Scandinavian Virtual Union Catalogue Project 1999-2003

Final report

Juha Hakala
Ole Husby
Tommy Schomacker
Bodil Dalgaard-Møller

January 2004

Helsinki University Library – The National Library of Finland

Table of contents

Introduction

European Union funded in 90s projects ONE and ONE-2, which built international library networks via implementing the Z39.50 Information retrieval protocol (http://www.loc.gov/z3950/agency/). Although these projects led to many tangible technical results (see for instance http://www.one-2.org/), the initiatives did not manage to develop business models which would have enabled mutual exchange of bibliographic data for copy cataloguing and information retrieval. There were two reasons for this; first, both projects were technically oriented and business aspects did not have a high priority. Second, the project groups were large and diverse, especially in ONE-2. The partners of the project were from 8 countries (Austria, Denmark, Finland, Hungary, Italy, Norway, Sweden and UK) and represented libraries, museums and R&D institutions. Apart from technical interests in developing Z39.50 applications the partners had little in common.

Because ONE projects were not concerned with "politics" of record exchange and the European Library project (TEL) concentrated on digital content and not on the bibliographic data representing it, there was a need for an international initiative which would continue the work started in ONE especially in the area of contract development. There were also some areas in which Nordic countries needed additional guidelines in order to establish efficient virtual union catalogue. The Scandinavian Virtual Union Catalogue project concentrated on these two topics.

The idea of establishing a Scandinavian Virtual Union Catalogue (SVUC) from existing physical union catalogues was presented in the Nordic digital library meeting in Helsinki, February 1998. Potential project partners had further discussions about the aims of the project later in 1998, and on the basis of these discussions Helsinki University Library drafted a project plan in Spring 1999. NORDINFO (Nordic Council for Scientific Information) board decided to fund the SVUC project in its meeting on 13 August 1999 with FIM 123.000, or approximately 20.000 €. This was half the sum the project had asked for. Thus substantial amount of the funding came from the project partners, Nordic national union catalogue hosts, which all participated in the SVUC.

The project group consisted initially of:

Juha Hakala	Helsinki University Library (project manager)
Ole Husby	Bibsys
Thorsteinn Hallgrimsson	Landsbókasafn Íslands – Háskólabókasafn
Poul Henrik Jørgensen	Dansk BiblioteksCenter
Jacob Hallen	Libris
Torstein Tjelta	Nasjonalbiblioteket

Once the new Aleph-based Gegnir union catalogue went into operation in Iceland, the license negotiations were held with the organisation responsible for the system, Landskerfi bókasafna (The Icelandic Library Consortium Inc.). After Jacob Hallen left the Libris department of Kungliga biblioteket, the Swedish contact person was Mats Herder.

The following databases were included:

Bibsys	Union catalogue of Norwegian research libraries
Danbib	Union catalogue of Danish research and public libraries
Gegnir	Union catalogue of Icelandic libraries
Libris	Union catalogue of Swedish research and public libraries
Linda	Union catalogue of Finnish university libraries, The National Repository Library and a number of other research libraries
Manda	Union catalogue of 20 largest Finnish public libraries
Sambok	Norwegian union catalogue of books

These databases contain tens of millions of records and contain the holdings of hundreds of Nordic libraries. They are important national services, which also provide a key access point for foreign users who want to find out more about Nordic library collections. Thus a virtual union catalogue of these databases is a one-stop shop of all important Scandinavian research libraries.

The SVUC project was dependent on local technical development of these systems; no partner had resources to implement a new system just in order to satisfy the needs of the initiative. However, three project partners implemented new integrated library management systems during the project period. Helsinki University Library and Libris acquired Voyager from Endeavor Information Systems, Inc, while the National and University Library of Iceland chose Ex Libris' Aleph. These systems went into production in 2001-2003; however, Libris continues to provide Z39.50 access via a proprietary system built for this purpose. Bibsys, DBC and Nasjonalbiblioteket have locally developed systems, which are constantly enriched with new features.

For Finland and Iceland new systems were essential from the SVUC point of view. The old systems used in Gegnir and Linda (Libertas and VTLS Classic) could not satisfy the Z39.50 requirements of the project. When the Aleph system in Iceland went into production in Spring 2003 Nordic countries finally had the technical infrastructure needed for an efficient virtual union catalogue, at the Z39.50 server level. In addition to providing accurate technical guidelines SVUC provides a solid contractual basis for co-operation. Organisations, which belong to a union catalogue consortium in one country, are able to use all (other) Nordic union catalogues for free for copy cataloguing, information retrieval and ILL localisation. Strengthening the Nordic co-operation via better access to key bibliographical resources was the main aim of the project, and in this respect the project has definitely been a success. Moreover, in the form of the ONE Association (see http://www.oneassociation.org/) similar co-operation has already spread beyond Nordic countries.

Project tasks

Creating a virtual union catalogue using Z39.50 is a complex task, both technically and from a contractual point of view. Although ONE projects and other past Z39.50-driven initiatives such ad Nordic SR-nett had created a solid technical platform for SVUC, and the old proverb "the devil is in detail" applies well to utilisation of a complex protocol such as Z39.50. Much time and effort was spent in SVUC to solve issues which are hard even to explain to those who do not have deep understanding of the protocol. But although their origin is shrouded in mystery for the non-initiated,

these problems could still render especially copy cataloguing non-operational or useless because the records delivered were not good enough for copying.

Project tasks were the following:

1. Create a Scandinavian Z39.50 profile
2. Improvement of MARC conversion capabilities
3. Improvement of Z39.50 server and client functionality
4. ISO ILL profiling in Nordic level
5. Contract negotiations
6. Documentation and project management

These tasks will be discussed below. Then the present situation in each country is given in more detail. Before the national status reports a review of Z39.50 implementation issues is provided. This section of the text is based on the experiences from SVUC and earlier Z39.50 initiatives.

Creation of a Scandinavian Z39.50 profile

Prior to the SVUC all Nordic countries except Iceland had developed national Z39.50 profiles. The main reason for this was a general understanding among implementers that a complex protocol such as Z39.50 can not be implemented successfully unless detailed implementation guidelines are provided. Lack of profiling led invariably to applications that could not communicate with one another; an ironic result but one which may happen if a protocol contains a lot of functionality. Alas, if the national profiles were different the dismal interoperability results did not improve a bit; therefore a great need for co-operation in the development of national profiles was foreseen.

As it happened, the SVUC project avoided to a large extent the need of co-ordinating the further development of national profiles. The reason for this was the creation of the Bath profile, the ISO-approved International Standardised Profile for the ANSI/NISO Z39.50 / ISO 23950 standard. After the acceptance of the Bath profile there was no reason for domestic profiles, except in the area of semantics which will be discussed below in the Using Z39.50 chapter.

One reason why the Bath profile was eminently suitable for SVUC was that two members of the SVUC team (Juha Hakala and Poul Henrik Jørgensen) participated actively in the early stages of the development of the Bath profile. This guaranteed that the profile was a good match for the Nordic libraries. The SVUC agreement requires support for the version 1.1 of the Bath profile, which was the current one when the agreement was written. Compared with the standard profile, some additional requirements have been made in the agreement concerning the use of other MARC format than MARC21 (which is obligatory) and Unicode instead of Latin-1. The partners may also choose to support the ONE-2 profile, which is a superset of Bath profile version 1.1.

As a part of their SVUC activities, Finland and Norway have created lists of the national Z39.50 Bib-1 attribute set use attributes. In Denmark, this work was co-ordinated by the Danish National Library Authority (Biblioteksstyrelsen); the Dan-1 attribute set contains 27 attributes and is available from http://www.bs.dk/danzig/dan-1.htm.

It can be claimed that whatever Z39.50 profiling remained to be done in Nordic level after the development and global acceptance of the Bath profile, has been done in the SVUC. Of course this applies, like the SVUC itself, to the basic bibliographic search and retrieval. More advanced profiling issues such as ILL and document delivery implementation guidelines are still missing, not

least because in some Nordic countries these protocols have not been implemented yet. But ONE Association will most likely be the forum in which this profiling – complementing for instance the ISO IPIG profile for ISO ILL standard and guidelines for implementing NISO Circulation Interchange Protocol, NCIP – will take place.

Improvement of MARC conversion capabilities

At the time the SVUC project was launched in 1999 MARC conversion was a major problem in more than one way. First, converting records from one MARC format to another is technically complicated. Since each Nordic country had its own national format a lot of conversions were needed for creating a viable Nordic union catalogue. Second, there were at least two public domain tools for record conversion, but there was not an obvious best choice; both MARCconv (see http://home.hio.no/~michaelp/english/Professional/MARC/) and USEMARCON (http://www.bl.uk/services/bibliographic/usemarcon.html) had pros and cons.

During the life time of the project format policies in Nordic countries have drastically changed. Three countries have made a decision to discontinue national MARC formats in favour of MARC 21. (More details about this in the Using Z39.50 chapter.) Thus the need for MARC conversion has diminished, but there is still a need for converting MARC records, and good stand-alone and/or embedded tools for doing this.

The SVUC project proposed evaluation of MARCconv and USEMARCON, the two freely available MARC converters available in the late 90s. This task was of little interest to DBC and Bibsys, since both had developed MARC converters for their own use. In Iceland MARC21 conversion was done by the software vendor, Ex Libris. So, Helsinki University Library, which needed a tool for converting the 15 million FINMARC records in VTLS databases to the new MARC21-Fin format as a part of the Voyager implementation project, evaluated the two MARC converters.

According to our quick analysis which was mainly based on software documentation and user guides, technically the two converters were more or less on the same level, although USEMARCON's flexibility and versatility was impressing. But choosing an application is never based on technical factors only; in this case the crucial factor was the British Library's decision to invest on further development of USEMARCON, and make the improved application and its source code available as shareware, just like the earlier USEMARCON version had been. Although the experts in Helsinki University Library did at that point discuss with the decision makers in the BL and tried to convince them that the new USEMARCON must remain shareware, in the end of the day there were two crucial factors which guaranteed that USEMARCON remained freeware. First, there was a possibility that other vendors might develop a free application further (which has actually happened). Second, due to lack of personnel resources the BL was not willing to commit itself into supporting the USEMARCON users. However, the library has established a discussion group via which the users of the converter may discuss with one another.

MARCconv has not been updated in recent years, although it is still available via the Web. But there has been a lot of investment on USEMARCON, which benefits those libraries and library system vendors who use the application. First the British Library funded development of a USEMARCON version that can be embedded into on-line applications such as Z39.50 clients and servers. Because it remained shareware, a number of software vendors, including Endeavor Information Systems, Inc and two Finnish library system vendors have incorporated this USEMARCON version into their Z39.50 applications. One of the Finnish library system vendors,

ATP, made substantial improvements to USEMARCON in 2003; Helsinki University Library participated in the tests of the improved USEMARCON together with the British Library. The present version, 1.4, is a robust and versatile converter which is more reliable and faster than the earlier USEMARCON versions.

