HPB.

Assisted Searching on the basis of the CERL Thesaurus has also been implemented in the CERL Portal which offers cross-searching of a range of databases of manuscripts and printed books.[8] From the *Advanced Search* interface, variant name forms can be included in the Meta search performed on several databases with manuscript material. Though developed in the context of the Hand Press Book Database, the CERL Thesaurus is able to support query expansion in other search environments in a similar way. From the technical point of view, the CT does nothing more than allowing remote searching and provision of a structured list of name forms from any record.

THE CERL THESAURUS AS A SOURCE OF REFERENCE

The second aspect of the CERL Thesaurus which we mentioned at the beginning is its role as a reference tool for the field of historic printing. The 2005 draft document on the Functional Requirement for Authority Records, which is currently under revision by the FRANAR working group, states 'serving as a reference tool' as one of the functions of an

authority file.[9] That implies in particular the differentiation of one entity from another, which is an essential task in the process of authority control.

As an aggregation of various authority files from different backgrounds, the CERL Thesaurus is well equipped to serve this purpose. It preserves all the information present in the source records and can therefore be used for reference in the same way as any national or local authority file. Even more, for the hand-press era it relieves the user of consulting a number of different authority files, because all the necessary information is already assembled in one source.

While differentiation of entities is an important task in cataloguing, the scholar using the CERL Thesaurus via its web interface has other interests: He might for instance be asking for the meaning of a rarely used Latin place name or for biographical details on a certain printer. Instead of looking up a number of printed reference works, he is likely to find the answer by a simple query on the CERL Thesaurus.

In addition to the information provided by the authority files contributed by the CERL member libraries, there are further details from printed reference works that have been added to the database during the process of manual editing. E.g. from 2002 to 2004 the entire segment of place-name records has been revised by librarians from the Bayerische Staatsbibliothek in Munich and all records have been checked against Graesse's *Orbis Latinus*, Deschamps's *Dictionnaire de géographie ancienne et moderne* and other widely used reference works.

To facilitate the user's access to the information stored in the CERL Thesaurus, a new web interface was made available in summer 2006. It allows searching by names within the four segments of the database (imprint places, imprint names, personal names and corporate names). It also allows searching for title words and acronyms of reference works cited in the records. In order to do justice to the international approach of the CERL Thesaurus, the character encoding within the database has been changed to Unicode. Any special characters from languages using Roman script can be displayed and processed by the search interface. To enable users without an appropriate keyboard configurator to enter characters from the Cyrillic, Greek or Hebrew alphabet, a virtual keyboard has been included. It is expandable to a number of further character sets like Arabic or East Asian scripts.

The database is primarily searched for names. The employment of permuting indexing allows finding names for printers, personal authors and corporate bodies with filing qualifiers or other additions to the name by

any word order. The user may either perform a direct search that retrieves all records containing his search term or may browse the index lists and pick out the name forms he wants to see.

By selecting more than one record from the index browse list or the search result list, the selected records are displayed in a list that contains the controlled headings and some information sufficient to distinguish one entity from another. From this list the record's full display can be accessed, which shows controlled and variant name forms, details on biography, occupation and fields of activity. Further it mentions the reference works and imprint sources on which the decisions for setting up the particular headings in the originating authority files are based.

The user may choose to have additional information displayed about the contributing libraries which are using the controlled name forms as headings. This might be useful to know for further searches in the respective bibliographic files (see Fig. 3).

In addition to this individual information, which the CERL Thesaurus provides in a similar way to other authority files or reference works, the CT is also capable of indicating different kinds of relationships between entities of the same type as well as of different types. Links within the records point for example to records of the predecessor or successor of a firm, or to persons being affiliated with this firm and so on. As the CERL Thesaurus is a database application, it is also possible to retrieve all links

Next >

display standard forms

Record Identifier	cnp00396685
Headings	Thomas <Aquinas, Saint> *Standard form of: English Short Title Catalogue(AACR2)* Thomas <de Aquino> *Standard form of: Die Deutsche Bibliothek / Personennamendatei(RAK-WB)* Thomas <van Aquino> *Standard form of: Koninklijke Bibliotheek - Nationale bibliotheek van Nederland* Thomas <von Aquin> *Standard form of: Die Deutsche Bibliothek / Personennamendatei(RSWK)* Thomas von Aquino *Standard form of: BSB,VD16*
Found in	[GK55=Thomas, Aquinas, Saint] [M] [Tusculum-Lexikon] [LThK] [CSGL, M] [LMA,VIII,706/11]
Imprint Sources	• De divinis moribus et De beatitudine aeternitatis.... / By Thomas Aquinas, 1474 • De rebuspublicis, et principum institutione, libri... / By Thomas Aquinas, 1651 • Logica beati Thome aurea et prof... Thomas von Aquino, 1505 • T228443: A hymn on the institution of the Eucharist: by St. Thomas of Aquin

FIGURE 3 CERL Thesaurus record full display with standard forms

that point backwards to a given entity. With the representation of the relationships, another layer of reference information has been introduced to the CERL Thesaurus. By providing this second level of information, the CERL Thesaurus evolves into something that might be called a *semantic network*.[10] As C. Hengel points out, the tendency to evolve into semantic networks is implicitly present in many authority files, beginning on the very basic level with recording see-also-reference tracings and on the next level adding cross-file links between authority files for different entities. This means establishing links from, for example, a corporate names file or a subject heading file to a personal names file and vice-versa.[11] Since the CERL Thesaurus combines such records of different nature within one file, implementing, making visible and using this network of relationships is rather easy.

Subsequent to the application of relationship information, one might ask if this should also have an impact on the information presentation and visualisation within the database's search interface. It might be worthwhile to reconsider the approach that has currently been taken to retrieval of and navigation through the records. The employment of a graphical visualisation of the entities' relationships similar to that found in topic-maps could probably not replace but surely supplement the conventional word-based searches and result sets.

THE CERL THESAURUS AS AN ACCESS POINT TO EXTERNAL INFORMATION RESOURCES

The third level of the CERL Thesaurus's information architecture is formed by the provision of access to information objects that exist outside the scope of the database. These currently include mainly digitised images of printer's devices and descriptions of them. The authority records for printers' names used in *Edit16*, for example have, when integrated into the CERL Thesaurus, been linked to the printers' devices which are present in the *Edit16* database, as well as to the *Edit16* authority records themselves. This enables the user to access descriptions and images of the printer's devices directly in the source database and also to execute searches for bibliographic information from the original authority record.

Further, printed reference works of relevance for the hand-press era are being digitised by the Royal Library in Copenhagen in close cooperation with CERL. In 2005 this undertaking began with the digitisation of Ronald McKerrow's *Printers' & Publishers' Devices in England & Scotland 1485–1640* and continued in 2006 with Philippe Renouard's *Les marques*

typographiques parisiennes. From the beginning the CERL Thesaurus was meant to form an access point to these digitised images and to descriptions of the printers' devices.

