

The authority control in the Academic context: a Hobson's choice

Guido Badalamenti
Università di Siena

The academic context

The control and standardization of cataloguing access is a subject that has always been present in the academic libraries' services organization. This control has been apparent in both the cataloguing context and in the setting up of OPAC tools that are coherent and adequate for the users, enabling students and researchers to conduct easy and thorough searches of the bibliographic collection.

As Liv Aasa Holm (1999) states, the problem of authority control is more apparent in larger libraries than in those of smaller size as it is easier to "control" the search results in smaller libraries. However the increased use of the net has highlighted the need to improve the quality of the results of bibliographic researches.

In the last few years this need has been felt more strongly because of the introduction of important projects involved in the conversion of cards catalogues and for the of retrospective cataloguing onto information mediums. In whatever way these projects have been conducted, through the use of optical scanning or retyping or through the derivation from databases, and also according to the evolution of cataloguing rules and the choice of headings, a need for an intervention in access standardization has been identified. The integration of the acquisition and cataloguing functions, coupled with the cataloguing orders, the chance to involve staff with less specific training such as the function of orders' cataloguing (visible in OPAC) and the incorporation of analytic cataloguing of the miscellaneous works has sharpened this need.

The derived cataloguing is by now a habit of many of our libraries, with a retrieval percentage in the humanities field of between 80 and 90 per cent of the researched data.

The retrieval comes from a variety of sources, both nationally and internationally, that often have different cataloguing standards which use headings that can vary according to the different standards.

The ability to rapidly correct cataloguing access and to standardize important data in a manner that is compatible with the catalogue's structure, is a need strongly felt both in terms of maintaining the catalogues consistency and as a tool to shorten the cataloguing time, thereby raising the overall quality of the catalogues.

The authority control, as Guerrini points out (2002, p. 35), must be considered as the first of the ten indicators (parameters) for measuring the catalogue's quality. The user doesn't have to "guess" the right form, but must be capable of using terms well-known to him, where he can rely on an automated system of translation of his search in a more appropriate form.

Until recently this kind of need could only be carried out by the checking, cleaning and the posterior merge of the cataloguing access, but the recent diffusion of library automation systems, that allow the configuration of authority databases to be integrated with bibliographic databases, has considerably changed the work's perspective.

But this technological development is hindered by the absence of important and already constituted authority archives in the national environment or the absence of an authoritative centre with specific duties, to which refer to and from which have the possibility of retrieving authority records that are coherent to the national cataloguing codes and with the international standards.

The Universities and other libraries therefore need to face this task from the inside and to commit considerable resources to this field, because a high level of specialization is required for the maintenance and updating of authority records.

Of course important bibliographic catalogues exist, among these SBN, that constitute a precious resource for the cooperation and for the derived cataloguing – especially in the humanities field – even if different levels of cataloguing expertise exist. In fact, sometimes the addition and substitution of old and new records can make records made by centres of excellence disappear.

The development in SBN of a specific project using authority file, with an output format UNIMARC/Authorities, will surely be an important answer to the needs of the Italian libraries.

Other important national points of reference are held by some specialized libraries, which are introducing databases with authority headings on specific subjects, structured according to the international standards. In particular, think about the work on the medieval authors in progress at the Library Franceschini of the Galluzzo (Florence) or the work in progress at the Museum of Science of Florence for authors of the historical-scientific field.¹

But the major problems that remain are that of the critical mass and that of the availability on line of these tools as they are the primary source of information for the other libraries, so it is necessary to save the resources and to avoid duplication of efforts.

The authority control as a search tool

The growth of the services offered by libraries, their growing complexity, the need to adequately reserve human resources for the integration of traditional library's services with those offered by the electronic resources, and the installation of the digital library have all contributed to the serious problems that exist in the organization of the cataloguing offices in the academic libraries, especially in the context of a general reduction in economic resources.

In the re-organization of the library's workflow, doubt has been raised on the possibility of maintaining adequate resources to the sector of cataloguing and to the control of catalogue access, with the added risk of a progressive decline in the quality of catalogues.

However, it has become evident that as the instruments for assisting searches are set up (such as the introduction of a controlled list of authors and thesaurus) even if it is time-consuming because of the high training level required by cataloguing staff, it constitutes an important tool not only for when consulting the bibliographic catalogues, but also as support for searches in the digital library and to help users search the internet more thoroughly.

The main function of the authority file will move from cataloguing towards searching (Weston, 2002, p.138).

It is widely recognized that there is a need to have the ability to define tools which help users to follow customized and controlled searches' paths, which normally escape from the commercial policy of the big editors or of the search engines. The search engines are always more powerful, but their results are influenced by the mechanism of information promotion which is linked to the practice of highlighting and emphasizing some electronic resources rather than others. Gori and Witten (2002) have analysed the mechanism through which it is possible to make the visibility of a web site arbitrarily grow, in a search engine such as Google, with the deliberate construction of a number of artificial pages; and also talk of the danger of a speculative bubble similar to those experienced by the Stock Exchange. The only way to avoid this danger is by the setting up of tools which allow a semantic analysis of the web (Burners-Lee, 2001)(Brooks, 2002).

Barbara Tillett (1996) suggests the idea that, from the existence of international authority files that are real or virtual, whereby versions of every entity with parallel forms in different languages and in

¹ Ezio Franceschini foundation (S.I.S.M.E.L.), Library of medieval culture: <http://www.sismelfirenze.it/>, Library of the History of science institute of Florence: <http://galileo.imss.firenze.it/biblio/icat.html>

different alphabets are recorded, it could be possible to develop a level of software interface suitable to the specific needs of each user. She also points out that the national and international cooperation, whereby institutions have the chance to benefit from the survey and investigation work done by other colleagues, can be the only valid solution to the economic problems that arise from an activity that is strongly dependent on the use of human resources.

Academic libraries can have an important role in this context because of the wide typology of collections that interest their institution's users, which includes the works of local, national and international authors covering a variety of subjects.

However the major problem that arises is that of the general work's organisation, which needs to succeed in combining the availability of internal resources with the ability to use general resources on line, in order to create an authority file compatible with the needs of the specific users' search and that can simultaneously create a common entry.

At the moment, even in the international context, there are not many authority files available on-line and accessible through the Z39.50 protocol.

From a recent survey (Kaiser, 2002) in which 568 libraries and archives from different nationalities participated, it was found that barely a quarter of them load authority records from external sources, while most of the structures work with local archives.

