Werk

Titel: Shared digital collection development: the first steps to the virtual library

Autor: Enderle, Wilfried

Ort: Graz

Jahr: 1996

PURL: https://resolver.sub.uni-goettingen.de/purl?514854804_0006|log65

Shared Digital Collection Development:
the First Steps to the Virtual Library

WILFRIED ENDERLE
Niedersächsische Staats- und Universitätsbibliothek Göttingen

I.

"Shared collection development" is actually an old, classical task of libraries, even if putting it into practice has been carried out with varying success in the individual states and regions. "Shared digital collection development" can in part certainly tie up with the experience and structures gained there, but the technical possibilities of digital communication techniques offer partly new possibilities of co-operation and distribution of tasks. Now that in the "virtual library" the boundaries of individual libraries have begun to dissolve, co-operation between libraries can and must play a key role in the development of a "shared digital research library".

In thinking about "shared digital collection development" it is of course not sufficient to concern oneself with the organisational problems of co-operative holding models only. One always has to take into consideration as well that with "digital collections" new structural tasks arise for the libraries. Taking the definition by Peter S Grahams as a starting point, whereby the "digital research library" is a ".....set of electronic information organised for the long

term"[1], it becomes clear that the term collection at a digital library is clearly more comprehensive and with that also less clear, as is the case in a conventional library; because - according to this definition - it is not restricted to documents only, but embraces electronic information in general which can also be databases and not only texts in the strict sense. One could quite rightly, in an extension of the definition by Graham, include conventional documents - which are digitised on request and, via the net, put directly at the disposal of the person making the request - in the material of a digital research library.

Despite the elevated status of co-operation and the technical possibilities of the virtual library, the problems to be solved locally always remain at the lowest level, particularly since digital libraries in general are not created as a new facility but form a new task area of existing, conventional libraries. "Shared digital collection development" can only function if sensible solutions for "local digital collection development" exist.

However, concerning the subject "shared digital collection development" , the following three subject groups should, therefore be addressed:

1. The organisation of the holding development of shared digital libraries: what needs to be done if libraries, in co-operation - be in the framework of co-operative systems or other models, - wish to jointly develop digital libraries?

2. The contents of specialised digital research libraries: what falls under the term "electronic information"? Which forms of digital information essentially belong to the holding development of a digital library? How can such a digital library be created technically?

3. The local holding development, that is the procedures which are to be organised locally at individual libraries: this

---

[1]   Peter S. Graham: The Digital Research Library: Tasks and Commitments, 1996 (http://csdl.tamu.edu/DL95/papers/graham/graham.html).

includes queries about internal organisation, but also the type of technique for acquiring electronic information.

So much regarding the subject: "shared digital collection development". "First steps to the virtual library" means that in the main these questions cannot be discussed in a theoretical manner but, first and foremost, an attempt must be made with concrete examples and planned projects - essentially from Germany - to show to what extend first concrete attempts at solutions have been made, which can form the basis for digital research libraries. With this, essentially the following projects and planned projects are addressed:

The WEBIS project; the SSG-Fachinformationsprojekt (SSG Specialized Information Project); the MEDOC project; the concept of new information structures for science through the associations of German Mathematicians, Physicists, Chemists and Computer Scientists as well as the WEBDOC-Project.

II.

The decisive problem in the realisation of a digital library is in the mean time no longer technical, but organisational and therefore also inner-librarian. Librarians have to agree on a concept how subject areas will be distributed to individual libraries, how the pertaining financial arrangements can be structured and, to me this appears to be an essential aspect, how mutual comprehensive access to these digital information resources can be organised for the user and ensured in the long run.

In Great Britain attempts are made to implement such a concept with project Catriona. In Germany are first beginnings which take up on the existing concept of the "Sondersammelgebiets-Plan" (SSG), special subjects field programme, sponsored and supervised by the Deutsche Forschungsgemeinschaft (DFG) (German Research Society). The

entire spectrum of knowledge is thus divided up into 110 main subject fields and supervised by a total of 36 libraries. There the relevant scientific literature is acquired as completely as possible. In compensation for the financial support by the DFG for the holding development, the libraries agree to make the acquired holdings available to the national loan service without restrictions[2].

