The DIEPER Project, DIgitised European PERiodicals[*]

by WERNER SCHWARTZ

BACKGROUND

Following technical developments digitising of printed information is being done on a large scale in many countries. Publishers and libraries are exploring a range of technical approaches for creating digital documents from books. In part this activity may be motivated by curiosity about the feasability of new electronic concepts. But in an increasing number of cases a very practical approach is underlying present projects of so-called retrospective digitising. I will concentrate on libraries here.

Today users request books, journals or photo-copies by the way of interlibrary loan or other document delivery services. Libraries with large holdings and with users well beyond the campus try to keep up with this ever increasing demand. However, this heavy use of library holdings is damaging books, which are out of print in most cases. In addition there is the threat of the brittle paper phenomenon. Although it is well known by now to every librarian and remedies have been sought intensely for more than two decades, it is still far from having been tackled.

Reformatting to microform has long been regarded as the best answer to this very real danger of ultimate loss of the world's literary and scientific heritage in printed form. Reformatting is to protect the original paper document from further damage by intensive use. While reducing direct use of the original a microfilmed work becomes available in service copies where-ever it might be requested. Indeed the longevity of microforms, as long as they are properly produced and handled, still gives the best perspective for long-term preservation.

But there are good reasons to use digitisation, not perhaps to replace microfilming but to add to its positive effects. Copying or lending out microforms to users beyond the campus in many cases is too slow to meet

actual needs, not to speak about the bureaucratic effort, that is required for placing a formal order. Only the user, who is in the library, may, by browsing the book, easily decide whether he needs to read it all, use certain parts or discard it. Should he be beyond visiting reach of the library he will normally have no choice but to request the whole publication or a copy of it before he may judge its value for his work.

Digitising can help to overcome these limitations if the digital document is made accessible in a local network or via the Internet. The user then will be able to have a look at the table of contents or browse pages of interest to him, before he goes on to request a hardcopy print of parts or all of the document or download it to his hard disk. In the case of a journal article he may even choose to read it all on his computer's monitor, to which the document is sent from a remote server.

It is precisely the latter point, that made the DIEPER consortium get together.[1] They believe, that at the present stage of technological development digitising periodical literature is of particular importance:

a) The transmission of single articles via the web, as image or as text, is economical even if the user should not have the latest hardware or the fastest gateway to the Internet.

b) Making periodicals available over the web, that are at the centre of users' interest will effectively reduce demand for the original paper version.

c) A library that has invested considerable effort to digitise such a periodical will care to preserve the digital version, because its loss could mean reducing revenue and replacing it would require renewed investment.

The DIEPER partners are aware of the fact, that agreeing on these three assumptions is only the starting point for efficient co-operation. A number of preconditions for gaining user acceptance for digitised periodicals have to be taken account of:

• A major part or best the complete run of volumes of a periodical must be available on the web.

• The number of digitised periodicals in any field of learning shall include core titles, that are of limited availability in their original paper versions or that promise to be of added value, when offered as a digital document.

• To enable easy use of digitised periodicals, that are dispersed over a number of servers at different places, a central point of access has to be built.

- To help the digital library of periodicals getting off the ground no general restriction of should be imposed.

It is no secret, that the DIEPER partners have been inspired by JSTOR,[2] which is supplying access to digitised American periodicals. Still they are not aiming at following this model in every respect, and indeed it would be impossible to do so under European conditions:

- The first difference is money. There is no funding agency in Europe at present, that would be willing to support digitising periodicals on such broad a scale as the Mellon foundation does for JSTOR. Investment into digitising in Europe is split up into a lot of different initiatives at the local, national and European levels.

- Following from this it is obvious that no uniform approach is followed by European initiatives, neither in the general set-up nor in the technical method.
    However, thanks to the almost unlimited communication made possible through the Internet a lot of exchanges are going on, which help to keep everyone aware of developments worldwide.

- Another disadvantage on the European side is, that selection of printed works, that are to be digitised, is often motivated by criteria, that do not take into account the interests of the wider research community. It is not astonishing then, that use of such digitised documents remains low.

- More serious than this is the fact, that the absence of concerted action has resulted in programmes, that may hardly be called a contribution to building the European digital library. This is true at least in so far as we can state the absence of any plan to bring such parallel programmes into line regarding the selected works.

While many European libraries are still experimenting, a lot of very useful work is already been done. The DIEPER project is about co-ordinating digitisation initiatives, that deal with periodicals, thereby making them more efficient and adding to their value for researchers and students.

DIEPER cannot, however, change the basic pattern underlying digitisation work done by or for libraries. It will try instead to build the infrastructure, that may help to overcome some of the pitfalls of European segmentation.

WHAT WILL THIS INFRASTRUCTURE LOOK LIKE?

At first DIEPER addresses the need in Europe for a central access point where all digitised periodicals shall be recorded. This shall be devised as a bibliographic database built on the model of the European Register of Microform Masters.[3]

Records in the Register of digitised periodicals will be linked to reliable and comprehensive archives of periodical literature filed on servers at different sites throughout Europe. Those archives will of course be equally accessible from other points (e.g. web pages and bibliographic databases) but the Register will be the site, where it is attempted to file *all* the relevant records. Searching the Register will help to avoid duplicating the effort of digitising one and the same periodical.

