

Adding value to Web-OPACs

Bernd Harmsen

Fraunhofer-Gesellschaft e.V., Germany

Abstract: *Web-OPACs not only offer simplified access to library holdings for end-users but also enable librarians to add value to their catalogue data. One feature which has become sort of standard for new library software is including links to fulltext or multimedia documents corresponding to a particular citation. Other features have not yet become common but will soon come into view: links to publishers, links to corporate sources, and links to journal titles. The latter usually provide tables of contents, and sometimes also offer direct access to fulltext documents for subscribers. Online database providers have acknowledged these new possibilities, librarians soon will. So when migrating library software to new products or new releases, such issues should be observed.*

The author is head of a support team who train and support the library staff of 47 scientific institutes of the Fraunhofer-Gesellschaft, the biggest German research organisation specialising in applied research. Recently they have organised the webward migration of more than 30 library systems.

Keywords: World Wide Web (WWW), scientific libraries, library software, virtual library

1. Introduction

The Fraunhofer-Gesellschaft is the leading organisation of applied research in Germany operating 47 research institutes with more than 9,000 employees, about half of them are scientists or engineers. Our internal service team "FhG-Online" offers library support, trainings, and also produces an online database covering all publications by Fraunhofer scientists.

Typical Fraunhofer libraries would have holdings between ten and thirty thousand books, and 150 journal subscriptions. Most institutes cite articles in their catalogues, at least those originating from their own institutes.

34 institute libraries have used TINlib as an integrated library software package since 1989 or the early '90s, respectively. With the advent of the World Wide Web TINlib became more and more criticised for its outdated ASCII interface, the OPAC users - i.e. the scientists - demanding a graphical user interface, preferably Web-based.

In 1996/97 we had begun screening the market and evaluating integrated library software systems. A working group discussed and defined the criteria, distinguishing KO criteria and further important or "nice-to-have" features. The provision of a Web-OPAC obviously was a "must", i.e. a KO criterion. Twelve software solutions were compared based on presentations and questionnaires. Next, six products were installed for in-depth evaluation phases.

We found that most library software products are meant for public libraries rather than special libraries. All had their strengths in cataloguing and circulating ordinary books but lacking strengths in regard to citing "grey" literature, journal articles, or conference papers. More flexibility was asked for to make the system - above all the catalogue module - adaptable to the requirements of our special libraries.

So in the end, our evaluation amounted to choosing a software package - Cuadra STAR - which originally was not a library package at all when first launched in 1982. STAR is a document management system, therefore meeting any special cataloguing requirements easily. As for the ordinary library transactions (aquisitions, circulations, cardex) the modules of a standard solution called "STAR/Libraries" provide well-fitted supplements to the flexibility of the basic system. Actually, we purchased STAR/Libraries and adapted it to our needs. (Note/Ref. 1)

18 libraries utilise the aquisition and/or circulation modules of the library system. Since our TINlib versions were not year-2000-compatible in regard to these modules these 18 library systems - plus additional two which were not Y2K compatible for other reasons - had to be migrated by the end of 1999. 16 libraries used only the catalogue module, so their data were converted and transferred in the first quarter of 2000. There will also be 7 or 8 first-time software installations or migrations from other library systems than TINlib so that we will have about 42 library catalogues by the end of this year.

2. Technical and organisational issues

While our former software solution was installed individually at each library this time we have decided that several library databases should share one database server. So eventually we will have more than 40 installations distributed amongst 8 cluster servers. Our wide area intranet connections have improved largely over the last years so we need not have servers at each location where a library software system is required. We have seen to it that the cluster servers are located where at least one major library is situated, though.

The cluster servers are most similar in regard to their configuration. Besides the operating system (Sun Solaris) and the STAR database software we usually have only installed a Web server (Apache) and a secure shell as replacement for telnet, rlogin, and ftp. By means of the secure shell we maintain all servers remotely from our office in Stuttgart. Furthermore, our cluster solution provides for easier and more efficient backup solutions.

Further software installations are required only for the library staff. About 100 proprietary client installations, running under Windows NT, are required. OPAC access is per HTTP (hypertext transfer protocol) which is commonly used within the Fraunhofer-Gesellschaft since it is our standard intranet protocol.

Access of the end-users to their institute's OPACs is organised by means of IP-mapping. The URL (uniform resource locator, used for identifying Web resources) is the same for all OPACs which share a common server. The IP (internet protocol) numbers of the workstations of the scientists - or their institute's proxy server - are mapped to the corresponding OPAC. "Illegal" IP addresses trying to access one of our servers are mapped to either a default error page or to our public database of Fraunhofer publications and patents which also resides on one of the cluster servers. (Note 2)

Since the URLs of the cluster servers do not have any meaning for the end-users we tend to "hide" them. Our team has developed two alternative Web frame solutions for the library start pages on the intranets. Thus the cluster URLs are hidden behind such Web frames which provide for all navigation aides outside the OPAC itself. The exit buttons of the OPAC typically would jump back to the homepage of the corresponding library (there is a default page, of course, reading "Thank you for your visit").