It is obvious that this collective order must also apply to digital media. Basically the concept of the special subject fields programme, especially with the new possibilities of the networks, will come really to its own with the technical basis for direct access to shared main collection areas. Then the user wouldn't be longer, as before, dependent on inter-library loan orders. There is no question that SSG-libraries have already started to offer their services via Internet: first Online-Document Ordering and - Delivery Services have already been implemented and some libraries meanwhile also offer access to their catalogues, Current-Contents-Services and new acquisition lists via the net and are beginning to develop their own document servers[3].

It is naturally of great importance for the success of these local endeavours that uniformly structured access is developed on the shared servers which - from the view of the user - combines all locally developed digital services. This uniform access is currently implemented with the WEBIS Project which is carried out by the Hamburg State- and University Library and is sponsored by DFG[4].

---

[2] Deutsche   Forschungsgemeinschaft.   Unterausschuß   für   die Sondersammelgebiete: Überregionale Literaturversorung. Index der Sammelschwerpunkte, Bonn 1985.

[3] See for example the services of the State- and University Library of Goettingen: http://www.uni-goettingen.de/sub/homepage.htm.

[4] For further information see: http://wwwsub.sub.uni-hamburg.de/.

What precisely is being done by WEBIS? A World-Wide-Web-Server has been developed which offers structured access to the entire subject range and the electronic services of SSG-Libraries. Users who are looking for information and literature concerning a certain subject are in each case steered directly to the holdings of individual libraries which have comprehensive collections, whereby the holdings in principle integrate conventional as well as new media.

Decisive for the practicability of this concept is that essentially uniformly organised central access is available, but this central access server can be administered remote by the individual SSG-Libraries. A special menu-driven editor permits the co-operative online-actualisation of the servers' HTML-files by the participating libraries without the respective employees having to have the necessary technical knowledge about the setting-up of such files. With this each library can - from a certain access level onward - decide whether information is to be stored on the central WEBIS server or links put on the own server. With this, a gradual transfer has been created between central entry which, in future, can also include enquiry in OPAC's or union databases with possible access to electronic documents.

WEBIS is a technical solution for entering a comprehensive future digital library which, from an organisational point of view, is based on the shared holdings of the special subject fields programme. Essentially the success of WEBIS is of course also dependent upon the fact which digital services the individual SSG-Libraries are able to offer. One example of what can and should in fact be developed in future offers the "SSG-Fachinformations-Projekt", the specialized subject information project carried out by the Göttingen State- and University Library which is also sponsored by DFG. In this project initially, electronic resources for a few select subject areas - the earth-sciences and geography, as well as pure mathematics, are being assessed and prepared for the user. The decisive conceptual approach of the Göttingen

project is not only to collect, and to organise access to the electronic specialist information and documents, but initially to establish guidelines for a qualitative assessment of electronic information for defined subject areas. The information collected in this manner is in accordance with a defined standard and offers the user who makes use of it the guarantee that his enquiry may lead to relevant results concerning the content and needs not lose himself in any innumerable sources of information as we have known it to happen only too well up to now on Internet. It is therefore a meta system which uses existing directories and search services, but extends these by a qualified component. Especially in the assessment of digital information the libraries have an opportunity to transfer their expertise which - up to now has been proven by the acquisition of printed media - to the new media and to create interesting additional value for the user.

At best, this model could then be taken over into the remaining special subject fields which would cover the entire spectrum of knowledge; and at the same time it could offer the beginning for subject related co-operation by librarians. The SSG-libraries could take on an editorial function for their subject fields and, technically speaking, they would function as information server whilst the other co-operating libraries could offer this joint service as an entry to certain specialist areas to their users.

