Maria Ines Lopes
Biblioteca Nacional, Lisbon, Portugal

Subject Authority Control in a Union Catalogue: Perspectives for a National Co-operative Project in Portugal*

Lopes, M.I.: Subject authority control in a union catalogue: perspectives for a national co-operative project in Portugal Knowl.Org. 20(1993)No.3, p.133-138, 7 refs.

In 1989, the existence of an online union catalogue, the National Bibliographic Database - PORBASE, raised the urgent need for mechanisms of co-operation in the field of subject authority control. A national co-operative project called CLIP (Compatibilizacao de Linguagens de Indexacao em Portugues) was launched. With CLIP special libraries will assume responsibility for the final content of the authority file with respect to terminology for subject access in different subject areas. Reference is made to the general situation in Portuguese libraries with respect to the availability and application of subject indexing tools, and also to the work being carried out towards standardization and harmonization of procedures in such matters.   (Author)

1. Introduction

The advantages of standardizing subject indexing vocabulary used by different libraries to provide access to bibliographic information pertaining to the same subject area are evident, and have always been recognized. From the point of view of developing a catalogue, it is a way of rationalizing the work of the subject indexers, at the same time allowing for the possibility of overall improvement in the quality of the catalogues involved in the project. From the point of view of the end user standardization could lead to better and easier use of the catalogues, particularly if they are intended for the same language community, and especially if they serve the population in a given geographic area.

Devising the means for achieving this ideal method for improving the quality and consistency of subject access is not a simple task. Efforts towards this goal exist at national and international levels, especially for those cases where a broad policy of bibliographical control in a given subject area has materialized through co-operative and internationally used databases. However, the use of a common language for subject indexing and retrieval remains a difficult option when different systems have already been established and applied. An alternative solution would be to develop some strategy for achieving compatibility. The development of such strategies in the field of subject indexing has been investigated, but this type of solution is very expensive and less feasible when talking about alphabetical subject indexing - at least from a pragmatic point of view. Both approaches, using the same indexing language or some method of compatibility among different indexing languges, become particularly difficult in the case of library catalogues.

2. Co-operation in Subject Indexing: Old Reasons, New Motivations

PORBASE[1] is the Online Union Catalogue of Portuguese Libraries, which gathers the present contributions from 75 libraries of all types, pertaining to many different subject areas, and serving different groups of users. Most of the libraries are university libraries. Others are special libraries including the libraries of public administration departments. There are also general public and private libraries of different sizes and types of collections. On the other hand, this same union catalogue, PORBASE, brings together all types of bibliographic materials, including current bibliography (Portuguese and foreign), manuscripts and rare books, maps, prints and photographs, other non-book materials, etc.

This fundamental source of bibliographic information is extremely important as it is unique in a country with scarce and scattered bibliographic resources[2]. Getting the most out of this resource requires much more than the use of the same format and the merging of records from different libraries in the same database. It requires the development of subject retrieval methods based on a strategy acceptable to all partners in the project. The main partner is, of course, Biblioteca Nacional (BN), not only because it assumes the role of the national bibliographic agency, but also because BN is the union catalogue manager, by government decree.

As is true in general for the whole community of Portuguese libraries, BN faces a lack of financial resources and has not enough staff, either in quantity or in quality, to properly carry out the task of managing the alphabetical index on the union catalogue. Nevertheless, the need for the control of that index has been recognized as an important issue from the beginning. This recognition allied with other existing needs, such as the improvement of the subject indexers' skills and their tools (which are non-existent in many libraries) and led to the launching of a national project called CLIP (Compatibilizacao de Linguagens de Indexacao em Portugues). This project is the subject of this paper.

Arising from real needs and dealing with the existing human resources, the Project must be viewed as an experience with strictly pragmatic goals, although with a basic theoretical background. However, it was not our objective to develop any kind of innovative methodology based on technological facilities especially devised for the system. BN has an OPAC and an authority file to control the access points, as well as people willing to work together. Thus, the special needs of the Project are to find ways of organizing the people and the workflow in the system, in order to provide an effective sharing of responsibilities in authority control.

3. Structuring and Developing the Project

By the end of 1988, some months after the public opening of PORBASE, it had been agreed, among BN and the co-operating libraries, that the question of subject access had to be worked out as a joint task. It was clear that the commonly adopted classification system, the Universal Decimal Classification (UDC) could not accomplish all the requirements for subject retrieval because of different levels of application which are difficult to control. Also, UDC is not a friendly search tool for the end user without an online version of its alphabetical index. Nevertheless, the usefulness of UDC has been clearly recognized for general subject approach to other databases and has proved so on numerous occasions.