An important point of reference in this context, which is able to provide answers to unsolved problems, is the European project LEAF (Linking and Exploring Authority Files)², which has as its aim the creation of a centralized system which will house localised records concerning personal and corporate authors existing in the databases of both small and large libraries in different countries.

These records will be gathered through recurrent procedures of 'harvesting' in the different databases, in order to assure a constant upgrade of data.

The project provides an aggregation of records, which refer to the same entity, though preserving the identity of the sources from which the records originate and the specifics of each database. Data would be easily accessible not only for professional staff, but also for the end user.

LACoBiT: a pilot project

Within the Siena Bibliographic Service – as in other Universities – it has tried to compensate for the lack of a reference's authority structure by creating a database of personal and corporate authors' headings starting from its own bibliographical catalogues.³ However large the catalogue is, built from the cataloguing activity of a library system, it could hardly answer – alone – to the demands of the current cataloguing system.

In this context and in that of the Tuscany Committee SBN – to which the representatives of the Tuscany libraries participate independently from the automation system that is used – the database has been built as an initiative to pilot a project which could be of use in the setting up of an archive of controlled headings, consultable in the context of the regional *Metaopac*,⁴ namely LACoBiT (Integrated List of controlled authors of Tuscany Libraries).

The decision has been to integrate and merge into one database data elaborated by a structure of excellence such as the Central National Library of Florence, which produces the National Bibliography, with those of two specialized archives: the Siena Bibliographic Service which gathers data produced by the academic libraries and municipal libraries and the database LAIT

² Linking and Exploring Authority Files (LEAF), <http://xml.coverpages.org/leaf.html>

³ To the net of the Siena Bibliographic Service (SBS) (<http://www.unisi.it/servizi/sab/welcome.html>) participate all the Libraries of the university of Siena, the principal Public and Private and Research Libraries of the town of Siena and the Libraries of the Municipality of Siena Province, that are part of the net ReDoS (Sienna Provincial documentary net): <http://redos.unisi.it/ALEPH/-/start/sbs02>

⁴ Metaopac of the Tuscany Region allows to consult the total amount of catalogues of all the Libraries of the Region which have a Z39.50 server. (<http://www.cultura.toscana.it/bibl/metaopac.htm>) e breve descrizione

(ancient books in Tuscany) produced by the municipal library of Pistoia, in which the data gathered is collected during the survey about the ancient books in Tuscany libraries.

The aim is to build a numerically consistent work's tool which, starting from the daily cataloguing activity of the libraries involved, connects the similar and the different forms of the names of a personal or corporate author, structured by records that cohere to the standard UNIMARC/Authorities. Within this project it has been decided to limit the database to personal and corporate authors' names.

The list of authors will undergo a merge of structured records from a selection of variants which are decided by the preferences of the participating institutions. Being a *work in progress*, the database could present records with different levels of analysis, with more or less deep degree of control, answering to rules of choice not uniform with each other.

The aim is to provide a numerically substantial on-line tool of comparison and analysis from which each library can access information that is useful for the organisation of its own database, answering the cataloguing needs of each library.

Barbara Tillett (1998) has pointed out the importance of moving the emphasis to 'access control' rather than to 'authority control' and of the necessity of building structures that are flexible and adherent to the users' demands. Linda Barnhart (1996) who defines the access control record as a "super authority record" has suggested various hypotheses in the configuration of the authority records (also using local fields to classify an author from a biographic point of view) in order to fully utilise the data gathered from external partners and then making adjustments according to the internal demands of each catalogue.

The project is based on limited resources, formed mainly from the work done in the structures which give their data, as well as through the support of the Tuscany region and the technological contribution and know how of the University of Sienna.

It aims to be a concrete occasion of reflection about the problems linked to the integration of records of headings of controlled authors (personal or corporate) built from the collections of libraries with different structures and aims.

As Alan Danskin (1997, p.33) highlights, in his attentive analysis on the relationship between costs and benefits linked to the management of authority control, massive investments don't automatically translate to massive results. Instead, it is necessary to pay major attention to the standardization of cataloguing codes, to the use of exchange formats and to the fact that elaborated data would be easily accessible on-line. With this in mind, we believe that we can offer everybody a huge archive of data through the Internet, even if it is incomplete and has internal inconsistencies due to the stratification of catalogues. Even so, being structured according to the major national and international standards, it can still make a large contribution to the development of common requisites for co-operation in a uniform cultural context, such as the national context.

In a limited amount of time, we propose to build a central database which will collect data from the catalogues of three libraries in the Tuscany area, and which will then be open to contributions from other libraries.

In this database, which has originated from the integration of three catalogues, it is therefore necessary to guard the identity and the specificity of each of the libraries that supply the catalogue with their data.

Operative/functional Model

The project consists of the loading and searching of all data through the software ALEPH500⁵, currently in use at the University of Sienna, through which a specific database for the management of authority control has been defined.

⁵ The software which is produced by ExLibris (<http://www.aleph.co.il>) allows the creation of databases of authority coherent with the MARC standards, independent and integrated with the bibliographic database.

Starting from the information in the bibliographic records, those of controlled authors in their variant forms will be created through specific conversion's programmes.

Such records will keep the system number of the original entry on the basis of which they have been created; this will allow the addition of periodically new records and the transfer of updated records from the original archives.

Through a web interface it will be possible to separately consult the three archives of controlled authors of BNI, LAIT and SBS, and a fourth integrated archive which enables users to search the three integrated catalogues: LACoBiT.

Records referring to the same entry and that present the same variants, even if a discrepancy between the accepted form and the variant form exists, will be gathered and represented in one single record, with reference to the database from which these records have been created.

In the absence of an ISADN number for the unique identification of the entry, this identification with all the complexities of the case, will be made according to the author's name.

The decision is not to reduce the complexity of each individual record (flatten or squeezing of records) as the choice of variants will depend on different perspectives, especially when considering the specificity of each library user. But it will be made possible at any time to verify the record in the context of the catalogue where it has been produced.