Libraries and library system vendors who needed a MARC converter were in less than optimal position when SVUC started; it was known that building a good tool from scratch would require approximately 1-2 man years. Alas, the two shareware converters had functional flaws; for instance, USEMARCON could only be used for batch conversions; there was no way of integrating it into on-line applications. BL-funded revision by Crossnet eliminated this problem, but there were still flaws in the application, which manifested themselves as memory leaks and other technical deficiencies. In USEMARCON 1.4 all known problems have been fixed, so any library system vendor, or a company building portals or Z39.50 clients has at least one good choice. Whether there is any need for format conversions between MARC formats five or ten years from now remains to be seen; a significant share of national MARC formats have disappeared during the last years, and the same trend seems to continue.

Instead of converting from one MARC format to another we may need to convert from MARC to other formats such as Dublin Core and ONIX (or vice versa), or from traditional ISO2709-defined MARC syntax to MARCXML or the new XML-based generic MARC syntax which is as of this writing being developed in ISO TC46. It is to be hoped that USEMARCON will be expanded in the future in such a way that it can carry out a broader choice of conversions.

Improvement of Z39.50 server and client functionality

SVUC project did not have much funds for application development. Most work done in this area was done by local funding and is described in country reports.

As a rule, implementing Z39.50 has not been easy for library system vendors. A quick proof of this is that most systems still have poor support of the Bath profile. In other words, the vendors have implemented only a small part of services specified in the standard.

In order to simplify the task of implementing the protocol vendors often choose to utilise a freeware Z39.50 protocol machine. At least three applications have been popular: Isite by CNIDR, Center for Networked Information Discovery & Retrieval (http://www.cnidr.org/isite.html), YAZ by Index data (http://www.indexdata.dk) and the ZedLib toolkit built by Crossnet Systems Ltd. (http://www.crxnet.com/) initially as a part of the ONE-2 project. Isite has not been developed since version 1.04 was made available many years ago and does not provide full implementation of the standard. ZedLib and YAZ do, and are compliant with the Bath profile. Alas, the former does not seem to be available as freeware any more.

Some U.S. library system vendors are still using Isite (including Endeavor); for them, the first step towards better Z39.50 functionality would be to replace this tool with a modern application such as YAZ. This would improve both reliability and interoperability with the Bath profile, and provide a good basis for implementation of Z39.50 version 3 services required by the interoperability levels 1 and 2 in the profile.

ZedLib is of major importance for the SVUC project since it has been embedded into Z39.50 implementations in Bibsys, Danbib and Libris (which does not provide free access via Z39.50 to its Voyager Libris database, but to a copy of it maintained in an another technical environment). Ex

Libris uses YAZ in Aleph, and Helsinki University Library has discussed with Endeavor about replacing the present Voyager Z39.50 protocol machine, Isite, with YAZ. Although Endeavor is committed to using YAZ in ENCompass, the company's digital object management system, they have not promised to incorporate YAZ into Voyager, partly because the company has invested a lot in refining Isite.

Z39.50 servers

From the software development point of view, SVUC partners fall into two categories:

- Finland and Iceland use an off the shelf commercial library system and its Z39.50 server; during the project both changed their application (from VTLS to Voyager in Finland; from Libertas to Aleph in Iceland).
- Bibsys, Nasjonalbiblioteket, Libris and Dansk BiblioteksCenter use locally developed systems, which include Z39.50 servers with some advanced features.

In Iceland, no special Z39.50 related demands were made to Ex Libris due to SVUC. In Finland, one of the crucial requirements was integration of USEMARCON into the Voyager Z39.50 server; this was successfully done by Endeavor. Alas, some other Finnish Z39.50 related demands such as implementation of Scan are yet to be implemented. Most Finnish requirements were Bath profile features missing from Voyager. Endeavor, like many other library system vendors, aims at making their product compliant with the profile, but it is not clear when this will happen. It is clear that poor support for the profile in Aleph and Voyager (see Manojlovich) is a problem for the SVUC.

From SVUC point of view, basic Z39.50 server-related requirements have been fulfilled via implementation of new systems by those partners who did not have a satisfactory system in the past. In the future those partners who rely on locally developed applications must keep them up to the level of the new versions of the Bath profile (version 2.0 is as of this writing being voted in ISO TC 46). And those partners who use commercial systems must continue telling their vendors that Z39.50 functionality is important to them and other users who have made their databases available via the Internet. In fact, the growing popularity of Information retrieval portals such as MetaLib makes Z39.50 more important than ever.

Z39.50 clients

All SVUC partners have just one Z39.50 server via which the services are available; the situation is less clear as regards Z39.50 clients since technical solutions differ from one another. Depending on the country, access to SVUC may be centralised (Denmark) or decentralised (Finland and Sweden). From the application point of view, they may be stand-alone clients such as BookWhere (http://www.webclarity.info/products/bookwhere.html) & Icone (http://www.crxnet.com/icone.php), or Z39.50 clients embedded into other applications such as IR portals, or different modules of the library system (acquisitions / cataloguing client, OPAC). These applications may be locally developed or built by a commercial vendor.

The central requirement for any Z39.50 client used in a virtual union catalogue environment is the ability to open a session simultaneously to multiple remote databases residing on different servers. The protocol as such does not define such a feature, so a client must be able to handle multiple parallel sessions at the same time; not a trivial requirement. Many Z39.50 clients can not do this. Another problem is that maintenance of the resource descriptions of the remote databases is time consuming. If this work must be done separately in each client installation, there is no way a user

could configure and maintain descriptions of hundreds of remote Z39.50 servers and databases. In Finland there was also a local problem: downloaded records had to be converted to the MARC21-Fin format, which made usage of the Voyager Acquisitions and Cataloguing clients cumbersome. Because of this crucial shortcoming, and the difficulty of maintaining database descriptions, Helsinki University Library had to select an another Z39.50 client tool for the libraries of the Linnea2 consortium.

After thorough testing of Icone (a freeware Z39.50 client built by Crossnet in the ONE-2 project) and BookWhere the decision was made to purchase the latter for the Linnea2 consortium. Compared with its freeware peer, BookWhere was more reliable and easy to use, and there was support available. WebClarity, the vendor developing the product, made in November 2003 a decision to support FINMARC in the next release of BookWhere on the request of Helsinki University Library. Beta version of this BookWhere version was delivered to the library in January 2004; beta testing has been completed successfully. The vendor is also investigating the possibility of integrating USEMARCON into BookWhere. This would enable further rationalisation of copy cataloguing process in the Linnea2 network.

In the SVUC project plan BookWhere was mentioned as one option of organising access to SVUC databases. Its downside is that unlike ICONE its not free. But stand-alone Z39.50 clients are as of this writing mainly used for copy cataloguing purposes by library staff. Therefore the cost of buying a commercial tool is small, even in a consortium such as Linnea2 which needs hundreds of copies of the Z39.50 client. Patrons and non-cataloguers among the staff generally use Z39.50 via IR portals, without even knowing that they dependent on the protocol when utilising remote databases efficiently. Thus for patrons' point of view it is important to make portals even more efficient tools for searching information. The key factor in this is, in addition to the good Z39.50 implementation, the quality of the portal metadata, which will be discussed later in Using Z39.50 chapter.

ISO ILL profiling in Nordic level

SVUC project did not develop a Nordic ILL profile, for several reasons. First, there is an international profile for the ISO ILL standard, the so called IPIG profile (see http://www.nlc-bnc.ca/iso/ill/ipigprfl.htm). Version 1.0 of the profile was published 15 September 1999 (shortly after the SVUC project was launched), version 2.0 was released 10 April 2001 and version 3.0 in 11 July 2002. There is a dedicated group of experts, which has spent a lot of time on extending the profile. Because the IPIG profile is approved by ISO and it was constantly being revised, there was no demand for Nordic superset of the profile.

Second, there were European profiling efforts in this area. In addition to the Z39.50 ONE-2 project dealt with ILL and electronic document delivery. Thus the ONE-2 profile, which is a superset of the Bath profile, contains also ILL related implementation guidelines. In most cases the ONE profile provides an alternative, Z39.50-based means for ILL messaging; this was an interesting choice for those partners who could not support the ILL protocol.

Third, while some SVUC partners had ISO ILL –based services in operation, some other partners had not implemented the protocol yet. For instance in the Finnish Linnea2 network, the university library directors did only in Autumn 2003 make a tentative decision to acquire the ILL module for Voyager, which had not lead to purchase by the end of 2003 since Endeavor changed the pricing of the product. ILL module acquisition for Linnea2 is therefore once more open to negotiation. Thus a truly Nordic ISO ILL network is yet to be established. And that will not be the end; there is a relevant new protocol, NISO Circulation Interchange Protocol (NCIP), which complements ISO

ILL by defining how information about patrons and items can be exchanged between systems. This protocol has been implemented by quite a few library system vendors, and it is to be hoped that the future SVUC complements basic ISO ILL services with NCIP.

Contract negotiations: the SVUC agreement

One of the main aims of the SVUC was to create a solid contractual basis for the Scandinavian virtual union catalogue. The central idea in preparing the agreement was that database access and usage had to be free for the consortium members. In other words, each partner would pay with its own data for the right to use all the other databases for information retrieval and copy cataloguing.

This was not a trivial thing to achieve, since three SVUC databases (Danbib, Linda and Manda) were available only for fee. Thus, in approving the idea of common free access, Dansk BiblioteksCenter and Helsinki University Library had to accept the fact that there would be no more income from non-domestic Nordic users. For DBC the financial impact of this change was substantial, for HUL moderate, but nevertheless something that had to be agreed upon in house. In the end it was clear both for HUL and DBC that efficient Nordic co-operation was more important than internal budgetary aspects.

Financial issues were not the only ones that had to be discussed thoroughly. The SVUC agreement – available via the project home page at http://www.lib.helsinki.fi/svuc/ - provides some details about the technical implementation and thus constitutes a simple SVUC Z39.50 profile. The agreement also specifies how new databases and host organisations can join SVUC. The project group saw SVUC as a platform that could be extended into a European initiative. This vision paid off when the SVUC principles were adopted by the ONE Association, where SVUC partners (with the exception of Nasjonalbiblioteket and Iceland) were founding members.

SVUC was the first international Z39.50 implementation project, which was able to solve the business aspect of the exchange of bibliographic records in a network where some databases where available for free and some for fee. Hopefully similar idealism will also pervade in the ONE Association which faces the same problem than SVUC: some, but not all target databases, are available for fee.

All SVUC partners signed in 2002 the agreement, which was in effect to the end of 2003. In December 2003 a new agreement, which extends the service for two years beyond the end of the project, was sent to the partners. This will guarantee continuation of SVUC, and enables seamless shift to ONE Association in 2006. On the other hand, SVUC may continue even if all its partners were members in ONE Association, since the service may incorporate services not available in wider European context.

The new agreement did not only extend the service to the end of 2005; there was one important change in the content. In Iceland Gegnir database has been converted from Libertas to Aleph. A new library consortium, Landskerfi bókasafna, is responsible of the system. This consortium replaces the National and University Library, which hosted the old system, as the SVUC service partner. The consortium has received green light from the library for participation in SVUC. More importantly, the consortium was willing to join SVUC; the earlier decision to participate was made internally in the national library, and did not bind the new consortium in any way.