3. Adding comfort

There are several advantages of Web OPACs as opposed to proprietary solutions. Firstly you use a standard interface - the Web browser - which the end-users are well accustomed to. So there is little need for end-user trainings. Secondly you use the Web's standard functionality hyperlinking text files as well as database reports or searches. The basic browse-and-navigate functionality of the World Wide Web from the end-users's point of view means ease of use.

In addition to searching and browsing the library catalogue end-users can transmit orders or requests directly from the Web OPAC as well as view their own borrower accounts. While they normally need not do anything but click on a hyperlink to access the OPACs of their local library they have to enter their personal user ID and password to access these services.

A STAR feature allows users to define their own Web reports. They can choose which fields shall be displayed and how the retrieved records shall be sorted.

Standard Web solutions have one disadvantage, though. The Web protocol (HTTP) is a stateless protocol. This means that each transaction is singular, there is no permanent connection comparable to traditional online sessions. On the other hand STAR originally was - and basically still is - a "traditional" online retrieval system which was accessed via telnet or a proprietary protocol. In regard to the Web interface database sessions are emulated, the database server keeping track of all transactions from each workstation or PC. Such sessions enable us to offer the following services:

- * OPAC-users can select items from different title lists retrieved after several independent search queries and later display them together;

- * there is an "expert search mode" which allows end-users to construct searches step by step and then combine them;
- * end-users can resume their Web sessions even after browsing other servers or having been idle for other reasons.

However, Web users are not used to end sessions properly but usually simply close their browser window or browse foreign resources. For license reasons (Note 3) and to reduce overload on the cluster servers we therefore time-out Web sessions. After 10 minutes without any user activity Web sessions are assigned an "idle" state. The STAR license "slot" is freed but all session files are stored so the users can resume their session if a new license slot is available at that time. After further 20 minutes of inactivity the Web session is killed, all temporary session log files being removed from the server. So, although it would be nice if all OPAC users exited their Web sessions properly on the other hand they need not care about.

4. Adding functionality

Hyperlinks need not be static but may start a new database search. For example, our end-users can identify other publications by the same author or corporate source by simply clicking on their name. Similar "link reports" are offered for classification codes, or controlled and supplementary terms. Such hyperlinks are only generated if there are further references within the database in addition to the one just being displayed. They provide for "horizontal" browsing, meaning "show me more like this".

Displaying hierarchical levels of bibliographic information - "vertical" browsing - is handled in a similar manner. If a query results in a reference to a series title, for instance, a box is displayed reading "monographs within this series". If a search results in the citation of a monograph or a journal title one can browse "downward" to view a list of articles contained within this book or journal.

We have also implemented a new database view for journal issues. It is mainly meant to include TOC (table-of-content) services in the OPACs. So implementing this additional feature results in a maximum of three bibliographic levels in the library catalogues which can be browsed in the OPACs: journal titles, journal issues, and single articles within an issue. Bound journals are handled like journal issues, TOC data for single issues can be combined automatically to form annual registers of contents.

5. Adding content

Meanwhile most new library systems provide for fulltext or multimedia links. We display such links as little icons. Since titles are hyperlinked in the title lists so that single database records can be selected for viewing we refrained from hyperlinking them once again in the respective full displays. Such different "meanings" of hyperlinking title strings might have confused the end-users. Furthermore, clicking on a fulltext link opens a new browser window and thus also "behaves" quite differently.

Our OPAC users can simply click on an icon next to the title of a citation to access the corresponding fulltext document. Fulltext documents often are provided for journal articles or conference papers. In regard to journal titles we also link to the tables of contents provided by their publishers. The latter sometimes again offer direct access to fulltext articles. (Note 4)

There are several "philosophies" in regard to integrating fulltext documents. You could duplicate the text and incorporate it into the database, for example, to be able to produce direct ASCII - or rather HTML - output and also provide for fulltext retrieval. Additionally you would have a link to the original document in its original format. We exclusively use Web links. Thus our databases remain "traditional" bibliographic databases. Since they usually contain abstracts and terms, anyway, this provides for consistency of the databases as well as relevance. Including fulltext retrieval for only part of the records would not improve the databases at all.

As for our database of Fraunhofer publications - called "Fraunhofer Publica" - we ensure consistency of the links by copying the fulltext documents to a dedicated server with a defined structure of file directories and corresponding Web URLs. The structure of this document server is pretty straightforward: we have 47 directories for the institutes, one directory per year within each institute's directory and directories for the authors underneath these. The fulltext documents are given filenames

like "001.pdf", "002.pdf", and so on. All fulltext documents cited in Fraunhofer Publica are converted to the PDF.

We remind our institute libraries to make sure that all links to documents and multimedia files remain the same over the years. Otherwise maintaining the databases could become very cumbersome, indeed. Often Web-URLs have a tendency to "wander". If URLs need to be altered this should be done in a way that you are able to update the links within the database by means of global changes. Fortunately Web URLs can be handled in a way that they can be kept the same independently of the underlying file structure. This is why we recommend using Web links instead of "hard" directory paths or file names.

