Consortium *of* European Research Libraries

Manuscript Working Party
Preliminary Report
October 2001

Contents:

1. Summary

The present report is based on a consultation in personal discussion and by e-mail of colleagues involved in current projects aiming to give electronic access to records of manuscript material, and on surveying a selection of such projects, conducted between May and October 2001.
The preliminary finding is that the current projects, although in principle all aiming at improved access to records and substitutes of manuscript material, and to giving access to information regarding the (Western) manuscript heritage beyond the confines of collections, their methodology and means of distribution vary to a great extent. There is therefore a clear need for a system capable of bringing together the information accumulated in each of the existing systems. This may be achieved by the modern technology of a web-based search engine for cross-file searching. An organisation independent of any of the current systems will be best placed to achieve this by co-ordinating the cooperation between projects that will be required. If its members agree, the Consortium of European Research Libraries may be in a position to initiate the process of co-ordination and organisation that should lead to establishing a facility for cross-file searching of manuscript material. The first recommendation is therefore to intensify the consultation process with a narrowing of its focus.

Members of the Consortium are therefore asked:

1.1. To accept the preliminary findings of this report.

1.2. To approve the continuation of the consultation process that has been set in motion.

1.3. Subject to further consultation, would they consider the Consortium the organisation best equipped to take on the role of co-ordinator and organiser of a project with the aim of providing cross-file access to the current projects?

2. Introduction

Following the majority approval by the Members at the Annual General Meeting of 12 November 2000 of the proposal to investigate whether the Consortium should take the initiative to co-ordinate access to automated projects recording (Western) manuscript material, the outgoing Chairman constituted a Working party, which was to bring out a report at the next Annual General Meeting in November 2001. Due to unforeseen circumstances this plan did not become operative until May 2001, when the present Chairman asked me to lead it. Given the shortened time in which to conduct a consultation, the present report can be no more than the result of an orientation in what is already available or in preparation in the form of electronic products giving access or information on manuscript materials; a range of colleagues working in manuscript collections have responded to my preliminary proposals and questions circulated among them..

The Working Party was given the following remit:
1. To survey initiatives relating to the written heritage in Europe, and to produce an overview of the main work in progress.
2. To liaise with appropriate organisations and bodies in Europe, and with the Digital Scriptorium, in order to establish the best means to build upon existing initiatives and avoid unnecessary duplication of effort.
3. To make recommendations on whether to extend CERL's scope in this direction and, if so, to produce proposals on how CERL's present remit can be extended through a phased development programme.
4. To consider the relationship (e.g. organisational structures, database development issues, finance etc.) between CERL and the Digital Scriptorium and other similar initiatives.
5. To take into account any other of the relevant issues that arise during the process of investigation, and to make recommendations.

Although the limited time allowed a preliminary survey of the main existing projects and valuable comments from colleagues were received, there has not yet been sufficient contact with the main players to produce a proposal for organisational structures and finance. Further consultation, broader but with a sharpened focus, is necessary. Nevertheless, a number of issues have become clearer in the past five months, and it is useful to summarise these for the Members as well as for the participants in the discussion, in order to give guidance to the consideration of options that must precede any decision whether to proceed, or not.
At this point the Members of the Consortium are asked to approve a continuation of the consultation.

3. Existing projects

There is a substantial number of projects currently taking place, all aiming to provide access to sections of the manuscript heritage, but only partially overlapping in their objectives and in their methodology. Projects addressing archival material, relevant for their technology as well as their advance in intellectual control, have so far been left out of consideration. They should be included if this survey is continued. Also, projects concentrating on 'subsets' of the heritage, e.g. SAGANET (large digitisation project of manuscripts of Nordic sagas and mythologies) have not yet been included. The main projects with which to date I got acquainted to a greater or lesser degree are (in alphabetical order):

