Gabriela Krčmařová and Ilona Trtíková,
The CASLIN Union Catalog

from / *aus*:

Union Catalogs at the Crossroad
Edited by
Andrew Lass and Richard E. Quandt
pp. / *S*. 141-172

Erstellt am 31. März 2005

Impressum

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.ddb.de.

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

This publication is also openly accessible at the publisher's website. Die Deutsche Bibliothek has archived the electronic publication, which is now permanently available on the archive server of Die Deutsche Bibliothek.

Diese Publikation ist auf der Verlagswebsite ebenfalls open access *verfügbar. Die Deutsche Bibliothek hat die Netzpublikation archiviert. Diese ist dauerhaft auf dem Archivserver Der Deutschen Bibliothek verfügbar.*

Available open access / *open acess* verfügbar:

Hamburg University Press / Hamburg University Press
http://hup.rrz.uni-hamburg.de
Die Deutsche Bibliothek archive server / Archivserver Der Deutschen Bibliothek
http://deposit.ddb.de/

ISBN 3-937816-08-9 (print)

Table of Contents

Part 6 South African Union Catalogs

Chapter 7
The CASLIN Union Catalog

Gabriela Krčmařová and Ilona Trtíková

Virtual union catalogs revolve around cooperating technologies, real union catalogs revolve around cooperating people.

1 Introduction

The CASLIN Union Catalog (Union Catalog or UC for short) is a centralized national union catalog. It is a single database that collects documents stored in Czech and Moravian libraries, which use a variety of library systems. Since 2000, it has operated in a tailor-made system called CUBUS, designed to fully meet the requirements of maintaining and operating exactly this type of union catalog. The launch of the Czech National Union Catalog was one of the tangible results of the CASLIN Project (Czech and Slovak Library Information Network) [17]. Between 1993 and 1995, the CASLIN Project gave life to all activities important for a national union catalog. Besides a clearly defined and detailed concept (modeling a union catalog in operation, gathering and maintaining data, user categorization, etc.), fundamental standards were established, and the Union Catalog administrator was identified.

In July 2002, the Union Catalog contained 1,578,868 records of printed monographs and special types of documents from 110 libraries. There were 60 active members regularly supplying the Catalog with records. In addition, the Catalog contains 84,683 records of serials from 550 participants. The directory of libraries and information institutions contains 2,947 records.

2 The Fundamental Strategy and Standards for the Union Catalog

The undeniable advantage of having to bridge a 20-year gap to the Western world lay in the opportunity to determine unified standards for record provision and exchange before launching the National Union Catalog (1995). However, this was the only advantage. The National Union Catalog was established in a similarly heterogeneous environment to that of its Western counterparts 30 years ago [23].

The following library systems existed in the Czech Republic:

- Academic libraries have been using TINLIB as their integrated library system since the early 90s;
- Public libraries have mostly started to implement a Czech library system called LANius, or later Clavius;
- The Czech National Library and other large libraries implemented the ALEPH integrated library system.

The heterogeneous nature of the library environment is further reinforced by other library systems, both Czech and foreign: KpSys, Rapid Library, Olib, ISIS and WINISIS, Daimon, etc.

The strategy of the Czech Union Catalog is the same as the fundamental conceptions of major union catalogs throughout the world. As a matter of principle, the Union Catalog is open to all libraries and information institutions able and willing to respect the established standards of record provision:

1. The primary exchange format is UNIMARC Exchange Format, with CDS/ISIS as secondary exchange format;
2. ISBD(G) is the basic standard for name processing;
3. The guiding rules are AACR2, 1998 edition;
4. UDC notation is binding for subject cataloging;
5. The Union Catalog Record determines the binding record format for particular document categories, as established by the administrator of CASLIN Union Catalog ČR. In its Standardization Series, the Czech National Library published the following union catalog instructions:
   - Union Catalog Record: UNIMARC. Printed Monographs (1996, Standardization Series #4);

- Union Catalog Record: Exchange Format. Printed Monographs (1997, Standardization Series #9);
- Union Catalog Record: UNIMARC. Special Types of Documents (1999, Standardization Series #16);
- Union Catalog Record: UNIMARC. Printed Serials (1999, Standardization Series #17).

From the very beginning, the Union Catalog has focused on general technical standards: TCP/IP, HTTP, FTP, etc.; the implementation of Z39.50 communication protocol has been planned for 2002.

3 The Functionalities of the Union Catalog

Currently the Union Catalog serves the following objectives [26]:

Information Function

The Union Catalog is a source for searching for and finding a particular document, or gathering information about documents concerning certain topics.

Document Location

The Union Catalog allows the user to locate the library that holds the document in question, and possibly also to obtain detailed data about the document's shelfmark, usually facilitating the borrowing of the document.

Document Retrieval

The Union Catalog makes it possible to act on a request to borrow a document or request its copy (Inter-library Loan Service—ILS). This service does not necessarily have to be a part of the Union Catalog. However, this type of service is often offered together with the ability to decide to which library this request should be forwarded, while keeping in mind the possibilities of any given library.

Shared Cataloging

The Union Catalog is a tool for shared cataloging, offers access to records and their copying, and is a tool for the formation and optimization of name and subject authorities.

4 Union Catalog Services

The fundamental principle in creating a union catalog is the controlled harvesting of data of the broadest possible scope, with the aim of creating a concentrated information base and a qualitatively and quantitatively rich source of secondary documents (records). This principle, if followed, allows for the introduction and development of additional services for the users of libraries and information institutions, as well as for librarians themselves [26].

The Union Catalog offers the following services:

- Searching, i.e. locating documents in Czech libraries;
- Provision of reference and inter-library services, i.e. sending a loan request to ILS, where the identifying data for a library are generated from the directory and the document data from the record in the Union Catalog;
- The Clipboard, used for storing located records for later use (printing, data export);
- Receiving document records for retroconversion of local library catalogs, for current cataloging, or for the national bibliography;
- Shared cataloging, in order to process current production in two ways: copy cataloging, and online shared cataloging by means of a preset input form; and
- Use of the input form to edit data in existing records and in the location data of the member library that owns a given document but does not send its documents to the Union Catalog.

The Union Catalog users are subdivided into three categories, based on the type of services they use. There is a fundamental prerequisite for any union catalog to achieve its goals—it must be filled with data.