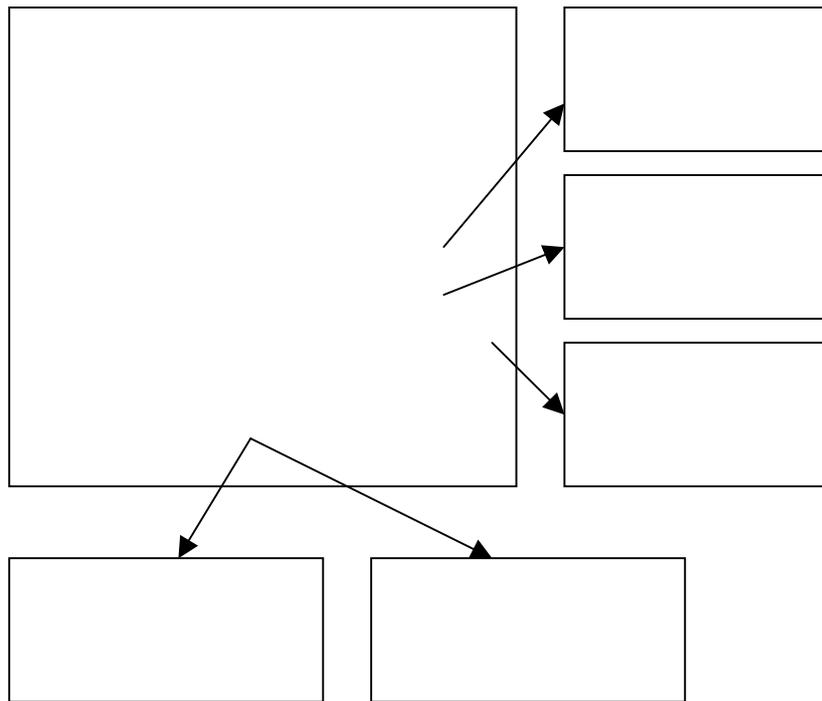
A suitable link will enable a user to move automatically from the record included in the integrated catalogue directly to the original record contained in the single database. Indeed, the link will include the reference of the original database, as well as the record's system number in the specific database and it will indicate the level of completeness of the record.

All the other records which present a structure with a wider number of variants or which include different interpretations of the heading, or that cross-reference to other connected headings, will at this stage be kept as independent records.

A second group of *links* will permit the navigation from the record that is currently being examined to other records that are marked as "to be checked" and included in the same shared catalogue:

- some of these can be records related to the same entity but that contain further variants to the form of the name; running regularly, these records will have no reason to exist as, from the inception, all the variants of an entity would be gathered in a single record (cfr. Serrai, 1983, p. 85). But in this initial stage, when much work still has to be done, we believe they can be useful as a warning for cataloguers and final users.
- other links can enable navigation among records which relate to different entities and can present a variant of the author's name similar to a record that has already been examined, or it could even present the same name's form as there isn't any qualification which permits the differentiation between authors.

This is shown in the example described below where, besides the variant forms of Latin, Italian and French names, remains ambiguous the philosopher Platone from the present-day author Platon.



When the records in question are controlled and integrated by the staff of the local database, we will no longer have the second group of navigation's links at the level of the shared database; we will only have the first group which will allow navigation from the record of all the variant forms of the name to the linked database.

The second record will now be unambiguous so it will not be part of the records "to check".
Example:

<p>001000334496 100 \$a 20020329aitay0103----ba0 106 \$a 0 120 \$a bb 152 \$a RICA 200 0 \$8 italat \$a Plato 340 \$a Filosofo greco 400 0 \$8 itaita \$a Platone 400 0 \$8 itafre \$a Platon 400 0 \$8 itager \$a Platon</p>	<p>001000110011 100 \$a 20020929aitay0103----ba0 106 \$a 0 120 \$a ba 152 \$a RICA 200 0 \$8 itafre \$a Platon \$c karuna \$f 1932- 340 \$a Autore francese di opere esoteriche nato il 07.03.1932, il titolo karuna significa 'compassione divina' 400 1 \$8 itafre \$a Platon, \$b Karuna \$f 1932-</p>
--	--

Obviously this does not take into account that in one or more local databases, records can appear altered between the first and the other variants:

100 \$a 20020329aitay0103----ba0
106 \$a 0
120 \$a bb
200 0 \$8 itaita \$a Platone
400 0 \$8 italat \$a Plato

Interventions of revision and integration of the records of controlled authors will take place in a peripheral manner, in the various native databases: BNI, SBS, LAIT, while the shared database will reflect the adjustments done to these archives.

This procedure ensures the autonomy and the coherence of each of the data's source database, without any danger of reducing the complexity of the record (flattens).

Interventions made on a single database can be evaluated and considered even in the context of other databases, because each subsequent amendment to a record after its initial loading, will be redistributed to the other libraries.

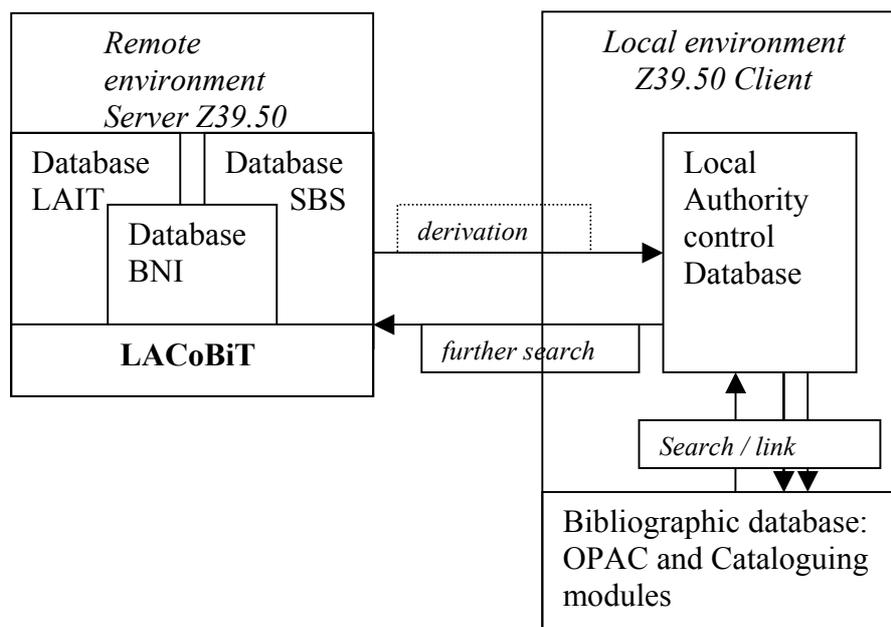
In this way the changes done on a BNI record (that is, for instance, the object of merging a single record with a record coming from SBS) will be highlighted in an electronic way to the users of SBS. The same will also happen for the other catalogues and it will therefore be an optional choice to align with changes that have been done in a database different from their own.

This aspect is particularly delicate as authority records, unlike bibliographical records which have a kind of stability, are subject to continuous change due to the necessity of modification in the variant forms, to the introduction of new bibliographic sources, to the necessity to construct clear and related forms, to the change of an author's name or simply the necessity of amending the biographic information to reflect dates of death etc.