3.1. EAMMS
Electronic Access to Medieval Manuscripts. Establishes cataloguing standards in SGML and in MARC, based respectively on the Hill Monastic Manuscript Library and the Vatican Film Library. Three-year project funded by the A. Mellon Foundation. Affiliated to Digital Scriptorium and liaising with MASTER.
3.2. Digital Scriptorium:
A prototype image database and visual union catalogue of medieval and Renaissance manuscripts. Based: Bancroft Library, Berkeley, CA and Columbia University Library, NY. Working with ten partners. records created in TEI DTD.
3.3. De Ricci project:
Supplement to the census of pre-1600 manuscripts in libraries in the United States and Canada, concentrating on the manuscripts not described. Based in California. Completed a Directory of Institutions with pre-1600 manuscripts. Plans to contribute cataloguing data to one of the current digital initiatives.
3.4. INITIALE:
Database of illuminated manuscripts in libraries in France, led by the Institut de Recherche et d'Histoire des Textes (IRHT) Gives access to some 100.000 digitised images, accessible through text as well as iconography. Distribution divided over two servers of different ministries.
Source: Élisabeth Lalou, 'Une base de données sur les manuscrits enluminés des bibliothèques: Collaboration entre chercheurs et bibliothécaires', in: BBF 46 (2001), pp. 38-42.
3.5. MALVINE:
An Internet gateway to ('modern', i.e. post-Medieval) manuscripts across many collections. Funded by the European Commission. Based in Staatsbibliothek zu Berlin. Developed with eight partner institutions.
3.6. Manuscripta Mediaevalia linked to Handschriftendatenbank der DBI.
Union catalogue of Mediaeval manuscripts in Germany giving access to scanned images of catalogues published after 1945 and to digital images held at Marburg University.
3.7. MASTER:
Manuscript Access through Standards for Electronic Records. Devised cataloguing standards for electronic access to collections, in partnership with especially smaller collections across Europe. Three year project funded by the European Commission.

Cataloguing projects in major collections:
3.8. British Library, London.
Electronic access to scanned images of published catalogues with to date some exceptions.
Link to digital images planned. Well structured access points following its own system.
3.9. Bodleian Library, Oxford.
Scanning of published catalogues with minimal EAD coding providing access points. New cataloguing XML encoded, EAD - DTD.

Projects primarily providing access to secondary literature:
3.10. Bibliotheca Neerlandica Manuscripta.
Conversion to electronic form of the existing documentation system on Dutch Medieval manuscripts.
3.11. Palaeography, Developing the National Resource.

Based in the University of London Library, in partnership with the University Libraries of Durham and Liverpool; gives primarily access to secondary literature, but is extending the project by developing a Manuscript Subject Portal.

The listing is at this stage very incomplete. A few more projects have been brought to my notice.
Austria:
Tabulae ÖNB- Datenbank.
Manuscripts in the Oesterreichische National Bibliothek.
Kommission für Schrift-und Buchwesen des Mittelalters der Oesterreichischen Akademie der Wissenschaften.
Katalog der Streubestände in Wien und Nieder-Oesterreich.
These two databases offer scanned images of catalogues with Mediaeval manuscripts, linked to digital images of samples of script and of bindings.
Italy:
MANUS database. *(No further information at this point).*
Portugal:
PORBASE. *National union catalogue combining printed material with manuscripts, catalogued in UNIMARC. Also: Database of illuminated manuscripts before 1500 in the Biblioteca Nacional, Lisbon.*
Spain:
Universidad Alicante, early catalogues encoded in XML, linked to digital images and to full text.

4. Objectives of the projects

A rough division into categories is already apparent in the list as given above. A distinction pervading all projects is: text-oriented, or image-oriented. The image bias is particularly evident in the databases of medieval manuscripts, whereas the conversion projects based on published catalogues are naturally more text-oriented. It is, however, also a distinction in the modern development of the study of manuscripts. Other distinctions in objectives and methodology may be:

4.1.Setting standards for description which are generally applicable:
MASTER, EAMMS, MALVINE.

4.2. Producing a national census of manuscripts:
Germany, Portugal, USA (De Ricci project).

4.3. Producing a union catalogue for a particular period:
Digital Scriptorium, MASTER, MALVINE

4.3.a. Producing a union catalogue for a genre in a particular period:
Initiale, also SAGANET

4.4. Primarily creating and providing access to digital images:
Initiale, Manuscripta Mediaevalia, Digital Scriptorium, the Austrian projects

4.5. Providing access to records produced by scanning published catalogues:
Manuscripta Mediaevalia, British Library, Bodleian Library, Austrian projects.

