Erik I. Vajda,
Principles of a National Union Catalog: Shared Cataloging in a Small Country

from / *aus*:

Union Catalogs at the Crossroad
Edited by
Andrew Lass and Richard E. Quandt
pp. */ S.* 327-338

Erstellt am 31. März 2005

Impressum

Table of Contents

Part 4 Hungarian Union Catalogs

Part 5 Baltic Union Catalogs

Part 6 South African Union Catalogs

Chapter 17
Principles of a National Union Catalog: Shared Cataloging in a Small Country

Erik I. Vajda

The main aim of this paper is to outline some specific characteristics and the background ideas for some decisions concerning the establishment of MOKKA[1,2] the Hungarian National Shared Cataloging System. We assume that the discussion of some of these ideas and of the resulting decisions may contribute (first and foremost, but not exclusively) in smaller countries to the development or improvement of similar systems in the national environment, i.e. shared cataloging systems with the participation of major libraries and national union catalogs as the product of the shared cataloging.

Some of the more or less system-specific characteristics and considerations leading to these decisions are related to the peculiar features of the Hungarian library environment. However, it might eventually also be useful

---

[1] MOKKA is the acronym for the Hungarian name of the Hungarian National Shared Catalog (*Magyar Országos Közös Katalogus*). See also http://www.mokka.hu.

[2] There are many papers available about shared cataloging and union catalogs in Hungary, including those dealing with MOKKA. However, these papers are all in Hungarian, and therefore no references are given, except for a single one about the problems of subject searches in a shared cataloging environment. For a general introduction to MOKKA, reference is made to the website of MOKKA (also in English) in general, and to the page http://www.mokka.hu/e-bemutat.html in particular. This page describes the history, aims, functional model, structure and possible future of MOKKA.

for some other smaller countries to get acquainted with these decisions and their background, whereas other decisions are based on considerations that seem to be relevant for most national shared cataloging systems, independent of the size of the country. In the first part of this paper, we try to give a survey of these characteristics, whereas in the second part we analyze in detail the question of physical versus virtual union catalogs; a broad question that has been discussed intensively in Hungary as well.

At the outset, one has to realize that the establishment of a shared cataloging system and of a union catalog starts in an environment of libraries with various traditions, habits, computerized library systems etc. However, it is the common interest of all libraries (whether participating in a national system or not) and of their users to have a tool for retrieval from the stocks of all the libraries, the holdings of which cover the majority of titles available in the country. In the paper, we discuss both the major problems and their possible solution.

1   Size of the System

The optimal size of a shared catalog and of the national union catalog system can be defined only on the basis of an analysis of goals to be achieved. These goals are:

- To create a tool for libraries and library users that enables them to determine the libraries in which they can find and borrow, or get a copy of, a given document available somewhere in the libraries of the country, but not available in the library in which this demand originated;
- To simplify the processing (cataloging) of documents by copying/ downloading items of existing records; and
- To contribute to the use of common standards and standard-like solutions for cataloging and retrieval.

Statistical investigations reveal that it is not necessary to include the catalog data of all libraries in a country, or even of the majority of libraries, in the shared cataloging system (i.e. in the national union catalog). In Hungary, it was proved by such investigations (based on an existing manual union catalog of documents published abroad) that these goals could be achieved

to a great extent by a shared cataloging system of fewer than 20 libraries. As a matter of fact, the 20 or so libraries in this set are, in any event, the main suppliers of documents in.inter-library loan and copying services, without taking into account the great county public libraries.

If, however, the coverage of titles needs to be extended even further, the inclusion of the catalog data of more libraries and/or virtual solutions—i.e. the near-completion of the physical union catalog by adding more data from other physical or virtual shared catalogs—can further improve the coverage.

In Hungary, 17 libraries (now actually only 15, because of the merger of four member libraries into two libraries) hold about 70% of all foreign titles available in Hungary and nearly 100% of Hungarian titles. These libraries are the members of the MOKKA system.

2   Coexistence—Not Always Peaceful—of Different Library Systems and Standards within One Shared Cataloging System

It is characteristic, with some exceptions, of most (smaller or larger) countries in Eastern and Central Europe that library automation started with the acquisition and use of different automated library systems. For example, the 15 member libraries of MOKKA even use different automated/integrated library systems. The central system of MOKKA uses.one of these systems, CORVINA, a version of which had been further developed for the purposes of MOKKA.