Documentation and project management

Helsinki University Library has established the SVUC home page at http://www.lib.helsinki.fi/svuc/. The page will be maintained after the project is finished at the end of 2003 as the home page of the SVUC service, but HUL will be dependent on other partners sending updates and e.g. revised Z39.50 server configuration information.

The final report of the project (this document) compiled by HUL is available only in an electronic form via the SVUC home page. No printed report will be published. If there is sufficient interest towards the project, a condensed version of the report may be published in an appropriate professional English speaking journal. SVUC partners may also report the project results in domestic professional journals.

HUL will contact the Library of Congress Z39.50 maintenance person and ask him to link the report and/or SVUC home to the Z39.50 home page.

The SVUC agreement will be maintained by Helsinki University Library. Each country will be responsible for maintaining the national Z39.50 attribute set; when published in the Web, HUL will link them to the SVUC home page.

Due to the limited funds the project group had just one single face to face meeting, which was held in Trondheim. The project manager has met most other partners during other meetings such as conferences and discussed the project issues.

The initial aim was to finish the project in 14 months, but this schedule was not realistic because national library system implementation projects especially in Iceland were delayed. Moreover, there were some national tasks that could not be accomplished within the assumed schedule; for instance in Finland Endeavor did integrate USEMARCON with Voyager Z39.50 server, but the work was completed late. This is one reason why Helsinki University Library has not been able to finish the development of the USEMARCON MARC21-Fin – MARC21 conversion table. There are also some political reasons; sometimes it is difficult to know which tag to use for a data element which does not have an obvious home in the target format. Because of this “last mile” problem HUL is not yet capable of delivering records precisely according to the SVUC requirements. In January 2004 the decision was taken to make the conversion available in February 2004, even though we can not guarantee that it is fully correct.

Using Z39.50 in a virtual union catalogue: generic issues

Pragmatics

Z39.50 is a complex protocol which defines a large number of services. Some of them are relevant in a virtual union catalogue setting, while some others may not be. And some have been widely implemented while some have remained unpopular.

Every Z39.50 application must support Init, Search and Present services. Init has a lot of functionality which would be highly useful (e.g. character set negotiation, access control) but are seldom supported. The issues related to search and present are discussed in the next chapters (semantics and syntax).

Scan service allows a patron to browse indexes of a remote database. In SVUC setting the value of this service is limited, because many target databases do not support Scan yet. And even if Scan had been implemented, a user would need to browse each resulting index listing separately, unless the client were able to merge the listings. This might be difficult, as the structure of the terms may differ, and finding out in each case how to sort them is not trivial. Thus the value of Scan in a virtual union catalogue seems limited.

The same applies to Sort. Even if each individual result set were sorted, the Z39.50 client would still be required to sort (and de-duplicate) the sum of the result sets. This would be difficult, if a) the individual result sets are large, or b) the record syntax is not the same. A MARC record and a SUTRS (Simple unstructured record syntax) record may look similar in card view, but for a computer sorting the former with e.g. author or title is easy, while the latter can not be dealt with since from the Z39.50 client point of view the record is just plain text with no internal structure.

In a virtual union catalogue it is important and at the same time difficult to learn to know the target databases, both in regard to the content and to the Z39.50 search parameters. Thus the Explain service would be very relevant, both for humans and for Z39.50 applications. Unfortunately there have been no new Explain implementations by commercial vendors, and interest towards implementation of this service seems minimal. This means that countless hours will be lost when librarians and library patrons are trying to learn the properties of remote databases via trial and error. Configuring one remote database into a portal well may take up to one week; if this remote system had a good Explain database, the same (or probably better) configuration data could be extracted from it automatically within a few seconds.

Bath profile Level 1 does require Scan, and many SVUC partners have either implemented the service or made it clear for their vendors that his service is relevant. This is especially true for Finland; complex morphology of Finnish language makes index scanning a useful feature in any information retrieval system. Some Finnish words have short or multiple roots, which means that usage of truncation is difficult if not impossible. In such cases browsing of indexes may be a useful feature.

Project such as SVUC must also pay attention to how the Z39.50 services are implemented. Stand-alone Z39.50 clients are traditional protocol implementations in that when a connection has been initialised, it is kept open until the user closes it, or timeout occurs. The downside of this is that the server must be capable of handling large number of sessions, most of which are inactive. This approach does not scale well to long time outs (beyond 10-15 minutes); a normal OPAC session lasts only a few minutes and extending that to an hour and a few minutes with 60 minutes would mean a huge increase in the number of operating system processes. This would eventually consume all memory and disk cache, and a server would grind to an abrupt halt.

Some portals, including MetaLib which goes into production in Linnea2 network in February 2004, solve this problem by disconnecting the server each time a query has been done, even during Z39.50 sessions (in HTTP session such disconnection is taken for granted). The status of a “session” is held within a portal database, independent of the access protocol. The portal is also immune to the timeout variation in the chosen target databases.

This approach enables long timeouts, since it is not difficult to store a lot of session information into a database and keep it for instance an hour, which for the user looks like a one hour timeout for every target system. The bad news is that if Z39.50 is used for accessing a remote system, there is a

lot of overhead each time a user makes a search during a “session”. From the server point of view each query comes from a new user which then disconnects after the query is done. In a typical Z39.50 server Init is a heavy load since the database server may need to spawn multiple processes (in Voyager one for the Z39.50 server, and another one for Voyager and yet one for keyword searching) and authenticate the user, possibly against a central identification system. If this kind of usage of Z servers gets common, there may be a significant load increase in library system servers.

NISO Metasearch Initiative (http://www.niso.org/committees/MetaSearch-info.html) will investigate the problems of virtual union catalogues, and propose solutions for them. According to the home page of the initiative:

> *One-search access to multiple resources will allow enabled libraries to offer portal environments so all library users can enjoy the same easy searching found in web-based services like Google. Unlike Google, however, a library's metasearch services must offer access to multiple database targets with a variety of content and accessibility. Metasearch services rely on a variety of approaches to search and retrieval including open standards (such as NISO's Z39.50), proprietary API's, and screen scraping. However, the absence of widely supported standards, best practices, and tools makes the metasearch environment less efficient for the system provider, the content provider, and ultimately the end-user.*
>
> *To move toward industry solutions NISO is sponsoring a Metasearch Initiative to enable:*
>
> - *metasearch service providers to offer more effective and responsive services*
> - *content providers to deliver enhanced content and protect their intellectual property*
> - *libraries to deliver services that distinguish their services from Google and other free web services.*

This initiative will be of vital importance for the future of SVUC and other virtual union catalogues. Two SVUC partners (Libris and HUL) are already closely involved with the Metasearch Initiative, and it is to be hoped that the others are able to join the initiative as well, so that Nordic know-how of implementing virtual union catalogues and portals can be utilised in this work, and we get better awareness of the work others have done in this area.

In the future virtual union catalogues may be based on both Z39.50 and ZING, Z39.50 international Next Generation (http://www.loc.gov/z3950/agency/zing/zing-home.html). As of this writing this protocol is still under development, and compared with Z39.50 it is very basic; it facilitates only search and retrieval of records via a single service. Other services will be added in the future. Although there are important similarities between Z39.50 and ZING (e.g. result sets and search attributes specified in the protocol) the latter is much more a mainstream protocol (see http://www.loc.gov/z3950/agency/zing/srw/background.html) . From the point of view of the Web community, the key benefit of ZING is that it is stateless and connectionless, and thus more acceptable Z39.50 which is both stateful and connection-oriented.

Moreover, instead of ASN.1/BER (Abstract Syntax Notation One / Basic Encoding Rules, ISO Open Systems Interconnection standard for representing data in hardware and software independent manner) encoding and multiple record syntaxes including MARC, ZING builds upon XML, both as

abstract syntax (encoding of messages) and as record syntax for bibliographic data. XML encoding of Dublin Core, Onix, MODS, and MarcXml records will be supported.

Some SVUC partners – most notably DBC – have been actively involved with the development of ZING. This close involvement is likely to continue, since the importance of ZING is gradually growing. Early participation in the protocol development will enable the Nordic countries to plan in advance the possible shift from the old search and retrieval protocol to the new one. This shift is not likely to be an abrupt one; most likely there will be a long period during which most protocols are supported in parallel.

ZING provides little added value for interoperability between integrated library systems; in this area Z39.50 at least for the time being does the job better than the present ZING version can with its limited set of services. But there are three reasons why ZING is nevertheless of major interest for forward-looking virtual union catalogue projects.

First, ZING is a better starting point for cross-domain (e.g a system covering libraries, archives and museums incorporates three domains) interoperability than Z39.50, because the additional burden of implementing ASN.1/BER and building a supposedly complex stateful server are avoided. Moreover, the partners can agree upon using XML as the syntax for exchanging bibliographic (or other) data, instead of implementing ISO 2709 or toying with antiquated record syntaxes defined in Z39.50 such as SUTRS (Simple Unstructured Record Syntax).

Second, rightly or wrongly Z39.50 has always been regarded as a library standard, and ZING may avoid such a burden. Whether this is indeed the case remains to be seen; for the time being the protocol has not been popular among the library system – or other - vendors.

Third, although support for Z39.50 in library systems is ubiquitous, some new systems being implemented by libraries – most notably digital object management systems – as a rule do not support MARC and therefore the vendors developing these systems (some of whom are not traditional library system vendors at all) will probably prefer ZING, especially if the application is XML driven. An example of such an application is ENCompass from Endeavor Information Systems, Inc (see http://encompass.endinfosys.com/). Endeavor intends to implement ZING in this system in order to guarantee interoperability with portal applications; there are no plans to implement Z39.50 since it does not fit well together with an XML-based system.

Semantics

Definition of use attributes

In Z39.50, it is possible to make structured searches; i.e. queries by author, title, subject, or any other search attribute defined in the Bib-1 attribute set or other attribute sets defined in Z39.50. The protocol defines 19 attribute sets (see http://lcweb.loc.gov/z3950/agency/defns/oids.html#3). It is not possible to make queries with something which is not defined as an attribute; thus at the time Z39.50 was published in the late 80s it was not possible to make searches with the notations of the classification used by the Finnish public libraries (YKL) since YKL had not been specified as an attribute in Bib-1.

Z39.50 Bib-1 attribute set (see http://lcweb.loc.gov/z3950/agency/defns/bib1.html) is constantly updated on the request of the standard implementers. As of this writing (December 2003), the

attributes have been last updated in August 2003. Over the years Bib-1 has become a complex set of terms needed in bibliographic databases; the set defines 223 use attributes, or bibliographic search terms. Most of them are global, but there are nevertheless several which are domestic by scope – good example of this are national bibliography numbers.

The term updates are indeed by now often based on domestic needs; any country using Z39.50 aiming at provision of exhaustive bibliographic search services must check if the current version of the Z39.50 Bib-1 attribute set covers all necessary search attributes. Outside the U.S.A, most often this is not the case, since the original Bib-1 was heavily based on USMARC, AACR2 and the needs of U.S libraries. For instance, if there are national subject heading lists or classification systems, the object identifiers for these search terms must be added to Bib-1. If the domestic attributes are not specified, it is not possible to make accurate queries with the notations and subject headings taken from these systems in Z39.50 context. It is possible to search these terms with Bib-1 attribute Any, which by definition incorporates everything. However, the semantics of such a query is not accurate and might lead to very poor precision.