4.6. Using early catalogues for the virtual reconstruction of collections:
Universidad Alicante.

4.7. Methodology:
MARC formats: EAMMS, Portugal
SGML (Standard Generalized Markup Language) : EAMMS, MALVINE
XML (Extensible Markup Language) : MASTER, Bodleian Library
TEI (Text Encoding Initiative) : MASTER, Biblioteca Real, Madrid
EAD (Encoded Archival Description) : Bodleian Library
Digital catalogue conversion with access points: Manuscripta Mediaevalia, British Library, Bodleian Library, Austrian projects
Digital conversion of documentation: Bibliotheca Neerlandica Manuscripta
Cross-file searching: MALVINE

5. Present users facing a diversity of systems

It is realistic to expect that for the researchers/ scholars these projects are on course for providing ever larger conglomerations of information regarding the subjects they may seek, whether text-oriented, production-oriented, or with a focus on early ownership and the history of collections. It is equally realistic to expect that even with a rapidly increasing number of electronic records accumulated in the various projects, it will remain necessary to consult a large number of systems, each with their own search protocols, in order to compile a survey of any given objects or subjects that can be defined by a search-term. Each project has its set of limits. One of the respondents to my initial questions wrote with feeling of the present 'chaos' in accessing manuscript material.

A further practical obstacle for the researcher is that many of the present systems operate in one language only, and sometimes in terms peculiar to an institution or a particular discipline. It is often difficult for the non-initiated or those not familiar with a language and specialist idiom even to read the helpfile, or to grasp the correct interpretation of the search terms. MALVINE is a brilliant exception in that it offers a choice of the main European languages.

6. Identified requirement

By and large, the manuscript projects are mutually incompatible. Although undoubtedly already a major development for any user relying on data regarding manuscripts, whatever their period or genre, the need is now obvious for taking a further step by creating a search engine that can search across these systems.

Also in its statement of its aims the MALVINE project offers a model for the type of project the Consortium may contemplate. Its aim is to produce a search engine for cross searching heterogeneous systems, independent of heterogeneous technical solutions and cataloguing traditions and which is accessible to end-users via web technology. Thus different kinds of local cataloguing traditions are allowed to continue. This statement has several familiar echoes to those who have created the Hand Press Book Database (HPB) for materials printed between 1455 and the middle of the nineteenth century. However, as in the MALVINE project and unlike the HPB, the creation of an interface, or a web-based search engine, would be the correct development, preferable to creating a new database accumulating records from diverse sources, as the respondents to my discussion document all agreed (with one exception). In the development of technology such a search engine can be expected to be a realistic aim.

7. Forms of organisation: CERL

The current projects are the product of work carried out by organisations with various administrative models, and with a variety of sources for financial support. The recently developed projects can potentially encompass large amounts of material, making them accessible at a very

sophisticated level of description. Their importance for scholarship needs no defence, although they may to date include only a modest number of actual records compared with the mass of material to be processed. The value of the aggregate, however, of a combination of projects, will exceed by far the sum of its parts. Since the projects are based in diverse countries, and subject to very different organisational and administrative structures, it is necessary to find an organisational form that can in due course, in a phased approach, produce a system that can give access across the files produced by the individual projects.

The Consortium's experience in the creation of the Hand Press Book database is that the project is - among other aspects - a catalyst in a crucial process of transition. Initially the project was entirely driven by the creators of the electronic data, in this particular case the file providers who are the same institutions that set up the organisation in the first place. Once data from a variety of sources became available to scholars through the consultation of a single system, the advantages to the user became evident, increasing exponentially every time a new source of data is put in place. In this ongoing process the Consortium identified obstacles still facing the users, and realised that further facilities should be developed in order to let the users derive maximum benefit from the data now assembled. Hence the Consortium's current development of the Assisted Searching facility which integrates a Thesaurus files of names. When completed, users will be able to search on any form of name (place-name, printer, author) without having to rely on the form primarily determined by a cataloguing tradition in a particular institution, country or language area.

The Assisted Searching facility is now in place for place-names, while a beginning has been made with the Thesaurus file which will steadily be expanded in the next few years. The Thesaurus file also functions as a stand-alone source.

This development illustrates the main principle underlying the Consortium's organisation and the HPB database: to give wide access to materials by integrating them into a single search system, while retaining form and contents of the original cataloguing. The Consortium is therefore not prescribing cataloguing rules. It also illustrates the transition from primarily organising access to a variety of data by obtaining files from contributing institutions (institution driven), to ensuring that the data can be used by a great variety of users (user-driven). The Consortium's main concern can be summed up as giving the maximum benefit to scholars and users, within the limits of technology and library organisation.

