

Subject indexing in SBN

Maria Lucia Di Geso
ICCU

Introduction

Over the course of the last ten years, the Servizio Bibliotecario Nazionale (SBN) (National Library Service) has been taking shape as the largest Italian network of services providing information, holdings details and document supply, based on cooperation and the sharing of bibliographical resources among thousands of libraries throughout the country. From the very beginning these were the objectives held in mind by Angela Vinay, first director of the Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU) (Central Institute for the Union Catalogue of Italian Libraries). In the process leading to the realization of such objectives – which we still cannot consider fully achieved – there was a long period during which SBN identified almost exclusively, at least at Index level, with the problems of author cataloguing and description.

For the development of services and of cooperation, the creation of the Index database and its enlargement through shared cataloguing were actually prerequisites – and of necessity the focus of attention during the first phase. Over time, to the objective of bringing the network into being, was added that of offering to the library world and its users a database of significance in terms of both quantity and quality, anticipating the development in the Index of data monitoring and “cleansing”, elimination of duplication and, relative to the author archive, the creation and management of an *authority file*.

For years then, discussion at least at Index level has been limited to questions of author cataloguing, subject cataloguing being excluded from cooperative activity. In a phase in which the problem of creating a consistent database of quality and without excessive duplication was paramount, it seemed premature to tackle the subject cataloguing issue – especially considering the difficulties of pursuing in the field of subject work, that uniformity and standardization that one can achieve reasonably, albeit with difficulty, in author cataloguing.

The SBN Index

Shared cataloguing in the Index did not consider functionality in relation to subject aspects, nor did it foresee the logic of sharing of information relative to the conceptual content of documents: it was decided not to tie the libraries which had joined the network down to the adoption of shared tools, initially because of the specificity and differing nature of the institutions involved. At the stage when the Index was being specified, therefore, the only functions defined as “protected” were those relating to the procedures for author and title cataloguing (bibliographic management) and for lending, while for subjects wide-ranging autonomy was left to the various installations of Polo software. This approach has made the practices of subject cataloguing and classification of documents appear somewhat aberrant in the Polo software with regard to both data and functions, above all in the part that relates to construction of the archive of terms and to the relationships between the terms themselves.

As a result of such early assumptions, when the Network System–SBN Index became a reality with “migration” activities from the various *Poli* at the beginning of the nineties, the choice was made to

copy to the Index only data relative to descriptive cataloguing (authors and associated bibliographic descriptions), excluding information of semantic type (subject headings and classification numbers).

The working group on subjects and classes in SBN

The absence of semantic data from the Index and the lack of common directives within SBN did not, however, prevent a process of growing consistency from being realized, even in the area of subject indexing, both through the increasing spread of the *Soggettario delle biblioteche italiane* (*List of subject headings in Italian libraries*), and through initiatives aimed at creating opportunities for librarians working in the subject headings/classification area within *Poli* to compare notes.

Thus was born, in February 1990, the “Working Group on Subject headings and classification”. This was established by the Institute on the recommendation of the “Joint Committee of SBN Experts”, with the brief of comparing and establishing the effectiveness of the software functions offered by the various installations present, with regard to the management of subjects, as well as the appropriateness of services offered in relation to user expectations. The working group was set up bearing in mind that its composition should be representative of all software applications and of the various types of library involved.

In accordance with the guidelines suggested by the Committee, the Group first prepared a questionnaire aimed at collecting information on the use of the various indexing tools among SBN *Poli*, which showed that 80% of libraries used the *Soggettario* of Florence as their main indexing tool.

Subsequently, a study was undertaken of the functions of Subject heading management and Classes within software then existing – BULL, UNISYS, IBM/ADABAS and IBM/SQL – a study which also made use of comparative analysis of data relating to the Subject archive file present in the four prototypes. It was agreed to attempt to define an “SBN Standard” in terms of functions and data for subject heading assignment and classification. On the one hand, this would be the “sum” of the best realizations present in the latest releases of the various items of software; and, on the other, it would take account of procedures and methods recommended for the control of terminology and for defining the types of relationship between indexing terms recommended by international standards such as UNI/ISO Standard 2788.