Thus, attention turned to the alphabetical subject index to the database, which was just the sum of access points exactly as established by each of the contributing libraries for their local catalogues, according to different systems and tools (thesauri, subject heading lists, etc.). As a matter of fact, it was not, as it is not yet, a good index. It is neither reasonably reliable in terms of the control of terminology or in terms of the syntax of strings. Gradually the idea of organizing inter-institutional working groups with special librarians working together in the same subject area developed. By July 1989 there were several groups established and the statutes and terms of reference for the Project were approved. In those terms of reference two main purposes were established:

a. The standardization and improvement of subject heading languages used by different libraries in the same subject area, in order to achieve the consistency of subject access in PORBASE; and

b. The establishment of regular procedures for co-operation that, in addition to the primary goal could lead to common criteria concerning all aspects of subject indexing.

According to the statutes, any special library is a potential partner in the Project, even those that are not yet co-operating in PORBASE. They allow for other types of institutions and even individuals, provided they have a recognized interest of skills in the field of specialized terminology.

The organization of the Project has three levels: A project Manager, a Coordinating Board, composed of the leaders of each growing group, and the working groups themselves. The main functions of the Coordinating Board are the planning and general control of the Project, the study of, and decisions on, procedures to be followed by all the groups, the evaluation and harmonization of the results and the readjustment of responsibilities in order to provide the best subject coverage possible.

The working groups needed to make the most of the participation of those institutions and/or persons that have recognized expertise or responsibilities in each subject area, and provide for all activities leading to the control of the terminology in their subject field. This co-operative effort led to the integration into the Project of some already existing working groups in specific areas. Thus we were able to profit from their past experience and achievements. In a further stage, about to be implemented, the aforementioned activities will be extended to direct work on the subject authority file of PORBASE. The working groups will share the responsibility for its content.

Almost three years have passed since the launching of the Project and only now do we have the conditions to be in the application of results. The reason for this is that it is dependent on the implementation of the authority module of the system which suffered some unexpected delay. At present the Project depends on the participation of about fifty individuals, involving more than fifty institutions distributed over twelve working groups.

4. Preliminary Achievements

The past three years have been a time of motivation, organization and preparation. As a result some fundamental preliminary steps have been accomplished. At the level of the Coordinating Board, we have achieved consenses on rules and standard procedures for the work on terminology control, which has been applied in the recent 2nd edition of SIPORbase. Although in its first edition (1988), SIPORbase[3] was just a code of rules for the subject indexing system prepared for use within BN only, it was reviewed and partially updated in order to extend its use at the national level. This general use of standard rules and procedures will be extended to the rules concerning proper names and the semantic structure at a later stage.

At the same time that guidance was being provided for the activities of the working groups, the strategy for the development was being clarified and the methodology for the implementation of results was prepared within CLIP by the Coordinating Board for the Union Catalogue. Although the Project had received general approval from the beginning, it was not until 1990 that its terms of reference and the option for the sharing of responsibilities in the subject authority file were introduced officially into the statement of principles for PORBASE. In establishing the strategy for the development of the system, the control of terminology was defined as the first and major priority and the main objective of the Project. On the other hand it was recognized early that because of the lack of human resources, it would be unrealistic to carry out simultaneously the standardization of terminology and the construction of the

semantic structure (hierarchical and associative relationships).

This being so, in the first stage of implementation the working groups are not required to include the semantic relationships in the authority records, but they are permitted to do so if they already have a good source for the structure in their subject areas.

Up to the fall of 1992, the working groups had developed a survey of the existing sources for terminology control in each subject area. These sources include all types of reference works (dictionaries, glossaries, terminologies, indexes to the major technical publications, basic manuals, nomenclatures, etc.) and documentary languages (list of subject headings or thesaurus descriptors; also classification schemes and their alphabetical indexes, etc.). Based on this survey, each working group produced a bibliography including book locations, together with a statement of guidance in the use of each source for purposes of terminology control. These bibliographies resulted from a comparative evaluation of the items.

The bibliographies are intended to be published as a first source of help for anyone wishing to do subject indexing in particular subject areas. This initiative not only established a basis for the further activities of the working groups - even allowing the consensus on primary sources in some of them - but also provided a source for presentation to external organizations and individuals approaching the working groups as focal points for their subject areas.

