



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 155-E
Meeting: 86. Bibliography
Simultaneous Interpretation: Yes

Future role of (electronic) national bibliographies

Juha Hakala

Helsinki University Library
The National Library of Finland

Abstract

Increasingly, the national bibliographies are available in the Internet and deal with the Internet (Web) content. This paper analyses the impact these changes have, or should have, on the roles of national bibliography databases. As regards the Internet access, support for Z39.50 is seen as essential in order to enable copy cataloguing. Cataloguing of Web resources is analysed from the point of view of traditional library systems, information retrieval portals and digital object management systems.

Introduction

In January 2004 Finland will become the third country in the world, which makes the national bibliography available only via the Internet. Publication of the national bibliography in print was stopped for a quite a few years ago, and in the end of 2003 also CD-ROM version of our Fennica database will become history. The CD-ROM was needed first and foremost for copy cataloguing purposes; since most of our domestic and foreign customers can copy on-line via Z39.50 connection there is no need to continue the CD-ROM publication any more. Many other countries will soon be in the same situation; apart from tradition there is no longer any need to continue publishing the national bibliography in any tangible form.

In this presentation I will first analyse the prerequisites of discontinuation of all tangible forms of publishing the national bibliography. Then, using the Finnish experiences and plans as the starting point, I'll discuss the ways in which the national bibliographies may develop, keeping in mind the increasing investment on cataloguing of electronic resources. In this context implementation of the Finnish national information retrieval portal and the university libraries' digital object management system will have important implications to the way the national bibliography is produced and to the services it can provide.

Access to the national bibliography databases

According to Unni Knutsen's survey made in 2001, 23 national bibliographies were available via the Internet. In addition, 15 national bibliographic agencies had plans to enable Internet access before 2006; of them, at least 9 had already accomplished this aim before IFLA 2003 (Knutsen). The number of national bibliographies connected to the Internet is growing very quickly, at the same time than Web access is becoming the preferred means of access.

This is of course very positive development, but something is still missing. As long as the national bibliography databases are only available for searching via the Web, we can't be fully satisfied, because copy cataloguing is impossible unless the Z39.50 information retrieval protocol (Z39.50) is supported. All the efforts we have done in order to harmonise cataloguing will become truly useful if we make our cataloguing data available to the others in an efficient way; this requires both technical and political agreements.

Next time the state of the art of electronic national bibliographies is analysed, we should find out if copy cataloguing is possible. This can be done by making some additional questions for those national bibliographic agencies which provide Internet access:

1. Is Z39.50 supported?
2. If yes, in which format(s) and character set(s) the data can be delivered?
3. Is Z39.50 access free, or is it necessary to get a contract? If so, is there a fee?
4. Would you be willing to agree on mutual data exchange contracts with other national bibliographic agencies, in which neither party would need to pay?
5. If Z39.50 access is not possible, is your national bibliographic data or at least part of it available via some other system/systems for copy cataloguing purposes?
6. If yes, what are the terms and conditions of using this database/databases?

The national bibliographic agencies who provide Internet access can and should make their bibliographic MARC records available for on-line copy cataloguing purposes with the help of Z39.50. There are of course quite exhaustive union catalogues out there; for instance WorldCat (OCLC) contained in late 2002 about 50 million records, of which roughly 60.000 records were about resources published in Finnish. But our national bibliographies have normally better coverage than international union catalogues, and the quality of records is not as high. Anyone seeking for really good coverage of Finnish publications or just high quality will be best served by the Finnish national bibliography. This situation will of course change if (and only if) Fennica records are loaded into e.g. WorldCat. It seems that few national bibliographic agencies have chosen this strategy for the time being; it would be interesting to know if NBAs have any plans for closer co-operation with e.g. OCLC.

