

Modelling authority data for libraries, archives and museums: a project in progress at AFNOR

Françoise Bourdon
Bibliothèque nationale de France

The Bibliothèque nationale *authority files*, created in 1985, were the first in digital format in France. Designed at first as simple tools at the cataloguers' disposal for the management of access points to the catalogue, they were promptly adopted by other libraries that used them in their own catalogues. But since in the catalogue there are digital links, both among authority files and between bibliographic descriptions and authority files, users quickly realized that authority records are a valuable aid to define a research and a means to enrich the bibliographic information retrieved.

Now that great bibliographic databases are expected to provide easy, relevant access, not only libraries, but also archives and museums are interested in authority files as tools for managing and providing consistent access points and also as means to better place a search in context.

In 2000 French professionals were ready for a detailed check of authority data in order to establish which authority data they needed as well as to improve interoperability among the different cultural sectors: libraries, archives and museums. AFNOR (Association française de normalisation) offered an adequate environment for this consideration.

1. THE CREATION OF THE AFNOR WORKING-GROUP ON AUTHORITY DATA

Why was an AFNOR working-group on authority data created in 2000?

1.1. to give a national basis to the considerations developed at IFLA with FRANAR

In 1999 a working-group on authority data was created at IFLA, with the acronym FRANAR (Functional Requirements And Numbering of Authority Records). [see the dossier state of improvement in Glenn Patton's lecture]. I chaired FRANAR for three years (1999-2001), thus I deemed necessary to share the considerations developed at international level with some French experts.

1.2 to develop interoperability among libraries, archives and museums

The BnF authority files were issued on microfiches (in a display format observing GARE recommendations) from 1989 to 1998 and on CD-ROM since 1991 (permitting display and download in INTERMARC, and starting from 1999, in UNIMARC). Since its publication the CD-ROM has been at users' disposal both at the national library and at other institutions, and it has quickly got its public. From 1998 the authority records have been accessible on-line *via* the BN-OPALE Plus catalogue, in local and remote mode. The diffusion of the BnF authority files has played a central role in awakening French librarians' attention to the issue of authority control in public and academic libraries.

Since the early 1990s, the idea to develop a single catalogue for all types of documents for the future "*Très grande bibliothèque*" has requested the development of a new format, an "integrated" INTERMARC format for bibliographic and authority records. To bring this project to a satisfying conclusion a thorough check has been made of all the needs of every department of the national library, for printed and audio-visual documents and for special documents too (engravings, medals, ancient and

modern manuscripts, etc.), similar enough to documents preserved in archives and museums. This blend of ideas and needs led to the creation of new types of authority records and to new types of links among authority records and between these and bibliographic records. This was a first step towards taking into account the needs voiced in archives and museums.

Then it looked feasible to widen the application field of authority records. Managers of authority files devoted themselves to finding out what was done in other libraries and cultural sectors. As regards libraries, the guide of the AUTHOR European project (1995-1998) by BnF was a first attempt at achieving interoperability. The creation of the AFNOR Authority Metadata group in 2000 was another step.

1.3.why AFNOR is a privileged meeting ground for the three interested communities

Within AFNOR we have the General Commission 46 "Information and documentation", corresponding to ISO Technical Committee 46, responsible for the sector on rules for libraries, archives and museums. After several years CG46 is presently aware that we no longer needed stating particular rules for the library and documentation sector only but we must develop a multi-sector normative process to avoid any sector being left marginal. That' s why CG46 recommends adopting general, open rules.

Within CG46 we have the Commission of standardization 357 called "Modelling, production and access to documents" comprising the Group of experts "Authority metadata". AFNOR/CG46/CN357 is the French equivalent to subcommittee SC4 "Computerization in documentation" and, in part, to ISO/TC46/SC9 "Presentation, identification and description of documents". At the end of 1999, CN347 decided to defer, for at least three years, revising traditional cataloguing rules, fixing new priorities:

- introduction to the SGML/XML encoding of documents, proceeding to the translation into French of DTD EAD (Encoded Archival Description) to code finding aids in archives and libraries.
- definition of the descriptive metadata automatically retrievable starting from electronic versions of theses and academic works, on the basis of Dublin Core Elements.
- training in techniques of modelling and study of the existing models
- draft of an authority data model based on the international projects now in progress.

The Group of experts "Authority metadata" was created in June 2000 and I accepted to chair it.