In this respect it is intended that the Project be developed as a truly national level intermediary for any kind of activity related to subject indexing. As such the Project will promote standardization, provide for training and consultancy and establish the basis for the link with other national and international bodies active in the areas concerned. These activities are considered to be of the utmost importance for libraries in general, and not only for those co-operating in PORBASE.

At the same time, almost all of the working groups carried out the compilation of lists of subject headings or descriptors already in use in the catalogues of their members. Where there was no separate list available it was necessary to compile such a list in order to begin to harmonize the terms. This type of work is extremely important as the majority of special libraries do not yet have appropriate tools in the Portuguese language. Most of them use one or more foreign thesauri or lists making the translation of terms or headings according to current needs. They do not have staff or time enough to develop, *a priori* or *a posteriori*, a proper thesaurus or subject heading list. There are exceptions such a medicine or agriculture, but there are large areas, particularly in the humanities, lacking any type of such reliable tools, even in languages other than Portuguese.

To help in the compilation and harmonization of terminology, and also to provide an easy means for the transfer of information, IDEIA was an application for thesaurus construction and maintenance, developed within the pro-

ject. This application is based on the Mini-Micro CDS/ ISIS Software (Unesco) which is distributed by BN to any person or institution wishing to manage a thesaurus or subject heading list in the computer, even if they are not connected either to the Project or to the National Bibliographic DatabasePORBASE. Some of the subject headings or descriptor lists resulting from working group compilations are already available, either as provisional tools or as material. In both cases, they are accessible to any person interested.

Of course the stages of development and the levels of achievement are not the same for all of the working groups, because there are many factors affecting these activities. Some groups are just beginning their work, while others are only being established. Moreover at its present state of development, we do not pretend that the Project allows for the appropriate coverage of all subject areas. Nevertheless, the principles, the strategies and some experience are there, pointing the way towards common goals that, even for the most advanced working groups, must be viewed as long term objectives.
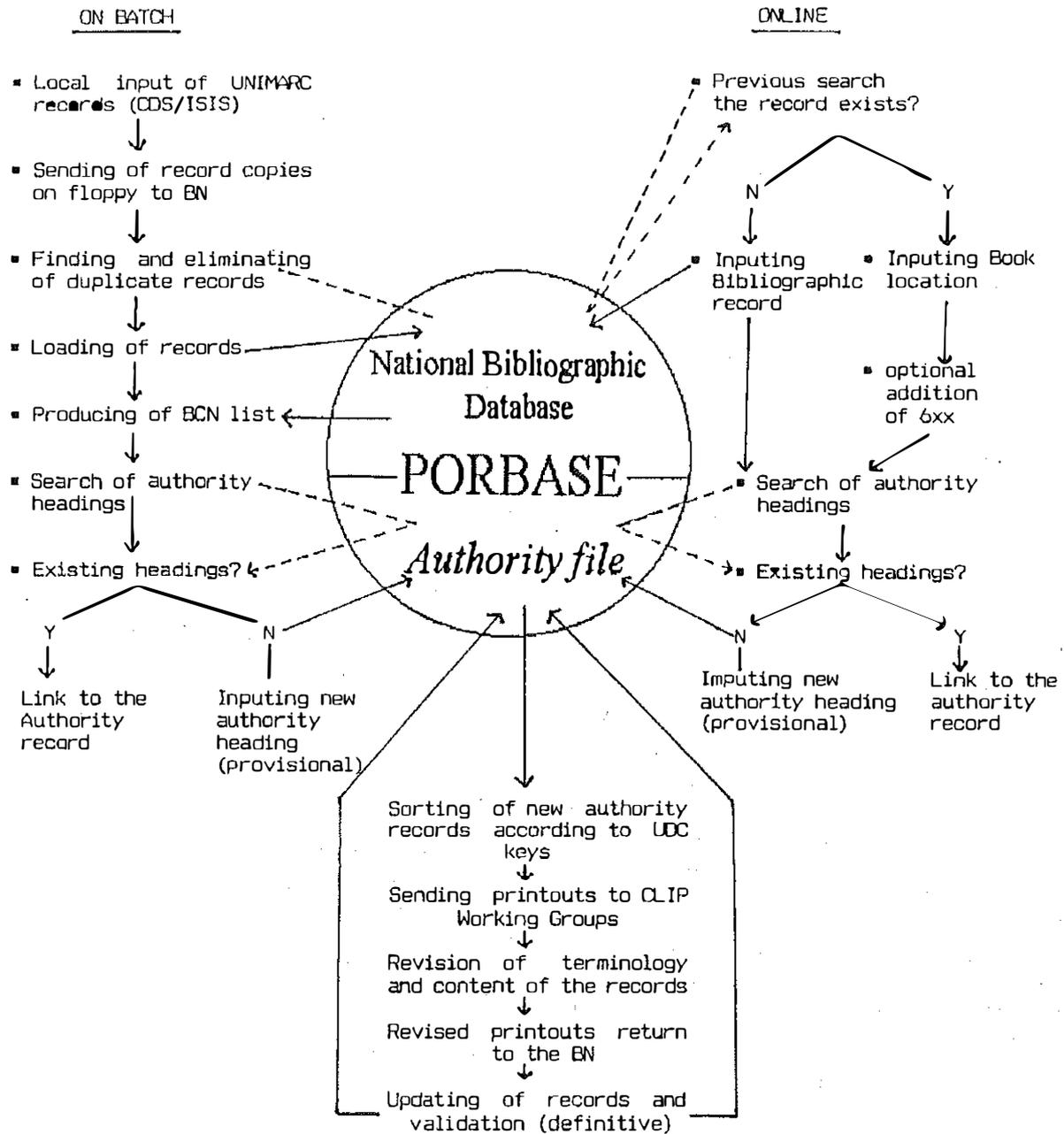
5. The Implementation of Results in PORBASE