In principle this possibility of staff co-operation cannot be emphasised enough. Due to the digitalisation of our communication system "shared digital collection development" may, for the first time in the history of libraries, not only mean that holdings are developed in a work-sharing form but also that regionally or nationally or at least internationally shared staff resources can be utilised jointly. Naturally totally new forms of organisation have to be developed for this. In Germany first approaches and considerations are evident on a regional level of library unions - for example in Northrine-Westfalia as well as on the national level of the special subject fields programme. This

also includes, of course, the necessity of technical solutions for shared assessment and acquisition of electronic information and documents. As it seems there can be two different approaches for that: Either one can use a model of cooperative cataloguing based on an union database or a model based on an editorial centre and an organised electronic communication system.

A further basic problem - which must be solved in the development of shared digital libraries - is the question of right of access. This then becomes relevant only if it is a case of documents from commercial publishers. Then the question arises how far shared digital libraries will possibly limit free access to information which - up to now - was guaranteed by conventional inter-library loans. Organisational concepts for shared holdings development must always take this political question into consideration as well.

Another basic problem which presents itself quite naturally is, whether this organisation must and should initially be carried out on a national if not regional level. It is obvious that global interlinkage not only enhances international co-operation but in fact also calls if not demands it, if one wants to work efficiently and cost effectively in view of ever decreasing public funds. Despite this, there are a number of reasons to begin with the development of shared digital libraries on a national, perhaps in some cases even on a regional basis. Here one of the most important reasons could be that the relevant organisational infrastructures already exist, such as the union system of libraries, special subject fields programmes and the like. It is in the genuine interest of libraries to further develop these structures in an evolutionary manner. In addition that, for example, in the "humanities" the digital research library should also include access to printed materials as these, in the foreseeable future, cannot be completely substituted by electronic publications. This integration underlines the necessity to take up existing areas of collection emphasis.

III.

However basic the organisational problems and the question of structurally uniform access in the development of shared digital holdings may be, one can only really speak of a digital library in the actual sense of the word, when a foundation of digital documents and information resources has been developed. With this the problem of content and the technical implementation of a digital research library presents itself. Two examples may illustrate the different methods of approach.

In the MEDOC-Project a digital library for the subject area Computer Science will be implemented which encompasses electronic documents, journals and books as well as electronic information. With this, not only publisher products subject to costs, but also free information sources from Internet should be integrated which are stored on different servers. MEDOC is a joint project of the association of German Computer Scientists, of the "Fachinformationszentrum Karlsruhe" (specialized information centre Karlsruhe) as well as the "Springer-Verlag". It includes numerous other project partners, among them the Hannover Technical Information Library which, as special subject field library for Computer Science, is involved on the librarian part in this project[5].

The main emphasis of this project is less on content but developing a technical model. It is planned to limit oneself to approximately 20 journals, 200 - 300 electronic books and approximate 1000 technical reports which will be made available within the framework of the project. The centre of the project is rather the development of technical solutions for a digital research library. The concept developed takes up on the approaches of the "Digital Libraries Project" of the University of Michigan in that, with the employment of so called "intelligent

---

[5]   See http://medoc.informatik.tu-muenchen.de/medoc/

agents", developments from the area of artificial intelligence and expert systems are taken up which are to organise access to the content of the entire spectrum of electronic information.

This concept reveals which technical problems can occur in the development of a complex shared digital research library. Integrated access to heterogeneous digital information is only possible by a technically sound level of conveyance, the "intelligent agents" which in themselves are again divided into different components: user agents, broker and provider agents. The precondition for the functioning of such a complex system is of course a general standardisation of communication between the individual components. In general, the development of shared digital libraries is increased by the need to adhere to technical standards and regulations. This also applies to the document format and mainly to standardised header information which actually makes the automated administration of digital information possible.

In the MEDOC project the development of a sophisticated technical concept for a limited subject shared digital research library is regarded as a precondition which can then be filled with further content. Project conceptions of German mathematicians which were developed at the Konrad-Zuse-Centre in Berlin, however use a different approach. Here one assumes that the existing technical possibilities are entirely sufficient for the implementation of a shared digital research library. What is important is that a suitable organisational solution is found. Their model envisages that, to a large degree, scientists organise their own publication method by themselves. Since each institute in the meantime has it's own web-server and in addition mathematicians have for a long time been using uniform standards and systems in the development of their documents - especially Tex and Postscript - and beyond that have an internationally recognised classification, they envisage a sufficient basis. What is considered

to be most important is the development of a system of reference and archiving which organises access to the documents[6].