Object identifiers (OIDs) are used in a variety of protocols. In Z39.50 context, OID is the means with which the Z39.50 clients and servers recognise search terms specified in queries. The server does not know that a user wants to make a title search unless there is a means of indicating this in the query. Object identifiers are, basically, strings of numbers. They are allocated in a hierarchical manner, so that, for instance, the authority for "1.2.3" is the only one that can say what "1.2.3.4" means. The formal definition of OIDs comes from ASN.1, or ISO/IEC 8824:2002 (see http://www.alvestrand.no/objectid/ and http://asn1.elibel.tm.fr/en/tools/tutorial/). There are three alternative ways of expressing OIDs: ASN.1 notation (e.g. {iso(1) member-body(2) us(840)}), dot notation (e.g. 1.2.840) and URN notation (e.g. urn:oid:1.2.840). Z39.50 uses the dot notation.

Z39.50 standard has OID value 1.2.840.10003 (in the OID hierarchy, 1.2.840 refers to US organisations which are ISO members). Within Z39.50, OID 1.2.840.10003.3.1 means Bib-1 attribute set, and OID 1.2.840.10003.3.1.4 identifies the fourth search term (Title) within Bib-1. So, Title has the use attribute value 4 within Bib-1. When a Z39.50 client sends a title query, OID and the search term are packed into the search string represented in ASN.1/BER encoding. In addition to the use attributes a search string may include some other attributes defined in Bib-1 such as truncation.

In Bib-1, 10.0000 use attribute values can be defined. Upper half (5000-10000) have been preserved for private usage; they will never be assigned by the Library of Congress, the Z39.50 Maintenance Agency. Within the SVUC project, Bib-1 use attribute values 6000-6099 have been reserved for Denmark, 6100-6199 for Finland, 6200-6299 for Iceland, 6300-6399 for Norway and 6400-6499 for Sweden. Hundred attribute values per country allows a large number of domestic attributes to be defined. In this way we can avoid "attribute collisions"; search with the Finnish YKL classification will not be interpreted as being something entirely different in a Danish Z39.50 server because the same OID has been assigned to a different attribute in Denmark.

The SVUC approach is not scaleable to global level, since there are only 5000 OIDs available in the private area of Bib-1. But there are enough numbers for CENL countries. The CENL Working Group on Networking Standards has expanded the SVUC strategy to European level and divided the remaining Bib-1 private OIDs between the European countries. For instance, Albania has got OID values 6500-6599, Belgium 6600-6699, and so on. The list, maintained by the Helsinki University Library, is available at http://www.lib.helsinki.fi/cenlwg/bib1OID.html.

Since Z39.50 allows an unlimited number of attribute sets defined, the best means of defining domestic attribute sets would be specification of domestic attribute sets. The problem with this approach is that Z39.50 clients and servers supporting Z39.50 version 2 only do not support combination of attributed from multiple sets into a single query. Therefore all relevant terms must be defined in Bib-1. Finland and Denmark have specified domestic attribute sets, so that when all applications support version 3 of Z39.50 a more advanced attribute architecture can be used. Until this happens all search terms must be added to Bib-1.

Semantic interoperability

Semantic interoperability, or the lack of it, is the greatest challenge of Z39.50-based and other virtual union catalogues. If the target databases contain structured metadata, often the common denominator between the targets is surprisingly small. For instance, the ONE I project found out in the early 90s that the only search term that could be used to query all ONE target databases at the same time were individual words from titles. Although the situation has improved since then due to emergence of the Bath profile – which pays a lot of attention to semantics – we are still very far from an ideal situation.

There are many reasons for the lack of semantic interoperability. The following list is not exhaustive, but includes most of the key issues.

1. Lack of data. Distributed searching may be impossible because the search term is not available. Because subject headings and classification systems vary from one country to another, missing data is a common problem for subject searches. A patron wishing to use YKL classification or Finnish subject headings for subject search, must limit the query to domestic databases. There are initiatives building multilingual ontologies; once these systems contain sufficient number of terms and portals can use them in order to translate the queries subject searching across national borders will become easier. In SVUC we did not investigate this topic any further. Neither did the project do anything to harmonise cataloguing; harmonisation of cataloguing rules and cataloguing must be done in IFLA (rules) and the national level (cataloguing practices). A project such as SVUC can only underline the importance of harmonisation for high quality virtual union catalogues.

2. Variation in data content. Although inter-indexer consistency is an important problem it is not the one discussed here, but lacking investment in authority control. Due to different national traditions (and languages) countries have different names for personal authors, organisations and uniform titles. Dostoevsky and Cehov have just one name written in Cyrillic, but there are a lot of different Romanised name forms for them, and none of them is wrong. The only viable solution for this is authority control; we need to create authority records which contain all valid name forms and – in case of organisations – their names in as many languages as possible. This requires a lot of work, and therefore libraries need to build mechanisms for sharing authority records on the global scale. Only then will a user find books by Plato also from Finnish databases, in which the name of the author is Platon.

3. Differences in search engines. Even if the bibliographic data is exactly the same two library systems may yield non-identical search results. One may have a search engine based on fuzzy logic, which enables the system to match the query on index terms which are not identical with the query, while the other system requires exact match. Or, a search engine may be just unable to support a feature defined in Z39.50, such as truncation from the left. The Z39.50 standard does not make any requirements concerning the properties of the

database management system and search engine underneath a Z39.50 server; thus we can not assume similar behaviour from all systems.

4. Differences in Z39.50 clients and servers. Building a Z39.50 application is a challenging task. Although some vendors have written everything by themselves, many library system providers have chosen to adopt and modify an existing Z39.50 protocol machine. YAZ from Index Data is a common choice, since it has a large and active user community and it incorporates basically everything from Z39.50 version 3. But YAZ does not dominate the market. The Z39.50 protocol machine applications used as the basis for implementation differ, and as the vendors have chosen to implement different sections of the protocol. It is therefore not surprising that there are significant variations in the behaviour of Z39.50 applications, in spite of persistent profiling efforts.

5. Differences in indexing. A common reason for variation in search results is, however, that different indexes are created from the raw bibliographic data in different libraries. Most modern library management systems are very flexible as regards indexing; the operations of the system itself requires only a few indices, for e.g. identifiers. In addition to these obligatory indices a library can build anything from one to even hundreds of indices, and choose freely which data elements are included in them. It is quite possible to not index at all a term which another library deems as vitally important.

6. Variations in mapping Bib-1 use attributes to the database indices. Even if the bibliographic data were exactly the same and even if the data had been indexed exactly in the same way there may still be differences between search results, if the Bib-1 attributes have not been mapped to the indices in the same way. Although there are guidelines for matching the Bib-1 use attributes and MARC fields, it may be difficult to follow the recommended principles if indexing does not allow exact matches. Or, the systems librarians may be unaware of the existence of these recommendations.

Z39.50 profiles, and especially the Bath profile, have paid a lot of attention to semantics. Early on it was assumed that default behaviour could be enforced; for instance, a Z39.50 server could always truncate from the right unless explicitly told not to do that. Alas, quite soon libraries realised that the vendors had not adopted the same defaults. Thus it was necessary to define in profiles exactly which Bib-1 attributes and attribute combinations had to be supported.

The first version of Bath profile made 41 identifiable requirements for conformance levels 0 and 1. Most of these were related to search semantics. According to Slavko Manojlovich (see http://www.sirsi.com/Sirsipdfs/ifla_whitepaper.pdf), in 2001 conformance of library systems to the profile was still quite poor (with the exception of Sirsi). Endeavor's Voyager, which was one of the best systems, fulfilled 12 requirements, while most of the other analysed systems supported less than ten of them. This explains why semantic interoperability between library systems is still quite poor, in spite of the fact that libraries are harmonising their cataloguing and authority control practices and copying records from one another.

SVUC project, unlike ONE, did not carry out a systematic test on semantic interoperability. But tentative results from copy cataloguers in Finland indicate that things have improved since the ONE days; for instance bibliographic identifiers such as ISBN and ISSN have been available for searching in all SVUC databases (and many other remote databases). Although there are less databases in SVUC than in ONE, perhaps the main reason for the better state of the affairs is that Z39.50 implementations have improved gradually during the last decade. Moreover, libraries are

aware of the requirements of the Bath profile and try to achieve conformance with its semantic requirements. Alas, good will alone is not sufficient if the application does not allow conformance with the profile.

As the analysis by Slavko Manojlovich clearly shows, we can not be satisfied with the quality and coverage of present Z39.50 implementations in integrated library systems. Libraries must convince their system vendors that additional investment on Z39.50 is necessary, because at present information retrieval and copy cataloguing from remote databases is not as efficient as it should be. Unfortunately, this problem has been much less critical for the libraries in the U.S.A, which due to OCLC WorldCat and other union catalogues seldom needed to copy abroad, or from multiple sources. However, implementation of portal applications and copy cataloguing directly from domestic and remote peers will lead to heightened understanding of the quality problems in Z39.50 clients and servers.

Information about remote services

One of the aims of the SVUC project was to improve the then existing Explain services. The time has not been merciful for this requirement, or Z39.50 Explain service in general. We have not seen any new Explain implementations lately, and even those Explain databases that still exist are not used. This is a paradox, since the emergence of information retrieval portals makes it necessary to find and retrieve information about remote databases. The patrons need to know about the contents of the remote databases and collections within them; search engines need to know how to connect to diverse remote systems, and which Bib-1 use attributes and attribute combinations are supported by the remote Z39.50 servers and databases. Portals should even understand the search terms available in non-standard systems, but specifying this may prove to be too complicated and/or time consuming, since mapping these terms to commonly used base such as Bib-1 must be done by humans.

In the absence of Explain services, most portal vendors and people responsible of maintaining these products have created portal metadata by hand. This is a time consuming work, and frustrating since access parameters of a remote system may – and most likely will - change constantly. What differs is the degree of centralisation of maintenance; for instance, BookWhere relies on a central configuration database from which every installed BookWhere client retrieves automatically the information about how to access remote systems described in the database. In case there are any updates these are updated every time the client is launched. With MetaLib the portal metadata is created by Ex Libris employees, and the updated information on remote systems is updated according to a pre-set schedule into the MetaLib systems run by customers. In other Z39.50 client products maintaining this information is the responsibility of local system managers.

Finding out which Z39.50 Bib-1 attributes and attribute combinations are supported in remote databases is a frighteningly complex task. Z39.50 client developers could automate acquisition of this information via implementing the ParaZ software developed by a Danish company Index Data (http://www.indexdata.dk), or by building themselves a similar tool. The operational principle behind these “Z spies” is simple. Once the “spy” application knows the access parameters of a Z39.50 database (Internen name/IP address, port and database name), it sends to the server all queries that can be built using the present Bib-1 attributes and their combinations. This results to a robot attack of hundreds of searches, but if carried out in the night time this should not be too much of a burden to the server.