1.4. the Group's objectives

The Group aims at examining in depth, from a multi-disciplinary perspective, the treatment of authority data to build up a dictionary of elements of these data. To accomplish such work one must control the authority data needed for managing databases of bibliographic, archival and museum data as well as their users, and assess the needs related to the management of intellectual copyrights.

The Group operates finding inspiration in the international activities in progress in the field of modelling within IFLA (works on FRANAR), ISO (works on CRM by CIDOC/ICOM) and in the European projects INDECS, INTERPARTY, and LEAF, on which some papers are presented at this conference.

1.5. the Group's composition

The Authority metadata group, that meets a full day a month, on average, is made up of about thirty participants of diverse origins :

- librarians and documentation experts representing: the Agence bibliographique de l' enseignement supérieur (ABES), the Bibliothèque nationale de France, the Bibliothèque universitaire de Nice, the École nationale supérieure des sciences de l' information et des bibliothèques (ENSSIB), the Bibliothèque de la Fondation nationale des sciences politiques, the Institut national des techniques de la documentation (INTD), the Médiathèque de la Cité des sciences et de l' industrie;
- some archivists representing the Direction des Archives de France and the Centre historique des Archives nationales;
- some members responsible for the museum databases of the Direction des musées de France, the Département des Estampes of the BnF, the Bibliography of the history of art ;
- some experts in audio-visual materials representing the Institut national de l' audiovisuel (INA), and, more precisely, its research and development department and its Département Inathèque;
- some managers of intellectual property rights for the Société des auteurs et compositeurs dramatiques.

This well diversified composition is a real ace in the sleeve for the Group that aims at implementing an assessment of authority data on a multi-disciplinary level.

2. A THREE-PHASE WORK PROJECT

The work methodology for these ambitious aims is based on alternating what we might call "reports from experience" and some "theoretical considerations".

2.1. Acquiring a common culture

Within the Group the "reports from experience" are an aid to acquire a common culture about authority data. What are the practices in the different sectors of activity represented in the Group? What have we in common? What are each sector' s specific features? In order to identify convergence and divergence points, each participant has been questioned on various themes like:

- *authority files* as tools for managing library catalogues: the BN-OPALE Plus authority files <www.bnf.fr>, the university system for documentation authority files <<http://www.sudoc.abes.fr>>, the role played by authority files in the management of OCLC, particularly in the CORC program (Cooperative Online Resource Catalog) <www.oclc.org> in progress at the time; have been presented;
- the vocabularies and authority lists in museum and iconographical databases were the subject of various papers on practices applied in the JOCONDE documentary database <<http://www.culture.fr/documentation/joconde/pres.htm>>, operated by the Ministère de la Culture et de la Communication, on the functioning and work organization on the thesaurus and authority files of the Bibliography of the history of art (BHA), on the management of the BnF Département des estampes et de la photographie (BN-OPALINE/Estampes base) iconographical database <www.bnf.fr>;
- the reference sources of the databases of the Institut national de l' audiovisuel <<http://www.ina.fr/inatheque>>;
- the role of authority data in archival description, presentation of ISAAR(CPF) and ISAD(G) and demonstration of the digital images ARCHIM databank <<http://www.culture.fr/caran/archim/>>
- authority data have been examined from two perspectives: the principles of thesaurus creation and the index languages as aids to help multi-language subject searching thanks to a presentation of the European project MACS (Multilingual ACces to Subjects) <<http://infolab.kub.nl/prj/macs/>>

- authority data useful for managing intellectual copyrights have been described by the representatives of the Société des auteurs et compositeurs dramatiques(SACD)) <<http://www.sacd.fr>>
- the presentation of the conceptual models of data, not confined to authority data, but bringing in a management of information helping the Group get familiar with the particular modelling technique:
 - * the FRANAR model in course of definition (see Glenn Patton' s paper)
 - * the *Functional Requirements for Bibliographic Records*, "entity/relation" model elaborated between 1992 and 1997 by IFLA, better known as FRBR;
 - * The conceptual reference model elaborated by the Groupe de normalisation documentaire du Comité international pour la documentation du Conseil international des musées (ICOM-CIDOC), better known as CRM (Conceptual Reference Model) <<http://cidoc.ics.forth.gr/>>. CRM is an "object-oriented" model approved at the end of 2002 as ISO/CD 21127 under the title "A reference ontology for the exchange of data related to cultural heritage".