These are all elements that obviously considerably complicate the problem of the catalogues' alignment and of the exchange of records; problems that cannot however be solved with the operation of the "flattening" of data.

It will be possible to consult the four catalogues via Z39.50, through one's own client or through those available on-line. This will allow the configuration of the cataloguing interface by the staff of any library so that it will be possible to consult its own database of authority first and of those at its disposal through Z39.50 Server of the University of Sienna, in addition to those available on line.

During the cataloguing phase, the cataloger could firstly search on his own authority database to link the heading chosen to the bibliographic record. If he is unsuccessful, he could do the search through Z39.50 on the integrated catalogue LACoBiT or singularly on the three catalogues which form it. The records found could be exported and loaded into his own database of authority, in order to integrate or modify them coherently with the cataloguing choices that are internal to each structure and to be later linked to the records of the bibliographic database.



The standards

The standard of reference for the creation of records within the databases follow the rules of RICA in the form of the authority's headings, GARR and ISAAR(CPF).

In the analysis and discussion of the rules, specific time has been devoted to the training and updating of cataloguing staff.

It has been considered the work done by the Working Group in charge of the revision of the RICA and have been faced with the themes concerning the choice of the heading and the coherence of the catalogues which have been also recently subject of a wide discussion (Guerrini, 2002, p. 79 and next.).

The comparison between different positions and the analysis of specific surveys has been particularly significant in the comprehension of the complexity of the choices that must be made by the catalogers and for the organization of future work.

In particular, the preliminary work of analysis has served to instil in the operators the belief of the necessity, often recalled by Gorman (1979) and Tillett (1990), to focus the concept of the authority record as an organized structure of variants of a name, rather than as a relation between accepted and rejected forms of a name. By following this viewpoint, the result was to easily acquire the principle – when in doubt opting for the first of the variants of an author or corporate's name to insert in a record – of abiding to the choices set by BNI and to follow the type of work which has been developed in that context.

Obviously this can't be immediately reflected on all the databases that, as we have mentioned, are in the first phase of execution and are based on the automatic creation of author's records derived from bibliographic catalogues; catalogues that are the result of a stratification of ages. It will only be through the progressive work of refinement and the increase of human resources that it will be possible to tend to a major internal coherence of the databases.

For the records format the reference standard is UNIMARC/Authorities. One of the first tasks has been to fix, together with the colleague Dina Pasqualetti from the National Central Library of Florence, a subset of reference according to what has been defined by the IFLA's Work Group which has established the minimum level of the authority records for the international exchange.⁶

Obviously the adopted scheme - that could be eventually reviewed during the application phase – serves as a reference for the creation of new records, but it cannot be applied in full to the automatic generation of records on the base of bibliographic databases, because the lack of a detailed analysis of the original data means that sketchy information could be integrated instead of the manual revision of records (scheme in appendix).

SBS database

The first database on which we have worked is that of the Servizio Bibliotecario Senese, for which the database SBS10 has been created and is now available on line at the web address <http://sbs2.unisi.it/ALEPH/>. The principle choice has been to use elaborate records for the authority control starting both from the shared database of the academic and civic libraries and from the cumulative database formed by the catalogues of the public libraries of Sienna's Province.

In these catalogues numerous duplications were present, in comparison with the main catalogue, but there were also many authors linked to local production and to children's literature.

A total of 1,200,000 bibliographic records have been processed, from which all of the tags 70x and 71x have been extracted to make a total of 1.328.278 entries for personal or corporate authors.

⁶Mandatory Data Elements for Internationally Shared Resource Authority Records, Report of the IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN: <http://www.ifla.org/VI/3/p1996-2/mlar.htm>

The outcome of the routine of de-duplication has been the creation of 460.168 records in the authority database enriched with the tags foreseen from the requested subset UNIMARC/Authorities.

The new records will maintain, in a 810 tag, the indication of the document held in the bibliographic database, according to which the authors' names have been created.

Programs of de-duplication have permitted the exclusion of strings which are an exact copy and a subsequent "normalization" of characters inside the strings with the definition of a table of equivalence between the ISO set of characters and the basic ones, concerning punctuation marks, special characters and abbreviations, capitals and small letters.

On this structure and contents equal, considering the characteristics of the bibliographic database, an algorithm has been defined for the selection of strings which have a bigger "weight" according to the criteria of complexity of the string: form (capital, small letter), numbers of subfields, majority of accents, chronology, staff's experience, level of completeness of the bibliographic record of origin etc.

As a result, the individualization of the rejected forms of the name have been used in an operation of cleaning the bibliographic catalogues.

The authority database, which is linked to the bibliographic database, is constantly updated, modified and integrated according to the daily cataloguing that takes place.

BNI database

The second step of the project, currently in progress, consists in the creation of an authority database, SBS11, where records from the database BNI will be loaded and immediately made available on-line.

The first part of the work consisted of the setting up and the control tests - together with colleagues of the National Library - of the format of downloading the data, that will happen with the coding UTF8.

From the database SBN all the authors of whatever type (ABCDEGR) of an accepted form, to which a record from BNI has been linked, have been extracted. The result has been 495.479 authors with these aforementioned features.

From the initial file (containing identification author, identification bibliographic record and BNI number) only the authors linked to BNI record from 1958-2002 have been selected, as in that range of time standards RICA and ISBD have been used.

The result is 247.740 authors in which, barring mistakes, there would be no duplications.

For each author with an accepted form the procedure generates an UNIMARC record which includes the label, indicator, tag 7xx, 300, 9xx, as well as the date of the creation of the data in order to allow the staff who have to use the record to contextualize them historically.

The structure of the record is ISO2709.

Now there are going to be defined programs for the conversion of tags 9xx to the correct tags 4xx/5xx and for the integration of data extracted from BNI with tags dictated by the subset UNIMARC/Authorites.

LAIT database

The LAIT⁷ database includes around 90,000 accumulated descriptions of ancient books edited between 1501 and 1885, the result of five years of work by 42 Tuscan libraries. The descriptions

⁷ TheDatabase, wanted by the Tuscany Region, has been created at the Forteguerriana Library on a database ISIS/EDAN and it has been reproduced on CD-ROM. Cfr. Angela Bargellini, Teresa Dolfi, Luciano Vannucci,

and in particular the names of 35.298 authors have been checked and normalized by the Forteguerriana Library of Pistoia.

The third phase of the project consisted of the exportation of the names of the authors and the variants of the author's name in ISO2709 format by the Library of Pistoia.