Each query creates either a result set of 0 to n references, or an error message stating that the attribute or attribute combination is not supported. From this response, the spy knows precisely what kind of searches are supported by the database. This information can then be loaded into a Z39.50 client, or be used to update the central configurations database such as the one implemented for BookWhere. Unfortunately, any robot such as this is only able to make a generic map of the semantic terra incognita. Two important tasks are still left to the humans: collection description plus mapping of use attributes between Z39.50 databases. The same query is often done with different Bib-1 use attributes in different databases; this is due to complexity of the attribute set. These differences can not be spotted by the present software products; a human expert is needed as the task requires evaluation of the result sets.

Collection description remains an even more serious problem, since there are no generally accepted cataloguing rules or formats for collection description metadata. There have been numerous attempts to develop such a format, including e.g. Dublin Core application profile for Collection description; it, an other recent specifications, will be used as a starting point in NISO Metasearch Initiative's Collection description task group. The group aims at developing a standard for metadata elements for collection descriptions, and defining one or more exchange formats (using e.g. Dublin Core and MARC) for them (see http://www.niso.org/committees/MS_initiative.html). If the group finishes its work successfully, libraries may one day be able to exchange collection metadata in the same way we have exchanged bibliographic metadata since MARC was standardised in early 1970s.

Syntax

When the SVUC project was launched each Nordic country had its own national MARC format, and no changes for this policy were discussed, although format experts were following the preparation of the MARC21 format, and the discussions on format harmonisation between the U.S.A, Canada and United Kingdom.

In just three years the situation has changed drastically. Iceland and Sweden have replaced their national formats with MARC21, and Finland is using MARC21-Fin, a slightly expanded version of MARC21 which allows loss-less conversion of records back to FINMARC. MARC21-Fin records can not be used for copy cataloguing in those libraries which rely on MARC21 because MARC21-Fin does not have ISBD punctuation in data. Adding it manually would be too time consuming. For the same reason Finnish libraries can only use MARC21 records efficiently for cataloguing after MARC21-Fin conversion, done by USEMARCON.

Although Denmark and Norway have no imminent plans to move from their national formats DANMARC2 and NORMARC to MARC21, there are constant discussions about the topic. In Denmark national cataloguing practices will be harmonised in such a way that conversion to and from MARC21 can be made as simple as possible.

Disappearance of some national formats has made the burden of format conversions in the Scandinavian virtual union catalogue lighter, but it is still by no means trivial. Specification of a format conversion is a task which requires not only a custom-made conversion tool, but also an expert who knows both the source and target formats well. And there are still quite a few formats to deal with: MARC21, MARC21-Fin, NORMARC and DANMARC2.

Those countries that have changed the national formats have chosen different approaches for carrying out the bulk conversion of records to the new format. In Finland Helsinki University Library took the responsibility of converting all bibliographic records in national databases and university library OPACs; in all these systems contained about 15 million records which were converted twice (test and production) with USEMARCON. In Sweden the conversion was also done locally, but in Iceland the software vendor Ex Libris did the job using the specifications provided by the National and University Library.

Within SVUC, USEMARCON will be used in Finland to create MARC21 records (see the Finland chapter for details). The application could also be used by Helsinki University Library to create NORMARC and DANMARC2 records. However, since Bibsys cataloguing module does not allow direct copying from remote databases for the time being there is little need for implementing NORMARC conversion in Helsinki. Some Danbib consortium members use library systems which enable them to copy records directly from Finnish databases, but for the time being there has been no discussions about implementation of DANMARC2 conversion in Helsinki to facilitate this. It has been estimated that creating a satisfactory conversion with USEMARCON would require approximately two weeks.

Both Bibsys and DBC rely on their own internal format conversion modules for the time being. There are no plans to replace these products with USEMARCON or any other third party application, as this would not bring any functional benefits according to the internal reviews Bibsys and DBC have done.

SVUC partners have no immediate plans to implement XML-based record syntaxes. Of course any SVUC partner whose system is capable of producing MARC21 can also deliver MARCXML off-line using the free tools developed by the Library of Congress. However, this conversion has not been integrated into SVUC partners' production systems. The Library of Congress is as of this writing developing a system which will enable them to deliver MARCXML via the Voyager Z39.50 server; once this system is in production Helsinki University Library may evaluate it, and if the result is favourable, implement it in their own Voyager system.

Moreover, MARCXML is not an ideal solution because it is entirely MARC21 based. There is a need for a more flexible XML syntax specification which does not dictate the procrustean MARC21 bed like MARCXML. Development of such a specification is already under way, and SVUC partners are involved in this effort. Dansk BiblioteksCenter is an active member in the working group which is preparing the next generation XML-based exchange format for bibliographic data as a new work item for ISO TC 46. This syntax will be generic so that any present MARC format which is compliant with ISO 2709 can be expressed with it. For instance, MARCXML will be one practical implementation of this abstract syntax to be developed, and XML version of DANMARC2 will be another.

Character sets

From Z39.50 profiling point of view, character sets have always been a problem. One reason for this is that MARC formats are less than clear in what they demand. For instance, MARC21 allows encoding of either MARC-8 and UCS/Unicode in the leader. Unicode encoding is not specified, but every vendor implementing Unicode has chosen UTF-8. But there is also tag 066 (Character sets present) which can be used if the character set is other than ISO 10646 or Unicode. It is not clear if anybody uses this tag and if so, which additional character sets have been encoded. But many

European libraries are using Latin-1 instead of MARC-8 as the character set for records; thus when receiving MARC21 records there are at least three character sets in which the data may arrive.

Because MARC21 character set is a thorny issue, a Z39.50 profile might differentiate between the character sets used in queries and data. The Bath profile does not make this division; it requires that "if a character set is not negotiated the server should assume that the character set is Latin-1" (Bath profile, p. 13). In profile level 0 (Basic bibliographic search and retrieval) clients should be prepared to negotiate Unicode but nothing is said about server requirements. Conformance level 1 requires that both Z39.50 clients and servers recognise character set negotiation. It is to be assumed that both applications should be able to negotiate both Latin-1 and Unicode UTF-8.

The Bath profile requirement for Latin-1 is in conflict with the MARC21 requirement for MARC-8 or Unicode. One way of solving this problem is to require Latin-1 support for queries, but make it clear that the character set of the bibliographic records will depend on the MARC format. The present situation is problematic; most SVUC partners send MARC21 records with Latin-1 encoding which means that any system using the MARC21 default character set MARC-8 must convert the so called Scandinavian characters. In Finland this is done as a part of the MARC21 – MARC21-Fin format conversion; libraries using MARC21 do not have any easy way of automating the character set conversion. This is a problem because manual editing of special characters may take a lot of time, which may render copy cataloguing useless.

Character set problems will persist at least until Unicode support is not ubiquitous in integrated library systems, and profiles are able to mandate Unicode only. In the short run the character set problems may get even worse; copying records between a Unicode compliant and "traditional" system may cause a lot of headache for future copy cataloguers. And even when everyone is using Unicode the problem may persist; MARC tag 880 (Alternate graphic representation) allows input of either Romanised or original data, depending on the library's preference. Thus for instance Russian libraries may use normal MARC fields for Cyrillic data, and use 880 for Romanised version, while the Finnish libraries are likely to adopt the opposite approach. It is not advisable to use 880 alternatively to Romanised or original data since this would make indexes messy. This means that the Finnish libraries must carry out a complicated format conversion in order to be able to use Russian records for copy cataloguing.

SVUC activities at Helsinki University Library

Helsinki University Library was a full partner of the SVUC project, and the co-ordinator of the initiative. In addition to the project management, substantial amount of time was spent on diverse technical tasks of the project. This chapter provides a summary of this work. Project management issues, and generic technical issues related to creation of virtual union catalogues using Z39.50 protocol are discussed in other chapters.

Profile considerations

According to the analysis made by Slavko Manojlovich, Voyager Z39.50 server provides moderate support for the Bath profile. Fulfilling 12 requirements out of 41 is not an optimal result, but it is still better than most competing products (Z39.50 serves in other integrated library systems) were able to provide in 2001.

On the request of Helsinki University Library, USEMARCON format converter software was integrated into the Voyager Z39.50 server, but unfortunately not to the client. This means that although we can not copy catalogue records with the Voyager Z39.50 client, MARC records can be delivered from Helsinki in basically any MARC format and character set mandated by it. Alas, the most serious problem Helsinki University Library has had in fulfilling the requirements of the Bath profile is still related to the presentation of records.

The Bath profile demands MARC21 support and strongly recommends UNIMARC. Default exchange format for Voyager systems in Finland is MARC21-Fin. However, Voyager system assumes that the internal format is MARC21, and its Object identifier 1.2.840.10003.5.10 has been hard-coded in the system; it is impossible to configure another internal format. Therefore MARC21 delivery is only possible by using some other, preferably dead MARC format such as CANMARC (OID 1.2.840.10003.5.17).

As of December 2003 this feature has not been tested yet, because the conversion table from MARC21-Fin to MARC21 has not been finished. The aim is to complete this work in December 2003 – January 2004, and start testing immediately with a client capable of requesting CANMARC. MARC21-Fin - UNIMARC conversion is missing because there has been no demand for it. The same applies to conversions to NORMARC and DANMARC2; if DBC and Bibsys could provide USEMARCON tables for converting from MARC21-Fin to these formats, Helsinki University Library could deliver records in these formats immediately.

Voyager character set is for the time being RLIN/Voyager, which is not fully compliant either with the Bath profile requirements or with ISO 8859-1 (Latin-1) supported by other partners. This problem is however transitory, for two reasons. Endeavor is implementing Unicode with UTF-8 encoding in Voyager. Testing of Unicode version has already begun, and according to the present plans it should go into production during 2004. Due to UTF-8, this version will support Latin-1 for Roman characters. On the other hand, USEMARCON conversion also incorporates character set conversion, and the plan is to enable record delivery in both MARC-8 and Latin-1.

Domestic use attributes

16 domestic use attributes have been specified in order to allow queries with “Finnish” search terms. In order to guarantee compatibility with applications supporting Z39.50 version 2 only, the use attributes have been defined both as the Fin-1 attribute set and as an extension of Bib1- attribute set, using the 6100-6199 attribute values assigned for Finland.

Fin-1 contained in December 2003 the following use attributes:

Fin-1 Use Attribute Value	Fin-1 Use Attribute Name
1	YSA Finnish general thesaurus
2	ALLÄRS Swedish general thesaurus
3	Finnish music thesaurus
4	CILLA Swedish music thesaurus
5	KAUNOKKI Finnish fiction thesaurus
6	BELLA Swedish fiction thesaurus
7	Geographical subject headings
8	Index terms for education and psychology

9	Legal subject headings
10	Thesaurus of the social sciences
11	Religion Indexes subject headings
12	Forestry subject headings
13	Agricultural subject headings
14	YKL Public library classification
15	Local item-number
16	Local patron-number

New attributes will be added to Fin-1 when necessary by the Helsinki University Library which is the maintenance organisation of the attribute set.

In addition to these use attributes, URN (Uniform Resource Name) was added to the Bib-1 proper on the request of the Helsinki University Library.