A search engine accessible from the Register will allow to do a fulltext search of the articles in digitised periodicals, of abstracts or at least of the tables of contents. This is fully dependent on three basic elements,

first, the existence of searchable text files linked to the image files,

second, such text files must be made available for searching beyond the local network and

third, they must adhere to certain minimal standards of encoding and indexing[4].

Another service to be offered by the DIEPER partners is output of a document in high quality print or on CD-ROM.

A few journals are to be scanned within the DIEPER project itself according to de facto technical standards. This shall serve to demonstrate the technical feasability of optimum online access and retrievability. At the same time the DIEPER partners will make accessible any already existing digitised periodical by including its record in the Register and by linking the search engine to it.

DIEPER is establishing contact to organisations outside the project, that have digitised or will be digitising periodicals to ask for this information to be included in the Register. As a rule this should be in the interest of the provider of the digital document, who will reach a wider market for his product.

In demonstrating the effective functioning of the DIEPER infrastructure it is hoped to develop a strategy for retrodigitising throughout Europe, that is based on a minimum standard for access and retrieval.

If access to digital documents is not free, licensing and accounting will normally be done by the system running the document server. To give a better service to users of the Register, however, tests shall be made to install a central Licensing and Accounting Module in conjunction with it. This would check the identity and rights of access of any user and give him access to the documents without secondary screening by the Licensing and Accounting Servers of multiple systems.

Here the rights issue comes in. In co-operation with other initiatives DIEPER will examine this issue and offer libraries models and advice, when they wish to start digitising a periodical. I am referring here to the European projects ECUP, ECUP+ and their follow-up TECUP who by concerted action aim at defining user rights in the electronic age as against rights of authors, publishers, collecting societies, subscription agencies etc.

After the end of EU support of the project the infrastructure set up by DIEPER shall remain intact. A model for its financial viabilitiy will be drafted within the project phase to enable starting with a permanent service without delay.

FIRST STEPS TAKEN

To give an idea of how the central European access point for periodicals could work, some partners have already prepared their systems. They may be accessed online to look at some of the components that are available already and that in part will be used in the DIEPER context.

A good example is the GBV bibliographic file,[5] based at Göttingen in Germany, which offers some of the technical features necessary for the Register of digitised periodicals. This file contains records of any kind of publication and offers simple as well as advanced searching facilities.

When searching for the well-known mathematical journal *Mathematische Annalen* a number of hits will display. Selecting the *electronic edition* the full display of this record[6] shows a URL. This is linked to a document server of the GDZ (The Göttingen digitisation centre based at the university library[7]).

When you follow this link, together with bibliographic information the volumes already available for online access are shown. You may choose a volume and browse its table of contents as if you had the printed volume at hand. The table of contents shall in future be created automatically from the

server's database of metadata. Selecting an article will bring the image of its starting page on you monitor. The reader may then turn over pages to read.

Output of a document in high quality print or on CD-ROM is technically feasible. In addition there is a facility for doing a simple search of this server's documents.[8] In future DIEPER hopes to offer fulltext search of the articles in selected or all periodicals via a search engine accessible from the central Register.

REFERENCES

* This paper has been read at the pre-conference to the 1998 LIBER Annual General Conference at Paris (30 June, 'DIEPER – The European Access Point for Digitised Periodicals'). A slightly evolved version was presented at the Nordic Conference on Preservation and Access 1998 at Stockholm (6 September, 'The DIEPER Project, Digitised European PERiodicals').

1 Project partners are Niedersächsische Staats- und Universitätsbibliothek Göttingen (co-ordinator), Universitätsbibliothek Graz, Springer Verlag Heidelberg, Helsinki University Library / The National Library of Finland, Det Kongelige Bibliotek Københaven, Centrale Bibliotheek of Katholieke Universitet Leuven, Bibliothèque de l'Université René Descartes / Paris V, University of Patras Central Library, Università di Siena - Facoltà di Ingegnieria and Tartu University Library.

2 For updated information visit <http://www.jstor.com>.

3 For updated information visit <http://www.gbv.de/eromm/gbvero-e.htm>.

4 SGML and its derivates HTML and XML have become such a standard. Indices in *ML can be generated by systems using RDF even if they don't use a *ML in their internal encodings.

5 From <http://www.gbv.de> go to *databases* and select the first file in the list.

6 In close cooperation with EROMM standard codes for describing technical features of an electronic document (to be used in tag 135 of UNIMARC or 007 of USMARC) will be defined and filed with the bibliographic record.

7 See <http://www.SUB.Uni-Goettingen.de/GDZ>.

8 Direct access at <http://134.76.162.14:8080/>. As Document Management System SAROS/Mezzanine is used, which is a product of FileNet.

The DIEPER Project, DIgitised European PERiodicals

Dr. Werner Schwartz
Staats- und Universitätsbibliothek Göttingen
37070 Göttingen, Germany
ORIENT@mail.sub.uni-goettingen.de