All these papers permitted us to know the objectives, the practices and the tools applied in the various fields of activity represented in the Group. Each one raised real interest and led to lively debates on the management praxis and on the terminology employed: referential, thesaurus, vocabulary, authority list, authority file, etc. This training about what others are doing is an investment that will yield fruit when the Group goes on from "reports from experience" to "theoretical considerations".

2.2. defining the elements of authority data

The fact that the group of experts spends half of its monthly meeting to get informed about what is going on in the field of authority data on a national and international level does not mean it is a permanent training course ! The Group works in this way to establish a vocabulary of authority data that enumerates, denominates and defines each one of the information elements that contribute to identify an entity, and may eventually specify the sources to be used. An unambiguous definition must be proposed for each term and agreement must be always sought among the representatives of the various sectors of activity (libraries, archives, museums, managers of copyrights).

The Group decided to devote itself, first of all, to defining the main entities met in the "reports from experience" sessions, that is Body, Person, Place, Medium, Material, Work, Technique, Object, Concept, Group, Temporary manifestation and Event. The identification card of each entity is compiled and the file includes the following sections: definition of entity, specification of what is not entity or of what must be excluded from the application field of entity, some examples, a list of unsolved questions, the reference documents used as definition sources and the text of definitions found in these diverse reference documents

The main sources used to compile the definitions of entities are the ones found in:

- ISO 5127: Oct. 2001 Information and documentation – Vocabulary
- ISAD(G): general and international rule for archival description <<http://www.ica.org/>>
- ISAAR(CPF) : International rule on archival authority records for bodies, persons and families <<http://www.ica.org/>>
- FRBR
- FRANAR
- The Dublin Core Element Set (ISO/DIS 15836 being voted at present)
- CRM <<http://cidoc.ics.forth.gr/>>
- Getty vocabulary and rules, <http://www.getty.edu/research/institute/standards>
- AFNOR cataloguing rules

- INTERMARC, MARC21 and UNIMARC formats (rather poor in their definitions!)

Just as an example, the cards for identifying PLACE, MEDIUM and MATERIAL entities are included (see Attachment). The definitions found in the reference documents cited below are not reproduced but these cards give an idea of the issues faced by the working group.

Debates are often lively! The more so since, collateral to this *stricto sensu* definition work, the Group tries to put at stake all data elements it wants to treat and meets serious difficulty in rendering all the wealth of information to be considered. The Group has tried many ways:

- to state immediately a hierarchy of identifying elements: e.g. for the PERSON identity we can subdivide some "biographical data" into "affiliation", "filiation", "artistic influences", etc.; affiliation, in turn, can be subdivided into affiliation "to a family", "to a corporate body", "to a school of thought", etc.
- to limit itself to an alphabetical list of identifying elements: "artistic influences", "initials", "inventions", "jingle", etc.
- to cross the identifying data elements and the above defined entities in order to point out the elements common to more than one entity

None of these methods resulted satisfying ... so the Group went on to the third phase in its program: to get familiar with modelling techniques.

2.3. Organizing data elements : modelling

Modelling is the last phase, the one in which entities are identified as well as their attributes and we can form the network of relationships among data elements.

Some prerequisites are imperative that require training the Group in modelling. This training must be progressive, even reiterated. Modelling techniques are usually addressed to informatics students, so the main problem was finding a trainer capable of understanding the information specialists' preoccupations and to get within their grasp.

A first awareness course was proposed to the Group in October 2001 by a doctor in engineering from the Centre national de la recherche scientifique. The aim of this first half-day course was to measure the contributions of a modelling method in the development of a project, to understand how analysis methods evolved, to discover the different approaches in the field of modelling and to master the principles of object oriented modelling. Since it was presented in this course as an aid for modelling and structuring ideas, as a help for reasoning and simulations and as a means for communication among different persons, the Group became quickly convinced they really had to make this effort.

The second phase of training took place in November 2002, for a whole day, during a Seminar on the start up of data standardization in UML [Unified Modeling Language], run, specifically for the AFNOR Group, by a professor from Institut national des sciences appliquées de Lyon (INSA-Lyon). The basic notions dealing with models and practical modelling were mentioned but the essential part of the presentation was based on UML diagrams and mainly on structure diagrams and the concepts of "class", "heritage", "properties" and on diagrams of usage cases. This second phase resulted more arid for the participants yet it should help the Group to choose a tool to give form to the results of its considerations. An assessment of this seminar takes place at the next meeting in January 2003.