Z39.50 client and server developments for SVUC

There are at least seven applications used in SVUC context in Finland: Voyager Z39.50 server and no less than six clients: WebVoyage Z39.50 client, Z39.50 client modules in MetaLib portal and BookWhere Z39.50 stand-alone client, plus Z39.50 modules of the three library systems (ATP Origo, TietoEnator Pallas and Innovative's Millenium) used by the largest public libraries, which are the data producers of the Manda union catalogue.

Voyager Z39.50 server

When in the Linnea2 project the university libraries were selecting a new library system to replace the outdated VTLS Classic, one of the central criteria was the quality of the Z39.50 server. Support for the Bath profile was required, and a possibility to convert MARC records and character set on the fly. Recognising the fact that developing such conversion capabilities from scratch, Helsinki University Library proposed the vendors utilisation of USEMARCON.

Endeavor Information Systems, Inc. won the contract for the new library system for the Finnish university libraries. The company has made numerous improvements to the Voyager Z39.50 server, including the USEMARCON integration. The ultimate aim is full compatibility with the Bath profile.

The most actively used Voyager Z39.50 server is the one in The Library of Congress. Heavy utilisation has brought up some novel problems, especially with malformed queries; the server has occasionally problems in recovering from them gracefully. LC is as of this writing experimenting with using the YAZ Z39.50 server and Voyager server back-to-back; if this type of usage improves the reliability of the service, then HUL will consider implementation of similar service.

LC and Endeavor have had discussions about implementing a new feature which would allow copy cataloguing of MARC authority records via Z39.50. This functionality has been defined by the Library of Congress into the version 3.0 of the Bath profile. This opens also a possibility for extending the future version of SVUC or ONE Association into this direction. Since production of authority records may require more time than traditional cataloguing, it is important that libraries will be able to exchange these records as well.

Comprehensive configuration information for Linda and Manda is available at http://www.lib.helsinki.fi/svuc/VoyagerZconf.htm. Record syntaxes and character sets currently available are:

- MARC21-Fin and RLIN/Voyager character set (the internal format and character set)
- FINMARC and ISO 6937/2

Delivery of data in MARC21 and both ALA/MARC-8 and Latin-1 character sets is as of this writing under preparation and should be available in early 2004.

Defining the relevant 45 Bib-1 use attributes into Voyager was straightforward, since the application allows specification of any Bib-1 Object Identifier value from 1 to 10000. No other attribute sets than Bib-1 can be defined; therefore usage of Fin-1 attributes is impossible for the time being. Endeavor has no plans to enhance the Voyager Z39.50 server implementation in this respect in the near future, but in this respect all vendors who responded to the Linnea2 project Request for Proposals were similar.

Bib-1 attributes have been mapped in Spring 2001 to USMARC and FINMARC (see http://www.lib.helsinki.fi/katve/toiminta/docs/bib1semf.html); if the Linnea2 libraries follow these guidelines in mapping the indexes of their bibliographic databases to Bib-1 attributes then the likelihood of semantic interoperability within the university library network is good, as long as the data contents are similar.

There is a need to update this specification so that it takes into account MARC21 and MARC21-Fin, plus the new attributes added to Bib-1 since May 2001. Helsinki University Library will take care of the update in 2004.

From the SVUC point of view, the most severe problem as of this writing is the lack of MARC21-Fin - MARC21 conversion table, which prevents Helsinki University Library from delivering records which could be easily used for copy cataloguing by other SVUC partners. A decision has been made to make the conversion available in early 2004, but even then copy cataloguing is complicated by the fact that Voyager thinks that the records within the system are in MARC21 format. Alas, in Finland this is not the case: our internal format is MARC21-Fin, which differs from MARC21 in many respects (including lack of ISBD punctuation). Therefore the Voyager Z39.50 server must be fooled into sending the records in the correct format via masquerading MARC21 into CANMARC, which is no longer in use and can be safely utilised for this purpose. In practice this means that anyone asking CANMARC from us will in fact get pure MARC21 records.

Z39.50 does acknowledge MARC21-Fin; Object identifier 1.2.840.10003.5.31 was approved for it in October 2001 Z39.50 Implementers Group meeting on the request of Helsinki University Library. As mentioned earlier, Voyager Z39.50 server does not allow changing the OID of the internal MARC format; MARC21 OID has been hard-coded into the system. Because some Z39.50 clients may be unable to demand CANMARC, or may process the Finnish MARC21 records as if they were CANMARC records and in the process of doing that demolish them, the CANMARC solution must be a transitory one.

Beyond the CANMARC trickery, there are two solutions to the format problem. The first is to enable in the Voyager Z39.50 server the specification of the system's internal format. Then Finnish libraries would be able to specify that their internal format is MARC21-Fin, and it would be possible to convert records with USEMARCON from that format to MARC21 on the fly. If

Endeavor can or will not do this, then the other solution is to change the internal format from MARC21-Fin to MARC21. Such a decision has political and technical implications way beyond Z39.50 usage, and has to be discussed thoroughly. Due to the fact that many countries have decided to abandon their domestic formats in favour of MARC21, format discussion is likely to continue in Finland, and other countries which have not yet adopted MARC21.

Z39.50 clients

WebVoyage

WebVoyage is the OPAC of the Voyager library system. It allows access to any Voyager system using Voyager's own search and retrieval protocol, and usage of remote Z39.50 systems. There is no limit on how many remote databases can be opened at the same time. All SVUC databases except Gegnir have been configured into WebVoyage parameters both in union catalogues and some libraries; those libraries that do not have SVUC databases locally use them via Linda and Manda WebVoyage implementations. Some university libraries use the SVUC databases actively, but primarily via BookWhere for copy cataloguing. Because SVUC utilisation is decentralised it is not possible to collect accurate statistics on Finnish SVUC usage. A small query among the users showed that Libris and Bibsys are the most popular databases, followed by Danbib. There has been little use of Sambok, and Gegnir utilisation via its native interface has been modest as well.

Any Bib-1 attribute can be defined for WebVoyage searching, but the systems librarian who makes the specification must know which attributes are supported in the remote Z39.50 server. For any use attribute there is a possibility for additional specification: usage for Boolean searching and truncation can be allowed or disallowed.

Every system librarian must input the configuration data for each remote database locally and manually. Endeavor does not provide good support for maintenance work, in spite of the establishment of eZconnect pages; libraries can not harvest configuration data from there.

Both Voyager Z39.50 server and client are based on Isite software, which has been much modified and improved. In spite of this both applications may misbehave; for instance the client does not always understand a negative response from a remote server, but repeats the same request over and over again. Larry Dixson from the Library of Congress has described the problem in the following way (Dixson):

> *When a Voyager client generates a result set at the target that only contains a single record, and for a variety of reasons that record is temporarily unavailable, the target server will send back a diagnostic informing the client about the system problem and to try again later, etc. While the Voyager client can process diagnostics when they are present in the searchResponse data unit (normally carries information about the result set), it appears that it cannot properly handle diagnostics that are present in the presentResponse data unit (the data unit that normally contains MARC records). The Voyager client keeps sending presentRequests (and getting back the same diagnostic) over and over. The Voyager client user is unaware of any of this activity and only sees the hourglass until the client times out and the Cataloging Client closes. However, behind the scenes, Voyager continues requesting that record . . . this can continue for days.*

This issue illuminates well the fact that developing Z39.50 applications is difficult, and that analysis of problem situations can be hard; in this case, the user sees absolutely nothing, except that the session somehow hangs and nothing seems to be going on.

Because the problem situations are rare, Linnea2 libraries use WebVoyage for providing patrons access to SVUC databases. In 2004 at least some of this usage will be replaced by the MetaLib portal.

Endeavor has embedded a Z39.50 client module into the Acquisitions and Cataloguing clients. Since these applications do not allow efficient centralised maintenance of database descriptions and because record conversions were difficult to accomplish with them these applications have been replaced with BookWhere in copy cataloguing usage. Helsinki University Library has discussed with Endeavor about integration of the USEMARCON converter into these clients, but up to now these talks have not led into any viable results.

MetaLib

MetaLib is an information retrieval portal developed by Ex Libris. One part of the system is a knowledge database, which contains configuration information for remote databases, supporting Z39.50 or other, less standardised search and retrieval protocols. Metadata in the knowledge database is maintained primarily by Ex Libris staff, although local changes and additions can be made.

MetaLib allows for specification of format (USMARC, UNIMARC, MAB, SUTRS, DANMARC or UKMARC), element set name (Full record, Full text, or Brief record), and attributes. It is not possible to use Bib-1 use attribute directly; the following search terms are mapped to the use attributes (one attribute per term, for each target database):

- WRD = word from any field
- WTI = word from title field
- WAU = word from author field
- WSU = word from subject field
- WYR = word from year field
- ISBN
- ISSN

In most cases this list is sufficient for a portal, but searching with classifications is a problem since the Bib-1 use attribute subject covers (only) subject headings. Thus, if mapping of Bib-1 and indexes is done according to the recommendations, classifications must be searched via MetaLib with the term “word from any field”, not “word from subject field”. There may also be specific local searching needs which do not match other MetaLib search terms and for which the any search is not granular enough. Therefore a possibility of adding new search terms / Bib-1 use attribute values locally into MetaLib would be useful.

The MetaLib searches may be combined freely with Bib-1 truncation, completeness, structure, position and relation attributes. Boolean operators AND, OR, NOT can also be used.

Z39.50 configurations for remote databases made by Ex Libris have contained some errors which can and should be fixed locally. Most often this is due to standard specification having been copied from one database to another. Unfortunately Ex Libris has not used an automated tool such as Index

Data's ParaZ for collecting the information about remote databases into MetaLib. In this respect all portal and Z39.50 client vendors seem to be similar: even in BookWhere the knowledge database is maintained by hand. This is a pity since ParaZ and other robotic tools can find out quickly which Bib-1 attributes are supported by a target system by sending to the target the full set of Bib-1 queries, and analysing the responses. Gathering the same information manually by sending the queries to the database by hand, registering the results and then typing them into a portal takes a lot of time. Hundreds of queries are needed to cover all Bib-1 use attributes and their combinations with other attributes (such as truncation) and Boolean operators.

The maximum number of databases that can be accessed simultaneously can be configured by the portal systems personnel. The default value is eight; in Finland this value is considered to be a good compromise and will not be changed until practical experience suggests that a change upwards or downwards might be both technically viable and useful for the majority of users.

It is expected that MetaLib will be the main SVUC user interface for patrons in Finnish universities already in 2004. Other Linnea2 libraries will continue using WebVoyage for this purpose, and the 20 public libraries participating in Manda will use either the OPACs of their library systems or the Manda WebVoyage for information retrieval.

BookWhere

BookWhere is a stand-alone Z39.50 client developed and sold by WebClarity, a Canadian software company. In a test carried out in Spring 2002 it outperformed other Z39.50 clients; as a result the Linnea2 consortium bought a 400-user licence of the software. By December 2003 all licences had been used and an extension to 500 users was negotiated with WebClarity.

WebClarity maintains a knowledge database from which the client retrieves new and modified configuration data each time the client is launched. It is possible to extract this resource metadata into a file host.dat and store it on a PC. Also, new target servers and databases can be added manually. In December 2003 the database contained 1600 servers and 1900 databases; the database has grown constantly and updates and modifications are frequent.