A specific program of conversion data will be converted to UNIMARC tags and they can be integrated with the tags dictated by the subset UNIMARC/Authorities.

The records that are obtained will be uploaded in the database SBS12.

The integrated database

The fourth part of the project consists in the setting up of a central catalogue (SBS19) where records will be gathered from the three catalogues of the controlled headings, and, if successful, with also those from the other libraries distributed in the regional area.

The day to day activity of cataloguing, of that of the cleaning and the integration of the existing records, remains at a local level and it is reflected on the central catalogue as a consequence of the batch loading of data.

The most complex aspect is represented by the definition of the procedures of de-duplication and by the setting up of an algorithm which will allow the 'weighting' of records in order to fix which, among similar records, will be displayed in the catalogue as a guideline for grouping the other records.

We are thinking about a dynamic procedure that also reflects the evolution of records in the native database and that will consider each time the fact that a record can be simply the result of an automatic generation or of a provisional/preliminary entry or rather of a record controlled by qualified staff and with a bibliography of support.

This will also allow the user immediate clarification of the status of the record, because it will be possible to establish in which order the links for the navigation to the native database must be created: from the most controlled record towards the less controlled one.

Besides the definition of the criteria of equivalence of characters and strings explained earlier (4.2) the following need to be taken into consideration:

- The authoritativeness of the source;
- The bare automatic generation or the manual revision of the record;
- The level of completeness of the record;
- The presence of a bibliography starting from when the entry has been checked,
- The date of creation or subsequent updating of the record.

CD-ROM

As we have mentioned, the project should be conceived as a *work in progress* that can be used as assistance for libraries during the process of original or derived cataloguing, to assist in their choice of headings of documents. To allow *off-line* queries of the databases, particularly for libraries with major problems of accessibility to the net, it is assumed that there will be a publication of a CD-ROM which has periodical updating.

Project development

The presence of three library realities in a specific geographic area makes contact among structures but especially among staff in charge of cataloguing and the maintenance of catalogues easier. In fact it is important in this kind of operation to arrange collective training times, a constant coordination of the involved resources and the formalization of moments of work in a group. The progressive aggregation - on a regional or thematic base - of a self-generated authors' database could also have significant implications for the future in the setting up of national or international integrated databases, such as the LEAF project.

Work Group

The management of the project has been undertaken by eminent figures from the libraries and the surrounding region,⁸ but an important development has been the constitution of an inter-institutional Work Group in which experienced librarians comment on all the structures involved.

Besides the work of normalization done on the respective catalogues, continuous common activities are scheduled, aimed at defining the setting up of a shared methodology of work, to make an analysis of the different levels of intervention, to make a precise evaluation of the resources necessary to the cleaning of the archives, to set up a schedule and a controlled procedure of intervention.

It is not sufficient to define operative manuals to give to staff. The complexity of the cleaning operation, the need for wide documentation, the discussion of the priority of the sources of reference, but also the discussion on specific cases, requires a closer involvement with relevant staff.

We strongly believe that this kind of approach can have significant effects from the point of view of the homogeneity of the catalogues that will be created, because they will progressively generate a common patrimony of knowledge.

Documentation

It has been shown that one of the major problems that are met in the authority control's work is that of the identification of bibliographic sources to consult. The Working Group has started to set up a bibliographic archive, purposely built, where the main bibliographies present in the involved libraries, and the bibliographies available on-line, are recorded and attention is drawn to the bibliographic databases that - for dimensions and authoritativeness - are appropriate to consult.

This will allow each cataloger to easily identify and locate bibliographies of reference and to be able to co-operate in an electronic way with the colleagues who hold the original material.

The bibliographic database has been structured in such a way that, with the progressive use of bibliographies, it is possible to record guideline notes about the characteristics of each source, on their interest from the point of view of the form and structure of headings, but also on their integration or superimposition with other paper bibliographies; and more impressively, with those available on-line that can be easily accessible from any library.

A key will be defined for the abbreviation of bibliographic sources, to be used in the notes of authority records.

Manuals about the use of the catalogues, the standard UNIMARC and the subset defined within the project, will be available on-line in a special section.

An other tool of great utility, on which we are working, and which has already been widely used for other operational modules, is the setting up of a database of FAQs.

⁸ Gloria Cerbai Ammannati (National Central Library of Florence), Guido Badalamenti (University of Siena), Mario Moretti (National Central Library of Florence), Dina Pasqualetti (National Central Library of Florence), Gian Bruno Ravenni (Tuscany Region), Paola Ricciardi (Tuscany Region), Maurizio Vivarelli (Forteguerriana Library of Pistoia).

It allows questions to be recorded that have arisen during the routine cataloguing activities of librarians in the different structures and to record the explanations of the experienced librarians who are involved in the Working Group.

The database, that has already been built, is filling up with data and it will be accessible on-line through a web interface.

Training

The process of training and updating is certainly one of the more important aspects of the project and must involve not only the staff who generate and update the records in the authority database, but also with those who perform a normal activity of cataloguing or pre-cataloguing in the bibliographical databases, as it is from their preferences that authority records will be generated.

A first step in the training has been done at the Siena Bibliographic Service and has involved – even if it has used different formative modules – all the staff from the academic and the provincial libraries (a total of 120 people), as well as some colleagues from the National Library in Florence who are involved in the Working Group.

The project has been accomplished with the contribution of internal staff, but has also turned to qualified external experts: Mauro Guerrini from University of Florence and Gloria Cerbai Ammannati from the National Library. Topics faced have been the authority control, the analysis of the principal experiences of the work, the analysis of the sources, the choice and form of headings, the evolution of the national code, the structure of UNIMARC/Authorities and the management and implementation of the authority database.

From this first step, others must surely follow, devoted to the staff in all of the libraries, differing according to the individual training needs: to the induction of new staff, as a moment of control of the work done and of the knowledge acquired by those who are already involved in the work.

For an analysis of the main subjects, frequent meetings must be held, with a different methodology of training.

Internal organisation

The specificity of libraries which participate in supplying data to the common database of the controlled authors and of the different information organisations determine the different procedures for the generation of records.

The LAIT database, specializing in antiques books, is the result of a regional project which has contributed staff from numerous libraries who conduct the control of sources and bibliography from the first phase of the cataloguing of documents. After this step the process of the control of headings is done at the Library of Pistoia and the generation of records of authority arrives successively as a specific program within this project.