This principle of linking to external resources can well be extended to other information resources that are available via the World Wide Web, like for example online biographies or projects on the edition of the work of a certain author etc.

The integration of external information objects into bibliographic databases has been discussed among librarians under the heading of 'catalogue enrichment' during recent years. Certain attempts have been made in this area, but adding such objects to authority files is an entirely new approach.

THE CERL THESAURUS AS AN ACCESS POINT TO PROVENANCE INFORMATION

The execution of searches in remote bibliographic databases under the aspect of provenance information is yet another approach to connecting the CERL Thesaurus to external resources. Provenance information is an area of research in many libraries with larger collections of old books. The answer to the question 'who has possessed which particular book?' is an invaluable contribution to the understanding of intellectual history.

Provenance data is strictly bound to bibliographic data – or to be more precise, it is residing even a level below, at the description of copy-specific attributes. Therefore provenance data is in most cases recorded in local library catalogues rather than in union catalogues or other collection-independent bibliographic databases. Since the CERL Thesaurus neither aims to resample a bibliographic database, nor is designed to serve this purpose, it does, of course, not provide for the storage of provenance information itself.

To make provenance information retrievable from a CERL Thesaurus personal name or corporate name record, the approach has been taken to redirect the user to the different library catalogues that contain the data in question. Unlike the method of linking outwards to external information objects, as we have discussed above, there is no direct linking to bibliographic records in library OPACs. Instead, similar to a metasearch provided by a portal system, the user may trigger a search in the remote system containing the provenance information related to the entity in question from within the CERL Thesaurus.

The reason for this remote search approach is simply the amount of information that otherwise would have had to be included in the CERL

Thesaurus's records, which would have made the records rather unreadable.

Currently the access to provenance information from within the CT is possible from 1,510 personal names records and 532 corporate names records. These records are derived from authority files for provenance information used by the member libraries of the Gemeinsamer Bibliotheksverbund (GBV) in Germany. Names known to be filed as former owners of books in a certain library catalogue are added to the CERL Thesaurus's record together with the information on the appropriate access method. A preformulated query can then be handed over to a client-side script that executes the remote search in a new browser window. The user will be presented with the remote system's record display and can continue his work in that system if necessary.

The information about the provenance information available in a remote system and the way to formulate appropriate queries is gathered as far as possible from specific authority records that serve provenance purposes. This has been done for the GBV provenance data and will be done in due course for the provenance data of the Bibliothéque municipale de Lyon.

Providing a centralised access point to provenance data from various library catalogues facilitates the reconstruction of historical book collections that have been scattered over the years and ended up dispersed among different libraries.

In the future, an even more portal-like approach, executing the remote searches and retrieving the search results by a server-side application rather than at client-side, might be worth considering. Responses could then be rendered in a consistent manner, as known from portal software used in subject gateways or virtual libraries.

THE CERL THESAURUS AS A PLATFORM OF SCHOLARLY COMMUNICATION

We have seen so far that the CERL Thesaurus can serve as a substitute for an authority file in data sources that lack consistent authority control. This applies to the Hand Press Book Database as a physically existing file that aggregates bibliographic records from various backgrounds as far as the cataloguing traditions are concerned. This also applies to the virtually joined collection of databases that are assembled in the CERL Portal. Here, the divergence of data formats, access points and retrieval techniques is even higher than in the HPB.

At another level, the CERL Thesaurus can be used as a reference source that provides individual information on each described entity (printing places, printers, persons or corporate bodies). In addition to this kind of individual information, existing relationships between entities can be represented, thus broadening the user's view on the universe of the book production from the fifteenth to the early nineteenth century.

Further, the CERL Thesaurus exceeds the limitations of a stand-alone database by centralising access to various information resources in a portal-like fashion.

By combining these different information services, the CERL Thesaurus can act as a central junction for navigating through the domain of historical printing. The CERL Thesaurus gives proof that the potential that lies in such a simple auxiliary tool as an authority file is much higher than commonly considered.

In addition to the aspects we have mentioned above, the CERL Thesaurus not only offers monodirectional access to existing information resources, but is also open to receive information provided by its users. Thus it not only establishes the link to codified knowledge but also to the living members of the scholarly community.

Integrated in the CERL Thesaurus web interface is the so-called *Scholar's Notepad*, an application developed by CERL and the Data Conversion Group, which provides a means for database users to comment on a specific record.[12]

By clicking on a button in a record's full display, the annotation area opens just next to the record. Here, anybody can read comments left by other users as well as insert their own annotations. These comments may contain

- hints on inconsistencies in a particular record,
- additions to the given information such as other name forms,
- indications of other interesting information resources such as links to external biographies, images etc.
- questions concerning the entity described,
- announcements of new publications on the person or corporate body in question
- as well as any other imaginable notice that is suitable for being publicly accessible.

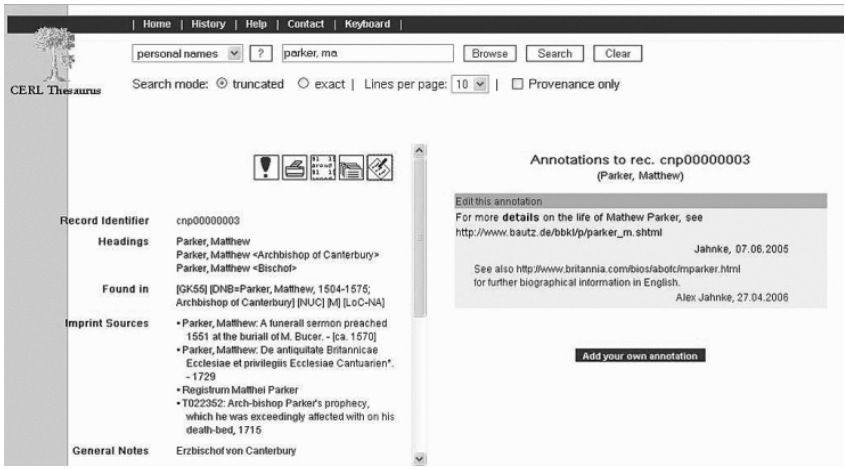


FIGURE 4 Annotation in Scholar's Notepad

To add an annotation it is not necessary to register beforehand or being granted special access rights. Clicking the *Add your own annotation* link opens a window, where the user may enter his comment (see Fig. 4). If he or she wishes to be contacted by CERL or be notified if an editor reacts on the comment, the user may provide further details such as name and email-address. These personal details will by no means be visible to the public. To ensure that a comment can be edited after it was submitted, it may be protected by a password of the user's own choice. Any password-protected comment shows a link labeled 'Edit this annotation', where, after the chosen password has been entered, the editing window opens again.

To enable CERL to keep track of the activity in Scholar's Notepad, the comments are accessible via a specific administration interface. This allows CERL to react on the comments, either by responding to a question or taking action, if, for example, a user mentions a mistake in the data.