8. Application to manuscript material

Whereas it has been feasible to create a single bibliographical database for pre-mid-nineteenth century printed materials, the description of individual copies belonging to the same printed edition incorporated into the knowledge about that edition, a comparable benefit does not exist for manuscript material. By definition, the production of every manuscript is a unique event, and every manuscript is a unique object.

There is an even stronger argument in favour of a cross-file searching system as against constructing a manuscript database from a variety of files. There is far less unity in the cataloguing of manuscripts, and in cataloguing in electronic form, than has been developed for printed materials. The parallel existence of MARC-based, SGML and XML-based projects is a given fact, although projects like MASTER are indicative of the recognition that consolidation is necessary. There are, however, no signs that consolidation between the main large projects will ever be achieved. Nevertheless, there are common denominators in manuscript cataloguing that in existing systems are exploited to provide access points, and that can be used in the same way to devise the access points in a search engine. It appears to me (as a researcher) that for an initial search the number of access points may be limited, the results leading to more refinement and detail as offered already in the existing systems (depending, of course, on the data being available in the search file). However, this should be open to discussion, and dependent on technical advice.

9. <u>Suggested minimal access points</u>:

TEXT   Author
                Language
                Title or supplied title
                Incipit

PRODUCTION
                scribe
                (or autograph)
                place
                scriptorium
                date
                supplied dating
                illumination / illustration
                artist

LOCATION
                present location
                shelfmark

PROVENANCE
                former owners
                former collection
                nickname
                binding
                combination with other items in volume

10. <u>Limitations / Demarcations of material to be included in a manuscript project</u>

As stated above, the current projects have as objective a variety of materials and demarcations, and as a consequence are led by a variety of experts, giving them a bias determined by respective disciplines. There is a concentration on Mediaeval, or pre-1600 manuscripts. Initially the Consortium envisaged that it should also concentrate in the first instance on pre-1600 materials, as we were aware of the projects that had this focus. Other major projects, however, do not make this distinction. Dr Rachel Stockdale pointed out that the British Library's manuscript catalogue encompasses materials from every period in a single system, and also includes items that elsewhere would probably be classified as archival and treated separately. Similarly, Dr Kaufmann informs me that the Bodleian Library includes the whole collection, albeit in sections, in a transparent system. There is no compelling argument, when creating a search engine, to exclude any manuscript materials if they form part of a consistent and coherent cataloguing system. Giving access to the widest possible range of material is a further argument for working with a limited number of access points.

The value in individual records is the expertise with which they have been created, in the traditions of an institution or organisation that lends its authority. It follows that the contents of individual files reflect the accumulated body of scholarship created under the aegis of an institution or project. This can be the result of work covering centuries – as in files created by the great libraries – or a relatively short period, as in MASTER and Digital Scriptorium, where, however, the tradition of the scholarly development of codicology is a determinant factor.

The consequence of this reflection is that the one demarcation that should operate – this goes more for the recent projects rather than those based on a long tradition – is that records should be created within the relevant discipline. This remark stems from my testing the MASTER and Digital Scriptorium database on the few incunabula recorded there. These records strike the incunabulist as very odd. Although they add valuable copy-specific features to the body of knowledge about these editions, their total avoidance of any reference to the extensive existing literature and their method

of dating (bizarre in an incunable context) is almost comical. Although modern scholarship has realised more clearly than ever before that the written and printed heritage are a continuum in the intellectual sense, it does not mean that their investigation requires the same expertise. Printed books cannot be described as if they are manuscripts. I herewith signal my advocacy of the exclusion of printed books in these projects.

The continuum can be achieved, however, if the HPB database is included in the files that can be searched through a search engine, especially for the textual access points, and with the Assisted Searching facility in place.

11. The Consortium of European Research Libraries as co-ordinator?

The Consortium can contribute:

practical considerations:

* Its permanent organisational structure.
* Its immediate contact in various form of membership with over 60 libraries, library organisations and institutions across Europe and the USA.
* Its further links and modes of operation.
* Its experience in setting up a large bibliographical project, working through consultants and organisations, and in constant consultation with the members and file-providers.
* Its recognition of the value of the original cataloguing traditions and systems.

At least as important, however, is its recognition that CERL's objective is to make historical data regarding the heritage of the Western world as transmitted in print available to users in the most efficient, user-friendly form attainable through current technology. The same principle can be applied to the heritage in writing.

12. The way forward

Sufficient interest in the Consortium's initiative has already been shown to warrant continuation of the investigation, with a narrowing of the focus on producing a well-founded proposal for a search engine and supported by technical information.

