EROMM and the Registry of Digital Masters

Werner Schwartz

Niedersächsische Staats- und Universitätsbibliothek, 37070 Göttingen,
Germany, werner.schwartz@mail.uni-goettingen.de

Looking at Information Technology of Another Age

Since the invention of printing the spread of ideas and knowledge has been in the mind of authors when publishing their work in the new medium. The aim to make a profit from printing and distributing was naturally in the mind of producers and distributors, i.e., printers, publishers and book traders. To achieve this they needed to make their products known as widely as possible. And they used the new medium itself for this purpose.

In 1628 the book trader Grosius in Leipzig reprinted and updated his list of books on sale. On its last page his *Catalogus Librorum* has an entry which reads, 'Nicolai Reymeris Geodæsia Ranzoviana, von Landrechnen und Feldmessen/in 4'. This is one of the earliest works by Nicolaus Reimers, well-known astronomer and mathematician, which was printed in the same town in 1583. Continuous advertising of the book has helped to sell it widely. Today, under the title 'Geodæsia Ranzoviana' we are able to discover 10 print copies by using just one popular search tool (KVK[1]). A few more copies can surely be found after more thorough searching. Still what we can find using bibliographical databases or search engines will certainly only constitute a partial selection of the copies existing to our day.

The KVK returned 6 hits (=10 print copies and 1 microfilm) found at

- Bibliothèque nationale de France, Paris
- Bayerische Staatsbibliothek, München
- Herzogin-Anna-Amalia-Bibliothek, Weimar
- Herzog-August-Bibliothek, Wolfenbüttel
- Österreichische Nationalbibliothek, Vienna

Fig. 1:

Nicolai Reymens Geodæfia Ranzoviana , von Landrechnen vnd Feldmeffen/ in 4.

- Staatsbibliothek zu Berlin–Preußischer Kulturbesitz, Berlin
- Thüringer Universitäts- und Landesbibliothek, Jena
- Technische Informationsbibliothek und Universitätsbibliothek Hannover, Hannover Zentralbibliothek, Zürich

One copy was probably lost through bombing in 1944 (Berlin), another is a likely loss caused by fire in 2004 (Weimar).

Will Your Digitised Book Survive for 425 Years

Just as the book printed in 1583 is found in libraries in our day? To raise this question today may seem as premature as asking whether someone will advertise it 45 years after production. But without doubt, you and those funding your work are not making all this effort to digitise a work in order to lose the digital version after a few years' time. So there clearly are questions that need to be asked:

- What do you need to do to make your digital copy last? First of all you have to choose a format that will be susceptible to future attempts of refreshing without losing content and quality. In particular this requires bridging the obstacles posed by ceaselessly renewed features of hardware and software. Second, there must be a reliable digital repository, safe and capable of dealing with a large quantity of data. Third and most important, you will need considerable funding not only to produce your digital document but also to keep things running and alive in the long-term.
- Will your funding come forward and subsist? (Although I am talking about libraries of the public sector primarily, some aspects are true for the private sector as well.) Today it is easy to convince your organisation and funding bodies of the attractiveness of digitisation, espe-

cially if you do not tell them of future costs for the upkeep of the digital library.[2] But by now, we can already see that digitising is rapidly becoming something of normality. More and more people tend to take it for granted that someone is offering a beautifully produced version for access on the web. Funding bodies will understand that somebody has to pay for production and maintenance. However, it is becoming more difficult than before to start new projects. One of the questions usually asked is whether the product is not offered by someone else. Some even want to know, whether you have checked possible parallel projects aiming at doing the same thing. Finally there is the question about proceeds that can be expected from offering the product to the general audience. It all comes up to the question, 'Will it be worthwhile to digitise a book yourself?'

It is relatively easy to secure funding for a project that is limited in time and will show nice results soon. Libraries and other memory institutions, however, cannot be satisfied with short-term support alone. They have to provide access to their digital holdings for this *and* future generations.

This requires a step-by-step approach. It is important first to secure visibility of what you have produced already and point to its quality and user friendliness to get a good response from local and national audiences and users beyond your country's boundaries. The uniqueness of your selection of digitised works on the one hand and the smooth integration of your digital service with the local library and information system on the other should be demonstrated. You may point out the possibility of exchanging information through European and international networks as a way to strengthen your position as a provider of digital content.