A user may define locally which Bib-1 attributes she/he generally wants to use; for instance, a copy cataloguer only interested in new materials may prefer identifiers. BookWhere may be configured in such a way that it hides for each target database those use attributes, which the database does not support. In addition to Bib-1 also GILS (Global Information Locator Service) attributes can be used; from the point of view of libraries this is not very useful.

BookWhere allows unlimited access to remote databases, but a user can configure a limit, plus set a time out for database connections.

BookWhere is the tool with which copy cataloguing is done. Libris and Bibsys are among the most popular copying databases. The records are downloaded in MARC21; with USEMARCON they are converted into MARC21-Fin and RLIN/Voyager character set, and then loaded into the Voyager Cataloguing client. Libris is a very popular copy cataloguing database, and Bibsys is a favourite as well. User identifiers and passwords needed for using Danbib were sent to libraries in late 2002, and some libraries have grown quite fond of this system as well. Sambok and Gegnir have not been popular; the first one because it provides little added value compared with Bibsys, and the latter because Z39.50 access to the new Aleph-based Gegnir has not been available yet.

In order to make BookWhere more suitable for Finnish libraries WebClarity has decided to add FINMARC support to the application. Beta testing of this feature was done in Helsinki University Library in January 2004. The company is also investigating the possibility of integrating USEMARCON into BookWhere.

Public libraries' Z39.50 clients

Both TietoEnator and ATP have embedded USEMARCON into their Z39.50 clients. This enables the user libraries to copy records from basically anywhere. Conversion tables developed by HUL are available to anybody for free.

Public libraries participating in Manda union catalogue have tested the Z39.50 clients of their integrated library systems against Voyager Z39.50 server and other SVUC servers. Helsinki University Library has assisted the libraries and vendors in this work.

According to the tests, copy cataloguing from Linda and the national bibliography database Fennica works well. Based on the comments from libraries, some changes have been made to the MARC21-Fin – FINMARC conversion table.

Public libraries have a need for copying both host and component parts records. In co-operation with TietoEnator, ATP and the Library of Congress, Helsinki University Library is developing guidelines for implementing this feature in Z39.50 context. The guidelines will specify e.g. the Bib-1 use attributes which can be used for searching component parts when the host record is known. If possible, the specification will be made so generic that it can also be used for searching work level record if a manifestation is known, or vice versa.

Access control

Access to Linda and Manda is controlled with IP numbers. Thus the technical difficulty of providing access to all consortium users depends heavily on how access to SVUC has been organised. At least three models have been used.

In Denmark, access to SVUC databases will eventually be organised via netpunkt (www.netpunkt.dk). This portal service is maintained by Dansk BiblioteksCenter and requires a login. All users who access remote databases via netpunkt receive the same IP address. Thus configuring it to the access files of Linda and Manda is easy.

In Norway and Finland there are different options for accessing SVUC databases. Two Bibsys-driven choices imply a Bibsys IP address and are easy to implement in Linda/Manda context. But libraries may also use stand-alone Z39.50 clients such as BookWhere. In order to allow this kind of access, Helsinki University Library has configured the IP addresses of Bibsys organisations compiled by the Bibsys into the firewall controlling access to Linda and Manda.

Large share of the Norwegian usage of Linda and Manda takes place via Bibsys. No statistics of this usage has been compiled, but the increase in Norwegian utilisation of Finnish union catalogues has been substantial. There is a simple explanation for this: in the past when the database was not available for free there were only a few Norwegian users of Linda, but now any Bibsys library can access the database.

In Sweden, Libris does not provide for the time being a central portal for accessing SVUC. Therefore all SVUC usage is decentralised to the level of organisations participating in Libris. Since Libris has not provided a comprehensive list of IP addresses, enabling access to Linda and Manda for Libris libraries has been a gradual process. Those libraries which were already users (there were quite a lot of them), could go on just like before; the only difference was that they no longer had to pay any fees. Other Libris libraries have been added on request. The net result is that the Swedish usage of Finnish union catalogues is now somewhat more active than in the past.

Dissemination of information

Finnish libraries which are entitled to use the Scandinavian Virtual Union Catalogue have been informed about the service, primarily via email. Helsinki University Library has taken the responsibility of keeping the users up to date.

Information about the project and SVUC databases have been compiled to the home page of the initiative (http://www.lib.helsinki.fi/svuc/), which is maintained by HUL.

SVUC activities at Bibsys

BIBSYS has been a full partner of the SVUC project, and spent a total of 43 man-days in the period 1999 - 2002 on project tasks. This country report, written by Ole Husby, summarizes the work and the results.

Profile considerations

The BIBSYS Z39.50 server is delivering records as well as other APDU's in the ISO 8859-1 character set only. This has always been a headache concerning MARC21 records, as it is in fact in violation of the formal specifications. However, we are behaving like many other European service providers, and this issue has been extensively discussed within the ONE and ONE-2 projects. The consensus reached in these projects has been reflected in SVUC, where ISO 8859-1 is explicitly allowed.

There has been no other problems with the technical requirements in the SVUC agreement.

Domestic use attributes

The needs for private use attributes for specific Norwegian search elements has been discussed within NorZIG. After a survey of needs as well as the possibilities offered by BIB-1, it was decided to propose to ZIG that a number of new USE values should be added to BIB-1 as *common* values, as these were not private Norwegian. These additions has been accepted by ZIG. In addition, 2 really national values were agreed upon, and added to the 6300-6399 range of BIB-1, set aside by CENL for private Norwegian USE values.

These 2 private values are 6300 (HUMORD subject headings) and 6301 (NTUB subject headings). See detailed information at http://www.bibsys.no/norzig/profiles/USE_norwegian.html

Z39.50 client developments

Norwegian users can access the SVUC services in 3 different ways:

1. **The BIBSYS integrated client.** This is built on the Z39.50 protocol engine developed within the ONE system, and allows searching and copy cataloguing directly, sharing the same data buffer as the BIBSYS cataloguing client. Records are converted from MARC21 to BIBSYS-MARC by private (non-perfect) algorithms. For SVUC purposes some adjustments were made concerning the search profile, mapping new BIB-1 attributes to the BIBSYS search engine as required by the Bath profile. New configuration files were made for the Finnish and Danish databases.

2. **The BIBSYS WebZ gateway ZSØK.** This is a third-party system, and although source code is available through OCLC's open source agreement, we decided not to make any changes regarding the functionality, as the main access points required by the Bath profile are supported. For SVUC purposes new configuration files were made for the Danish and Finnish databases.

3. **Any standalone (personal) Z39.50 client.** The most popular of this type in Norway is probably EndNote. No effort has been made concerning the use of such clients, apart from dissemination activities and the publishing of the BIBSYS IP ranges for the other SVUC partners. One could have spent some work on securing availability of configuration files for the other SVUC databases, but this was considered to be outside the SVUC scope.

Z39.50 server developments

The BIBSYS Z39.50 target is built on the same protocol engine as the integrated client. The implementation has been running for several years, and have been tuned to align as good as possible to the Bath profile and the ONE-2 profile. Some minor adjustments were made for SVUC purposes in the attribute set mappings. The server does support INIT authentication, but this is in fact "turned off", so that any authentication tokens will be accepted. The server will offer access to the BIBSYS union catalogue, as well as a number of subsets of this. It is possible to search the catalogue of any BIBSYS library separately, and also periodicals, dissertations, analytics etc.

There have been some stability problems with the service lately. The "Z39.50 robot" at Index Data reports an "uptime" of 97 %. No analysis has been done concerning the use of the BIBSYS database from SVUC partners, and it would be a non-trivial task to do it.

Access control developments

Access *to the BIBSYS database* from the partners system did not require any attention within the project, as the BIBSYS policy is to offer free and anonymous access both to search services and to record download. This applies to web as well as Z39.50 access.

To support the access *from BIBSYS users* to the partners databases, three different use scenarios mentioned above had to be considered: The BIBSYS integrated client, the BIBSYS WebZ gateway, and standalone personal clients (like Bookwhere, EndNote etc.). The first two are governed by control mechanisms, allowing only authorised users the access of restricted services. Both have built-in support of the Z39.50 standard INIT authentication. So far it has not been necessary to employ this authentication for SVUC purposes. This means that the other partners either allow anonymous access (like BIBSYS), or use IP recognition of the clients in question, both operating within the bibsys.no domain.

For the purpose of informing the partners of the IP ranges of the BIBSYS user community as required by the SVUC agreement, a special XML file was developed. This is freely available on the web at http://www.bibsys.no/svuc/bibsysinst.xml

The same file is used for a number of other access control systems at BIBSYS.

Dissemination of information

Web:
Information page about the SVUC services, targeted at Norwegian users:
http://www.bibsys.no/svuc/bibsysinst.xml

Presentations:
FinElib: Seminar on Nordic Digital Libraries, Helsinki, October 2000
NordInfo: Nordisk Forum for Forskningsbibliotekschefer, Reykjavik October 2000
BIBSYS: User meetings 2001 and 2002

Publications:
Short notices about SVUC in BIBSYS-NYTT 99-1, 99-2 and 00-3
Article about SVUC in BIBSYS-NYTT 02-02: http://www.bibsys.no/bibnytt/02-2/02-2-jun.htm#P105_10992

Handouts:
BIBSYS user meeting 2002.

SVUC activities at Landsbókasafn Íslands – Háskólabókasafn

Due to the delay in implementation of the Aleph system, which was to the large extent related to the complexity of converting records from Libertas to Aleph and into MARC21, National and University Library was not capable of participating actively in SVUC.

The following description of the Icelandic project has been extracted from a Web page available at http://www.landskerfi.is/english.php. The page is maintained by The Icelandic Library Consortium Inc., a company owned by the Icelandic government and municipalities.

> *The contract between Ex Libris and the Consortium of Icelandic libraries assumes that the new library system will be used by all or almost all of the approximately 400 libraries in the country, and that legacy data will be added to the new system in steps, based on their existing systems.*
>
> *Aleph system opened for production (for 15 out of 400 libraries; those who used Libertas) on May 19, 2003. Approximately 70 Dobis/Libis libraries are expected to join the system by the end of 2003. Micromarc systems, Metrabók systems and Alice/Oasis libraries will switch to Aleph during 2004; therefore the whole system is expected to be fully implemented by the end of 2004.*

There are five library systems in use in Iceland whose data will be converted into Aleph: Libertas (used by 11 libraries), Dobis/Libis (80 users), Mikromarc (30 users), Oasis/Alice (30 users) and

Icelandic Metrabók, which has 70 users. After the project is over Iceland will be the first country in the world in which there is only one library system and database in use; a true union catalogue. In a country such as Iceland this centralisation provides substantial economical and functional benefits.

Although the SVUC project will be formally over by the end of 2003, Helsinki University Library has contacted The Icelandic Library Consortium and asked whether the consortium is and willing to join either or possibly both SVUC or ONE Association. The response was positive, so a SVUC agreement was prepared and sent to the consortium.

Technically the Aleph system and usage of MARC21 provide a good starting point. The only potential problem is that the Icelandic consortium is already using UNICODE; other SVUC consortia are only moving towards this direction.