Technically speaking, Scholar's Notepad runs as a separate application with its own underlying database. Comments and records are linked by the CERL Thesaurus record identifiers, which are persistent. Physically separating the annotations from the records follows on one hand from obvious security considerations and on the other hand from the aim to use this feature also in combination with other databases, particularly with the HPB.

Scholar's Notepad is still under development. The final version will be able to handle annotations to different databases simultaneously. Additionally, it will provide a customizable, probably XML-based, output to enable the integration of this service into applications based on newly emerging technologies like AJAX. Already available is an RSS-feed that provides the most recent annotations in chronological order.

PARTICIPATING IN THE FURTHER DEVELOPMENT OF THE CERL THESAURUS

The CERL Thesaurus is a living file. CERL and the Data Conversion Group are continously working on the integration of new authority files and secondary information such as external resources and provenance information to connect to. Intellectual revision and deduplication of records is undertaken in cooperation with the Bayerische Staatsbibliothek, Munich. Both activities are aiming to improve the quality of the data recorded in the CERL Thesaurus and to increase the benefit of the database for the end user.

Participating in this process is possible in several ways. Users of the web interface are invited to share their knowledge on specific entities by adding annotations with the Scholar's Notepad. Giving feedback on the application's usability and proposals for enhancements are most welcome as well.

Libraries, particularly CERL members, may contribute their authority files – together with or separate from their contributions to HPB. This may happen as a one-time contribution or on a regular update basis. Currently the CERL Thesaurus receives the weekly update of the German national authority file on personal names, the *Personennamendatei*. It has been agreed that in return, the Deutsche Nationalbibliothek can demand a data extraction to make use of CERL's editing work for the improvement of their own files.

Provenance information that is concealed in library catalogues could gain a higher visibility if it was accessible from a central point, together with provenance information from other libraries.

The CERL Thesaurus is meant as a public service to libraries, scholars and all persons with an interest in the European history of printing, rare book collection management, antiquarian book trading or similar. The database is freely available from the World Wide Web without any restriction. A library or an institution that provides a database related to the period up to approximately 1830 may in principle agree with CERL to make use of the CERL Thesaurus for applying Assisted Searching to its

own database. Providers of other information resources such as full-text databases or online biographies might also benefit from allowing access to their data through the CERL Thesaurus, since this means an increase of the visibility within the international scholarly community.

We have seen, how the CERL Thesaurus has been continously emerging from an auxiliary tool that substitutes authority control in the Hand Press Book database to a central access point for information relating to the European printed heritage. The CERL Thesaurus provides access to information in several dimensions: there is individual information on names, biographical details, occupation or other entity-specific information. On another level, existing relationships between the described entites are made visible. The third level comprises linking to external resources and providing access to provenance information, which is present in a number of library catalogues. On top of all these it provides access to the insights and non-codified knowledge of fellow scholars who are engaged in research on historic printing. At the same time it also functions as a means to share one's own ideas on certain printers, authors, publishers etc. of that time period.

The aggregation of these different approaches exceeds the classical concepts of authority files and makes the CERL Thesaurus a unique resource for navigating the domain of research in the history of books and printing and in connection with these, the history of European culture.

NOTES

1. ISO 2788 defines a thesaurus as 'the vocabulary of a controlled indexing language [...] formally organized so that the a priori relationships between concepts (for example as broader and narrower) are made explicit.' *Documentation – principes directeurs pour l'établissement et le développement de thésaurus monolingues = Documentation – guidelines for the establishment and development of monolingual thesauri*, élaborée par le comité technique ISO/TC 46, ISO, 2788, 2nd ed. (Genève: Organisation Internationale de Normalisation, 1986).
2. Charles A. Cutter, *Rules for a dictionary catalogue*, Special report on public libraries, 4th ed., rew. (Washington: Government Pr. Office, 1904; repr. London: Library Association, 1953), p. 37. Cf. *Statement of Principles*, adopted by the International Conference on Cataloguing Principles, Paris, October, 1961, annotated ed. by A. H. Chaplin, provisional ed. (Sevenoaks, Kent: IFLA Secretariat, 1967), c. 5.2, 6.1 etc.
3. A comprehensive list of all CERL Thesaurus source files is available from: *List of authority files included in the CERL Thesaurus* (2006) <http://cerl.sub.uni-goettingen.de/ct/help/AuthorityFilesInTheCT.html>

4. *Unimarc/Authorities: universal format for authorities*, recommended by the IFLA Steering Group on a Unimarc Format for Authorities; approved by the Standing Committees of the IFLA Sections on Cataloguing and Information Technology, 2nd rev. and enlarged ed., UBCIM publications: new series, 22 (München: Saur, 2001).
5. The Unimarc authority format permits the repetition of a heading field only if the heading is also used in an alternative script. (ibid.*, p. 68*).
6. A similar system of assisted searching will be implemented when the HPB Database is migrated to OCLC's server in mid-2007.
7. Given the fact that the municipality of Budapest was formed from the cities Buda, Pest and Óbuda in 1873 only, it is not at all astonishing that there are only a few records holding this name as imprint place. On the contrary the fact that imprints with this place name are found in a database is (except for those record that went into the file by mistake) is a manifestation of a particular principle of authority control applied by some libraries.
8. The CERL Portal is available from *http://cerl.epc.ub.uu.se/sportal/*.
9. *Functional requirements for authority records: a conceptual model*, IFLA UBCIM Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), draft 2005-06-15 ([S.l.], 2005) <http://www.ifla.org/VII/d4/FRANAR-Conceptual-M-Draft-e.pdf> (p. 4)
10. Cf. Christel Hengel, Normdaten und Metadaten: die Idee eines Internationalen Authority File, *Zeitschrift für Bibliothekswesen und Bibliographie*, 50 (2003), 210–214 (p. 212).
11. Ibid.
12. The implementation of Scholar's Notepad within the CERL Thesaurus web interface is still a prototype, but the release of a full version is planned for 2007.

Old books in new libraries: democratisation of access or a digital divide

KRISTIAN JENSEN

A digitalizálás elterjedésével eddig elképzelhetetlenül nagy számú olvasó férhet hozzá a régi szövegekhez. Egyik gyűjteményből a másikba kerülve a könyvek új kontextust és új jelentéseket kaptak, míg a régi jelentéseket és funkciókat egyre nehezebb azonosítani. Ehhez hasonlóan, az elektronikus szöveggé alakítás, amely minden fizikai közegtől megfosztja a könyveket a régi szövegeket új jelentésekkel ruházza fel, a régi jelentések pedig elhomályosulnak.

Ez a példa nélküli szellemi befektetés a társadalom és a magánemberek részéről a régi szövegek elektronikus hozzáférésének megteremtése érdekében mind a könyvtárosok, mind a tudósok számára hatalmas lehetőséget jelent. Együtt kell biztosítaniuk, hogy ezek az új források folyamatosan újrafogalmazódjanak, hogy az olvasók új generációi új keresési stratégiákat alakíthassanak ki. Ehhez az szükséges, hogy a régi könyvek fizikai valóját és tartalmát egymástól elválaszthatatlanul kezeljék, beleértve ebbe a tulajdonosi, használói, könyvelőállításra vonatkozó és kereskedelmi adatokat is.