Z39.50 and national bibliographies

Z39.50 support is nowadays a standard feature in every good integrated library system and IR portal such as MetaLib. The problem is that in spite of the existence of internationally approved guidelines for implementing Z39.50 – the Bath profile (Bath) – the integrated library systems still do not “talk” with each other too well. There are many reasons for this, including the following:

- Not all national bibliographies are using MARC21 (the format mandated by the Bath profile) internally, and some of those who are not, can't convert the records into MARC21 on the fly for Z39.50-based data exchange. Since manual format conversion is not feasible for a copy cataloguer, this is a serious problem. The best means of solving this problem is usage of a good MARC format converter such as USEMARCON (USEMARCON) as a part of the copy cataloguing process. Then, even if a copied record arrives in "wrong" MARC format, conversion to the local MARC format can be done easily. All you need for this is a set of conversion tables, which in the case of USEMARCON a format expert who is familiar with both the source and target formats can build in a few weeks.
- Our MARC formats use a wide variety of character sets, and sometimes for instance MARC21 records are encoded in different ways. UNICODE will – or may - eventually change this, but a lot of water will flow in the river Spree before every library system supports UNICODE fully. Z39.50 has a brilliant solution to this problem: character set negotiation. But only a few library systems support this feature of the Z39.50 Initialisation service. The quick solution to this problem is usage of USEMARCON or another application capable of converting the character set in an appropriate manner.
- Different Z39.50 targets provide different semantics (search terms). The Bath profile gives explicit guidelines for how to support the standard-defined query terms and term combinations in Z39.50 context. There is no reason for the national libraries to deviate from the profile recommendations; the only exception from this is that in the national bibliography database, given the richness of data elements in it, it may be possible and recommendable to provide additional access points. But even then the Z39.50 standard, and especially its Bib-1 query attribute set will define the ultimate limit for what can be done. Luckily those limits can be pushed further; as it is quite easy to define new terms into Bib-1.
- Different Z39.50 targets and clients support different sets of services and service features (parameters) defined in the standard. Again, adherence to the Bath profile is the best means of avoiding trouble in this area. Luckily copy cataloguing does not require anything exotic; the basic services Init, Search and Present are sufficient.

In order to provide a practical example, let's review the situation in Finland. Our national bibliography database is based on MARC21-Fin, the Finnish extended and modified version of MARC21. University and polytechnics libraries use the same format, but public libraries still rely on the old FINMARC format and ISO 6937/2 character set. However, public libraries have no problems using Fennica for copy cataloguing.

The largest Finnish library system vendors have integrated USEMARCON into their cataloguing clients (and made some improvements into the application while doing that). When a public library using either one of these systems copies records in MARC21-Fin format via Z39.50 from the national bibliography, USEMARCON converts these records on the fly into FINMARC and ISO 6937/2. Predictable outcome of this was that on-line access became very popular, while subscriptions of CD-ROM Fennica diminished. Therefore the decision to discontinue the Fennica CD-ROM was easy; in our situation only the on-line version of the national bibliography is needed. Other enabling factor is the phenomenal advances in computer technology; 10 years ago it would have been impossible to buy a server

which would have been sufficient for copy cataloguing from every Finnish public library; now this is not a problem at all.

For the foreign users we will in 2004 enable Z39.50 access with MARC21 delivery. USEMARCON has also been embedded into the Voyager Z39.50 server, and this allows us to send the bibliographic records in any MARC format and its character set, at least in theory. In practice only a handful of format conversions can be supported by any NBA, since maintenance of conversion tables requires detailed knowledge on both source and target formats. By using MARC21 as the common denominator the NBAs could provide a large set of conversions, enabling basically any national library to copy MARC records directly from any national bibliography via Z39.50.

Usage of Z39.50 supplemented by a format & character set conversion software such as USEMARCON enables efficient information retrieval and copy cataloguing from national bibliographies. Since technology is now to large extent in place there is an urgent need to revise recommendations on national bibliographic services so that Z39.50 guidelines are included. It is odd that this standard, or its ISO version, ISO 23950, are not even mentioned in the final recommendations of the International Conference on National Bibliographic Services from 2002.

Once the technical infrastructure is in place it is necessary to create contracts needed for data exchange between NBAs. International co-operation such as Scandinavian Virtual Union Catalogue (<http://www.lib.helsinki.fi/svuc/>), linking the Nordic national union catalogues into a single virtual service, or ONE Association (<http://www.oneassociation.org>) are aiming at making bibliographic data, often including national bibliographic records included in union catalogue, available internationally.