For the Italian National Bibliography, produced by BNCF, the daily activity of the description of monographs given to a specific section is connected to that of the revision, done by a group of control of headings which operates frequent checks and validations

A new database UNIMARC/BNI will soon be available which will be independent from the bibliographic database, where structured records will be connected to authors and titles according to the principle of authority control.⁹ At this stage, records of the controlled authors will be created with a specific procedure originating with the bibliographic records.

⁹ As Gloria Cerbai Ammannati has explained in : *La Bibliografia nazionale italiana e il controllo dei punti di accesso*.

Within the Sienna Bibliographic Service, next to several bibliographic databases, one database for the authority control has been separately set up but remains strongly connected with those bibliographic, which updated daily by the cataloguing activities of the different libraries.

In the organization of the daily work, in order to guarantee a certain control on the quality of catalogues, a different password has been designed for the staff who work on the bibliographical database:

- the first level only allows them to derivation and capture the headings from the authority database to import them into the bibliographical catalogues and create a link between the controlled heading and the cataloguing access;
- the second level - besides the derivation of records – allows the automatic generation of a record in the authority database according to tags that contain the name of the authors included in the bibliographic records. Records automatically created in the authority database must be checked as they will have a low level of completeness and they must be controlled and validated by the staff with a password of level four;
- the third level allows access to the database of authority for adjustments, but only on the accepted form and for the creation of bibliographic notes;
- the fourth level permits the operation of adjustment, amendment and validation of the records of authority, including the introduction of the variants of the authors' names and allows the cataloguing of the record as "complete". This password is reserved to the responsibility of the cataloguing staff of each library who's work is primarily cataloguing;
- with the password of the fifth level there is no limitation and it is reserved for the staff of the inter-institutional Working Group, who will operate the whole database. Besides records of authority, they can generate cross-reference records for complex variants of the accepted form or records containing general notes about the use of some terms within the authority database.

Cataloguing staff and to the Working Group will be kept informed of developments through a procedure which arranges the automatic creation of electronic messages starting from the contents page of the authority record. This will facilitate the exchange of information, the act of communication and will keep people informed of any problems that have been reported.

Messages can be addressed to a specific person or to a group and they can include descriptive notes or alerts on changes to be done on records.

Librarians in charge of the authority will find on their desk, at the opening of the system, information on how to deal with problems and any suggestions proposed. These records can be extracted and controlled at any time.

Bibliography:

Barnhart, Linda (1996) *Access control records: prospects and challenger*, in *Authority control in the 21st century: an invitational conference, March 31-April 1, 1996*,
<http://www.oclc.org/oclc/man/authconf/barnhart.htm>.

Berners-Lee, Tim, Hendler, James, Lassila, Ora (2001) *The semantic web*, <<Scientific American>>, 284 (2001), n.5, pp. 28-37.

Brooks, Terrence A. (2002) *The semantic web, universalist ambition and some lessons from librarianship*, <<Information research>>, 7(2002), n. 4,
<http://InformationR.net/ir/7-4/paper136.html>.

Danskin, Alan (1997) *International standards in authority data control: costs and benefits*, <<International cataloguing and bibliographic control>>, 26 (1997), n.2

GARR: International federation of library associations and institutions, *Guidelines for Authority Records and Reference*, 2. ed., revised by the IFLA Working Group on GARE Revision. Munchen: Saur, 2001.

Gori, Marco and Witten, Ian (2002) *The bubble of web visibility*, Working papers. Siena: Università degli studi di Siena, Facoltà di Ingegneria. To be published in <<Communications of the ACM>>, (2003), <http://www.acm.org/cacm/>.

Gorman, Michael (1979) *Cataloging and the new technologies*, in *The nature and future of the catalog: Proceedings of the ALA's information science and automation division's 1975 and 1977 institutes on the catalog*, edited by Maurice J. Freedman and S.Michael Malinconico. Mansell: Oryx, 1979, pp127-136.

Guerrini, Mauro (2002) *Il catalogo di qualità*. Firenze: Regione Toscana, Pagnini e Martinelli, 2002.

Holm, Liv Aasa (1999) *Authority control in an international context in the new environment*, <<International cataloguing and bibliographic control>>, 28 (1999), n. 1, pp. 11-13.

ISAAR(CPF): *International Standard Archival Authority Record for Corporate bodies, Persons and Families, Final ICA approved version*, prepared by the Ad Hoc Commission on Descriptive Standards, Paris, 15-20 November 1995. Ottawa, 1996.

Kaiser, Max (2002) *LEAF User survey*, <<LEAF Newsletter>>, 1 (2002), n.1, http://www.crxnet.com/leaf/news_online.html.

RICA: *Regole italiane di catalogazione per autori*. Roma: Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, 1979.

Serrai, Alfredo (1983) *Biblioteche e cataloghi*. Firenze: Sansoni, 1983.

Tillett, Barbara (1990) *Access control: a model for descriptive, holding and control records*, in *Convergence: proceedings of the second national library and information technology association, October 2-6, 1988, Boston, Massachusetts*. Chicago: American Library Association, 1990, pp.48-56.