Cease, ignorant sophists, to find pleasure [in museums]; the ruins caused by time inspire respect; those which are the effect of barbarism excite horror . . . Tell us no more that works of Art are observed in these depots. The matter to be sure is there; but where is the train of tender, deep, melancholy, sublime or affecting emotions, which once surrounded them? Have you been able to transfer to your storehouses that union of ideas and associations which spread so powerful an interest over the work of the chisel or the pencil? These objects have lost all their effect in losing their motive.

. . .

To displace all these monuments, collect the scattered fragments, class methodically the remains, and compose from such an assemblage a practical course of modern chronology, is for a nation in existence to place itself in the state of a nation that is extinct; it is a living being attending its own funeral; it is killing Art in order to write its history; it is not writing its history but its epitaph.[1]

These are the words of Antoine Chrysostôme Quatremère de Quincy (1755–1849), here in the English translation from 1821 of his work *Considérations morales sur la destination des ouvrages de l'art*. He is perhaps better known for his letters to General Miranda, which he wrote during the Italian campaigns of the French revolutionary armies in the 1790s, already then arguing against the spoliation of art.[2]

No-one has expressed more forcefully than he the importance of the original cultural context of works of art for our ability to understand them. The *Considérations morales* were written in response to the creation of the big public museums in Paris which first grew out of the Revolution. Objects were gathered together and displayed, deliberately deprived of the context of the repugnant oppressive past which they had originally been made to glorify, now placed in a chronological order to show the progress of mankind. These museums continued, often with nationalistic aspirations, after the restoration of the monarchy. Quatremère de Quincy took equal, if not more, exception to the great but private collections which were being created in Britain at the same time. They too deprived the objects of their context, and in addition made them less accessible: 'After Italy, no country is richer in antiquities than England … And what is the result? Riches are scattered in all castles. We have to travel through all counties, hundreds of leagues, to see these partial collections. I know nothing less useful for Europe, and even for the arts in England themselves, than the objects of this nature possessed by England.'[3]

There is a striking analogy between the fate of objects of art and that of early printed books, especially of incunabula, from say 1789 to 1815. They too were collected from their former owners, to be inserted into new interpretative contexts. In Great Britain they mainly entered the vast collections of a few extremely rich men. Given that the British Museum bought little in those years, the only British institutional purchaser of any significance was the Bodleian Library.[4] In France and Bavaria these were the years which saw the formation of large public collections of early printed books.

Collections of books and of art were physically affected by the same political events so, not surprisingly, the intellectual treatment of books transferred to new institutions bears comparison with that undergone by art objects. This is perhaps especially true of collections of incunabula. Librarians did to their newly acquired books what Quatremère de Quincy denounced when it came to art. They set out to 'collect the scattered fragments, class methodically the remains, and compose from such an

assemblage a practical course of modern chronology.' These were the years in which the Royal Library in Munich began classifying its immense incunable collection, newly acquired from dissolved religious houses, chronologically by year of printing,[5] inserting them into a sequence which reflected early nineteenth-century interests, but obscured their earlier functions and meanings.

This reinterpretation was not restricted to the way in which books were organised on the shelves. Old books underwent radical physical transformations which had the same effect of inserting them into new aesthetic and interpretative strategies. They would be disbound by librarians, or by book sellers, in order to separate individual editions so that they could be 'classed methodically'; they would rebound in a modern style to form a coherent part of the visual presentation of the owners' aspirations with their collections.

George III and his librarian created a vast and intellectually ambitious collection, since 1828 an important part of the collection of the British Museum and now of the British Library. His incunabula were part of an Enlightenment collection – and that inevitably was given physical expression by the way they looked on his shelves in Buckingham House. We know that he bought many books which were still in their fifteenth-century bindings. For instance he acquired through the London bookseller G. and W. Nicol numerous lots from a sale at which the books were specifically stated to be 'in the finest preservation and in the original monastic bindings.'[6] They were all disbound and with a few exceptions separated into what was felt to be their constituent parts. They were of course rebound in styles which were deemed appropriate to their specific status within the royal collection of books. 'These objects have lost all their effect in losing their motive' were the words of Quatremère de Quincy. The same is true of these books, although of course, seen from our perspective, they have acquired another effect from their new motive.

While not unparalleled in the nineteenth century, Quatremère de Quincy's sensitivity to the importance of context for the true appreciation of art is remarkable. It may seem the more astonishing, therefore, that this sensitivity did not at all extend to books. He himself created a significant collection of valuable early printed books in the wake of the disruptions which followed the French Revolution, evidently with no concern about detaching them from the cultural spheres in which they had been produced, used and owned. Like so many other collectors of the grand eighteenth-century taste, Quatremère de Quincy benefited from the rapid

changes in institutional structures which took place in the late eighteenth century, throughout the Habsburg empire, elsewhere in Italy and Germany and, of course, in France. From the institutions where books had been kept, and sometimes used, for centuries they came to his – private – collection in France, where they remained as a group until 1858, some ten years after his death, when they were sold to the Königliche Bibliothek in Munich, whose successor organisation is the Bayerische Staatsbibliothek.[7]

However, for my purpose it is important to understand that, in his own terms, he was entirely consistent. He divided objects into two grand categories. One was art, the product of genius. This category of objects was so far removed from other things that the opposing category was very large, included both 'useful productions' and 'the frivolities of luxury'. Several intellectual misconceptions had, in his view, led to the creation of museums: 'But the essential error', he said, 'is that of assimilating the art of genius to those of laborious ingenuity.' And here we have the explanation why he treated books in exactly the way which he had rejected so forcefully for objects of art. Books evidently fall into the category of those objects which were the creation of 'laborious ingenuity', whether they were useful or luxurious.

With its emphasis on context, Quatremère de Quincy's view of art as the product of genius owes more to Winckelmann than to romanticism; his indifference to books as objects fits into this Enlightenment background. Books provided physical support for texts with their meanings, which transcended any specific physical object; the materiality of the book was required but each instance was accidental to the text. In this they were completely different from works of sculpture and painting, although they might contain works of art which have at least as high an artistic or intellectual value in the view of the world of learning. The authors of the article on 'Imprimerie' in the *Encyclopédie* emphasised the absolute preeminence of the text over its vehicle:

> The advantage which authors have over these grand masters [of painting and sculpture] derives from our ability to multiply their writings, by printing, renewing for ever the number of copies which one desires, without the value of the copies being inferior to the originals. What would one not pay for a Virgil, a Horace, a Homer, a Cicero, a Plato, an Aristotle, a Pliny if their works were confined in one place or in the hands of one person, as may be the case for a statue, a building or a painting?[8]

Old books in new libraries

Books are – or at least they were – nonetheless objects, and like any object therefore produced within highly specific economic and intellectual frameworks, traded and owned as objects which have specific but changeable cultural meanings. In recent years the history of reading and of the ownership of books has emphasised the integrality of text and the physical nature of the book: a text produced in duodecimo has not got the same function and meaning as the same text produced in a luxurious large quarto. And the history of books has increasingly focused on the economy of production, based both on archives and on a physical examination of the books produced. The two related disciplines have seen the book as text and object jointly part of the same system of communication.