Since national bibliographic agencies usually have relatively little use for each other's bibliographic records, sharing of this data should indeed be done in a wider context. ONE Association is, I think, a good model for record sharing initiatives. On the other hand, sharing of authority data between NBAs is very relevant, and we must do our best to build so called Virtual International Authority Files. But even though the version 2.0 of the Bath profile does specify how authority data can be made searchable via Z39.50, library systems do not yet have Z39.50 implementations that would support this functionality – not to mention other technical and political advances missing for the time being.

As the example of exchange of authority records illustrates, the national libraries seem to have special Z39.50 related and other requirements that other libraries as a rule do not have, and we would do wisely to form our own vendor-specific user groups in order to foster the development of features essential to us. At least one such group exists already, National Libraries' Endeavor Advisory Board, NLEAB (see <http://www.infotoday.com/IT/nov01/news18.htm>).

And what about the ZING protocol?

Many of you may have heard about the ZING protocol (ZING). This acronym means Z39.50 International next generation. Technically it is quite different from Z39.50, although for the users things may look pretty much the same. The ZING maintenance agency is the Library of Congress, which makes sure the peaceful coexistence of these standards.

The ZING protocol will, to put it simply, take Z39.50 to the Web. Using for instance ZING SRU service, it is possible to send queries encoded into URLs, like this:

<http://www.loc.gov/sru.cgi?query=dc.title=cat&maximumRecords=10&recordSchema=DC&sortKeys=title,dc&startRecord=1>

Although the data looks awful the technology works. For instance Google uses similar technical solution, and also in OpenURL bibliographic data about the resource sent to the OpenURL resolver is packed into a URL. ZING server sends the records back not in ISO2709 exchange format but encoded in XML using Dublin Code schema.

Since most library system vendors have already implemented Z39.50 which has many more services defined in it than the present version of ZING protocol, it is not a good idea to replace Z39.50 with ZING in the library systems. But it may be useful to support both Z39.50 and ZING in order to make library systems more accessible from outside the library domain. For instance, if a national bibliography database were available via ZING, it would be easier for publishers and other memory organisations such as museums and archives to connect their systems into the national bibliography.

According to my opinion, the NBAs should foster the use of national bibliographic data by making our records available in XML syntax via ZING and possibly also via the Open Archives Initiative (<http://www.openarchives.org/>) harvesting protocol (<http://www.openarchives.org/OAI/openarchivesprotocol.html>) will be important milestones in achieving this aim. In a few years our library systems will contain this functionality; by that time, we must be ready to use these technologies.

Future roles for the national bibliographies

According to Unni Knutsen's review presented earlier in this session, national bibliographies may or may not be merged with the national library's OPAC or other bibliographical systems, but there is one thing we have up to now taken for granted: national bibliography is built using an integrated library system and from technical point of view is indeed just one database.

The integrated library system as it is currently known is as of this writing being seriously challenged. Library system vendors and other companies are developing information retrieval portals – Ex Libris's MetaLib being probably the best-known example of these applications – and digital object management systems, or DOMS, such as ENCompass from Endeavor Information Systems.

IR portal can be seen as a major extension of a traditional OPAC. A portal usually enables simultaneous searching of multiple target databases, using different query protocols, record syntaxes and character sets. It allows the patrons to store queries and tell the system to carry them out periodically. Lately it has also become a norm that portals support dynamic linking with OpenURL. Although these services can be integrated into a traditional library system, the vendors have usually built separate applications for them. There are some technical and lots of commercial reasons for this; a library using ILS from vendor A may buy a portal from vendor B, if they do not need to change the library system.

The same technical and commercial background applies for DOMS; a library system might incorporate the required functionality. But since a good DOMS must deal with a wide variety

of document and metadata formats and enable controlled delivery of protected content, a traditional library system needs to be extended heavily to fulfil all requirements.

