Study on the feasibility of multilingual subject cataloging

at the Swiss National Library

Working report 1

First thoughts on issue (i): Suggestions for the organization of subject cataloging as the automated cataloging system is introduced.

Prof. Dr. Rainer Kuhlen - Universität Konstanz
Prof. Dr. Dagobert Soergel - University of Maryland

Postfach 5560, D-7750 Konstanz   4.6.1993

The new organisation of subject cataloging should meet the following goals:

1. Processing of incoming materials without undue delay in spite of the shortage of personnel (which seems unlikely to change).

2. Improved searchability for the users, exploiting the possibilities offered by automation.

3. Integration of all catalogs now maintained in the library, including catalogs dealing with special materials.

4. Access to the exsisting collection (called old collection in the following) so that the user finds both old and new materials in one search (long-range goal). At the very least, the user searching the new system should be referred to the appropriate places (keywords, class numbers) where to look in the old system. The access to the old collection is of particular importance in the Swiss National Library, since materials in area studies (Landeskunde) and history do not become outdated.

5. Linkage to other relevant collections, especially to the Parliament Library Bern, to historical libraries, to Swiss national and provincial archives, and to university libraries. This is a long-range goal, but it should be considered in the overall planning.

6. Production of printed information services, such as Schweizer Buch and Schweizer Buecherverzeichnis.

Goal 1. Processing without delay

Goal 1 requires the following measures.

1.1   Avoid duplication of work through cooperation

Swiss books in the areas of science and technology, medicine, social sciences, etc. are being cataloged by specialists in university libraries; these cataloging data can be shared. The personnel of the Swiss National Library can then concentrate on cataloging materials in its areas of specilization, area studies and history, and share the results with other libraries.

Starting such a cooperation poses both organizational and technical problems. The technical problems fall into three areas: software, cataloging rules, and thesaurus and authority control in general (e.g., for corporate names).

!   *Software*. All participating libraries must be able to read and to produce records in the MARC format (or another common data exchange format).

!   *Cataloging rules*. All participating libraries must use the same or compatible cataloging rules and must use an agreed-upon mimimum depth of cataloging.

!   *Thesaurus and authority control in general*. All participating libraries must use the same subject headings (including place names and ethnographic names), personal names, corporate names, etc., or at least designations that can be mapped to each other automatically. This can be achieved through all participants using the same authority files or at least authority files that are harmonized. The problems of a common subject thesaurus or harmonized subject thesauri are treated separately below.

Implementation of such cooperation requires a substantial investment in the beginning, but in the long run it is cost-effective for all participants and increases user friendliness through standardization.

1.2 Efficient cataloging

The new system will make cataloging more efficient even without special measures. However, one should check whether VTLS has the following labor-saving functions or whether they could be easily added.

!   Identification of words (especially place names and personal names and perhaps at first limited to these) in the title (in all four languages), look-up in the authority file, and automatic insertion of the controlled form in the appropriate data field. This is preprocessing for the cataloger, who must check the result and, if needed, correct it.

!   Application of linguistic methods in conjunction with the thesaurus for automatically assisted catloging (see 2.2).

!   *Copy* and *paste* between any data fields.

!   Input of subject descriptors/subject classes either as notation or as text (descriptor or synonym, in any of the four languages).

Furthermore, efficient cataloging requires the development of a unified system integrating shelf classification and subject headings so that subject cataloging need not be done twice.

Goal 2 Improved subject access

2.1   Free-text searching

VTLS allows for free-text searching (for example in the titles and subject headings) and thus provides the user with an additional option. However, the user is confronted with the difficulties of an uncontrolled vocabulary. It is unfortunate that VTLS does not allow the **user** to specify the data fields to be searched; only the library can specify the data fields, and the users must live with that specification. One should check with VTLS whether the system can be modified as follows: The libraries specification of data fields to be searched in free-text searching is used as the default which the user can modify in the specific search.

2.2   Enhancement of free-text searching

Free-text searching should be enhanced through simple linguistic methods, such as stemming and decomposition of composite words (very important in German), and through an option for query expansion using synonym relations (across all four languages) and hierarchical relations in the thesaurus.

2.3   Use more subject descriptors per document (exhaustivity of indexing)

With an automated system higher exhaustivity of indexing is feasible, improving subject access. While on the one hand assigning more subject descriptors requires more effort, on the other it saves effort since the catologer is freed from deciding between descriptor candidates, for example, which three of six obviously relevant descriptors should be assigned.