But with mass digitisation of historic books a fundamentally new situation now challenges this integration of text and the object which carries it, as a unified source for the historian's engagement with human communication.

It is in my view urgent for us who engage with the evidence presented by historical books, as groups of objects and as individual objects, to understand the function of the information which we provide, within the field of tension created by the two functions of books, between the many meanings of books as texts and as objects. The reason for this urgency arises from the current state of this very tension. Even more strongly than the *Encyclopédistes* the digital environment asserts texts as having a separate existence which transcends any physical carrier.

A very high proportion of British early printed books are now available on-line in digital form, detached from all their previous material existence. Books printed before 1701 are extremely comprehensively covered by *EEBO*, Early English Books Online, published by Proquest, providing digital images based on the microfilms which over many decades have been created from specific copies and sold commercially. Eighteenth-Century Collections Online, *ECCO*, published by Gale Thompson, likewise based on previous microfilm collections, covers the eighteenth century somewhat less comprehensively, but the collection is growing fast.

Whereas *EEBO*, overwhelmingly, presents the user with plain images of the digitised pages, *ECCO* offers the opportunity for very sophisticated full-text searches, on the basis of optical character recognition, and to some extent manual keyboarding of certain parts of the works. It is a magnificent resource and I use it very much for my own work.

Microsoft's project with the British Library and a number of North American libraries, along with Google's project, which includes the

Bodleian Library and the Bayerische Staatsbibliothek, are already now opening the vast nineteenth-century book production to consultation and search as electronic text on the internet.

Gallica, the magnificent resource of the Bibliothèque nationale de France, is for the moment not searchable as text, but I should be very surprised if that were not soon to change. And it has the great advantage that it is created with public funding and is freely available world wide, through the catalogue of the BnF, in the true spirit of the open nature of the internet. Again I am an avid user of Gallica.

There is in my mind no doubt that these electronic resources have improved access to texts and also, to my mind even more importantly, immeasurably broadened access to texts. People who could never have hoped to have any form of access to historic texts can now consult them with greater freedom and ease than could specialist scholars of previous generations, with their privileged access to physical collections.

All these resources present users with images of pages detached not only from the environments where they were produced, sold, bought, used and collected, but also detached from the physical objects themselves. The dissociation from context is taken to an even further point when users, as a result of a search for words or phrases, can gain access to specific passages, without having to locate them within the wider context of the book for which they were written. Texts identified like this are dissociated from the textual context, from the physicality of a specific copy and from physical reality which makes it clear that the copy was part of an edition. Thus, as currently configured, the digital format discourages users from appreciating the nature of all those aspects of book history which have most engaged scholars for the last decades. This is reinforced by the needs of the new technology. Digitisation and especially optical character recognition work best if it is done from clean copies, books which carry as few marks of use as at all possible.

Users of these resources are increasingly people who have no familiarity with any manual production process. The physical production of the objects which surround us is geographically as well as conceptually remote and alien. Manufacture increasingly does not take place in Europe. Few students even have a relative who takes part in the manual creation of commercial objects. Gone are the days when one could take one's radio apart and repair it. And gone are the days when students, avid creators and consumers of texts, for instance on mobile telephones, associated the multiplication of texts with any mechanical process.

At the same time as the economic, social and physical conditions for the production of books is ever more distant from the daily experience of students, they are increasingly not confronted with historical books as objects which might lead them naturally to be inquisitive about the complex social conditions within which they were created. At a conference held at the British Library, on 30 October 2006, we launched on the British Library's website the *ESTC*, the English Short Title Catalogue as a free public access web resource, no passwords, no charges, a major improvement in the provision of information about the British historic book. One of the papers at the conference was given by Ian Gadd, from the University of Bath Spa, a new university which has next to no historical books in its collections. His talk was about the new type of guide to material bibliography which he is preparing for students who are only familiar with hand-press printed books from electronic facsimiles, and who have difficulty in interpreting what they see on their screens as the representation of marks on paper made by physical means.

Another London conference took place on 4 November 2006, organised jointly by the Royal Historical Society and The Bibliographical Society, to encourage greater integration between the two disciplines. One bibliographer talked about the methodology of her work, referring exclusively to the digital facsimiles which she had used. As currently configured, the very format of digital facsimiles of historic books discourages even an expert user from engaging with the text as an integral part of a physical object which is the product of specific economic and cultural considerations.

If even an expert feels that she can bypass the physical books in her work, it is easy to imagine how hard it will be for students to overcome the barriers put in place between them and the physical object, by the disproportionate ease with which they can get access to a vast, but highly selective body of information. This distance will be compounded by the growing difference between the communication systems of their own world and those of the world which they study.

At the same time the resources of on-line digital facsimiles offer us tremendous opportunities to reach a geographically and educationally diverse audience which never before had access to these texts. The bibliographer in question, and most especially her students, have access to historical texts in a way which is completely unprecedented. More students now engage with texts produced in the hand-press period than ever before.

For me this contradiction highlights the importance of our ability to place provenance studies as an important tool for a number of specific disciplines, if I may understand provenance, the theme of this conference, broadly as the information which individual books contain about themselves as manufactured objects, merchandise and as conveyors of texts. By addressing the issue of the increasingly alien nature of the book as a physical object resulting from a manual process of manufacturing, provenances studies may have an important role to play in exploiting the opportunities now offered to us as people with a professional concern both for contemporary communication and for historical books.

To achieve this, we need to clarify for ourselves what intellectual purposes we believe can be served by the information on distribution, ownership and use which we provide, and we need to enable others to use the information easily, for those purposes, as well as for the multiplicity of purposes which we have not thought of ourselves.

In a Hegelian view of the world a discipline, Wissenschaft, can only be considered as such if its subject matter in itself constitutes a unity. It is the objectively existing unity of the subject matter of a science which enables us to address it with a coherent intellectual approach, reflected in a unity of methodology.

It is immediately obvious to anyone who has engaged with the practice of establishing the provenance of a book that, with such a definition of Wissenschaft, provenance studies do not stand a chance of qualifying. Fortunately, there is no reason why anything has to qualify as a science in Hegelian terms for it to be either important or interesting. But a consideration of the status of provenance information as a means not an end in themselves may make it clear that there is a necessity to formulate explicitly how information about provenance relates to the disciplines within which such information in used.