5 Union Catalog—Data Administration Maintenance

In the Czech National Library, the Union Catalog Department is in charge of the Union Catalog's administration. Libraries mostly contribute to the Union Catalog offline, i.e. from time to time they upload batches of newly processed records or records formed in the retroconversion process.

In January 2000, the Union Catalog began to operate in the CUBUS system, and all processes involved in union catalog database administration, including input data analysis, were automated, making maximum use of existing software tools [19]. Record processing is automatic, and its steps are

- receipt and identification of a data file (including conversion);
- formal logical data control: a UNIMARC test and a duplicates test;
- data import;
- statistics for the participants and the administrator, and
- problems left for the administrator's decision.

A participant library places its data in its allocated space on the FTP server (if the data are delivered on a floppy disk, they are transferred to the FTP server by the administrator). The program periodically checks whether there are new data on the FTP server. If so, the program downloads the data and, using the name convention (see below), identifies their owner, their format, and the character set used. For import, it is important to name the file in compliance with the naming convention.

Name Convention (Name Format)

The data filename may have up to 8+3 characters (i.e. 11 in total). The first 6 characters identify the institution, characters 7 and 8 stand for the code page used, and characters 9, 10 and 11 identify the data format (e.g. aba001kg.vfi).

Table 1. Character Set (Characters 7 and 8)

um	ISO 646 or ISO 5426
gi	all accent marks are recorded by means of the GIZMO notation
lg	PC Latin 2 (Microsoft Code Page 852) + GIZMO
kg	Code Page Kamenicky + GIZMO
uc	UNICODE UTF 8
sg	ISO 8859-2 + GIZMO
an	ANSEL

Table 2. Data Format (Characters 9 through 11)

dat	file exported from ALEPH
rum	textrow UNIMARC
uis	UNIMARC ISO 2709
vfo	ISO 2709 exchange format
vfi	exchange format, file exported from the CDS/ISIS system
dtt	file exported from ALEPH 500

An example is provided by aba006lg.uis, where the institution identifier is ava006, the Code Page is PC Latin2 + GIZMO and the data format is UNIMARC ISO 2709. All records, both the new and the edited ones, are tested before they are imported into the Union Catalog.

The automatic test is set up

- to test the file for UNIMARC compliance: an application automatically tests individual records for their compliance with UNIMARC field format;
- to weigh the record quality: there are six quality weight grades (4, 9, 10, 11, 12, 20), and the higher the grade, the better the record;
- to test the record for duplicates: the records are automatically tested for duplicates, and, if necessary, the duplicates test is accomplished in two

stages, the duplicate record with higher weight replacing the one with lower weight.[1]

Having tested a given record for duplicates in the current Union Catalog, the CUBUS system imports it into the Union Catalog database and then compares it with all the records within the currently processed batch, i.e. all records are also tested for internal duplication within their own batch. The resulting report on non-complying records is sent (by e-mail) for correction with appropriate commentary to the library that provided them. In case of a 5,000-record batch, the whole processing takes about 40 minutes and the administrator can set the start date and time. A member can also read the import result from the statistics available after entering the password at the CASLIN website.

6 The Array of Libraries Cooperating with the Union Catalog

As a matter of principle, the Union Catalog is open to all libraries and information institutions in the Czech Republic, regardless of the library system they are using. Libraries cooperating with the Union Catalog include

- central, universal, and specialized libraries;
- university and college libraries, and libraries of the Czech Academy of Sciences;
- public libraries in statutory and district towns;
- other libraries whose collections comply with the qualitative standards of inter-library services.[2]

---

[1] The UNIMARC testing algorithm, weight calculation algorithm and the algorithm for primary and secondary record comparison are described in detail in the document "CASLIN—Union Catalog ČR". The data necessary for programming the ORACLE-based applications are available at the CASLIN website (URL http://www.caslin.cz)[13].

[2] The list of libraries contributing their records to the Union Catalog, including the number of records supplied, is available at the CASLIN website (http://www.caslin.cz).

The growth of the database shown through the number of records since 1995 is shown in Figure 1:

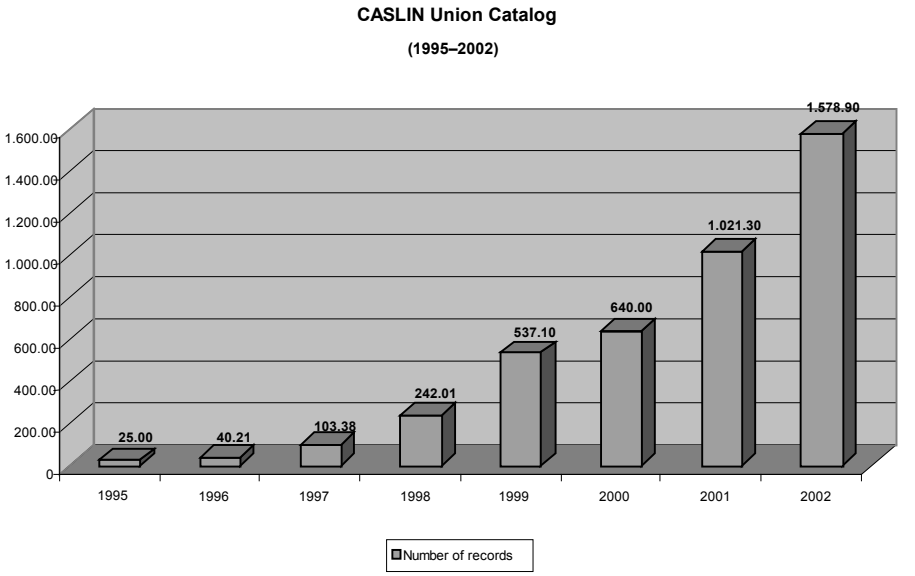

Figure 1. CASLIN Union Catalog, 1995–2002

7 Union Catalog Hardware

There are no special hardware and software requirements for the Union Catalog users. All services provided by the Union Catalog work well with the Netscape browser version 4.04 and higher and with the MS Explorer version 4.01 and higher. The Union Catalog is operated on an Alpha Server 1200 with 1GB RAM. Disk capacity is around 100GB. Digital Unix is the operating system used. The database server is provided by Oracle.