SVUC activities at DBC

This country report is written by Bodil Dalgaard-Møller and Tommy Schomacker.

Z39.50 profile

The Z39.50 service at DBC is supporting the danZIG profile, which aims to be a superset of the Bath profile. The profile may be viewed and downloaded from the Danish National Library Authority at http://www.bs.dk/ (choose “vælg emne = danZIG”).

For search elements that are specific Danish, and therefore not could be included in BIB-1, DanZIG has developed the DAN-1 attribute set (OID: 1.2.840.10003.3.15) with 27 Use attribute values, which maps to fields in danMARC2.

Services: Init, Search, Present, Scan, Close

Formats in Present: SUTRS, XML, danMARC, MARC21

Search syntax: RPN, CCL (it is recommended to use RPN searches)

Record syntaxes are danMARC2 (http://www.kat-format.dk/danMARC2/default.html) or MARC21.

MARC conversion

Long before the start of the SVUC project DBC had developed its own conversion tool (DBMF) as part of the DBCLib-system. This tool is used for all kinds of transformation of bibliographic records, e.g. conversion between different MARC records, between MARC records and non-MARC records and between different record syntaxes (e.g. ISO 2709 and XML). It is also used for all kinds of presentation formats and even to produce paper-based products.

As a consequence a great knowledge in using this conversion tool is established. This implies that there has to be very heavy weighting benefits to make a shift to a new conversion tool. Such benefits might be access to an open toolbox in an XML environment.

Z39.50 server and client functionality

Server

The Z39.50 service at DBC is established through the NEP (Neutral Entry Point), which is built on the Z39.50 protocol engine from the ONE project.

On the rear of the NEP a number of back ends are giving access to different sources of information. There are back ends for DBC's own databases, and especially the national union catalogue DanBib. Other back ends gives access to external Z39.50 targets, e.g. WorldCat and ArticleFirst from OCLC and the SVUC databases (for the time being only BIBSYS and LIBRIS). Also back ends for accessing databases via http are established.

At the front the NEP acts as a Z39.50 target covering the services, which are presented by the back ends.

Access control, search administration, accounting, etc. is handled inside the NEP.

For the moment the NEP is able to deliver only the formats from the databases, but the DBMF conversion tool is now implemented in the NEP, which will implicate that conversion may take place on the fly.

The NEP uses "V3-style" authentication (userID/groupID/password).

The NEP exists in a number of incarnations for different purposes. One of them is Zpunkt, which is DBC's external Z39.50 Target.

Clients

There are developed a number of web-to-Z gateways that uses the services of the NEP.

Speaking of the SVUC databases, access is done from netpunkt (http://netpunkt.dk), a web based user interface for professional use in libraries. Netpunkt is password protected and gives access only to the services the user has ordered.

Libraries with their own client may access the services directly from Zpunkt.

Access for Danish libraries to SVUC databases

Access to SVUC databases via netpunkt will require a subscription fee to cover hotline, support and maintenance – DKR 900. For the moment netpunkt only gives access to BIBSYS and LIBRIS.

There is no access to SVUC databases via Zpunkt, but libraries with their own client may access the SVUC targets directly.

It is not practical for DBC to maintain updated information about IP-addresses used by all of the relevant Danish SVUC users. This should not pose any practical problems for our SVUC partners, however, as all Z39.50 traffic managed by DBC will be associated with the IP-addresses of DBC.

Access to the other SVUC databases and access via Zpunkt are listed on the DanBib development plan for 2004, which is now in a hearing process in the libraries. The use-value of the databases will rise substantially when the NEP can convert to danMARC2 on the fly.

Access for Nordic libraries to DanBib

DBC can offer access to records from DanBib – the Danish National Union Catalogue. DanBib includes records of foreign origin. SVUC users are not allowed to reuse these foreign records, although it is technically possible to copy and reuse those records.

Z39.50'url to DanBib via Zpunkt: z39.50s://z3950.dbc.dk:210/danbib

Access control is based on userID/groupID/password supplied by the Z39.50 client. For administrative reasons, access by IP addresses is not offered.

For the time being passwords has been given to Finnish libraries. BIBSYS and LIBRIS have access for test purposes.

References:

Bath profile. Electronic resource, available at: http://www.nlc-bnc.ca/bath/ap-bath-e.htm. (Referenced at: 2003-12-01).

Dixson, Larry: Voyager client behavior with OCLC Z39.50. Email message sent to the Sapphire-Z email list on 25 August 2003

Manojlovich, Slavko: Bath profile Z39.50 server compliance test results: preliminary findings. Electronic resource, available at: http://www.sirsi.com/Sirsipdfs/ifla_whitepaper.pdf. (Referenced at 2003-12-04)

Appendix 1: SVUC Agreement

AGREEMENT ON THE MUTUAL ACCESS TO THE NORDIC UNION CATALOGUES FOR INFORMATION RETRIEVAL AND COPYING OF RECORDS

Partners and an aim of the agreement

This agreement is made between BIBSYS, Dansk BiblioteksCenter, Helsingin yliopiston kirjasto (Helsinki University Library; the party responsible for governing this agreement), Kungliga biblioteket, Landsbókasafn Íslands – Háskólabókasafn (see Appendix 2) and Nasjonalbiblioteket for mutual access to the union catalogues maintained by these partners for information retrieval and copying of catalogue records. The databases covered by this contract are Bibsys, Gegnir, Libris, Linda, Manda and Sambok. Dansk BiblioteksCenter (Danbib database) may join the Scandinavian Virtual Union Catalogue (SVUC) consortium at any time in the future. NOSP database may also be included in the future.

The national consortia maintaining each of these databases and their parent organisations have the right to access all the databases listed in this agreement. The national consortia covered by this agreement are listed in Appendix 1. The consortia are obliged to maintain in the Web lists of their member libraries. The consortium member libraries and their parent organisations have the right to access all the databases mentioned in this agreement at all their offices.

The duration of the agreement and amendments to it

The agreement is effective as of 1st March 2002 until 31st December 2003 minimum.

To include new domestic libraries in a consortium maintaining a union catalogue is agreed on national level. The other SVUC hosts must be informed of these changes via e.g. making available in the Web a list of all libraries participating in the union catalogue.

To include foreign libraries in a consortium all SVUC union catalogue hosts must be consulted separately.

To include a new database (maintained by a host organisation already involved

with SVUC) and the library consortium maintaining it in the SVUC service the database host must consult all other SVUC host organisations separately. A new database can be included if the majority of the SVUC hosts approve of it. The libraries belonging to the new consortium will be able to use the other SVUC databases once their IP address information has been updated.

To extend this agreement to include new union catalogue hosts and the databases they maintain all SVUC hosts must be consulted separately. A new organisation can be included only if the majority of the SVUC hosts approve of it.

If the new host organisation is from a country already involved with SVUC, the discussions must be initiated by a SVUC partner from that country (e.g. for new Finnish hosts the gatekeeper is the Helsinki University Library; in Norway both Bibsys and Nasjonalbiblioteket can suggest new hosts). If the new host organisation is from a non-SVUC country, any existing SVUC host may propose the extension of the SVUC group.

To add new functionality and services to the SVUC all union catalogue hosts must be consulted separately. A new service will be added if the majority of the SVUC hosts approve of it.

The technical implementation

Access to the databases is implemented using Z39.50 applications. These applications should be compatible with the release 1.1 of the Z39.50 Bath profile (http://www.nlc-bnc.ca/bath/bp-current.htm). WWW access to the databases can also be provided. The applications used by the SVUC hosts may extend the functionality specified in the Bath profile by supporting the ONE-2 profile, version 3, revision 2 (http://www.portia.dk/pubs/one2/Profiles/ProfileV3R2/ONE-2ProfileV3R2.pdf).

Any stand-alone or embedded Z39.50 client program or search portal incorporating Z39.50 client functionality can be used as a SVUC interface.

The records are transferred from one system to another in MARC21 format and in Latin-1 character set. The participants can agree on transferring the records also in some other character set, MARC format or in XML syntax in addition to MARC21. In the future, the character set may be changed to UNICODE, but only if and when all SVUC hosts agree on this.

Access to the databases is limited to computers connected to the local area networks of the consortia member libraries and their parent organisations. The staff, researchers, students and visitors to the libraries have the right to use the databases.

Access to the databases based on user ID and password from computers not

connected to the local area networks of the consortia member libraries and their parent organisations will not be supported.

Access is determined by IP identification using the IP addresses of the consortia member libraries and their parent organisations or the IP addresses of individual computers if necessary. Each union catalogue host is responsible for providing the other hosts the list of the relevant IP addresses. An updated list, or an announcement that no changes have been made, must be sent at least four times a year. Updates may be made more often if necessary. The other hosts should configure the announced changes into their own systems within two weeks of the notification or at a specific time mutually agreed.

If the network system used by a consortium member does not enable the use of one or several of the SVUC databases the member organisation is responsible for finding a technical solution to the problem. The union catalogue hosts provide technical support, but they do not commit themselves to changing their production environments so that the use of their databases would be possible without any measures taken by the customer.

A guide to the use of the SVUC databases is accessible online via the SVUC Project website (at www.lib.helsinki.fi/svuc/) and also elsewhere if necessary. Each member library of the national consortia is responsible for implementing and configuring the necessary Z39.50 and WWW browser programs and for providing instructions and user training.

The costs

The use of the SVUC databases for searching and copy cataloguing is free of charge. Not applying the charges is based on the mutual access to the other consortia databases.

The costs for the libraries outside the consortia are not defined in this agreement. To access the databases they have to consult each union catalogue host separately.

Additional services, such as training or batch transfer of authority records, may not be free of charge.

Cancellation of access and termination of agreement

A SVUC union catalogue host can terminate this agreement unilaterally by serving notice six months prior to the termination. After this the consortia member libraries and their parent organisations will lose the free access to the remaining SVUC databases via the immediate update of IP number information.

If a library membership in consortium maintaining a union catalogue is

cancelled permanently or temporarily for a restricted period its access to the other union catalogues will be cancelled when the update of the IP address data is made.

Signature

[*The Partner Organisation*] approves the attached Agreement on the mutual access to the Nordic Union Catalogues for Information retrieval and copying of records.

[*Place]*, on the of 200

[*Signature*]

Appendix 1

The Consortia in 2002

Bibsys Libraries

Danbib Libraries

Sambok Libraries

Gegnir Libraries

Linnea Libraries (http://www.lib.helsinki.fi/kirjastoala/linnea/linneah.html)

Manda Libraries
(http://www.lib.helsinki.fi/kirjastoala/linnea/MANDA_kirjastot.htm)

Libris Libraries

Appendix 2

Landsbókasafn Íslands – Háskólabókasafn agrees to the terms stated in the agreement with the following caveat:

When the present union catalogue of the GEGNIR libraries will be dissolved (according to the project plan, this will happen in the middle of the year 2002), all the records will be transferred to a new Union Catalogue that will be operated by the Landskerfi bókasafna. This organisation will thereafter take over both the responsibilities and the rights of the GEGNIR libraries as they are stated in the agreement.