It can be extremely difficult to establish the identity of an early owner of a book. It requires a wide range of skills, not merely palaeographic and linguistic, if that were not enough. One will need knowledge of the history of institutions across Europe, how they structured their collections, a knowledge of heraldry, geography, art history, binding decoration, binding structures, paper history and much more. One needs to know the history of the book trade, as well as the reference sources for biographical information of professions, religious orders, and authors. And most of all perhaps one needs the imagination to combine these skills in practice and to judge the various types of evidence and their plausibility. So I am not

underestimating either the difficulty of much provenance work, nor the immense gratification which one derives from finally identifying an owner whose marks for a long time defied all attempts at interpretation. But there is a distinction between data and information, as one might put it in the language of modern information management, or between knowledge and understanding, as our medieval and renaissance Aristotelian predecessors might have said.

To establish how provenance information can reach beyond the accumulation of facts to become useful within a range of different disciplines, we must engage with researchers who work in all these disciplines. We need to ensure that our work continues to assist them to enrich their understanding, by providing appropriate access to the whole range of information held in our collections, far beyond the images provided on-line.

I suppose we can roughly divide these disciplines into three although, in many of the most fruitful studies where provenance data is used, the three aspects are studied as part of one unified system of communication:

1 Those which are concerned with texts and their transmissions. They may take a specific text, genre or author as their starting point.

2 Those which are concerned with the economy and structure of the book trade. They make take the production of a particular place or press as a their starting point

3 Those which are concerned with the readers of books, who they were, where they lived and but also what they used their books for and how they used them. They may take the books owned by a particular person, or group of persons, however the group may be constituted, socially, geographically or chronologically.

It is worth bearing in mind that most provenance information relates only to one aspect of my third category, biography. Provenance indexes are typically organised by name of a private or institutional owner.

I will give you three examples from my own studies of how work on texts and their transmission can benefit from being studied with provenance in mind, drawing on considerations falling under all three points. I have worked on three different types of texts produced in the fifteenth century, Bibles, Latin grammars and certain classical texts, seeking to understand the complex relations between the use of texts, the generation and modification of texts, and the economics of their production and their distribution.

The economic structure of production of books in Venice has been well studied especially in so far as some of the major producers and their consortia are concerned. The presence of Venetian books in many North European collections has also been noted in numerous provenance indexes. Much less work has been done on analysing what books were exported to where, and to whom. The knowledge of numerous individual instances of export has not been linked up with an analysis of texts, genres or even geographical distribution within present European countries. In my article on the production of Bibles in the fifteenth century I have tried also to place the Venetian export of Bibles in the context of Venice as a producer of luxury goods for Europe and to see how the intellectual need for specific texts and the commercial need to compete in an international market place worked together to make that which one might think of as a one single text, the Bible, into highly diverse corpus of texts.[9] My conclusions relied in equal part on a detailed examination of the texts and on the examination of the ownership and distribution of thousands of copies of Bibles printed in the fifteenth century.

The absence of such analyses has perhaps made it more difficult to gain a proper view of how distribution towards Northern Europe from other Italian printing centres differed from that of Venice. The distribution of books from Rome, whose economy in no small part relied on its status within the church, seems to have been very different from that of Venice for instance, but this has never been studied in detail.

It is well known that the new Latin of Italian humanists had a big impact on Northern Europe. In several articles on Latin grammars in the fifteenth century I have looked at how that knowledge was actually spread, how Humanist Italian grammars of Latin made their way into German schools and reached the clear conclusion that Italian grammar books made their impact on Germany in two different ways.[10] In southern Germany, not surprisingly, they were exported across the Alps. In his Latin grammar of 1508 Aldus Manutius, the publisher, used this as an example of the gerund: *Libros meos onerandos multis mulis a mulione tuo mittam in Germaniam*; 'I will send to Germany my books which are to be packed on to many mules by your muleteer'. Numerous Latin grammars printed in Italy can be found whose earliest owner can be shown to have been have been based in southern Germany. Here schools and teachers would seem to have relied on Italian importation, probably because it was not viable financially to produce local competing editions, although there is a clear indication in a Venetian book produced for the Viennese market, that books produced

for Italian schoolboys were not ideal for teaching Latin in German class rooms. However, it turns out that the situation was completely different elsewhere in Germany. In the west and the north we find that a number of Italian Latin grammars were reprinted locally, often with modification to suit a German readership. Hardly any surviving copies of Humanist Latin grammars printed in Italy can be shown to have had a fifteenth-century owner from the Rhineland or the north and west of Germany. In terms of the reception of the modern Latin language Germany thus fell into at least two quite distinct areas – and that can only be seen if you examine the distribution of locally printed texts jointly with the distribution of imported texts, on the basis of provenance information.

My third example relates to the use and distribution of editions of classical authors. The dense marginal and interlinear notes in Leipzig imprints are well known, but have rarely if ever been studied for their contents. I have done some work on marginal notes by anonymous students in Leipzig in their editions of certain classical texts and will give you just one brief part of the argument I built up around these annotated copies. Notes in several copies of the same text could be shown to derive from the lectures of individual professors and it could be shown that the lectures depended on books printed in Venice, owned by the professors. By studying the marginal notes and the relationship between locally produced books and imported books we can gain an insight into the way Leipzig teachers engaged with Italian humanist texts, and how they communicated the new learning in the class rooms. We can also gain an insight into the textual and functional interaction between a local book production and imported books.

My three examples of studies of the fifteenth-century book also indicate that although different researchers may ask different questions, the objects which we study were not designed to answer any of them. In the case of the Leipzig humanists I began with a question relating to their use of texts – and only got an answer when I engaged with the nature of the trade between Venice and Leipzig and the commercial viability of a local book production in the face of the pressure of goods imported from Venice. So questions about textual transmission may also be answered by examining evidence provided by individual volumes about the economics of the trade and the distribution of books, by examining decoration, bindings, and by studying the ownership and the use of books.

The objects with which we deal contain evidence of many different natures, and because they are embodied in one object, or group of objects,

they cannot be properly understood without an engagement with the object as a whole – including its texts. The very nature of the evidence contained in a physical object with a complex history tends to reinforce the need for an interdisciplinary approach. The information which we extract from the books must, to make full sense, be interpreted as part of that whole complex of data which the book presents. It is a much more difficult choice to engage with this type of evidence as a unity, if the text which you first read is detached from any physical reality. While the book as a complex object encourages interdisciplinarity, it is far from clear that the same is true of digital facsimiles.

Scholars will work on the materials which are made available to them. Funding structures and power structures will, now as in the past, determine what those materials are. The vast amounts which are already available in electronic form enable certain research strategies but will make others less obvious.

The history of the printed book has by book historians increasingly been understood to be an international history. This has been strongly emphasised in Britain where the recent multi-volume history of books was not the history of the British book, but the *History of the Book in Britain*.[11] Mass digitisation programmes are most often funded and undertaken within national frameworks.