The prospect for the practical application of results in PORBASE, assuming the availability of the means for activating the shared responsibility, is a strong motivating factor in the short term. Implementation will require all working groups to fulfill current tasks, no matter whether they have more or less achieved preliminary results or not. Naturally, it will be easier for the more advanced working groups, but *a priori* achievements are not a condition for becoming active in the subject authority routines of PORBASE, provided there are the right people willing to do the work *a posteriori*.

The routines to be implemented will cover all the terminological subject access points regardless of the source of the bibliographic records that are carrying them. Until the end of 1989, all subject access points were admitted into the database, generating an immoderate increase in the subject alphabetical index, whose size was revealed to having an inverse ratio to the desired quality of retrieval. Keeping in mind the emerging Project, a very difficult decision was made - to stop the growth of the alphabetical index by cutting the subject access points carried by records coming from libraries contributing in a batch mode. This was the majority of the contributors. We continued to accept complete bibliographic recorods from libraries co-operating online, because those libraries do not have their own databases. This measure was, of course, very controversial but we believe that it has contributed to clarification of the fact that PORBASE, besides being a co-operative database, is also a catalogue to be managed. On the other hand, this decision stressed the importance and urgency of the Project, inducing special librarians to participate and making the contributing libraries more aware that there must be basic agreement and discipline in the field of subject access.

PORBASE

COOPERATION PROCESS

AND AUTHORITY CONTROL

ON BATCH   ONLINE

- Local input of UNIMARC records (CDS/ISIS)
- Sending of record copies on floppy to BN
- Finding and eliminating of duplicate records
- Loading of records
- Producing of BCN list
- Search of authority headings
- Existing headings?

Y   N

Link to the Authority record   Inputing new authority heading (provisional)

- Previous search the record exists?

N   Y

- Inputing Bibliographic record   - Inputing Book location
- optional addition of 6xx
- Search of authority headings
- Existing headings?

N   Y

Imputing new authority heading (provisional)   Link to the authority record

National Bibliographic Database

PORBASE

Authority file

Sorting of new authority records according to UDC keys
↓
Sending printouts to CLIP Working Groups
↓
Revision of terminology and content of the records
↓
Revised printouts return to the BN
↓
Updating of records and validation (definitive)

The new routines for the control of subject headings will stop that restrictive procedure, but a similar decision has to be made regarding the existing subject index, where data introduced up until December 1989 remains. The new routine will start a new index. This probably means that the backlog will not be controlled, but an exception will be made for those libraries that are online. The reason for this is pragmatic. There will not be enough staff for the task, and even for future current work, so far we have no guarantee of being able to accomplish the work at a reasonable rate.