8 Union Catalog Software

Since 1995, the Union Catalog has been operated in three different systems: CDS/ISIS, ALEPH, and CUBUS.

CDS/ISIS (1995–1996)

In the years 1995–1996, the Union Catalog was operated under the CDS/ISIS system as a union database with duplicates testing, containing records on foreign documents only (Union Catalog CEZL—*Centrální evidence zahraniční literatury*). The records were regularly converted to the ALEPH system and made available on the Internet, while only the search function was possible on the Union Catalog. In late 1996, the Union Catalog contained more than 40,000 monograph records.

The ALEPH System (1997–1999)

In the years 1997–1999, the Union Catalog for monographs was operated in the ALEPH system and included domestic as well as foreign documents. Duplicates testing was available, although it was cumbersome and performed by external programs. The Union Catalog was used for searching (location of a document), and offline record sharing was possible only between the members that used the same version of ALEPH.

ALEPH made it possible to solve two key problems:

1. Online shared cataloging;
2. Duplicates testing, including the preservation of better-quality records and formal logical tests.

Problem #1: Online Shared Cataloging

Even earliest strategic analyses for the future union catalog maintained that "the target principle of the CASLIN Union Catalog is online shared cataloging." However, ALEPH does not allow any manipulation of database records from outside itself, which, as a practical matter, prevents any online shared cataloging for members using different library systems, since they cannot carry out primary cataloging in the Union Catalog's own

database [21]. This fact was officially communicated to the Union Catalog members at the Union Catalog Task Team meeting on June 3, 1997 [28].

Problem #2: Duplicates Testing

ALEPH is a high-quality library system which, however, does not allow for the import of records with duplicates testing that matches the needs of a real Union Catalog operating in a heterogeneous environment. When it became obvious that Ex Libris was not able to satisfy the special requirements of the Czech National Library and modify the programs supporting Union Catalog administration (the ULM module), a Czech company developed a duplicates testing program. The program is of very good quality, but since ALEPH does not permit record manipulation from outside ALEPH, the necessary program modifications were somewhat cumbersome and had to be carried out outside the database. This solution meant that the duplicates testing procedure had to be activated outside the ALEPH database, and hence it was only possible to process data offline [3]. Another highly restrictive factor was that the duplicates testing and logical inspection procedure required the administrator to start seven support programs manually. The procedure for processing one batch of data provided by one library (regardless of whether the batch contained 100 or 10,000 records) took two days. And the time requirements were increasing in direct proportion to the size of the database. As a result, it was not possible to import the records supplied by the ever-increasing number of Union Catalog participants in real time [19].

These issues triggered a discussion on developing a system of administering and operating the Union Catalog with our own resources. In September 1997, the Union Catalog administrator presented a document entitled "A Potential Path of Future CASLIN Union Catalog Development with Regard to the Up-grade to ORACLE 7" [12], and in October 1997 we applied for a grant from the Mellon Foundation, which would allow us to purchase the ORACLE database system and develop our own union catalog system.

ORACLE became the tool for developing a system for the administration and operation of union catalogs, and the new system was called CUBUS. For system development, we chose a smaller software company with which we had cooperated since 1994 and which also created

the duplicates testing program mentioned above (fully implemented and improved in the new system). The company was paid by funds awarded by the Ministry of Culture of the Czech Republic to the project "CASLIN Union Catalog Development". In November 1997, a representative of the software company presented the basic philosophy of the new system to members of the Union Catalog Task Team and to members of the Union Catalog Research and Development Team. Also invited were the directors of all major libraries and representatives from the Slovak Republic. The crucial task was to draw up the requirements for the development of a new system.

The CUBUS System (since 2000)

During the first quarter of 1998, ORACLE was installed on our ALPHA server, and in March 1998, the "Requirements for Application Development under the ORACLE System" [13] was published on the CASLIN website. The members of the Union Catalog Task Team and of the Union Catalog Research and Development Team were invited to submit their comments to the Union Catalog administrator. After processing the comments in mid-1998, the development of CUBUS was launched. The resulting product contained the following improvements [17]:

1. Work procedures were optimized by eliminating human intervention wherever possible and effective;
2. Data control was improved by supplementing "human-based" control with automatic formal logical control;
3. The comparison keys for duplicates inspection were expanded and became more subtle, resulting in a reduction of unwanted duplicates in the Union Catalog [9];
4. Statistical monitoring was introduced to cover the movement of data, the administrator's performance and the work of Union Catalog members;
5. The user interface appearance and functionality followed the access rights setup and configuration changes,
6. Shared cataloging and editing of old records already in the database was improved;

7. Data security was improved; and
8. The possibility was added to modify existing applications and to develop new ones for better performance in the future.

The beta version of CUBUS was provided to the Union Catalog Research and Development Team in September 1999 and to the Union Catalog Task Team in November 1999. Most of their comments were implemented and, early in 2000, the Union Catalog under the new system became accessible to the general public.

Since 2000, the Union Catalog has operated under CUBUS, which fully meets the requirements of the Union Catalog administrator as well as the needs of its users. CUBUS is owned by the Czech National Library. It is equally open to all Union Catalog members regardless of the system used by the member in question, and thus solves the problem arising from the heterogeneous nature of the library environment in the Czech Republic.

CUBUS offers solutions in critical areas:

1. Real-time data import in supported formats and code pages without compromising any further operations within the Union Catalog database;
2. Formal logical data inspection;
3. Duplicates testing using both primary and secondary keys with further amendments for series;
4. A search interface reflecting UC member requirements and capable of parallel processing of any number of simultaneous search requests from different users;
5. A direct link to the Library Directory is provided;
6. It distinguishes between users and active UC participants;
7. It provides batch-based data export in the supported formats and character sets;
8. It allows the user to place ILS loan requests;
9. It uses online shared cataloging by means of an entry form; and
10. A link exists to MetaLib, established in cooperation with Ex Libris (Uniform Information Gateway) via the HTTP protocol by means of the XML format.

Several problems remain at the time of writing:

1. The UC records are to be linked to the authority records of the cooperative authority database under construction in the Czech National Library;
2. The Z39.50 communication protocol is to be implemented; and
3. The Directory is to be transferred to CUBUS.