All this will support your request for continuous funding as a precondition for sustained availability of digital content. A funding body can be assured that investment in your digitisation activity will not be lost.

How to Achieve Visibility and Why this is Indispensable

Publications not known to the relevant audience will rarely be used. Unused works tend to be neglected physically and are susceptible to getting lost. This is true for

library holdings of any kind. Thus, any effort to make a digitised work visible and accessible involves an aspect of preservation. Visibility in a way can be seen as indispensable for survival of the item.

The electronic catalogue in its modern technical setting is the best tool when it comes to integrating multiple functions and achieving the broadest possible visibility. This sounds rather old-fashioned after all the trendy tools that we are making use of today. But in fact, many of the systems that allow us to trace a publication on the internet with some reliability originate from bibliographic databases of some kind, most of them catalogues.

A digital library may offer the most sophisticated options for retrieval, such as full text searching of OCR'ed, digitised books. It may also have a superb layout and user-friendly tools for browsing and reading. However, you will be aware that the content of the digital library is nothing but a very minuscule portion of what the conventional library holds. The latter has the mass of material and it will take quite some time until digitisation will have covered only the most important works it contains.

The user, as a rule, is not just browsing for fun, but is looking for items he needs in his research or study. This is why he must start from the resource that offers the broadest information about the library's holdings. There he will discover that the item he is looking for is available in digitised format as well. A click should carry him to the item in the digital library, where he will be able to enjoy all the advanced features of this setting. To inform the user, you have to make sure that bibliographic reference to the digitised work is present in the OPAC. Easy access from here is just as important as is its quality as a digital object.

There are also more general benefits of recording digitised works in the library catalogue. Through the networked OPAC you will not only reach the local, but also the national and international audiences. The uniqueness of your digitised works will become transparent when the library catalogue's information is made available through union catalogues and search portals. Electronic catalogues are the dominant user interface when it comes to precise searching. A lot of other services of the local library and information system relate to it. So there is no reason why the digital library service should not do the same.

Exchanging catalogue information through European and international networks is a well-established practice. Information about digital content provided by your library or project will be carried along without any additional effort on your part. Intelligent use and integration of digital content from other providers can be supported by the OPAC. Increased usage of the digitised material will support any requests for funding the sustained availability of the digital library.

Many Digitised Works are Still Absent from Catalogues

In spite of modern technology, the content of European digital libraries is only partly accessible through catalogues. This is easily visible when searching EROMM, the *European Register of Microform and Digital Masters*.[3] This database holds 15,966 records of digitised books and periodicals (995 of the total) from European sources (October 2007). This is less than half of the estimated production to date. The total number of digital masters in EROMM comes to 46,248 because of the contribution by the American partners, RLG and OCLC.[4]

Although the European hesitation towards recording digitised items in library catalogues is not visible in America, there is still a certain delay in supplying catalogue information. The *Digital Library Federation*[5] observed that it is at present impossible to tell with certainty what has been digitised, what the technical features of digital masters are and who will take care of their long-term preservation. Experts called upon by the DLF proposed to create a *Registry of Digital Masters*[6] as a specialised resource, open to everyone. The criteria for what the RDM should record were defined, and OCLC created this bibliographic database, which is online since last year.

EROMM and OCLC Cooperating to Increase Coverage

There is no sense in creating a national resource for American digitisation activities alone. Indeed, libraries world-wide shall contribute to the RDM. In 2006, OCLC and EROMM signed an agreement of cooperation with LIBER as the primary signatory, which supports this initiative pledging to work for

cooperation throughout Europe.[7] The agreement provides for the mutual exchange of records of digitised works.[8] This is the continuation and expansion of the exchange of data that went on between RLG and EROMM for the last ten years. A few technical issues remain to be resolved, but RDM records are visible in the EROMM database already.

In future, networked implementations shall ensure real-time updating between the two nodes.