The methodology to be followed has three sequential stages, after the records are loaded into the database, no matter whether they are received online or in batch mode. (see Fig.1). The first stage is intended to ensure that all subject access points are controlled in the system through the online authority module. That is to say, there must be an authority record for each heading, and for each subject element of a pre-coordinated heading. An authority record must be created whenever a heading appears for the first time. The same routine will include the linking between authority records and the fields carrying the relevant headings in bibliographic records. When creating a new subject authority record, the UDC keys corresponding to the concepts/subjects represented by the headings will be established and input to the appropriate field in the record. This work will be done by a special team at the BN, concerned with the control of subject headings pertaining to records loaded in batch mode. All libraries working online must follow these same procedures during online bibliographic processing.

With this first state the authority control of terms input to the system is ensured but not the intellectual level. It is in the second state that the working groups of CLIP will be active. Periodically, they will receive printouts of the new subject authority records extracted from the database using the UDC keys that identify the subject areas of each group. Then their task will be to approve or modify the terminology of the heading, to supply the appropriate *see* references and optionally, *see also* references, or other authority data pertinent to the record. Those working groups that have the possibility of using the online bibliographic processing of the central system will be allowed to use that facility instead of working with printouts. In those cases they will receive a list of just the identification numbers of the records to be reviewed.

At the third stage, the reviewed printouts must be returned to the BN for the input of the revised content to the authority records. Revised records will then be validated at a higher level. That is, further updatings that do not come from the same source will not be permitted. Of course, there are fringe areas which may cause some overlap in responsibilities among groups. In such cases there will be further joint decisions required if there is no agreement at the first round.

This is an abridged overview of the methodology to be used, which may also include the input of subject authority

lists prepared in advance by some groups, whenever there is available staff for the work. This routine would contribute to faster growth of the authority file with great advantage to the whole process. One can imagine that this methodology may be slow. Probably it will be, but only practice will demonstrate problems and difficulties. From the Project point of view, this issue is less relevant, because in the meantime the bibliographic records are searchable through the subject access points they carry when originally loaded into the database. The only concern is that it will take some time to get the subject access points controlled from the intellectual point of view. Moreover, as the linking of bibliographic records to the authority file has been established from the beginning, any further change in those headings will be performed automatically by the system. Thus bibliographic records and indexes will be updated without further intervention of the staff.

Nevertheless one concern remains at this point. This has to do with keeping alive the motivation of the participants in the Project because global effectiveness at the level of the index will take time to appear. But this is, after all, a natural issue arising from any co-operative project, where the partners should not expect everything to be clear, linear and immediate. Co-operative work requires much generosity and understanding in order that readjustments and delays be accepted, especially in the case of projects such as PORBASE and CLIP, where all the required resources from the Project are not allocated. In such circumstances one must be pragmatic and keep in mind that ultimate goals are sometimes a long way off.

Notes
* Paper presented at the Joint FID/CR and ISKO Meeting, Madrid, 10 Oct.1992, preceding the 46th FID Conference and Congress, Madrid, 21-30 Oct. 1992.
1 The management of the Online Union Catalogue of the Portuguese Libraries - PORBASE - has been committed to the Biblioteca Nacional (BN) from its beginning in 1987, simultaneously with the process of the automation of BN. The database is maintained in a GEAC 9000 system, with the following configuration: 3 processors, 16 MB memory RAM, 16 disks (6GB), 2 tape drives, 1 printer (600lpm), 4 UPS and approximately 100 peripheral units (micro-computers emulating terminals of the system).
GEAC is an integrated system and the modules, installed since 1988, are the BPS (Bibliographic Processing System, including authority control) and an OPAC. The acquisitions module is about to be implemented.
The size of the database is about 400,000 bibliographic records and 50,000 authority records. These are concerned mainly with personal and corporate names. The bibliographic database content is about 65% Portuguese bibliography, 32% foreign bibliography, and 3% periodicals and nonbook materials.
There are 56 libraries co-operating in a batch mode sending bibliographic records on floppy disks, prepared on the PORBASE application of Mini-Micro CDS/ISIS, tailored in UNIMARC according to the same parameters used for the GEAC system. BN is responsible for this application, being at the same time the National Distribution Centre for Mini-Micro CDS/ISIS which has already delivered 400 copies of the application. The other

127 libraries from all over the country are co-operating directly online (via X.25) with the GEAC system.

PORBASE is available online via the public data network (either by X.25 or by telephone line). The access is free of charge, and searches can be carried out either in Portuguese or in English. Within the PORBASE environment, the BN also provides a number of types of services: training in UNIMARC and Mini-Micro CDS/ISIS, the development of new applications of the software, consultancy in library automation and the supplying of UNIMARC records on tape or floppy disk (ISO 2709).