In this concept the library is primarily allocated the function of continuous archiving and the accessibility of digital documents in the long run. The shared holdings development itself is, for the time being, the task of scientific institutes which publish the works of their own members. First concrete organisational steps haven been taken insofar that a co-operation agreement by four specialist associations, that of mathematicians, physicists, chemists and computer scientists has been concluded which is to form the basis for concrete co-ordinated projects.

Whilst the mathematicians' conception basically reflects conventional publication structures simply and directly on digital technical possibilities, it essentially being a case of organising access to digitised texts, the MEDOC-Project reveals on the other side that digital libraries can be more but in return display clearly more complex structures. In the digital age, the term 'information' itself becomes more complex and multi-layered as was the case up to now, as the transition between individual forms often is no longer clearly defined. Electronic information includes documents, that is texts, as well as electronic discussion lists, databases and expert systems. The conventional term of 'collection' may therefore in future change, as users will not only be searching for certain texts or journals in electronic form, but for thematic servers which permit access to a broad spectrum of different information; and beyond that, own expert systems will carry out for the user the enquiry and, selected from the available holdings, offer the results of the enquiry directly to him.

---

[6] A short description of the project concept gives: W. Dalitz, M. Grötschel, J. Lügger, W. Sperber: Verteiltes Informationssystem für die Mathematik. Kurzfassung eines Projektplans der DMV, 1996 (http://elib.zib-berlin.de/0x82496c0_0x00008b73).

The possibly end of a clear separation between reference databases and information itself means for the libraries that they either themselves, or in co-operation with each other and with subject information centres and research facilities, set up and organise such servers and thereby move more and more into task areas of computer centres; or that they only organise access to such servers for their users and concentrate on the collection of digital complete texts and guarantee their continuous availability - in other words - transfer their classical tasks to the area of digital documents.

IV.

Even if libraries restrict themselves to the latter, namely the integration of digital documents in the strict sense, they have to develop technical and organisational solutions for the respective procedures for the development and administration of their local digital holdings, which at the same time, as a tie-up of a shared digital research library, must in turn be integrated into central reference databases. A project which currently develops solutions especially for the integration of digital documents in the strict sense into the available holdings of a library, is the WebDOC-Project. WebDOC is a joined project, organised and coordinated by Pica, of several Dutch and German University Libraries, together with scientific publishers. Meanwhile a cooperation agreement exists also with the Research Libraries Group (RLG) in the USA. The general purpose of the WebDOC project is to provide a network infrastructure where library users are presented with adequate localisation facilities of digital documents and where they will have a guarantee of the document maintenance. Therefore a central database of electronically available documents, the WebCAT, contains the necessary bibliographic information together with abstracts and subject classification. The participating libraries and publishers build and maintain their own

document servers that are referred to the central catalogue. A licensing and accounting mechanism allows the integration of documents of publishers[7].

What I wish to emphasise here is that with the WebDOC-Project a solution has been found which permits libraries an evolutionary development of their existing structures. Electronic documents are recorded in a central catalogue analogous to printed books; access is steered via this catalogue and instead of the librarian at the loan counter, an accounting module ensures that only entitled users may have access to the electronic documents. With this it is also possible, via one system, to integrate free digital documents as well as those subject to licence from publishers, and to offer them to the user - naturally on different conditions.

One viewpoint of the WebDOC-Project which I would like to emphasise here, is the local procedure of a library. Because development of shared digital holdings must of course commence first and at the lowest level, which means with the individual local library. This in turn means that for the different tasks in the holdings development of a local digital library concrete organisational and technical solutions have to be found. Among others the following aspects have to be observed:

- enquiry and assessment of electronic information and documents
- acquisition of electronic information and documents subject to costs
- control of the completeness of acquisition, e.g. in the case of electronic journals
- technical dealing with digital documents
- indexing of electronic documents
- providing end-user access to electronic information and documents

[7]   For further information see: http.//www.pica.nl/

- storing large quantities of digital documents
- long-term archiving of electronic documents

Essentially, within the WebDOC-Project a solution for the recording of electronic documents is being offered which takes place in a conventional manner in the central union catalogue, as well as for the provision via Web-Server. The central development is of particular importance insofar that it represents the precondition for sensible use of share digital holdings. At the same time it permits libraries at a local level to continue their previous practice as it were also for digital documents. Local solutions still have to be developed for selection and acquisition and the technical dealing of digital documents.