Obviously, the use of a variety of systems by the libraries that supply data to the central database causes a lot of problems. Solutions might be the application of the Z39.50 standard, the up- and downloading of MARC records, or the use of, or conversion to, other common standards. MOKKA decided on a solution based on MARC export and import, since the cataloging modules of the overwhelming majority of the library systems used by the member libraries are MARC-based or are at least able to export and

import MARC records (in the case of MOKKA, either HUNMARC[3] or USMARC). In some cases, MOKKA supported the development of in-house tools to facilitate MARC export and import. As a result of these developments, all library systems are able to upload (see the reasons below) and download either HUNMARC or USMARC bibliographic and authority records.

There are also a lot of other conventions (resulting in the use of many conversion programs), because of the diversity of practices and rules among member libraries. They include the following:

- An USMARC and HUNMARC conversion program was needed to convert the record of the uploading library to the internal format of the system, and conversely, conversion programs were needed for downloading in order to convert from the internal format to the MARC-format used by the downloading library;

- The member libraries use various coded character sets, and therefore a conversion of the input to ANSEL (used as the character set of the central MOKKA database), and a conversion of the output to the character set of the downloading library was needed;

- MOKKA (the central database) uses a standard record-linking technique for volumes, and for the whole document in the case of multi-volume documents. Some member libraries use repeatable fields for the volume data, and therefore conversion programs were needed for uploading and downloading record(s) of multi-volume documents, if the library did not use the standard record-linking techniques.

Experience has shown that the above-mentioned problems can be solved, although not easily. Without these solutions, however, consistency cannot be ensured, because it was and is impossible to force a retrospective change of the systems and standards used by the member libraries.

---

[3] HUNMARC, the Hungarian standard exchange format, is USMARC-based, but—mainly because of specific features of the Hungarian language, such as the form of names of persons, but also for other reasons—it deviates from USMARC. Its newest version also takes into consideration the developments included in MARC21. The MOKKA system allows conversion from and to HUNMARC and USMARC.

3   Uploading vs. Cataloging in the Central Database

The 'classic' method of shared cataloging applies the following model. When a new document arrives at the cataloging department of the library,

- A search is executed in the database of the shared (union) catalog;
- If the search result is positive, the relevant record and the corresponding authority record(s) are downloaded to the catalog of the cataloging library and completed by local data;
- The name (code) of the downloading library is marked in the union catalog;
- If the search result is negative, the library executing the search catalogs the item in the central database and downloads the record that has been prepared.

The regular MOKKA procedure deviates from this well known practice. Of course, the process in MOKKA also starts with a search of the central database of the union catalog for the item to be cataloged. If the bibliographic record of the item is available in the central database, the cataloging library downloads the record, edits it by adding the contents of fields of local significance (e.g. subject headings, indices of classifications, notes, uniform titles, if not present in the downloaded record, etc.) and uploads the record 'back' into the central database. The uploaded record will be eliminated by a duplication check mechanism except for the identification data of the record-supplying library and its record identifier, as well as for the contents of some fields/subfields (e.g. subject headings, classification indices, uniform titles, country code, notes etc.). These will be added to the records, if different from the content of the given field present in the existing 'central' record.

   If the record of the item to be cataloged is not present in the central database of the system, the member library does not catalog in the central database of the union catalog. As mentioned already, the member libraries use different automated library systems. Therefore a number of special cataloging clients would be needed in the member libraries for cataloging in the database of the union catalog, and all catalogers in the member libraries would have to learn the rules of the cataloging modules of both their home library systems and of that of the central system. To avoid the

additional costs and the additional workload on their catalogers, the member libraries of MOKKA decided that their catalogers should not catalog the new item in the central database of the union catalog, but catalog it 'at home,' and upload their records (of course in MARC format) by placing the item in a file designated for this purpose, the content of which is regularly checked by a program for new records in the file. This program uploads the new records to the central database. The duplication check mechanism provides for the elimination of duplicates (although there is usually no duplication, because the cataloger is obliged to check before cataloging whether the given record is not already available in the central database) and if the uploaded document was a duplicate, then only the record identifier, the identification mark and name of the uploading library (and the new contents of some fields mentioned above) are added to the record existing in the central database.

4   Authority Control

The existence and variety of authority files vary from member library to member library. The central database of MOKKA includes authority files for names of persons and names of corporate bodies (including the names of conferences, other meetings, fairs, etc.). In addition, there are formal authority files for titles and subject headings (but not for standardization of the data, only to facilitate their global change if necessary).
For the 'real' authority files, MOKKA uses the following procedure:

1. Libraries that maintain authority files (a minority of cases) have been asked to upload these files prior to uploading the related bibliographic record;
2. The uploaded authority records are placed in the given authority file of MOKKA, following their duplication check and then linked with the relevant bibliographic record;
3. If the given authority data in the uploaded bibliographic record 'find' their authority record, they are linked with each other;

4. If the authority data of uploaded bibliographic record are not present in the authority files, a so-called 'skeleton' authority record is prepared and linked to the given bibliographic data of the bibliographic record.