6. Adding service

Many of our libraries have maintained Web resource lists as an additional service for their end-users. Typically their "virtual libraries" would contain links to societies, companies, or publishers related to the research activities of the institutes. We now have provided data fields in our new library system to integrate such information into the Web OPACs. We aim at establishing the OPACs as a major starting point for browsing scientific and technical information on the internet.

Just as the link reports mentioned above may broaden the scope of a particular query, so may browsing publishers catalogues and announcements, or the information provided by organisations or societies. By simply appending their URLs to the normal catalogue data we have sort of a relevance filter as a by-effect: our own OPACs. Publishers and societies relevant to a particular scientific community are supposed to be represented in the library catalogue. Thus the catalogue itself guides the search for relevant resources (Note/Ref. 5). The queries of the OPAC users again grant the relevance of their results. The same applies to links to journal titles, i.e. links to TOC data, which I mentioned before.

In the end we have different types of links in our library catalogues. On the one hand there are links which identify related records within the catalogues - link reports or cross searches as well as bibliographic hierarchies -, on the other hand there are those which refer to external resources, i.e. fulltext links, links to journal titles (usually TOC services), links to companies or scientific societies, and links to publishers. So we had a database design problem: how exactly should which type of link be displayed?

As to the cross-searches the decision was quite straightforward, we simply hyperlink them in the ordinary manner. In regard to hierarchical "downward" information we implemented a standard solution recommended and provided by Cuadra STAR which displays hyperlinks in extra information boxes. Links to sources (the "upward" hierarchy, i.e. journal titles or monographs) are again displayed as ordinary hyperlinks. As I mentioned above we have decided to use icons in regard to fulltext and journal TOC links rather than hyperlinking the titles themselves. Links to corporate home pages are displayed in the same way, but with a different icon.

7. Future plans

In the future we will also allow to search several library catalogues simultaneously. Not all libraries may want this, and it does not make sense for all, either, since the research areas of Fraunhofer institutes differ very largely. We intend to provide access via the Z39.50 protocol (Z39.50 is the standard for retrieving bibliographical data; a Z39.50 server is delivered as part of the STAR software package).

Which may seem like a disadvantage - using Z39.50 instead of our own Web OPACs which would offer more flexibility - actually provides extended opportunities on the other hand. Z39.50 enables us not only to combine our own institute catalogues covering similar research areas but also to include public services. So, from an end-users point of view we combine our own holdings with holdings of the relevant loan libraries. "Subject orientation" as well as "availability" being the focal criteria. (Note 6)

We intend to develop several report formats which enable the libraries to generate resource lists directly from their catalogue data. Thus their "virtual libraries" which formerly were static lists of URLs might not only be integrated into the OPACs but also be generated dynamically from the same data pool. Such reports will be provided for links to corporate sources and companies, publishers, and

journal titles. The software - although session-oriented in principle - provides for so-called "quick-links" which enable us to generate the required reports without our end-users having to start database sessions.

8. Conclusion

By utilising the options which the World Wide Web offers OPACs can be made the starting points of choice when searching the Web for scientific and technical information. Including analytical data like references to journal issues with the corresponding tables of contents or references to articles and conference papers provides for in-depth bibliographic information. Including fulltext links provides for easiest and fastest access to the source information. And including links to external resources provides for comprehensiveness.

End-users are not interested if information is available in print or electronically, they are interested in the information itself. Therefore special libraries will have to become hybrid, and so will their OPACs.

Bernd Harmsen
Fraunhofer Informationszentrum Raum und Bau
FhG-Online
Nobelstraße 12
D-70569 Stuttgart
Germany
E-mail: har@irb.fhg.de

Notes and References

- [1] Our software evaluation process is described comprehensively in:

Sattelmayer, C. (1999) "Das Auswahlverfahren für ein integriertes Bibliothekssystem in der Fraunhofer-Gesellschaft", *B.I.T. online*, Vol 2 No 4, pp. 469-472

- [2] Our publication database "Fraunhofer Publica" utilises the same software and principles. Just the design differs. "Fraunhofer Publica" is publicly accessible on the Web:
<http://publica.fhg.de>

As a comparison one might check out one of our OPACs - the only one which is accessible from outside the Fraunhofer-Gesellschaft:
<http://biblio.zv.fhg.de/cgi/starfinder/0>

- [3] We have licenses for 280 simultaneous sessions distributed amongst 8 servers and about 40 library systems. 80 session licenses are reserved for library staff, 200 remain available for OPAC users.
- [4] At the moment we are negotiating the site-wide electronical access to articles of journals our institutes have subscribed to.
- [5] We actually used such catalogue data as a relevance filter when providing a virtual library for all Fraunhofer scientists. See:

Harmsen, B. (1998) "Tailoring WWW resources to the needs of your target group: an intranet virtual library for engineers", in Raitt, D.; McKenna, B.; Graham, C.; Kerr, J. (Ed.), *Online Information 98. 22nd International Online Information Meeting, Learned Information Europe Ltd., Oxford*, pp. 311-316
- [6] We will make this service look like the OPACs and obviously integrate the corresponding links into the Web frame solution described above. But we will not be able to offer quite the same functionality as for single OPACs.