The experience of the consultation by e-mail is that this is not the ideal medium for a discussion across language barriers, and between people not well known, or not known at all to one another. I am nevertheless very grateful to those who have responded, and whose opinions I hope to have represented in this report. A more formal discussion list may be asked to respond to any proposals. Face to face discussions with a smaller number of colleagues have been much more productive. This suggests that, provided funding can be found, a working conference may be the most effective way to arrive at a concrete plan. This would mirror the process which led to establishing the Consortium, which was preceded by two broad-based conferences. With hindsight we can perceive that these meetings did not address to any extent the technical requirements to achieve what was envisaged as a principal objective. Only after the Consortium was formally established, technical proposals were brought forward and alternatives discussed. Taking advantage from this experience, it should be realised that all proposals should be firmly grounded in technical advice. This should also ensure that the best possible use is made of what has already been achieved in individual projects.

13. The remit of an initial conference

If Members agree that a working conference is the best way to move forward (provided funding can be found), its remit should be:
a. to formulate requirements within the bounds of what technology can offer,
b. to devise a development strategy
c. to consider organisational forms.
Not until clarification on these issues is obtained is it possible to estimate the financial implications and the ways to raise the necessary finance.
Further consultation and a conference should therefore aim to include:
Specialist users of manuscript material (from a variety of specialisms, e.g. mediaeval and modern, text-oriented, codicology oriented, from the library world as well as the scholarly community of users).
Specialists in database technology .
Representatives of the major manuscript automation projects
Representatives of members of the Consortium.
Ideally a working conference should have a small number of active participants (not to exceed some 15).
The Seminar /Table Ronde, to be held on 8 November 2001 preceding the Consortium's AGM in Lyon may be considered a modest beginning of a further consultation process.

14. Conclusion and Recommendations:

It is up to the members of the Consortium to decide whether theirs is the organisation best equipped to take on providing the organisational and administrative structure to support a cross-file searching system between the large resources for the recording of manuscript material that are currently in existence, and that are mutually incompatible.
If the members decide in principle that the Consortium should indeed investigate further whether to undertake this role,
the recommendation that can be drawn from the consultation that has now taken place is that
a. the Consortium should not align itself with any single organisation,
b. but instead should concentrate on the creation of a search engine that eventually will link the data in existing projects by cross-file searching;
c. it should initiate a close consultation with existing projects,
d. in order to devise a strategy, or an order in which projects should be 'processed' (i.e. mapping of data structure) to show benefits in the shortest possible time;
e. it should intensify the consultation process that has now begun in order to arrive at decisions on meeting users' requirements by the most effective use of technology.
f. The involvement of (independent) scholars as well as technical experts in addition to participants in the current projects is therefore crucial.

15. List of participants

On the mailing list were: Dr I. Boserup (Copenhagen KB), Mr L. Burnard (Oxford University), Dr Luisa Buson (Centro Ateneo, Padua), Dr Melissa Conway (De Ricci project), Dr T. Delcourt (Troyes BM), Dr Consuelo Dutschke (Columbia University, New York), Dr Lisa Fagin Davis (De Ricci Project), Professor Charles Faulhaber (Berkeley, CA), Dr Rosa Marcelino Galvao (Lisbon BN), Dr Monique Hulvey (Lyon BM), Dr M. Kauffmann (Oxford, Bodleian Library), Dr Anne Korteweg (The Hague KB), Dr Maria-Luisa López-Vidriero (Madrid, Biblioteca Real), Dr V. Mazulis (Riga, Latvian AcL), Dr Eva Nylander (Lund UB), Dr E. Overgaauw (Berlin, StB), Mr J.-L. Rodriguez (Madrid, Biblioteca Real), Miss Rachel Stockdale (London BL), Dr Bettina Wagner (Munich BSB), Dr Mirna Willer (Zagreb, NUL).

One library declined to take part. A further three did not respond to my invitation. In addition I had very useful discussions with Dr Peter Robinson (De Montfort University and leader of the MASTER project), and with the members of the Consortium's Advisory Task Group. I am very grateful to all who responded to my attempts to initiate a discussion, in the form of replies to all and in individual comments. I am particularly grateful to Dr Claudia Fabian, Dr Rachel Stockdale and Dr Bettina Wagner, who provided a wealth of information.

Lotte Hellinga
London, 30 October 2001