5. The internal and external staff of MOKKA edits the 'skeleton' authority records and the links.

This process is just at the starting point. However, it is considered by MOKKA to be one of the most important tasks to improve the results of searches and to standardize access points for retrieval in MOKKA and, through MOKKA, in the member libraries. Because of many inconsistencies in the catalogs of member libraries, this is a huge, but nevertheless important, task. To help and accelerate this process, MOKKA acquired the Library of Congress Name authorities file and is eagerly awaiting the preparation of the authority files of the Széchényi National Library, to be based on the existing index files and their cross references.

5   Subject Approach

Views concerning the role of union catalogs for subject searches are highly variable. One extreme opinion considers the union catalog merely as a tool for finding document data, about the existence (and subject) of which customers are clearly informed. This means that their aim is only to find the library that is able to deliver the given item. The background of this view is that the real tools for subject searches are not the library catalogs at all, but subject bibliographies, citations, etc., and so the task of a union catalog is only the delivery of the document, although the existing retrieval access points (e.g. title keywords or subject headings and classification indices) can obviously be used by the customer.

An other argument against attributing great importance to the subject approach in shared cataloging systems or union catalogs is that in most cases (at least in Hungary and many other countries similar by size and by tradition), the various different subject heading 'systems' (if they really are systems, and not merely natural language keywords used as subject headings) prevent the establishment of a consistent, common subject heading vocabulary. The same can also be true for classifications, although some classification systems, like Universal Decimal Classification (UDC),

are widely used or even standardized in many countries of Central and Eastern Europe and are used by most potential members or record suppliers of a shared cataloging/union catalog system. However, there are libraries that do not use classification schemes or use different ones, and even if they use the same system they are often use different updated versions of the system.

On the level of MARC fields, values belonging to different types of vocabularies or schemes for designating subjects can be represented and specified by indicators and/or by subfields. However, retrieval is only possible via the relevant indexes, and in the CORVINA system and in many other systems, there is only one common index for subject headings and keywords and one other for all classifications used. This means that from the point of view of subject search techniques, MOKKA cannot offer solutions for the use of individual subject indication languages.

In spite of all the weaknesses of carrying out subject searches in MOKKA, it is nevertheless possible to use it for that purpose. There are plans for improving the existing procedures, among others by the use of an all-subject thesaurus as a kind of authority file, which can offer a link from various terms to others and can be used for the retrieval of a given subject.

As mentioned above, classification indices and subject headings recorded in the relevant fields and indices of MOKKA are not only those supplied by the library which uploaded the given record, but also include classification indices and subject headings in uploaded duplicate records. As a result, information different from that recorded earlier is added to the relevant fields of existing records, and through this the recall ratio (and of course also the noise) may be increased.

There are also other approaches to subject designation in shared catalog databases. For details, see work by Klára Koltay.[4]

---

[4] Klára Koltay, "Az ODR adatbázis új szolgáltatásai" (New services from the database of the National Document Delivery System), *Tudományos és Műszaki Tájékoztats*. (48) 2001, No. 8: 315–132.

6   Financial Considerations

It has been clear from the very beginning that MOKKA could not be a self-supporting system. Some financial principles have already been agreed upon, whereas other principles and rules will be fixed only after MOKKA has functioned normally for some time (probably at the end of 2002). It has also been finally decided that the record-supplying member libraries will not get any payment for their records. On the other hand, member libraries can download records free of charge. It has not yet been decided under what conditions other libraries can download records. There are two contradictory views about these conditions. According to one view, the system was developed and is maintained first and foremost from national and international resources, and it is accordingly not justified to demand payment for the downloading of its records. Those who support the idea of payment for supplying records refer to the cataloging expenses spent by the record-originating library.