On the basis of comparison and past experience, the Subject Working Group recommended that at Polo level there should be employed a single indexing reference tool, without ruling out the possibility of maintaining other tools of a more specialized nature on the side. The suggestion for subject specialization among the various *Poli*, whilst undoubtedly fostering rationalisation and standardization in the practice of semantic cataloguing in SBN, was however held to be premature at that stage, which coincided with the launch of the network. Actually it would also have had to provide for the creation of lists under centralized control and a complex organization of authority control for the maintenance and coherence of the whole process. The Group therefore came down in favour of a more modest plan, one that envisaged on the one hand the possibility of off-line loading of subjects and classes into the Index, solely by the Biblioteca Nazionale Centrale di Firenze (National Central Library of Florence) (BNCF), with periodic updates – and, on the other hand, the possibility of displaying in a supplementary file subject headings/class numbers linked to BNCF records and the possibility of storing data to be used subsequently at a local level, with the creation of the relevant link to the description of the document.

Such a solution, even though only partial, was meant as a response to continued requests to use the Index functionality to capture, along with individual bibliographical records, the matching semantic data, with considerable savings in time and costs for the libraries which intended to use the subject headings assigned by the National Library of Florence; furthermore, the presence of such information in the Index would have offered cataloguers a form of “cross-check” and a guide to the formulation of subject heading terms.

The development of the OPAC and the creation of the archive file of subjects and classes in the SBN Index

The Subject Working Group's idea could only be put into effect several years later, in 1996/97, thanks to the project "Accessibility of databases residing in the SBN Index", finalized at the creation of the Index OPAC. The realization of the latter represents an important stage in the evolution of SBN towards the development of the network of services, which is the principal objective of the whole project. The decision to enrich the information residing in the Index with subject data was taken by the SBN Management Committee at the suggestion of the ICCU: in relation to the *Libro moderno* database it was planned to develop in the OPAC functions which would permit subject searching of documents through interrogation by subject and/or class. To this was added the goal of offering subject cataloguing support for librarian users, by copying this information into the management database of the Index. In the scope of this project two phases were identified:

- ❑ firstly, formation of the Index database of subject headings and classes and its enlargement through batch loading of subjects/classes from 11 *Poli*;
- ❑ a subsequent phase, in which the Index would be enlarged with the assistance of online procedures to be developed by those same *Poli* that were identified in the first phase.

To guarantee consistency and coherence to the database as it was taking shape, it was resolved to input only subject entries created by *Poli* that used the *Soggettario* of Italian libraries; once the BNCF Polo had been given priority, the choice of the other 10 *Poli* was based on two principles: quality of data and level of 'coverage' in the Index, interpreted as a percentage of records possessed by one *Polo*, compared to the whole national database.

This first phase ended in February 1997 with the copying to the Index of approximately one million subject entries covering about 860,000 titles; subsequently the data was loaded into the SBN OPAC.

The second phase concerned the development of Polo operations for both the supply and the capture of Index semantic data; after the BULL software, gradually other software has acquired such functionality, such as for instance the IBM/ADABAS and UNIX applications in client-server architecture.