When libraries install portal and DOMS systems, one integrated system may be replaced with two or three applications, interconnected via standard interfaces. In the Finnish Linnea network, we will from 2004 use Voyager as the library system, MetaLib as the IR portal and ENCompass as the DOMS.

From the cataloguing point of view, the emergence of this new technical environment is interesting for two reasons. First, link information which is vital for description of (electronic) resources should not be put into the traditional library system at all, but into the portal. On the other hand, efficient utilisation of portal and DOMS will require new kind of metadata for which we as of yet do not have cataloguing rules or formats. These changes will have an impact both on OPACs and on national bibliographies.

Let me provide a more detailed description of the first problem. A common solution for providing a link from a MARC21-based system to a Web resource is that of putting Web resources URL into the 856 tag of the MARC record. Then, every time the URL changes we must change the record in every database into which the record has been copied. Unfortunately at least in Finland bibliographical databases do contain quite a few URLs which do not work any more, or worse still, link the MARC record into a different resource than was originally catalogued.

856-based solution is also incapable of solving the appropriate copy problem; that is, they may have multiple copies of the resource in different locations, and one user may be allowed to use one of them, the second an another copy, and the third yet another. Static, MARC-record based linking, simply can't deal with these situations, which will be increasingly common due to licensing arrangements.

The appropriate copy problem can be solved, and the maintenance of URLs simplified, by implementing OpenURL. In this approach, link and license information is stored in OpenURL resolver database; in the Finnish case, into the SFX system incorporated into the MetaLib portal. After this has been done, we must remove the link data from the national bibliography because it would only be counterproductive there. It makes no sense to maintain both static links in the national bibliography and dynamic ones in the OpenURL resolver.

Our conclusion therefore is that national bibliographic data will be divided between traditional library system and OpenURL resolver application such as SFX. Some metadata related to e.g. long time preservation may only be placed into digital object management system. A logical follow-up question is, what metadata in OpenURL resolvers should be seen as belonging to the national bibliography¹?

¹ Of course, OpenURL resolvers may be integrated to traditional integrated library systems, and management of digital assets may become a routine activity in some ILS. At that point we may again have a single integrated system, maybe even a single format or family of them, which is capable of accommodating all the required metadata. But whether such integration happens or not, our concept of the national bibliography needs to be expanded.

Most cataloguers would probably agree that if existing data elements are moved or copied from the library system into OpenURL resolver system, that information is still part of the national bibliography. The contents of the 856 tag are an example of this. Then only those data elements which are not used yet remain problematic. Put in different terms, do we think that the national bibliography should for instance contain license information for electronic resources? My answer to this is conditional yes; when our systems develop in such a way that they can store and utilise license information, we should use this data at least when it is reasonably stable. As an example, terms and conditions of use applicable to deposited resources should definitely be included in the national bibliography.

So, in the near future our national bibliographies may consist of a traditional library system and OpenURL resolver. From the end users point of view these two applications enable a single service, providing dynamic links to electronic contents. These contents may be stored into a digital object management system, which will form the third cornerstone of our future services. We do not yet understand fully all implications of the new application environment; for instance, we do not know how to deliver bibliographic records with link information if this data has been split into two applications. But these open issues will be discussed and eventually solved in such for as IFLA.

References

Bath profile home. The National Library of Canada [on line]. <http://www.nlc-bnc.ca/bath/> (2003-07-10)

Knutsen, Unni: Electronic national bibliographies: state of the art review [on line]. <http://www.ifla.org/IV/ifla69/papers/109e-Knutsen.pdf> (2003-07-10)

OCLC WorldCat home [on line]. OCLC Online Computer Library Center. <http://www.oclc.org/worldcat/> (2003-07-10)

USEMARCON home [on line]. The British Library. <http://www.bl.uk/services/bibliographic/USEMARCON.html> (2003-07-10)

Z39.50 Maintenance agency home page. Library of Congress [on line]. <http://lcweb.loc.gov/z3950/agency/> (2003-07-10)

ZING home page. Library of Congress [on line]. <http://www.loc.gov/z3950/agency/zing/zing-home.html> (2003-07-10)