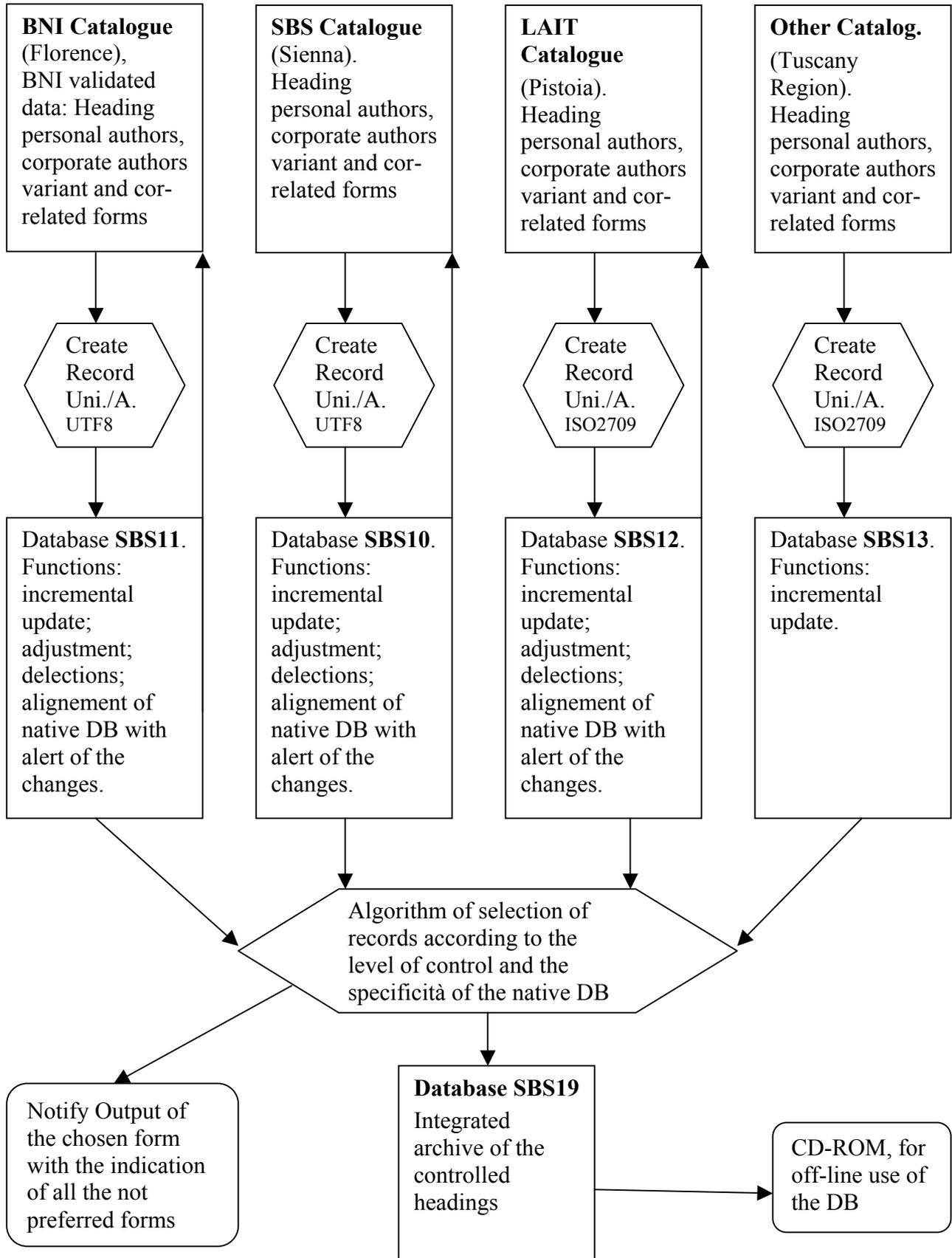
Idem (1996) *International shared resource records for controlled access*, in *Authority control in the 21st century: an invitational conference, March 31-April 1, 1996*, <http://www.oclc.org/oclc/man/authconf/tillett.htm>.

Idem (1998) *International shared resource records for controlled access*, <<ALCTS Newsletter>>, 10 (1998), n. 1, http://www.ala.org/alcts/alcts_news/v10n1/gateway.html.

UNIMARC/Authorities: International federation of library associations and institutions, *UNIMARC Manual: Authorities Format*, 2. ed. rev., enlarged ed. Munchen: Saur, 2001.

Weston, Paul Gabriel (2002) *Il catalogo elettronico: dalla biblioteca cartacea alla biblioteca digitale*. Roma: Carocci, 2002.

Project Schema



Subset of the UNIMARC Authorities standard defined for the LACoBiT project¹⁰

by Guido Badalamenti, Dina Pasqualetti

Data	Definition	UNIMARC/A	Mand. Repe.	Notes for the Automatic generation of the authority records.
Status of the record - c = corrected - d = deleted - n = new	New or corrected record. The value 'd' in the exchange of data means 'has to be deleted'	Pos. 5 of the Label	Yes Not repea.	Automatic records will have value 'n'
Type of record - x = authority entry - y = reference entry - z = general explanat.		Pos. 6 of the Label	Yes Not repea.	Automatic records will be only authority entry records, value 'x'
Type of entity - a = personal name - b = corporate name	The code indicates the type of entity identified in the 2xx fields	Pos. 9 of the Label	Yes Not repea.	In the automatic creation of records the value could be defined because of the type of 2xx tags
Encoding level - # = complete - 3 = partial	Show the degree of completeness of all the record	Pos.17 of the Label	Yes Not repea.	Automatic records will have value '3', it could be changed to # when it is checked manually
Record identifier	Record control number	Tag 001 No indicators	Yes Not repea.	Assigned by the system when the data will be loaded in the LACoBiT Database
Last update	Date and time of the latest record transaction	Tag 005 No indicators	Yes Not repea.	Automatically done by the system
ISADN	International Standard Authority Number	Tag 015		Reserved for ISADN
Other system control numbers	System number of the Agency where the record originated	Tag 035 Indic.: ## \$a, \$z	Yes Repea.	A code for the Agency in parentheses will precede the system number
Date entered on file	Date of loaded records in the LACoBiT database	Tag 100 No indicators \$a pos. 0-7	Yes Not repea.	Historical data, not to be changed. (ISO8601) aaaammgg
Status of authority heading code - a = established - c = provisional - x = not applicable	Show if a deeper investigation is necessary to establish the heading, when it is next used	Tag 100 \$a pos. 8	Yes Not repea.	Automatic records will have value 'a'
Language of cataloguing	Code of the language used in cataloguing	Tag 100 \$a pos. 9-11	Yes Not repea.	Value 'ita' automatically assigned by the system
Transliteration code - a = ISO scheme - y = no transl. scheme	Show if and which transliteration scheme has been used	Tag 100 \$a pos. 12	Yes Not repea.	Automatic records will have value 'y'
Character Set	Codes related to the character sets used in the	Tag 100 \$a pos. 13-16	Yes Not	Character set Unicode: 50-----, used by the Agency that

¹⁰ The LACoBiT project is limited to the headings personal name, corporate body name. Accordingly to that, the subset is limited to the related fields. So the scheme is not complete and namely we didn't include – for the moment – family name, place name, uniform title, subject, etc.