Let me give you a few examples of the impact of the separation of information nation by nation. Even if we wished to focus on the history of our national countries we would tell a very partial story if we only were to look at the books which will be included in more or less nationally based digitisation schemes. *La Bible des poètes*, a version of Ovid's *Metamorphoses*, is a very French book, but the British Library's copy[12] could hardly be more English: it was decorated for Henry VII, king of England. Yet it will be out of scope for most digitisation programmes with a clear British focus. The *Pragmatica sanctio* of 7 July 1438[13] is a text which relates to a matter of specific French importance and for obvious reasons all fifteenth-century editions were produced in France. So it is a very French book, but the copy now in Bibliothèque municipale in Lyon is an eminently Italian book with the arms of Francesco Todeschini-Piccolomini, as cardinal, the Piccolomini who was to become pope for a few days in 1503. The decoration is also Italian – even without the coat of arms the book could have told us something about where this Lyonnais book was used. It is not likely to be included in any Italian digitisation programme. There are many early German books in the Bibliothèque nationale de France, the

results of confiscations and purchases during the last two centuries. But many German books have early French connections, for instance a copy of the *Institutiones* printed in Mainz in 1468, with the arms of a French abbot who was a member of the Roquette family. But this copy of a German book with French importance is now in London, in the British Library.[14] A somewhat more unusual example is the British Library's copy of Alexander de Ales printed in Nuremberg by Anton Koberger.[15] Its typography and its layout reveal it as a typically German book, but it has North Italian decoration and is in a North Italian binding. Books printed in Germany of relevance for French book history but found in the British Library . . . the possible permutations are as numerous and as complex as European history itself.

Printed books crossed continents, and so do digital facsimiles, but our administrative structures and the way in which we get financed, nation by nation and institution by institution, will affect what gets digitised and what gets left out. Perhaps even more significantly, they will affect how the material is made available.

These considerations take me back to the important challenges which we face in the new digital environment. It is beyond doubt an unprecedented opportunity for the study of the past that early texts can now be read more easily and that they can be read by many more people than ever before. Simultaneously it is no small challenge for librarians to design strategies, with teachers and researchers, to enable a new generation of readers to appreciate that the electronic images which they see, have been abstracted from the physical evidence for the social, intellectual and economic realities which were the conditions which made their creation possible. We must enable them to find out that there is more information and different types of information to be had about the texts, their uses, and the world in which they were produced, than what is presented electronically in digital images.

The answers to this challenge have to be sought in the very technologies which our users will choose to gain access to data. We need to make our information as readily available as we can within the search strategies which our users deploy. We cannot present in electronic form all the evidence which our institutions hold, but we need to make it both possible and easy for our users to find out that the evidence is there, even when they do not know that they are looking for it.

CERL could play a role here as an international body, reaching beyond the restraints of individual institutions and individual nations, to create

the over-arching view which will enable the rich resource of digital facsimiles of historic books to be used in the even richer context of other electronically available information on physical objects, eventually enabling users to identify questions which will require the exploration of the physical objects themselves.

If CERL can develop into a nodal point which facilitates intelligent contact between resources which funding streams have separated, it may also be able to go facilitate an approach to our information about the shared European history which mines deeper into our joint resources, which combines information of a disparate character held in different environments, by enabling user-generated and user-interactive technologies. Wiki technology is currently the best known example of these, but it will probably soon be superseded by even more flexible approaches which enable users to bring together information which is now held separately. Enabling such technology to be used in the same environments as our digital collections might help bring the information which we now hold in various forms into the environment where users increasingly expect to find it.

If we do not grasp the opportunity posed by the unprecedented expansion in access to texts, and fail to enable our teaching and researching colleagues to go beyond digital facsimiles, to use them as a marvellous gateway to further information, the study of the disembodied, decontextual historical text may be the future for all but a handful of privileged researchers. It would then be our responsibility if mass-digitisation fell short of its potential of being a great force for widening access. We could instead be seen as colluding in the creation of deeply split research environment where a small, exclusive group of researchers could understand the original material, whereas others would only be encouraged to engage with that with which has been pre-selected for them, a very partial view of the rich evidence which we store.

If a future Quatremère de Quincy were to ask us: Have you been able to transfer to your storehouses that union of ideas and associations which spread so powerful an interest over the work of the printing press? I for one would be unhappy if the answer were to be 'These objects have lost all their effect in losing their motive'. Nor is it necessary: it is our responsibility to see to it that digital information about our past becomes part of a democratisation of access to information not part of a digital divide.

NOTES

1. *The destination of works of art and the use to which they are applied. Considered with regard to their influence on the genius and taste of artists and the sentiment of amateurs*, translated from the French by Henry Thomson, R. A. (London, John Murray, Ablemarle Street, 1821), p. 55–57. The first French edition was *Considérations morales sur la destination des ouvrages de l'art, ou de l'influence de leur emploi sur le génie et le goût de ceux qui les produisent ou qui les jugent, et sur le sentiment de ceux qui en jouissent et en reçoivent les impressions* (Paris, impr. de Crapelet, 1815).
2. *Lettres sur le préjudice qu'occasionneroient aux arts et à la science le déplacement des monumens de l'art de l'Italie, le démembrement de ses Écoles, et la spoliation de ses collections, galeries, musées, etc. Par A. Q.* (Paris, an IV [1796]).
3. Here quoted from Quatemère de Quincy, *Considérations morales sur la destination des ouvrages de l'art Suivi de Lettres sur le préjudice qu'occassionneraient aux arts et à la science le déplacement des monuments de l'art de l'Italie, le démembrement de ses Écoles et la spoliation des ses collections, galléries, musées etc. [1796]*, ed. Édoard Pommier (Paris, 1989), 5th letter to general Miranda, p. 227: 'Après l'Italie, il n'est aucun pays plus riche en antiques, que l'Angleterre … Qu'en résulte-t-il? Des richesses sont éparses dans tous les châteaux; il nous faut aller dans tous les comtés, faire plusieurs centaines de lieues pour voir ces recueils partiels: aussi, je ne connois rien de moins utile à l'Europe et aux arts même en Angleterre, que ce que l'Angleterre possède en ce genre.'
4. See *A catalogue of books printed in the fifteenth century now in the Bodleian Library*, 6 volumes (Oxford, 2005), the result of a project which I began in 1990, and which was published last year. As well as detailed textual analysis it provides highly detailed copy specific information on decoration, manuscript notes and provenance information from the earliest owners down to the price paid by the Bodleian Library when the item was acquired, giving a very full price series from 1789 to 1900. The provenance index provides detailed information about owners, institutions, their collections and their dispersals, and details of auction sales and of booksellers and the catalogues of their sales.
5. On the introduction of this classifiction system in 1814, see Bettina Wagner, 'Von der Klosterbibliothek zum Gesamtkatalog der Wiegendruke', *Gutenberg-Jahrbuch* (2006), 168–178, here at p. 174.
6. *A catalogue of a valuable collection of books printed in the fifteenth century, consigned from abroad containing specimens of most of the early printers; in the finest preservation and in the original monastic bindings* (London, Leigh and Sotheby, 15 June 1799).
7. They only kept part of the collection, however, selling those which were considered duplicates at a number of auctions held in Paris in 1859: *Bibliothèque Quatremère. Catalogue d'une collection de livres précieux et importants provenant pour la pluspart de la bibliothèque de feu M. Étienne Quatremère de l'Institut …* rédigé par C. H[alm], (Paris, A. Franck: 27 November 1858 – 7 November Feb. 1859). Students of provenance beware – the sale also contains duplicates from

the Königliche Bibliothek in Munich which came from other sources, as was clearly established by Bettina Wagner, now at the Bayerische Staatsbibliothek, in her work on Bavarian provenances for the Bodleian Library incunable catalogue, see note 4 above.