Technological Parameters of the CUBUS System[3]

From the very beginning, it was clear that, once developed, CUBUS was going to have a single installation and would accommodate only the Union Catalog. This is why its development was entrusted to a smaller software company, whose programmers were better motivated to deal with the client's requirements. This option also brought about a reduction in the development costs. On the negative side, this choice involved higher risks related to the long-term stability of such a company. To minimize this risk, we adopted the following important steps:

1. To minimize dependence on a particular implementation team, we used only widely available technology (Oracle, PL/SQL, Java, HTML, XML); and
2. Professional software was created to document in detail the functions and structure of the system, which allows for rapid acclimatization of new staff (analysts and programmers) into the team.

The Openness of the CUBUS System

Openness is understood as the ability to provide the necessary interfaces to interlinked systems. Such interfaces should accommodate the established de facto standards. From the very beginning, CUBUS was designed to focus on established de facto standards (TCP/IP, HTTP, FTP etc.), which has made it possible for the system to expand rapidly and widely. This also eliminated the high risks involved in using de jure standards (i.e. standards defined by a relevant commission) that had not been previously verified in real implementations.

---

[3] The following description was provided by an independent analytical company [1].

The Flexibility of the CUBUS System

A system's flexibility consists in its ability to absorb modifications aimed at changing or extending its functions. A system's flexibility depends, first, on the technology used and, second, on the way it has been implemented (e.g. a system's modularity increases its flexibility). To achieve a sufficient level of flexibility, it is also important to provide consistent and easy-to-maintain documentation.

CUBUS provides sufficient flexibility because it employs widely available technology. This applies both to the database itself, built in Oracle, and to other tools used such as PL/SQL, Java, and servlets. Hence the system is not tied to a narrowly specialized development team, since the technology utilized is widely used and known. Considerable effort was also focused on a coherent conceptual approach to functionality requirements, which now makes it possible to easily expand the functions of the system. The system is continuously monitored by special software, a tool that allows quick and thorough analyses of the potential impact of any planned changes.

The Scalability of the CUBUS System

Scalability is interpreted as the ability to increase a system's performance without having to modify it, or, in other words, the possibility to boost performance through a mere hardware upgrade and administrative operations. CUBUS' scalability is chiefly assured through using a robust relational system for database administration, Oracle, which is capable of absorbing several times more data than it currently holds without affecting the current speed of responses to users' queries.

During its development, CUBUS has not encountered any limitations. Its basic development has been completed both conceptually and practically. It is a system based on modern and widespread technological solutions (Oracle, Java, and servlets), and any improvements and modifications for changing users' needs are easy to carry out. In 2001, we had purchased a professional tool that made it possible to create a detailed data and process model of CUBUS, thus obtaining a high-quality description that is easy to understand. Since 2001, a full-scale copy of CUBUS is operated on a separate server that may be used as a backup when implementing an upgrade.

There is no serious reason for abandoning the system and having to face again the obvious limitations that ALEPH imposes when it is used for administering and operating a real-life heterogeneous union catalog.

CUBUS as free software

The presentation of our paper at the Tallinn conference inspired one of the guest lecturers, Stefan Grandmann, to inquire about the ownership of the CUBUS system and whether there was any possibility of making the CUBUS system available as free software under the GNU GPL (General Public Licence) provision.

The author of the application, who owns the moral rights, and the National Library of the Czech Republic, which holds the economic rights, share the copyright of the CUBUS system. If the author of the application gives his written authorization, the National Library of the Czech Republic can consider making the CUBUS software available as free software under GNU GPL[4].

ALEPH + CUBUS (2002–?)

In April 2002, without identifying any factors that would mark the CUBUS system as an inadequate application, the Czech National Library decided to reverse course and operate the Union Catalog under ALEPH again, with the proviso that some modules (data import with duplicates testing, and online

[4] Software licences are mostly designed to take away the right to share and change the program freely. By contrast, the GNU General Public Licence is intended to guarantee freedom to share and change "freeware." This GNU General Public Licence applies to most of the Free Software Foundation's software. A short quotation from the GNU GPL preamble: "When we speak of free software, we are referring to freedom, not price. Our General Public Licences are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things. "

updating of series records) would nevertheless keep running under CUBUS, because otherwise they would have to be terminated.

The following key reasons were given for this decision:

- The Z39.50 protocol has not been implemented in CUBUS;
- There is no link from CUBUS to the authority files;
- It is not practicable for the Czech National Library to operate two systems (ALEPH and CUBUS); and
- CUBUS was produced by a small software company.

As for the first two issues, suffice it to say that both features should have been provided by the end of 2002 (see below for more detail) and had been planned in the projects submitted to the Ministry of Culture where we had applied for funding.

The third reason was dropped as soon as it had become obvious during a discussion at an Expert Council meeting that CUBUS cannot be entirely abandoned after all.

As to the fourth reason, it is evident that the small size of a company raises concerns about its long-term stability. But that concern needs to be contrasted by the fact that during the period in which the Union Catalog was operated under ALEPH, it proved impossible to induce Ex Libris, definitely a major and stable company, to implement any of the requested improvements to the Catalog. The reason is evident: an international software company which has several large installations in the United States, among others, is unable to deal with a single requirement of a minor customer somewhere in the heart of Europe. This may make one ponder what is more beneficial for the Union Catalog in the Czech Republic: to be an insignificant minor customer of a major software company, or to be a major customer of a minor but stable software company?

How Will the Union Catalog Function under the ALEPH-CUBUS System?

The transfer to ALEPH will take place in two stages:

In the first stage (expected completion by January, 2003), CUBUS will process the members' batch imports during the daytime, while at night it

will export the new or updated records in the RUX format.[5] The file will be imported into the ALEPH database, the records of which will be updated according to the system number. Users will be able to search and export records under the ALEPH system.

In the second stage of the CUBUS–ALEPH transition, after installing the Z39.50 protocol containing the Update function on ALEPH,[6] only those libraries which have implemented the Z39.50 protocol will be able to catalog their monograph and special document records online in the Union Catalog. Every day at 7 pm, ALEPH will close the data editing function in the Union Catalog and will export the new or newly edited records. This file will be imported into CUBUS by a standard method, thus assuring the congruence of the two databases. Then CUBUS will turn to the batch record imports that arrived during the day. By 4 am, the newly imported records will be exported in RUX format and they will update the ALEPH database according to their system number.