CONCLUSION

At the end of 2002 the Group is at a turning point in its existence. Sharing experiences on the treatment of authorities in various cultural sectors has been definitely fruitful, an asset for the future. Librarians have come to know ISAAR(CPF) and CRM ... Archivists and professionals from museums have come

to know FRANAR and learned a lot on normative tools at librarians' disposal. Each one was made richer by the other one' s point of view, and has acquired new knowledge on modelling. Then, are we ready to meet our commitment and propose a common conceptual model? The Group must examine again its primary objective that meant proposing a general model for authority data in order to allow interoperability among libraries, museums and archives. It must measure its actual ability to fulfil this task within a reasonable lapse of time, eventually curtail its aspirations identifying some more manageable subgroups and choose the type of model to develop: entity/relation model, object oriented model, or semantic model ? The Group next meetings will be decisive ones.

ATTACHMENT

- ENTITY n°3 : **PLACE**

- Definition:

Real or fictitious part of space

- Examples:

- place where Nelson died: it is on a ship, but what are the precise coordinates of the place? On the ship, but somewhere on the Ocean. (CRM)
- place of abode of a person, a family. (Archives-ISAAR (CPF))
- seat: place where the body acts. (Archives-ISAAR-(CPF))
- places of existence: place of foundation, of expiration, of birth, of death. (Archives-ISAAR(CPF))
- location of a municipality, a city, a congress. (FRBR, Getty Thesaurus of geographical names)
- geographical attribute: continent, rivers, mountains (Getty Thesaurus of geographical names)
- oil platform (example made by the Group of experts)
- the 140 altitude (example made by the Group of experts)
- the 52nd parallel (example made by the Group of experts)
- the Atlantis(example made by the Group of experts)
- the isle of Avalon (king Arthur cycle) (example made by the Group of experts)

- Comments

This part of space can be taken into consideration in order to locate another entity or in itself.

- CRM differentiates "place" as mere topographic indication and "site", a place whose boundaries are vague and that can be represented on an iconographical object. Place is used to qualify something else (an event, a body, etc.). Site can be defined in relation to a place.
- ISAAR-CPF specifies that it is necessary "to qualify the place name with a word or an expression specifying its context and use".
- The Thesaurus of geographical names (Getty) defines "place" but the Guide for the description of architectural drawings (Getty) defines "location".
- FRBR similarly differentiates "place" and "location"
- corporate bodies are excluded (e.g. the Louvre Museum as a corporate body)
- the horizon is not a place

- Unsolved issues:

- confusion, sometimes, between a place and a building, or a body, e.g. n° 10 Downing Street. How can we distinguish them? It depends on points of view ...
- difficulty in differentiating body and place: e.g. "Paris"; the "Louvre" (building, place, body); the Picasso Museum at Antibes is located in the Grimaldi castle.
-

ENTITY n° 4 : **MEDIUM**

Definition:

Concrete element designed to contain written, sound, electromagnetic, numeric, graphic information.

Examples:

- audio-cassette, video-disc, microfilm, slide (FRBR)
- archival medium, recording medium, magnetic medium, optical medium, physical medium, primary or secondary medium, chemical photosensitive medium (Vocabulaire de la documentation)
- papyrus, paper, metal, parchment, fabric, wood, film, magnetic tape (Archives)
- canvas
- parchment (see also material)
- movie film (FRBR)

Remarks:

- CRM: the concept of medium is not dealt with by CRM, perhaps because the English term for "medium" is "material" is therefore the concept dealt with under that class? Also the English term "format" designates the medium.
- Archives: medium is defined as material but there is no definition for material.
- FRBR: the concept of "medium" implies the concept of "dimension". "medium" and "material" are not differentiated.
- The Vocabulaire de la documentation: associates medium and process (a technique, then?)

Unsolved issues:

- to create a Physical features entity including "medium" and "material"?

ENTITY n°5 : **MATERIAL**

Definition:

Material an object is made of or part of its components.

Examples:

- ink
- microfilm (medium) silver or diazo (material)
- parchment (medium and material): calf or goat
- painting, oil
- plastic (Vocabulaire de la documentation)
- glass (Vocabulaire de la documentation)

Remarks:

- except the case of sculpture in which medium and material cannot be distinguished, the material is what we put on the medium. Concept of "injective" and "subjective" (!)
- all media need a material, conversely we may have a material without medium (e.g. a sculpture)
- Archives: medium is defined as material but there is no definition for material.
- Vocabulaire de la documentation: it uses the term "substratum"

Unsolved issues:

- to create a Physical features entity including "medium" and "material"?