8. *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers, par une société des gens de lettres*, vol. 8 (à Neufchâtel, 1765), p. 609: 'L'avantage que les auteurs ont sur ces grands maîtres, vient de ce qu'on peut multiplier leurs écrits, en tirer, en renouveler sans-cesse le nombre d'exemplaires qu'on désire, sans que les copies le cèdent en valeur aux originaux. Que ne payeroit-on pas d'un Virgile, d'un Horace, d'un Homère, d'un Cicéron, d'un Platon, d'un Aristote, d'un Pline, si leurs ouvrages étoient confinés dans un seul lieu, ou entre les mains d'une personne, comme peut l'être une statue, un édifice, un tableau?'
9. 'Printing the Bible in the Fifteenth Century: Devotion, Philology and Commerce', in *Incunabula and their readers: Printing, selling and using books in the fifteenth century*, edited by Kristian Jensen (London: The British Library, 2003), pp. 115–138.
10. In particular Kristian Jensen, 'The Latin grammar of Aldus Manutius and its fortuna', in *Aldus Manutius and Renaissance culture*, edited by David Zeidberg, Villa I Tatti, The Harvard University Centre for Italian Renaissance Studies, 15 (Florence, 1998), pp. 247–285 and Kristian Jensen, 'Elementary Latin grammars printed in the fifteenth century: Patterns of continuity and of change', in *Von Eleganz und Barbarei: Lateinische Grammatik und Stilistik in Renaissance und Barock*, edited by Wolfram Ax, Wolfenbütteler Forschungen, 95 (Wiesbaden, 2001), pp. 103–123.
11. The international aspect is perhaps most relevant for the volume covering the earliest period of printing, namely *History of the Book in Britain, Volume III, c. 1400 to 1557*, edited by J. B. Trapp and L. Hellinga (Cambridge, 1999).
12. Paris: Antoine Vérard, 1 Mar. 1493/94. The British Library, shelfmark IC.41148
13. Lyon: Guillaume Balsarin, 1488.
14. Mainz: Peter Schoeffer, 24 May 1468. The British Library, shelfmark, C.10.d.13.
15. Alexander de Ales, *Summa universae theologiae*. Printed Nuremberg: Anton Koberger, 1481–82. The British Library, shelfmark, IC.7251, IC.7238, IC.7272, and IC.7260.

List of Contributors

Anthony G. Curwen has worked as a part-time consultant on formats and other matters for the Consortium since 1993. He worked in public libraries in England (City of Westminster and Kent County) before moving to Aberystwyth where he taught at the College of Librarianship Wales for 22 years. MA in Library & Information Studies (University of London); Fellow of the Library Association (afterwards CILIP). For many years he was Hon. Secretary of the LA's Cataloguing & Indexing Group. He is an active Corresponding Member of the Permanent UNIMARC Committee.

Ádám Hegyi is the Project Manager of the Hungarian National Shared Catalogue – Early Prints (MOKKA-R) and assistant lecturer of the Library and Information Science Department of the University of Szeged.

Alexander Jahnke has been a staff member of the Data Conversion Group at the Göttingen State and University Library since 2003. He is responsible for the technical development of the CERL Thesaurus and related services provided by CERL. Further he teaches at the Faculty for Media, Information and Design at the University of Applied Sciences and Arts in Hannover. His subjects are Library Management Systems, Digital Libraries and Web publishing.

Kristian Jensen is head of the British and of the early printed collections at the British Library. He has worked and published extensively on historic printed books.

Gunilla Jonsson has recently retired as director of the Department for Collection Development & Documentation at the National Library of Sweden, where she worked for 35 years. She was deeply involved in retrospective bibliographic projects, e.g. as bibliographer and later project leader of *Svensk Bibliografi 1799–1829*, and was responsible for the Swedish official publications 1533–1833 project in the late 1990s. She was chair of the Swedish committee of cataloguing 2002–2006, and of the Cataloguing Section of IFLA 2003–2005. She has taught the history of the book at the Stockholm Graphic Institute and has written on cataloguing, typography and the history of the book. She has been involved

in CERL's work from its inception and served on its Executive Committee and Advisory Task Group.

István Monok received his Doctorate from József Attila University, Szeged, Hungary in 1983, and was awarded the C.Sc. in Literary Sciences, by the Hungarian Academy of Sciences in 1992. Having worked in the Central Library of József Attila University and in Schools of Library Science, Szeged, he was appointed Director General of the National Széchényi Library in 1999. He has taught and published widely on the history of culture and the book and was visiting Professor in 2000–2001 and 2006–2007 at the École Pratique des Hautes Études, Paris. He is currently President of the Committee of the History of European Civilization of the Hungarian Academy of Sciences. He was made Chevalier dans l'Ordre des Arts et des Lettres de la République Française and Cavaliere dell'Ordine al Merito della Reppublica Italiana in 2004.

Dr Béla Rozsondai read chemistry and taught physical chemistry at Roland Eötvös University, Budapest. Then he joined a research group of the Hungarian Academy of Sciences, studying the structures of molecules by gas-phase electron diffraction. Following his retirement in 1995, he undertook to solve software problems in the Library of the Academy, Department of Manuscripts and Rare Books, to devise the data structure for the special collections catalogues, applying modern standards for bibliographic description. For the last two years, he has been engaged in cataloguing the incunabula.

Dr Marianne Rozsondai read Latin philology and librarianship at Roland Eötvös University, Budapest (1963–68). Since 1969 she has been working at the Department of Manuscripts and Rare Books of the Library of the Hungarian Academy of Sciences; since 1997 head of the Department. Her interests are book history, the history of bookbinding, and the cataloguing of rare books; has published in these fields since 1971. Book: *A könyvkötés művészete*, 1983 (The Art of Bookbinding). Research scholarships in Wolfenbüttel, Leipzig, Oxford. Teaching activity at Roland Eötvös University and in conservator training. Member of the editorial board of *Magyar Könyvszemle* (Revue pour l'histoire du livre et de la presse) and the Internationale Gutenberg-Gesellschaft.

Róbert Zawiasa is the programmer of the Hungarian National Shared Catalogue – Early Prints (Mokka-r) and programmer of the University Library of the University of Szeged.