The serials records will continue being updated online in CUBUS without having to close the database to users [11].

The Union Catalog administrator was not invited to attend any of the policy discussions of the Czech National Library's top management about the Union Catalog platform change. Instead, she made use of a Union Catalog presentation to inform the management about the potential problems of operating a Union Catalog under ALEPH in combination with CUBUS.

A functional connection between the two systems would present the Union Catalog administrator with the following problems [25]: the National Library's need for programmers will increase; the number of ALEPH licenses needed will have to grow[7]; hardware requirements for ALEPH will grow in excess of planned system upgrades; data administration (especially

---

[5] RUX is an internal format for data import into ALEPH and is based on the structure of textrow UNIMARC in UTF-8 + RecID.

[6] Although it appears that Ex Libris has totally abandoned the development of a Z39.50 protocol containing the Update function.

[7] CUBUS is licensed for an unlimited number of access points.

imports) will continue to be provided under CUBUS; and the processing of series under different library systems will be highly problematic. Hence, series processing including their online updates is to remain under CUBUS, which will still have to be maintained and developed. Finally, problems that exist in local library systems may compromise the functionality of the national Union Catalog and vice versa.

Users will not be offered new functionalities or benefits; on the contrary, the change will cause service to deteriorate in many respects. Thus, the Union Catalog's handling of series, ILS, and the Directory is bound to suffer, and the existing unlimited license for access with no time restrictions will be replaced by one with time restrictions. Furthermore, the benefit derived from the existing connection with the Universal Information Gateway (UIG) through HTTP and XML will be lost.[8] CUBUS-to-ALEPH data imports will have to be carried out overnight, for which the National Library's databases may be closed to users, although statistics show that the Union Catalog is accessed at night as well (e.g. by users from distant time zones, or by those waiting for a freer and faster Web). Under CUBUS, imports are conducted in the background, and so there is no need to close the database. With ALEPH, users will be provided only with an out-of-date copy of the Union Catalog, which is not only a non-standard approach, but is counterproductive and to be used only for very serious objective reasons.

The reasons that made the Czech National Library start to develop its own Union Catalog system in 1997 still exist. Even in combination with CUBUS, ALEPH does not have the tools to successfully operate a heterogeneous Union Catalog. ALEPH is a very good system designed to administer and operate a library, but not suited for a real-life heterogeneous union catalog.

The Union Catalog run under ALEPH will again be limited by the capabilities of ALEPH, which—although broader than in 1997—are still unable to react flexibly to the requirements and demands of its customers, in contrast to the functionality of CUBUS.

---

[8] The existing UIG connection does not impose license requirements on the use of the Z39.50 protocol within UIG.

9 The Connection between Authority and Bibliographic Records in the Union Catalog

The National Authority Department at the Czech National Library accomplishes two basic objectives:

1. The administration of local authorities at the Czech National Library; and
2. The formation of a national authority database for libraries of the whole Czech Republic.

In December 2000, representatives of Czech libraries and library system vendors met at the Czech National Library to discuss the cooperative creation of a national authority database. This discussion ended with a clear and unambiguous recommendation to build the national authority files within the Union Catalog system, i.e. under CUBUS, after which the Union Catalog administrator drafted a document that identified the fundamental policies in creating a cooperative national authority file. This document was posted on the CASLIN website by the end of 2000 [14].

Simultaneously, the administrator presented a proposal to the management of the Czech National Library for a reorganization [20] that included the transfer of activities related to the creation of a national authority database to the administration of the Union Catalog. This proposal was rejected.

In January 2001, the top management of the Czech National Library announced that the authority files would be built under ALEPH [5]. This denied the opportunity to build a single system (CUBUS) for both bibliographic and authority records within the framework of the Union Catalog. Subsequently, alternative paths were explored.

At first, the viability of linking the bibliographic records from the CUBUS-based Union Catalog with the ALEPH-based authority database was examined. It turned out that even ALEPH 500 does not permit record modifications using tools other than its own.[9] Hence, it was impossible to

[9] Such modifications have to be carried out outside the ALEPH database, as was the case with ALEPH 300.

set up an online link between the bibliographic records from the CUBUS-based Union Catalog with the ALEPH-based authority database. The Union Catalog administrator proposed and provided a detailed description of a solution whereby a copy of the authority database would be defined within the CUBUS system. The database copy would be updated every night, so as to achieve congruence with the ALEPH master [15]. This would have enabled CUBUS to establish a link between the bibliographic records in the Union Catalog and the authority records, thus making the records in the Union Catalog, and the authority database—all very straightforward authority records—available to all Union Catalog users regardless of which system they were using. The Union Catalog administrator's objective was to expand CUBUS with functionalities for gathering and administering authority records, closely linked to functionalities for work with bibliographic records. By storing the bibliographic and the authority records in one database, the overall complexity of the system would be reduced, since measures to treat duplicates and inconsistencies created during data transfer between different systems would become unnecessary. It would also provide an opportunity for establishing a link between the data. The increase of complexity under the alternative solution also increases the probability of errors, and hence of costs. The biggest benefit of storing the bibliographic and the authority records in one database would stem from the reduction of processing costs. It would be possible to make use of the existing applications for bibliographic record administration, which would also allow us to create a two-way link between the bibliographic and the authority data. Such a link would make it possible to display all relevant bibliographic records for a given authority.

10 The Union Catalog and the Z39.50 Protocol

As late as 2000, there were extremely few libraries in the Czech Republic that had a truly functional Z39.50 protocol implemented in their systems [24]. Basically, only the ALEPH libraries, probably four in number, had it, which is why the implementation of Z39.50 protocol was low on the list of priorities during the development of CUBUS. At the same time, the Union Catalog administrator was aware of the fact that due to the inherent

openness of CUBUS, the Z39.50 protocol would not provide the system with any new functionality and would not help other entities to join the Union Catalog and enhance its expansion. The pressure on the part of the Czech National Library's management to implement the Z39.50 protocol was not, at the time, based on any existing needs of the Union Catalog members.