The structure of data and the functions provided for in the Index presented, and still present, characteristics that differ from those in Polo: the subject is registered as a link to the bibliographic record, like an 'appendix' to it; this causes a duplication of identical subject strings, when these are sent to the Index by different *Poli* and thus with different identification (CID); from an operational viewpoint, the choice has been not to manage the subject headings archive in accordance with shared cataloguing principles and therefore:

1. there is no subject query function: a subject heading is seen only at the stage of searching/capture, from the entry to which it is linked;
2. when a record is captured, *Poli* are free to choose whether or not to capture a subject heading that is linked to it;
3. *Poli* supplying the Index with their subjects have available to them software developed to allow them to upload data online, in batches at the end of the day's work; or they may upload directly online, which also permits choice as to which subjects to send;
4. it is not possible to link subjects or class numbers coming from more than one Polo to the same record: an entry always represents subject headings and class number from a single *Polo*;

5. in the case of ‘competition’ between data from more than one *Polo*, the Index selects **one** in accordance with a “priority” scale for subject cataloguing among *Poli*, with the BNCF in first place;
6. there are no updates to *Poli* for corrections or for substitution of a subject string; therefore it can happen that a *Polo* that has captured a subject heading subsequently modified in the Index may remain out of step.

Present situation

The last total load of subjects and class numbers to the Index was performed in October 2001; probably that will be the last batch upload, since now almost all the SBN software is equipped with procedures that permit online uploading.

At present the Index contains approximately 5,585,000 document descriptions,¹ of which more than 5,124,000 represent monographs, 3,150,000 of them published in Italy.

Of the total entries present, 1,280,000 have subject headings – some of them more than one, which explains the presence of 1,610,000 subject links to entries.

In absolute terms, the percentage of entries with subject headings in the Index is less than 23%; if, however, we bear in mind that many publications do not receive subject headings because they have no subject (for example, fiction, the classics, general encyclopaedias, etc.) and that many entries coming from retrospective conversions will probably never have subject headings applied, we can perhaps reduce the Index entries which could potentially have subject headings to a figure between 4,500,000 and 5,000,000; in which case the percentage of records provided with subjects would be somewhere between 28.4% and 25.6% of the archive file.

On the other hand, studies show that, as the software applications of *Polo* acquired functions which would enable them to capture the Index’s subject data, there was a gradual increase in the activity of subject capture: at present, with about 6,500 records captured daily, between 2,700 and 3,000 transactions are carried out that are subject searches for capture and from about one thousand subject captures a day in the early years the present figures show an increase of 270–300%.

From the analysis, even if summary, of the data quoted some considerations emerge: from the first load of subjects into the Index, which occurred at the beginning of 1997, six years have passed; then publications present amounted to about 2,800,000 and now, as already mentioned, they are 5,585,000; the *Poli* which then were 33 in number, at present are 51: so we have an Index which has almost doubled in that space of time. It therefore seems obvious that the quantity of records with subjects is slender with regard to an archive file of such size, in that about one quarter of the database has subjects; on the other hand, it can be seen that cooperation has certainly taken advantage of these new functions, in order to facilitate and speed up the cataloguing activity of the *Poli*, which have displayed a great interest in capturing subjects. Today, as regards current material, about 50% of entries captured are captured with a subject and surely such a percentage would rise, if there were a greater number of subjects in the Index.

If the intention is for the subject archive file to have a broader coverage and eventually to encourage cooperation in that sector, it is necessary to review in part the initial choices, if possible extending the function semantic cataloguing in the Index to other *Poli*; the selection criteria could be the same as applied in the case of the previous eleven: use of the *Soggettario* and a well-established and accurate practice in the indexing of documents; the new *Poli* could be identified on the basis of their contribution to the subject analysis of documents that are not present in BNI, for example, foreign material, the highly specialized or local material; with reference to the latter type of material it must be remembered that some publications in the Index will never have a subject entry, unless it is done by the one *Polo* that owns it.

¹ (The following data refer to a report dated 15 December 2002)

Furthermore, above all among *Poli* that display a greater consistency in their subject work, for instance those of the two National Central Libraries, not to mention others, one could consider dividing up the publications requiring subject work by type of material or on the basis of other criteria which can emerge only from a comparison between the libraries involved.

The SBN Index Evolution program

The Index Evolution program, on which the ICCU has been engaged for about two years, arose from the need to make SBN evolve into a new system, which would become a broad point of reference for the supply of widespread services.