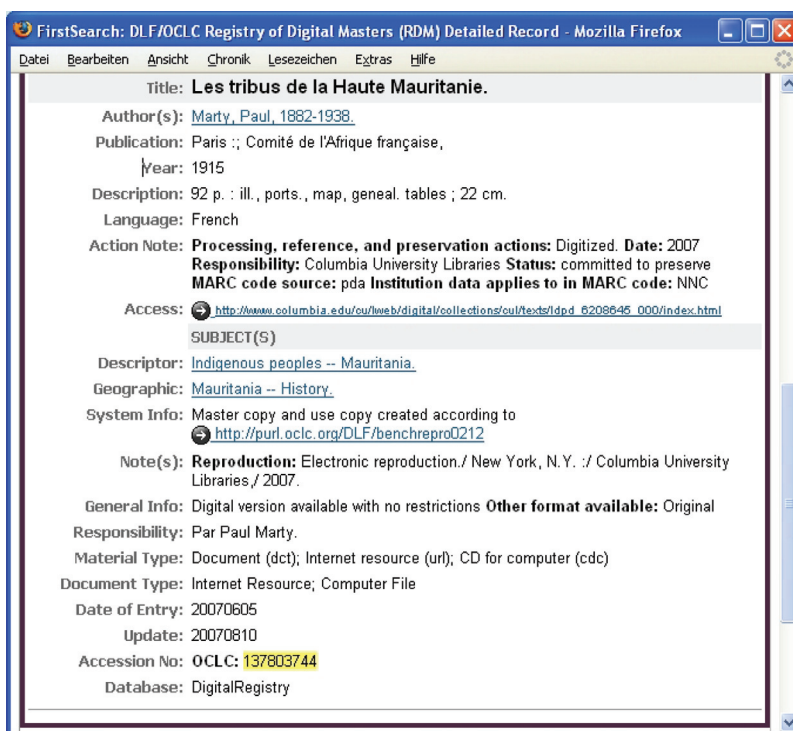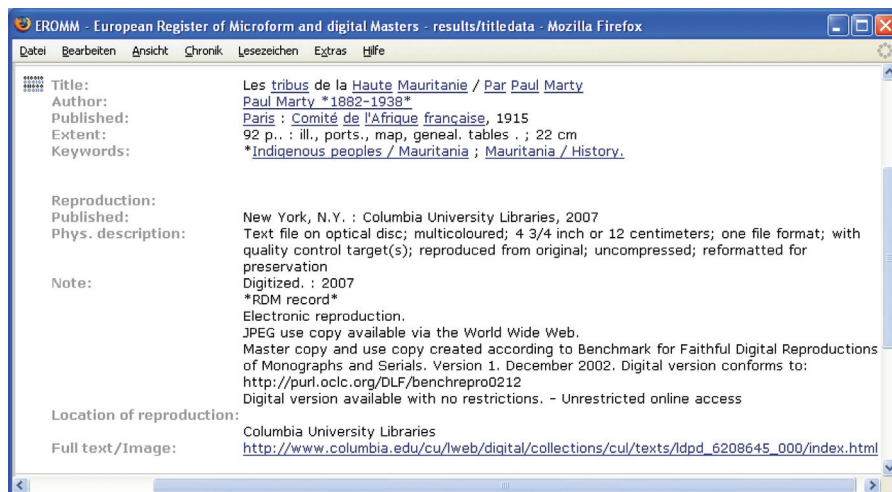*Fig. 2: Example of a record from the Registry of Digital Masters* (RDM).

Fig. 3: The same record as shown in EROMM.

How EROMM Collects European Records

At present, EROMM collects data through its 14 partners from twelve European countries. The EROMM partners in turn contribute their own records and those from library networks in their own country. In addition EROMM receives records from North and Latin America. Files are retrieved through ftp or through harvesting the file owners' systems.

The present scope of data harvesting in Europe is still insufficient, because it does not yet include all countries on the continent. EROMM partners are trying to collect records from their country to gather information from all digitisation projects. With the support of LIBER and organisations like the Consortium of European Research Libraries (CERL) EROMM is helping its member libraries to expand this network. In a growing number of cases and wherever this is feasible EROMM has established direct contact with projects

and libraries other than the country's official EROMM partner.[9] The main objective is to include updates as frequently as possible.