2.4   Having a unified system for shelf classification and subject headings will also make the catalog more user-friendly

2.5   Subject search in the authroty file(s) for personal names and corporate names (long-term goal)

VTLS should enable the user to search authority files. Assigning subject descriptors to persons and corporate bodies provides the user with an additional access path to documents.

Goal 3. Integration of all catalogs

From a software standpoint, this goal can be reached easily with VTLS, since VTLS supports all MARC formats. However, this goal requires detailed data modeling which accounts for the different structures of various specialized catalogs that have developed historically. The result should be a data model that covers the entire library and all its functions. Integrating the catalogs is essential for a user-friendly library.

Goal 4. Access to the existing materials (the old collection)

The ideal is, of course, the complete integration of the old collection into the new system. However, for reasons of cost this can only be a very-long-range goal. Therefore, interim solutions should be implemented.

4.1   Thesaurus references to old classification notations

While searching the thesaurus, the user should be referred to the old classification numbers, which he can then use to search the old system. Therefore, the old classification system must be

used as a source in constructing a new thesaurus (or modifying an existing thesaurus). If it is possible in VTLS, the user should be given a reference to the old classification number whenever he searches under a given descriptor in the new system.

4.2   Selective incorporation of the old collection into the new system over time

Books that are checked out can be cataloged for the new system as part of the process, provided there is personnel capacity at the time. The bibliographic data will often be available in machine-readable form elsewhere. Subject cataloging should be based on the old subject headings which are input and mapped automatically to the corresponding new descriptors. The book is not reshelved to its old location but placed in its new location. If personnel capacity is available, one could convert old materials into the new system, using a priority order of subject sections, beginning with materials in Swiss area studies and history and biographies of important Swiss personalities. For space reasons it may be necessary to separate the change in cataloging from the physical movement of the book in the stacks. However, from the point of view of the user it would be much clearer if these two processes were always simultaneous. The problem of a piecemeal movement of books from their old location to their new one is this: Both the old location and the new location must have enough space for these books since the space freed up in the old location cannot be actually used until after a general reorganization.

Goal 5. Linkage to other relevant collections

This goal can be achieved by establishing a record for each such location, assigning descriptors about which there is material in the collection, and including these records as pseudodocuments in the catalog.

Goal 6. Production of printed information services

Classifications and subject headings needed for the printed services should be assigned in the process of subject cataloging or should be derivable automatically from the assigned descriptors. For example, a specific descriptor assigned in subject cataloging may be mapped to a more general descriptor used in a printed service.

Use of a thesaurus

The problems of developing/modifying and using a thesaurus are closely linked with subject cataloging. They will be discussed in a separate report. A few remarks follow.

!   The ideal would be a Swiss Thesaurus which would be developed and maintained cooperatively and used by all participants in the envisaged cooperation. This would distribute the significant effort of thesaurus building/modification and maintenance.

!   The following variant may be a bit more realistic. The thesauri, subject heading lists, or classifications now used by the participating libraries would be combined into a master thesaurus data base which would harmonize the individual systems. Maintaining this master thesaurus would be a collaborative effort in which each participant would participate according to subject competence. The Swiss National Library would deal primarily with the subjects area studies, history, and perhaps Swiss law. Each participant would transfer changes in subject areas of interest from the master thesaurus into their own system.

!   Owing to the Swiss situation and requirements the thesaurus must be multilingual. The hierarchy of concepts should be designed independently from any language. The goal should be to have a descriptor text for each concept in each language. In addition there will be synonyms in each language.

!   The following are possible sources for the development of such a thesaurus:

        a.   TERMDAT (all translations needed should be available there)

        b.   The thesaurus of the Parliament Library Bern

        c.   The existing classification of the Swiss National Library

        d.   Thesauri, subject heading lists, and classifications used by the cooperating libraries.

!   Thesaurus development must be supported by a program that makes the best possible use of the classificatory and terminological data available in these sources and thus frees the lexicographer from a lot of work.

Thesaurus problems will be addressed in a detailed comparative and systematic report in fulfilment of tasks (ii) and (iii).

Attachment

Report no. 2 by Sabine Langohr: Description of the present state of subject cataloging at the Swiss National Library.