In the fourth column, the choice of mandatory and repeatable fields for LACoBiT project is explained. In the fifth column we specially give notes about the automatic generation of the authority records from the bibliographic records and the related default value. During the manual check of the records the default value of the mandatory fields have to be completed by librarians.

	record	e 17-20	repea.	maintains the records
Script of cataloguing		Tag 100 \$a pos. 21-22	Yes N.rep.	Value 'ba' assigned by the system
Direction of script - 0 = left to right - 1 = right to left		Tag 100 \$a pos. 23	Yes Not repea.	Value '0' assigned by the system
Language of the entity	The field contains indication of the language/s used by the entity	Tag 101 Indic.: ## \$a (repeatable)	Yes Non ripet.	Automatic records will have tree fillers ' ' in subfield \$a. Value to be completed.
Nationality of the entity	The field contains the nationality of person, corporate body	Tag 102 Indic.: ## \$a \$b	No Not repea.	Automatic records will have value 'XX' in subfield \$a. Value to be completed.
Coded data fields: heading used as subject heading	Show if the 2xx heading could or couldn't be used as subject heading	Tag 106 Indic.: ## \$a	Yes Not repea.	Automatic records will have value '0' (more frequent), it could be changed when it is checked manually
Coded data field: personal name	Show if the name is differentiated or not and what is the gender of the entity	Tag 120 Indic.: ## \$a pos. 0 (a,b,c,u,x) pos. 1 (a,b)	Yes Not repea.	Pos. 0 = filler automatically assigned by the system Pos. 1 = if there is a subfield \$c, \$f o \$g in the 2xx tag, the default value is 'a', in the other circumstance the default is 'b'
Coded data field: corporate name	Codes for the level of the government organization	Tag 150 Indic.: ## \$a pos.0 (a,b,c,d,e,f,g,h,u,y,z)	Yes Not repea.	Pos. 0 = filler, automatically assigned by the system
Rules \$a cataloguing rules \$b subject system	Rules used to formulate the heading/reference structure	Tag 152 Indic.: ## \$a	Yes Not repea.	Code 'RICA' Automatically assigned by the system for the subfield \$a
Heading – Personal name	Heading chosen according to the RICA rules	Tag 200 Indic.1: #; Indic. 2: 0,1 \$a,\$b,\$c,\$f,\$g,\$8	Yes Repea.	In the Automatic generation of the record, the control subfield will have the value: \$8ita . Value to be completed
Heading – Corporate name	Heading choosed according to the RICA rules	Tag 210 Indic.1: 0,1 Indic.2: 0,1,2 \$a,\$b,\$c,\$d,\$e,\$f,\$g,\$h,\$8	Yes Repea.	In the Automatic generation of the record, the control subfield will have the value: \$8ita . Value to be completed
Information note	Explain the relationship between the 2xx heading and the other entities	Tag 300 Indic.1: 0,1; Indic.2: # \$a	No Repea.	The first indicator will have the value '0'
Textual see also reference note	Explain in details the relationship between the 2xx heading and the 5xx headings	Tag 305 Indic.1: 0,1; Indic.2: # \$a,\$b	No Repea.	The first indicator will have the value '0'
Biography and activity note	It records biographical details or activity notes about the entity	Tag 340 Indic.: ## \$a	No Repea.	
See reference tracing – personal name	The field records the variant form of a heading – personal name	Tag 400 Indic.1: #; Indic. 2: 0,1 \$a,\$b,\$c,\$f,\$g,\$8	No Repea.	In the Automatic generation of the record, the control subfield will have the value: \$8ita . Value to be completed
See reference tracing – corporate body name	The field records the variant form of a heading – corporate body name	Tag 410 Indic.1: 0,1 Indic.2: 0,1,2 \$a,\$b,\$c,\$d,\$e,\$f,\$g,\$h,\$8	No Repea.	In the Automatic generation of the record, the control subfield will have the value: \$8ita . Value to be completed
See also reference tracing – personal name	The field records the correlated form of a	Tag 500 Indic.1: #;	No	In the Automatic generation of the record, the control subfield

	heading – personal name	Indic. 2: 0,1 \$a,\$b,\$c,\$f,\$g,\$8	Repea.	will have the value: \$8ita . Value to be completed
See also reference tracing – corporate body name	The field records the correlated form of a heading – corporate body name	Tag 510 Indic.1: 0,1 Indic.2: 0,1,2 \$a,\$b,\$c,\$d,\$e,\$f,\$g,\$h,\$8	No Repea.	In the Automatic generation of the record, the control subfield will have the value: \$8ita . Value to be completed
Parallel heading – personal name	Parallel form of the heading in a language other than the language of the 2xx chosen according to RICA	Tag 700 Indic.1: #; Indic. 2: 0,1 \$a,\$b,\$c,\$f,\$g,\$8	No Repea.	At the moment it is not planned to have headings in alternative scripts, that demand the use of the subfield \$7
Parallel heading – corporate body name	Parallel form of the heading in a language other than the language of the 2xx chosen according to the RICA	Tag 710 Indic.1: 0,1 Indic.2: 0,1,2 \$a,\$b,\$c,\$d,\$e,\$f,\$g,\$h,\$8	No Repea.	At the moment it is not planned to have headings in alternative scripts, that demand the use of the subfield \$7
Originating source	The field identifies the agency responsible for the creation of the record and the date of the entry. Repeatable for transcribing, modifying or issuing agency	Tag 801 Indic.1: # Indic.2: 0,1,2,3 \$a,\$b,\$c,\$9	Yes Repea.	- 801 0 name of the agency originating the record - 801 3 name of the agency issuing the record \$9 used for the BNI number of the record where the heading was last used.
Source data found	Reference sources containing useful information about the heading. The first 810 field usually contains the citation of the bibliographic work for the cataloguing of where the heading was established.	Tag 810 Indic.: ## \$a,\$b,\$9	No Repea.	The first 810 will have a subfield \$9 containing the BID of the record in the SBN-BNCF database (or in the other databases) related to the bibliographic work for the cataloguing of where the heading was established..
Source data not found	Consulted reference sources in which no information about the heading was found	Tag 815 Indic.: ## \$a	No Not repea.	
Usage or scope information	Information for differentiating persons or bodies with similar names	Tag 820 Indic.: ## \$a	No Repea.	
Example under note	To highlight that the heading has been used as an example in another record	Tag 825 Indic.: ## \$a	No Repea.	
General cataloguer's note	Biographical, historical or other information about the heading	Tag 830 Indic.: ## \$a	No Repea.	
Deleted heading information	It records why the heading in 2xx is being deleted from the authority file	Tag 835 Indic.: ## \$a,\$b,\$d	No Repea.	
Replaced heading information	It records why the heading in the 2xx is being replaced by an authority file	Tag 836 Indic.: ## \$b,\$d	No Repea.	It appears in a record with value 'c' or 'n' in position 5 of the Label
Electronic location and access	It contains information available electronically about the entity for which the record was created.	Tag 856 Indic.1: 4 Indic.2: # \$a,\$g,\$u,\$z	No Repea.	It contains information required to locate an electronic item via HTTP
Data not converted from source format	It contains data (of the source format) for which there is no specific UNIMARC field	Tag 886 Indic.1: 0,1,2 Indic.2: # \$a,\$b,\$2	No Repea.	