Just as up to now national bibliographies, review organs, brochures and other materials have served as a selection basis for the acquisition of literature, in addition specialists at libraries now have to regularly and systematically assess comparable sources of information for electronic documents. Besides specialist electronic discussion groups relevant news servers can be part of this, as well as the regular new acceptance of electronic journals in NewsJour-Server of the Association of Research Libraries. For such purposes automated solutions can and should also be employed; robots which regularly scan relevant servers for new information.

If relevant electronic documents have been selected, the question arises, what the technical form of acquisition can look like. Apart from the possibility to only put one link on certain documents the question arises for all other documents to be stored locally, how to carry this out in a technical manner: documents can be downloaded individually from the net. This is without doubt the most cumbersome and worst solution. It is better to work with the mirroring technique, which means via FTP to automatically mirror regularly from a defined server all new documents into the local server. From a technical point just as practicable are off-line-data deliveries.

For documents such as electronic journals which are published regularly, it must of course analogous to the printed periodicals be checked that all documents are regularly received and catalogued. For this task for example the acquisition module of an integrated library system can be employed. This applies in particular when the electronic documents are also catalogued in the own catalogue. The development and administrative form of acquisition would then take place with the conventional modules of the library system.

More complex than the individual tasks to be solved, are of course the integration of these tasks into a stringent process. Only then a problem becomes really obvious: one has to deal with data which can no longer be handled physically. Organisation and comprehension for the necessary process require therefore a higher degree of abstraction than the conventional, in part already sufficiently complex, procedures have required. Insofar that the digital library does not exceed a certain structure one can surely work with such means as email and a well defined directory system for the various work procedures. In the long run this may of course not be sufficient. In future workflow-systems will have to be developed and employed for this which, based on a database, steer the entire work procedure. It is also conceivable that existing integrated library systems will be expanded by a relevant function.

V.

If one tries to present current developments collectively, new digital libraries in various projects are currently being developed; whereby the projects can be divided into two broad categories: in the one group subject related digital collections and concepts are implemented which - so to speak - can form a technical node in the net of digital libraries. In the second group attempts are made to organise entry to this net of digital libraries.

If one should attempt finally to classify the mentioned projects in an overall concept for the development of a shared digital research library, among others, the following basic aspects must be considered:

1. The area of information of a digital library will be more complex than that of a conventional one. It will not only encompass digital texts but electronic information in a general sense.

2. For the development of shared digital holdings an organisational infrastructure together with the necessary technical concept must be used or, if necessary, newly developed which organises the technical distribution of tasks between libraries. The respective responsible libraries have to function as information server for a subject area, which can certainly mean, that they partly organise only central access to information which, in turn, is stored on separate shared servers.

3. For the development and care of such information servers the libraries must also organise the co-operation of staff resources. This means, specialists for a subject area who work at different libraries, have to co-operate for assessing, acquiring and indexing electronic information.

4. For access to shared digital holdings it is necessary to have central integrated entry for the user. This applies to specialist information as well as to digital texts in a literal sense. Union catalogues may be used for this, but also entry servers.

5. Further, in developing shared digital holdings free access for users of all libraries, who are registered within the structure of this co-operation, must be guaranteed. A relevant solution has to be found for commercial publishing products.

6. Especially the shared development of digital holdings requires strict standardisation of the document format and the meta information about digital documents and information. In this manner only may local administration, but also the central

collection development and the indexing for reference databases, be automated.

However important each requirement may be, decisive for the success of libraries in the development of shared digital research libraries will be not the technical solution but whether and to which degree sensible and efficient organisational solutions for co-operation can be found.