In 2001 and 2002, rather than developing its own Z39.50 protocol, the Union Catalog administrator was planning to implement one designed by an independent software company. For successful implementation of a protocol from a different company, it was crucial to describe the existing CUBUS system, and so the Union Catalog Department staff created a document called "A Functional Model of the CASLIN Union Catalog" [15].[10] Although the model was correct and effectively described the state of affairs and anticipated CUBUS developments, it did not explicitly differentiate the process and data models of the system, and the data flow diagrams were not based on standard description tools (which were not available to the Union Catalog administrator at the time), which is why this functional model was not sufficient for the Z39.50 vendor. The only logical solution was to entrust the creation of CUBUS' data and process models to an outside group of consultants. The group completed the analysis of the current state of CUBUS, including a forecast of its future connection with the authority database and of the implementation of the Z39.50 protocol. The group also prepared a documentation of CUBUS, which included a rational approach to recording changes. This tool is still available and at the Union Catalog administrator's disposal.

11 The Union Catalog in the Uniform Information Gateway

MetaLib, the Uniform Information Gateway (UIG) software by Ex Libris, may be connected to sources through the Z39.50 protocol or through the HTTP protocol. To set up a connection through the HTTP protocol is more

[10] Available at the Library and Information Sciences' Reading Room at the Czech National Library.

difficult, and so Ex Libris is usually willing to provide it only for world-class information sources. Connecting local sources accessible only through a HTTP protocol is quite expensive, because it requires more programming on the part of Ex Libris [18]. The Union Catalog belongs to the latter class, and we are grateful to Ex Libris for connecting it without additional charges. The Czech Libraries appreciate the presence of a national information resource in the UIG, and UIG, in turn, adopted the CASLIN logo, which is well-known not only in the Czech and Slovak Republics, but elsewhere as well.

The Union Catalog was connected to UIG through an external program assuring the conversion of a query and its result between the two systems [26].

A query placed by a UIG user is sent to the Union Catalog through the HTTP protocol[11] coded in UTF-8 and in the format http://server address?access file=value&access file=value. The output is in the form of a record exported in XML. The extent of the record is a compromise between the data displayed in MetaLib and the specifics of the Union Catalog in question, i.e. the record contains the basic identification data of the document plus the name of the library that owns it, and for periodicals also the year.

For the output in XML, the Union Catalog administrator created a DTD[12] that complies with the current Union Catalog requirements for a connection to UIG. In the future, it will pose no problem to expand the proposed DTD or modify it in order to comply with the worldwide accepted definition of a document type for library formats.

The method of establishing a connection to UIG via HTTP and XML will be also used to create a connection to other sources that neither work with the ALEPH system nor use the Z39.50 protocol.

---

[11] See http://www.caslin.cz:7777/caslin/ENG/parameters.html.

[12] DOCTYPE CaslinMeta. For more information see http://www.caslin.cz:7777/caslin/ENG/dtd.html.

A Minor Excursion into Real and Virtual Union Catalogs

The introduction of computers, library systems, and the MARC format represented a quantum leap in the accessibility of information in union catalogs in comparison with their card-based predecessors. And while the tasks and problems in administering a union catalog[13] have not changed, the cost of administering and operating electronic union catalogs has probably changed for the worse. The high cost of operating a real union catalog and the proliferation of the World Wide Web are probably the main reasons for the appearance in the early 1990s of the first information gateways. In the mid-1990s, two nationwide virtual union catalogs emerged in the Czech Republic:

1. The homogeneous TinWEB, established as a system for parallel searching of library catalogs within the TINLIB system; and
2. The heterogeneous ATpar (ALEPH-TINLIB Parallel Query System), designed and implemented for single-query transparent searching of a selection of library catalogs with an ALEPH-based and TINLIB-based WWW interface [2].

In the library community, access points to Z39.50-based sources are established via information gateways based on WWW, where the end-user interface is a browser (in 2001 in the Czech Republic, it was UIG).

In general, an information gateway may be considered a virtual Union Catalog only when it complies with the following requirements (from the end-user perspective):

- Access takes place through a single user interface;
- The format for search queries uses a unified syntax, using a single set of search attributes;
- The records are provided in a single, common format allowing a single record to be displayed;
- Duplicate records are not displayed;

[13] For example data import, duplicates testing, accounting for system heterogeneity.

- All records are available with location data and the full set of data about the library collection;
- All copies are provided with up-to-date loan status and availability data; and
- A virtual union catalog is interlinked with several integrated loan and order services [8].

Ordinary users trying to identify and locate a document particularly appreciate information gateways, functioning as virtual union catalogs. For inter-library loan services and shared cataloging, professional librarians prefer real union catalogs with records from tens or hundreds of libraries that have been classified and evaluated and made available in one database.

It should be noted that real union catalogs provide their users with feedback in the form of error messages, i.e. information about deficiencies in delivered records. Thus the participants have the possibility to improve the quality of document processing in their home institutions. At the same time real union catalogs support the implementation of uniform standards. Virtual union catalogs use a technology that moderates and compensates for the differences among the various participants' systems. Virtual union catalogs do not provide any feedback, its participants do not have the need to improve the quality of its records and there is no need to implement uniform standards.

12 Union Catalog Terms of Payment

Both access to and use of the Union Catalog are free of charge. The members do not obtain any fees for the records they provide, and conversely, they do not pay anything for importing Union Catalog records, which they use for cataloging or retroconversion of their collections [22].

The Czech National Library funds the administration and operation of the Union Catalog, i.e. the Union Catalog is fully state-funded through the Ministry of Culture of the Czech Republic.

13 Union Catalog Organization and Control

Union Catalog Administration in the Czech National Library

In connection with the launch of CASLIN and based on the assumption that the CEZL Department (National Registry of Foreign Literature) would become the administrator of the CASLIN Union Catalog, CEZL staff articulated a detailed "Strategy for the CEZL Union Catalogs' Transformation into the CASLIN Union Catalog" [4]. Based on this strategy, on July 1, 1994, the CEZL Department became, for a short but very useful period of 11 months, a part of the Cataloging Division, where CEZL staff learned the details of the newly discussed standards. As early as June 1, 1995, the CEZL Department became an independent Union Catalog Section (with CEZL and CASLIN as the union catalogs to work with) and it became directly answerable to the Director of the Czech National Library. Currently, the Deputy Director of the Library heads the independent Union Catalog Division [16].