7   Development Trends

The most important development tasks are described below:

1. Errors and mistakes detected by internal examination of the system, by the staff of MOKKA, and, last but not least, by the end-users of the system, should be eliminated;

2. The editing of existing (real and 'skeleton') authority records should be started, and this should become a regular maintenance task;

3. Plans for the expansion of the system should move in the following directions:

   - Libraries now outside MOKKA but having a special importance for inter-library loans (public libraries of the counties, further academic libraries and some research libraries) should be invited to join MOKKA as member libraries;

   - Links and direct access to electronic union catalogs for kinds of documents not included in MOKKA (primarily, but not exclusively, serials) should be established, and also the establishment of

    interconnected union catalogs for specific types of documents should
    be encouraged;

- An electronic inter-library loan system should be created within MOKKA, enabling the users to send inter-library loan requests immediately after the identification of the library where the requested document is located and available (MOKKA already offers a link via the Web to the electronic catalogs, holdings data and circulation modules of the library systems used in the member libraries). This enables the user to find the holdings data and the circulation status of the document to be requested); and

- Links to existing virtual union catalogs should be created.

8   Virtual or Physical Union Catalogs

The idea of creating virtual union catalogs emerged more than ten years ago. At the very beginning, this was only possible for libraries using the same electronic library system. With the advent of the Z39.50 standard, this possibility became, in principle, a reality for any group of libraries. Nowadays, Z39.50 gateways and other—usually Z 39.50-related—software solutions (METALIB, LibriVision etc.) offer further possibilities for searching in the databases of many libraries by using a single user interface.

Obviously, these technical solutions offer possibilities for establishing virtual union catalogs. However, one could also speak of virtual shared cataloging systems if the system not only searches, but also downloads and—in the case of libraries using different library systems—enables the conversion of records.

The question that emerges from the above technical possibilities is whether, and to what extent, virtual union catalogs can replace the physical (real) ones. It seems that it is easy to answer this question if we reduce the function of a union catalog to executing simultaneous searches in catalogs of various libraries. While it is worthwhile to discuss this question, it must be made clear at the outset that virtual solutions have a lot to offer in comparison with a situation without union catalogs.

Before investigating this question, we quote a paragraph from the executive summary of the Feasibility Study for a National Union Catalog (in the United Kingdom):[5]

> Moving from vendor systems to a comparison of physical and virtual catalogs, it was evident in all cases that the physical catalog architecture offered a more reliable, faster and consistent response than any of the virtual systems tested. Comparison of identical searches confirmed the supremacy of the physical model at present, particularly in relation to the user requirements identified in both the conceptual model and the questionnaire survey: for all possible search points the physical catalog showed superior consistency and performance every time …

It would be easy to close the discussion about real (physical) versus virtual catalogs by referring to the experience gathered by the authors of the above-mentioned Feasibility Study via questionnaires and experiments. However, one could object that the cited opinion is based on a situation in which well-developed physical union catalogs were compared with less developed virtual catalogs. It seems that a further analysis of the possibilities offered by the two solutions is justified.

Let us start with the most important question. A physical union catalog can exist only if it applies a high degree of standardization. One document is represented in the physical union catalog by one single record (in the case of MOKKA there exists a ranking of libraries based on the quality of their catalogs, and if the duplication check finds a duplicate, the record of the higher-ranked library is always kept). The catalog data of other libraries are represented only by the identification data of these libraries. In the case of a virtual union catalog, many slightly or substantially different catalog records of the same document are the result of the search.This means that the physical union catalog offers the same information as the virtual one, but in a uniform way, whereas the use of data available through multiple hits in virtual catalogs can impair the quality and compatibility of catalogs.

---

[5] Peter Stubley, Rob Bull and Tony Kidd, *Feasibility Study for a National Union Catalog. Executive Summary*. April, 2001. http://www.uknuc.shef.ac.uk/NUCrep.pdf: 6.

Another important milestone of standardization is the existence of authority files. Their establishment and maintenance is completely impossible in the case of virtual union catalogs. Of course, this is not an easy task in physical union catalogs either, but it can be managed. Without the existence and use of authority control, search results can also have a high noise ratio, but it is even more important that information loss can be very high.

Nevertheless, virtual catalogs can offer the possibility for simultaneous searches in the catalogs of many libraries, in spite of all the problems mentioned above. It is also possible to organize an electronic system for inter-library loans from all libraries, the catalog data of which are available through the virtual union catalog. There is also the possibility to copy retrieved catalog records for cataloging purposes. Taking into account all these possibilities, it cannot be denied that virtual union catalogs can fulfill the functions of shared cataloging as effectively as those of union catalogs. It is also possible to use the solutions offered by the software tools for virtual union catalogs to build the links between various physical union catalogs and/or between physical and virtual union catalogs. It is also obvious that a physical union catalog requires much more effort, manpower, and financial resources, and that a virtual national union catalog or a virtual catalog of any group of libraries offers much more than the searches in scattered electronic catalogs can. However, it should be stated unambiguously that the price to be 'paid' because of the lower quality of virtual union catalogs is too high for a 'core' national union catalog.