2 The total existing bibliographic resources in the country are roughly estimated at 5 million titles, scattered in approximately 600 libraries. According to the LIB2 Study (1991) the majority of libraries have small and medium sized collections. Only 6.6% of libraries surveyed have collections larger that 50,000 items.

3 Biblioteca Nacional, which has applied UDC (Universal Decimal Classification) for many years, has only recently prepared and published a code of rules for subject cataloguing (1st edition 1988) as the foundation for a new alphabetical subject catalogue. The main feature of the subject indexing system designed is precoordination, adopted to ensure the integrity and specificity of the subject information to be indexed.

The manual for the system, SIPORbase Manual, provides rules and procedures for all stages of the subject indexing process, from the analysis to the representation, including all steps necessary for the control of the terminology and construction of semantic structures, according to ISO 2788.

A high level of terminological control was considered necessary for the consistency of the precoordinated system. The rules provided either for common or proper names, are intended to generate a thesaurus within the subject authority system. The rules for precoordination are intended to avoid inconsistency in the syntax for the representation of complex subjects. The development of the subject indexing language will be the results of the application of rules to the current needs in subject cataloguing. To give a more precise idea of the results, it can be said, broadly speaking, that the subject index entries will be LCSH-like, or similar to the main PRECIS strings, which systems and practice have been the theoretical basis for the development of SIPORbase. Furthermore, SIPORbase tries to provide solutions for the lack of consistency, either terminological or syntactical, and for a proper semantic structure that is traditionally applied in pre-coordinated systems, namely LCSH.

4 IDEIA is an application, developed on the CDS/ISIS Software, and especially designed for the construction and maintenance of a subject indexing language, with the following facilities: Construction of one or more databases; automatic checking of terms already existing in the database; automatic generation of reciprocal relationships (RT; BT; NT); searching on all fields; production of several types of printed outputs; and the possibility for use with the PORBASE application in order to move search queries between two databases. The format used in this application is not UNIMARC, but a special format devised to fulfill the needs of a thesaurus or subject heading list. It also includes specifications concerning Project CLIP, which can be ignored by other types of users.

Update Note
At present (1993) Project CLIP involves about 70 institutions and 16 working groups. The current figures of the PORBASE Database are: 90 contributing libraries (including the National Library), 485,000 bibliographic records and about 80,000 authority ones.

Further Reading

(1) Associacao Portuguesa de Bibliotecários, Arquivistas e Documentalistas: Ponto da situacao da aplicacao das novas tecnologias de informacao às bibliotecas e o seu impacto na actividade das bibliotècas em Portugal: LIB2 - Actualizacao. (State of the art of the application of new information technologies in libraries and their impact on library functions in Portugal: LIB2 Update). Lisboa: BAD 1992.

(2) Cabral, Maria Luísa: The use of CDS/ISIS to prepare UNIMARC records for PORBASE. In: UNIMARC in Theory and Practice. Papers from the UNIMARC Workshop, Sydney, Australia 1988. The Hague: IFLA 1989. p.47-50

(3) Cabral, M.L.: Fulfilling the National Library's functions through automation. Paper presented at the IFLA/Unesco Presession Seminar, August 1991, Moscow (to be published by Saur).

(4) Cabral, M.L., Lopes, M.I.: The modernization of Portuguese libraries: Five decisive years. Program 26(1992)p.249-258

(5) Lopes, M.I.: Projecto CLIP: melhorar, compatibilizar et cooperar em indexacao. (CLIP Project: to improve, make compatible and co-operate in the subject indexing process. Paper presented at the 3rd Congress of Portuguese Librarians). Lisbon 1990.

(6) Lopes, M.I.: Situacao e perspectivas de compatilitizacao de linguages documentais em bibliotecas portuguesas. (Achievements and perspectives of the compatibilization of subject indexing languages in Portuguese libraries). Paper presented at the 4th Congress of Portuguese Librarians, Braga 1992.

(7) Portugal. Biblioteca Nacional: PORBASE: Carta de principios. (PORBASE: Statement of Principoles). Lisboa: BN 1991.

Sra.M.I.Lopes, PORBASE Manager, Biblioteca National, Campo Grande, 83, 1751 Lisboa Ccodex, Portugal.