Union Catalog Task Teams

As a part of CASLIN, the CASLIN Task Team for the Union Catalog was formed jointly with Slovak librarians in January 1994. This task force produced high-quality strategic documents dealing with union catalog administration, architecture, construction, and member typology. Due to the ever-increasing scope of its activities, the Task Team's decision-making became too cumbersome, and so in November 1995, the Union Catalog Task Team was established, this time without Slovak participation. This Team has been in operation to the present day.

The cooperation between the Union Catalog and libraries is strictly regulated on a contractual basis. Since October 1996, the agreement concluded between the Czech National Library and the cooperating libraries has been called "Cooperation Agreement on CASLIN Union

Catalog."[14] Consulting bodies available to the Union Catalog administrator are

1. The Union Catalog Task Team, with staff consisting of representatives from libraries that supply the Union Catalog with records, since November 1995;
2. The Union Catalog Research and Development Team, and expert body established in June 1997 by the Union Catalog administrator, whose main task was to overcome technical difficulties arising in the Union Catalog. In late 1997, the R&D Team discussed the issues related the development of a new tailor-made Union Catalog system;
3. The Union Catalog Expert Council, formed in December 2000, which absorbed most of the R&D Team's members.

The members of the Union Catalog Expert Council played a decisive role in formulating the changes in the Union Catalog system. The work of the Council has been adversely affected by a variety of conflicts of interest—to wit, the ALEPH sales representative for the Czech Republic is also a member.

Despite the diversity of opinions and interests, the first vote was in favor of CUBUS [6], and the second ended in a 6:6 draw [7]. It is also the case that that vote was characterized by certain irregularities; for example, the absentee vote of a member was counted, the ALEPH sales representative's vote was not disqualified, etc. The Union Catalog administrator participated in the Council only as a non-voting secretary. This was the vote on the basis of which the management of the Czech National Library decided to replace the current platform of the Union Catalog, although the Union Catalog administrator expressed her disagreement with the decision.

14 The CASLIN Consortium

The cooperation underlying the CASLIN Union Catalog is legally based on the Cooperation Agreement on CASLIN Union Catalog. During 2000, the

[14] Its text is available on the CASLIN website.

Union Catalog Administrator undertook specific efforts to create a CASLIN Consortium.[15] The item on the Association's agenda was to be the Union Catalog.

However, conditions in the Slovak Republic did not permit the creation of a functional international CASLIN Consortium based on union catalog cooperation. The very name of the consortium proved to be a contentious issue, since its acronym contained the letter "S" standing for Slovak, although none of its would-be members was Slovak. Although imprecise, the existing acronym acquired such a familiar status in the Czech library environment that it seemed counterproductive to abandon it. In this situation, during the preliminary discussions, the Union Catalog members themselves suggested a new name that would fit the original acronym: CASLIN would stand for the Czech Association for Services in Library Information Network.

Developments on the Slovak side seem to indicate that the their union catalog initiatives have been following a different path from their Czech CASLIN counterpart, both in relation to the speed of development and to its strategy. Nevertheless, the crucial orientation established at the beginning of the joint CASLIN Project in 1993 has been preserved. There is always a real possibility that the CASLIN Consortium will expand and become, once again, a functioning international body, thus returning to its roots and forming a truly Czech and Slovak library information network. The idea of founding a union catalog-based CASLIN Consortium has been abandoned, since it is now part of the legal mandate of the Czech National Library (the new legislation was passed in 2001) to build and run a national Union Catalog.

15 Conclusion

There are no signs that suggest that, on a worldwide scale, real union catalogs are beginning to fall into disuse, not even in countries where

[15] A draft of the Statutes of the CASLIN Association was drawn up, and the Cooperation Agreement was amended.

information technology is very advanced and virtual union catalogs are relatively easy to build. Real union catalogs represent a unique tool for value-added services expected and required by both users and librarians. Virtual union catalogs provide additional functions (especially those of location) suitably complementing their real counterparts.

The Czech National Library has not set out to build its virtual union catalog (i.e. the Uniform Information Gateway) with the aim of providing a supplementary service to the existing real Union Catalog. It is meant to become the primary union catalog format, which is corroborated by the decision to stop developing CUBUS as a real union catalog system. This approach has de facto compromised the ability of the union catalog to provide equal service to all users and, above all, to libraries with different library systems.

References

[1] *Analýza systému CUBUS*. LBMS, c2002, p. 16.

[2] Bartošek, Miroslav. "ATpar: Aleph-Tinlib Heterogeneous Virtual WWW Catalog." In *Union Catalogs: their organization and services. Zadov (Czech Republic), 30th–3rd 1999*. Praha: Národní knihovna ČR, 2000: 47–53. ISBN 80-7050-354-8. Also at http://www.caslin.cz:7777/caslin99/s2.htm.

[3] Dvořáková, Helena. "Budování souborného katalogu s ohledem na možnosti dané programem na řízený import s kontrolou duplicit." Unpublished paper. Praha: Národní knihovna ČR, 13.11.1996.

[4] Dvořáková, Helena, Gabriela Krčmařová, and Lydia Vítová. "Koncepce transformace CEZL v souvislosti se souborný mi katalogy CASLIN." Unpublished material. Praha: Národní knihovna ČR, 21.3.1994.

[5] *Expertní rada pro Souborný katalog ČR: zápis 1.jednání dne 24.1. 2001*. Praha: Národní knihovna ČR, 2.2. 2001. Available at http://www.caslin.cz:7777/caslin/experti1.html.

[6] *Expertní rada pro Souborný katalog ČR: zápis 2. jednání dne 27.2. 2002*. Praha: Národní knihovna ČR, 11.3.2002. Available at http://www.caslin.cz:7777/caslin/exp0202.html.

[7] *Expertní rada pro Souborný katalog ČR: zápis 3.jednání dne 20.3.2002*. Praha: Národní knihovna ČR, 26.3.2002. Available at http://www.caslin.cz:7777/caslin/exp2003.html.

[8] Hajná, Kateřina. "Testy na duplicitu klíčů v souborném katalogu." Unpublished paper. Praha: Národní knihovna ČR, březen 2002:4.