With this in mind, the program offers an enrichment of services, integrations and extensions of the database and the opening-up to new types of usage with varying profiles.

Besides guaranteeing – for systems currently in existence – maintenance of the rules already in force in the present SBN, the future system will allow:

- ❑ an extension of the services provided by the central catalogue
- ❑ an extension of SBN cooperation to other types of material
- ❑ a simplification of interactions between central system and polo in the new shared cataloguing (including through the introduction of mechanisms for derivative cataloguing.
- ❑ receptivity towards other LMS by means of international exchange formats (*import–export of bibliographic descriptions*)

Within the scope of the program, new functionalities and new data relative to the area of subject cataloguing have also been envisaged.

In the new database structure data there will be implemented an archive file for subjects and one for classes, which will sit side by side with the other two, those for titles and authors: the subject will no longer be an “appendix” to the bibliographic entry – a record visible only from the analytical display of a title, as it the case at present, as is now the case – but an autonomous entity, able to be interrogated separately.

Such a structure, besides the existing functions of creation/correction and capture of subjects and class numbers in the Index, which are possible with the present SBN protocol, will need to permit the following new functions to be available to the *Poli* to be connected via the new SBN MARC protocol:

1. Interrogation of the Index for semantic data, subjects and classification numbers, used both as channels of access directly to the files of subjects and classes, and as channels of access to the records;
2. “navigation” from the subject file to the title file and vice versa;
3. “monitoring” of the subject file, in order to detect possible duplicates, coding and keyboarding errors, etc.’
4. facilities for “cleansing” the archive file, functions of which will depend on the outcome and results obtained by monitoring activities and will be directed towards creating a consistent database;
5. statistics of activity and the production of the *Poli* relative to the subject area;
6. extension of subject cataloguing operations to new types of non–book material, which the Index is preparing to manage (e.g., prints, photographs, maps, etc.)

Feasibility study for the New *Soggettario* of Italian Libraries

Almost simultaneously with the “SBN Index Evolution” program, a project of the National Central Library of Florence was developed in relation to the Feasibility study on the renewal of the *Soggettario* for the catalogues of Italian libraries. The ICCU welcomed with considerable interest the invitation of the BNCF to participate in the working meetings of the relevant Study Group, and shared with it the outcomes. These were that the choice of arriving at the construction of a new, consistent documentary language, organic and constructed on the basis of a few clear, simple and easily applied rules, is the complete answer to the need for renewal expressed many times both in many parts of the library world and within the SBN; and that the three principles by which the new language would be guided – unity, predictability and specificity – have by now become part of the cultural baggage of those involved in indexing; we also saw as valid notions concerning the principle of the co-extension of the string, and the imposition of a citation order no longer founded on the heading/subdivision principle, but on the analysis of syntactic roles. Besides reference to the application of Standard UNI/ISO 5963 for the phase of conceptual analysis of documents and to Standard ISO 2788–1986 for aspects relative to vocabulary control, they are going in the direction of that activity of diffusion of international standards which the ICCU has been urging for years. The realization of such a project seems, then, to lead towards creation of an “authority archive file” in the field of semantic cataloguing, which would be a tool of excellent assistance to SBN *Poli*. It is necessary, in my opinion, however, to make one final effort to be able to understand whether it will be possible to achieve coordination between the two projects, individualizing the tools, methods and the levels of integration of the New *Soggettario* system within the framework of cooperation and of the SBN System. A great deal of work still awaits us, for the fulfilment of both projects; but we are optimistic that the two roads we have set out on can one day meet, so that Italy, too, can in future offer the community of users and librarians not only a large bibliographic database, but a complex, articulated system open and endowed with quality tools, such as for instance a subject authority file, as is already the case in other countries more advanced in this regard. The Central Union Catalogue Institute of Italian Libraries is working in this direction.