There are new technical options for RDM and EROMM alike. They allow access to remote systems without (at least in theory) requiring the file owner to make any arrangements other than to document by which protocol[10] the system can be searched and records be retrieved.

So why don't we – OCLC and EROMM – set up a search engine to retrieve information from library systems saving them the effort of collecting records and running a dedicated database? Indeed this would bring the great advantage of being up to date in a way that can never be achieved through the present setup with EROMM and RDM, which work as repositories of records extracted from other databases.

But while we agree that the technical option of cross-file searching is very attractive, it is too early to go that way for most of the library systems accessible today. The reason is quite simple: the records that describe digital masters lack the required uniformity and quality.[11] For more than a decade, EROMM together with relevant bodies has been working on improving the quality of cataloguing surrogates. Progress has been made, but so far, EROMM is still forced to adapt to the special features of internal format usage of every single library system. Only by doing this, we can be sure to understand,

- who has digitised a book,
- at what time,
- according to which standards, and
- who will care for its preservation.

In the present discussion about new technical possibilities for retrieving digital content, the overall focus is on user access. By far less attention is paid to the four basic questions above, which must be answered if sustained access is to be provided. The role played by cataloguing, persistent identifiers and library catalogues is still being disregarded by many.[12]

Catalogue Information Will Increase the Use of Digitised Works and their Chance to be Preserved

Once your library system has established contact to EROMM, your records will become visible in this international database. Through EROMM they will be carried to the RDM at OCLC, which is also open to the international audience. Both systems link back to your digital library so that the user will be directed to your local or national system. If there is no free access, your system will reply with information as to how access can be obtained through subscription, pay per document or other. Even in cases where the item is still protected by copyright or other restrictions, the user will be informed about this.

EROMM has a requesting facility in place that enables the general user as well as other digitisation projects to inquire about digitised works at the responsible library in Europe, even when the record's URL is deficient or absent. This facility is playing an important role not only in the context of international requesting of documents. Because of requests by users, it may alert a library to the fact that it owns a surrogate of a specific item. In fact it is not rare at all that some digital surrogates are getting no attention any-more within a library after the end of a digitisation programme. There is a clear danger that these will not receive the necessary attention for keeping them up and going. Requests coming via international systems can do a lot to bring the necessity of digital preservation back into focus.

Notes

[1] Karlsruher Virtueller Katalog (Karlsruhe Virtual Catalog), http://www.ubka.uni-karlsruhe.de/kvk/kvk/kvk_en.html

[2] The term 'digital library' is used here to designate all the various collections of digital content created by libraries or by digitisation projects.

[3] http://www.eromm.org/

[4] This figure includes 1,457 periodicals.

[5] http://www.diglib.org/

⁶ http://purl.oclc.org/DLF/collections/reg/OCLCservice. For background informa-tion go to http://www.diglib.org/collections/reg/digregguide200705.htm

⁷ At the invitation of OCLC PICA and the Digital Library Federation, LIBER has accepted responsibility for co-ordinating European activity in the collection of records about ongoing or completed digitisation projects. All such metadata records will be submitted through the EROMM database to the global Registry of Digital Masters (RDM).

⁸ At the beginning, the focus will be on digital masters, but other preservation surrogates such as microform masters are going to be included later.

⁹ This approach can be successful only if after preparatory work the regular harvesting of records can be configured to run more or less automatically. The staff working for EROMM is limited and cannot handle data deliveries that require adjustments for every new file.

¹⁰ Z39.50, SRU/SRW and the OAI-PMH are the protocols to be used here. Most library systems support at least one of them.

¹¹ The situation is better for microform masters. But even though defined bibliographic formats (UNIMARC and MARC21) as well as rules for cataloguing do exist, many have implemented them only in part or even not at all.

¹² This is true in particular when we look at digital libraries and the bibliographic information they offer in their own setting. Recent testing done by EROMM has shown that in many cases bibliographic informa-tion is insufficient to identify the digitised work with precision. On the other hand it is fragmented due to the needs of retrieval and browsing within the digital library. To identify the print edition of the digitised work, a back-up from the library catalogue would be needed, which for its part would require unambiguous referencing.