[9] Hajná, Kateřina et al. "Časová a finanční rozvaha vývoje systému CUBUS v roce 2002 (2003)." Unpublished paper. Praha: Národní knihovna ČR, březen 2002.

[10] Hajná, Kateřina et al. *Zadání pro vývoj aplikací v systému ORACLE: tištěné seriály*. Praha: Národní knihovna ČR. Available at http://www.caslin.cz:7777/caslin/caslin-ser.html.

[11] Hajná, Kateřina and Gabriela Krčmařová. "Záměr rozvoje CASLIN—Souborného katalogu ČR v souvislosti s přechodem na provoz v systému ORACLE 7: materiál na poradu vedení NK ČR dne 24.9.1997." Unpublished paper. Praha: Národní knihovna ČR, 5.9.1997:5.

[12] Hajná, Kateřina and Gabriela Krčmařová and Danuše Vyorálková. *Zadání pro vývoj aplikací v systému ORACLE: tištěné monografie.* Praha: Národní knihovna ČR, 1998. Available at http://www.caslin.cz:7777/caslin/zadani2.html.

[13] Hajná, Kateřina and Eva Svobodová. *Budování báze autorit v prostředí CASLIN—Souborného katalogu ČR (SK ČR)*. Praha: Národní knihovna ČR. Available at http://www.caslin.cz:7777/caslin/autor-zad.html.

[14] Hajná, Kateřina and Eva Svobodová. *CASLIN—Souborný katalog ČR: funkční model*. Praha: Národní knihovna ČR, únor 2001: 36.

[15] Husby, Ole. "Real and Virtual Union Catalogs." In *Union Catalogs: their organization and services, Zadov (Czech Republic), 30th–3rd 1999*. Praha: Národní knihovna ČR, 2000: 112–120.

ISBN 80-7050-354-8. Also available in HTML version at http://www.caslin.cz:7777/caslin99/a2.htm.

[16] Krčmařová, Gabriela. "35 let historie CEZL a současnost v rámci CASLIN." *IT lib: informačné technológie a knižnice*, 4 (2000). ISSN 1336-0779. Available at http://www.cvtisr.sk/itlib/bc2000_4/krcmarova.htm.

[17] Krčmařová, Gabriela. "CASLIN—The Union Catalog of the Czech Republic," In *Union Catalogs: their organization and services, Zadov (Czech republic), 30th–3rd 1999*. Praha: Národní knihovna ČR, 2000: 158–164. ISBN 80-7050-354-8. Also available in HTML version at http://www.caslin.cz:7777/caslin99/a7.html.

[18] Krčmařová, Gabriela. "CASLIN—Souborný katalog ČR: materiál na poradu vedení dne 21.11.2001." Unpublished paper. Praha: Národní knihovna ČR,19.11.2001:3.

[19] Krčmařová, Gabriela. "Kam kráčí CASLIN- Souborný katalog ČR." In *Automatizace knihovnických procesů—VII, sborník ze 7. ročníku semináře pořádaného ve dnech 9.-10. června 1999 v Ústí nad Labem*: 59–69. ISBN 80-238-4233-1.

[20] Krčmařová, Gabriela. "Návrh na organizační změnu oddělení pro souborné katalogy." Unpublished material. Praha: Národní knihovna ČR,18.12. 2000+:2.

[21] Krčmařová, Gabriela. "Od systému ALEPH k systému ORACLE." In *INFOS´98. 2001: zborník z 28. medzinárodného informatického sympózia, ktoré se konalo v dňoch 27.-29. apríla 1998 v Starej Lesnej: 66-73*. ISBN 80-85165-67-8. Also available in HTML version from http://www.caslin.cz:7777/caslin/wwworac.html

[22] Krčmařová, Gabriela. "Rozvaha a návrh financování CASLIN— Souborného katalogu ČR." Unpublished paper. Praha: Národní knihovna ČR, srpen 1998:6.

[23] Krčmařová, Gabriela. "Souborné katalogy v heterogenním prostředí." In *INFOS 2001: zborník z 31. medzinárodného informatického sympózia,*

ktoré se konalo v dňoch 2.–5. apríla 2001 v Starej Lesnej: 310-318. ISBN 80-85165-84-8.

[24] Krčmařová, Gabriela. "Z39.50—protokol z minulého století." *IT lib: informačné technológie a knižnice*, (4) 2001. ISSN 1335-793X. Available at http://www.cvtisr.sk/itlib/itlib014/z3950.html.

[25] Krčmařová, Gabriela and Eva Svobodová. "Prezentace CASLIN—Souborného katalogu ČR na PV dne 5.4.2002." Unpublished paper. Praha: Národní knihovna ČR, 2002:3.

[26] Krčmařová, Gabriela and Ilona Trtíková. "Funkcionalita CASLIN—Souborného katalogu ČR jako centralizovaného zdroje pro informační profesionály a katalogizátory." In *INFORUM 2002: 8. konference o profesionálních informačních zdrojích, 21.-23. května 2002, Nová aula Vysoké školy ekonomické, Praha*. Available at http://www.inforum.cz/inforum2002/prednaska15.htm.

[27] *Zápis 3. jednání Pracovní skupiny pro souborný katalog při NK ČR*. Praha: Národní knihovna ČR, 3.6.1997.

Links

http://www.caslin.cz—Website of CASLIN Union Catalog.

http://psi.nkp.cz:2400/r/SKK/p210/pcz—Access to CUBUS:

CASLIN Union Catalog—Monographs and special types of documents.

CASLIN Union Catalog—Serials.

http://sigma.nkp.cz:4525/ALEPH0/~/START/adr—Access to Aleph:

CASLIN Union Catalog—Directory of libraries and information centers of the Czech Republic.

http://www.caslin.cz:7777/caslin/historie/document.html—Documents about CASLIN.

http://www.caslin.cz:7777/caslin/dtd.html—Description of CaslinMeta.

http://www.caslin.cz:7777/caslin/parameters.html—Description of query CUBUS.

http://jib-info.cuni.cz/dokumenty/dokumenty_tech.html—Specification of http interface for connecting to Metalib.

http://dior.ics.muni.cz/hales/atpar/—ATpar (ALEPH-TINLIB Parallel Query System).

http://sd.ruk.cuni.cz/tinweb/sd/k6—TinWEB.