The other half of cataloguing: new models and perspectives for the control of authors and works

Alberto Petrucciani
University of Pisa

The title of this paper recalls the one of an article published exactly 30 years ago (1973) in *Library resources & technical services* that aroused considerable interest at the time; in it two librarians from Florida analysed, with great clearness and wealth of examples, the complex set of activities required for a correct and functional integration of the new cards into the catalogue. It was the time of the card catalogue but also of the spread of OCLC automated services essentially based on the Library of Congress MARC records. In fact, the described procedures included the retrieval of the Library of Congress MARC records on the OCLC terminal, their control and adaptation, for the following arrangement of the printed cards to be inserted by hand. The authors wrote: «Catalog building consists of two phases: (1) the creation of cataloging copy representing the works being added to the collection; and (2) the integration of that copy into the existing catalog. Phase One simply involves the creation of a record; Phase Two determines whether or not the reader will be able to retrieve that record once it has been dropped below the rod among a million others»

This warning should be rewritten today in slightly different terms, but its relevance is unaltered.

More exactly, today's electronic catalogue makes retrieval of specific records very simple and quick in most (but not all) cases. We may use many different elements, combined or truncated too, so that, practically, difficulties in this phase arise in two cases: when the information used in a search is not exact or when the available clues are not sufficiently distinguishing (that is, they are all elements very frequently occurring in the database). But what is easy is *only* the retrieval of one or more records including certain elements, that, in a card catalogue, would have implied very long, tiring, or even not feasible, searches. On the contrary, searches aimed at the full, reliable retrieval of material answering a well defined research need (author, work, theme, form, etc.) are still long and tiring, and sometimes not feasible – particularly in ever richer and crowded bibliographic databases. Therefore the activity usually included under the label of *authority control* or *control of access points* is essential, or even, in many ways, more and more important.

In spite of its great relevance, both for the catalogues of the past and for today's ones, authority control has been and still is the "poor relative" of cataloguing, the often neglected or overlooked "other half" if we compare it to the more striking one: the creation of bibliographic records by document description.

On the one hand the study on *Functional requirements for bibliographic records* (FRBR), on the other hand the new versions of authority control standards, *Guidelines for authority records and references* (GARR) and *UNIMARC manual, Authorities format*, with the related documents, are important steps towards a thorough consideration of access points, a not merely instrumental and managerial consideration but one stimulating us to think about future perspectives.

The analysis model for *entities* and *relationships*, cannot be regarded as new, because it has already been applied widely not only on a theoretical level but also for links among bibliographic records in large databases (e.g. on a particularly ample scale in the Italian National Bibliographic Service). Yet, the relational model has not had a deep impact on author cataloguing, on the contrary, the logical schemes we are accustomed to, the traditional ones of twentieth century codes, are flat, based on a stiff, monodimensional approach. An example: in the present cataloguing rules – from RICA (§ 17) to AACR2 (§ 21.9 and following) the treatment of works based on other works is still seen as the addition of another access point to the same bibliographic record (the main entry of the original work, used as an added entry). An intricate as well as unsatisfactory solution, because the user searching under access points other than the main entry of the original work, will not get any information about the related work.

Not making clear the differentiation between relationships among works, irrespective of specific publications, and the access points related to elements in the publication has its equivalent in the lack of actual development of a catalogue with "two-dimensions", the one of access points to the documents' records and the one of the network of relationships among records of entities that are not publications (works, expressions). Analogically, this lack of a clear differentiation, reminds us the "prehistory" of alphabetic subject cataloguing, a time when it was not yet clear and accepted – even obvious as it was later–, that references link subjects, not books; it follows that access points to specific documents, though multiplied and manipulated, cannot perform the linking function among subjects. In subject indexing, then, the basic feature of the distinction between paradigmatic or *a priori* relationships, independent from specific documents, and *a posteriori* ones, implying more concepts being present in one subject (distinct from the still diverse case, of more subjects being present in the same document), is universally accepted. I believe that in this phase, when we are basically reassessing the catalogue structures, and developing logic, simple, powerful searching models to help us control extremely rich archives, the continuous comparison between the author and title cataloguing tradition and modern semantic indexing is really helpful and necessary. These two traditions have lived almost completely isolated one from the other, developing fully disconnected conceptual tools and even vocabularies, while, at the same time, the physically separate catalogues of the past were becoming an electronic catalogue that is more and more converging with bibliographic databases.

Anyhow, also the author and title cataloguing tradition offers various hints for a new assessment, provided we look from an "estranged" viewpoint not taking the traditional methods for granted. For example, we take for granted, at the same time, that a bibliographic record must include proper headings under persons and corporate bodies, represented by *only one* form of their name (not by all the variants).

Even better, represented by automatic link to a single authority record, so that the form of name can be easily controlled and, if needed, changed. But we take equally granted an opposite treatment for titles (proper title, original title, uniform title, etc.), and we consider them access points to a single record instead of different forms to refer to the same, single entity. It certainly isn't the same situation, but anyway we have here situations similar enough to arise at least a doubt about the advisability of treating them in the same way, instead of an opposite way.

Following FRBR we might take this analogy much further. As we consider obvious that the issue of the form of heading for Dante Alighieri is not an issue to be dealt with in the description of a single publication but at the authority file level, couldn't it be equally obvious – to those who come after us – that the issue whether or not Dante is the author of the *Divina commedia* or of the *Fiore* has equally nothing to do with the compilation of the bibliographic record, but must be dealt with in an archive treating the relationships between *authors* and *works* (not *publications*)? For the document description the only pertinent element is whether or not it is an edition of the *Divina commedia*, or of the *Fiore*. We would consider primitive or amateur any person who might consider enumerating on the same level *Dante* and *Alighieri, Dante* among the headings of a bibliographic record, those who come after us might give the same judgement on our praxis of repeating the original title and the names of the two authors of, let us say, the *Manifesto of the communist party* in Italian in the record for each single edition. Three elements among which, obvious *a priori*, constant relationships exist, independent from the single document, an indexer would say. It looks as if the control of authors and works might usefully be distinguished, much more than it presently is, from the catalographic treatment of publications and developed autonomously with the necessary – essential – relationships with bibliographic records.

If these are, as I believe, the incentives and trend indications that FRBR and related works arouse, what relationships can we perceive in them with the most relevant features of the recent evolution of standards for authority control? I think the most innovative elements in the standards for authority files are two:

- the development of authority records in order to include *information about the entity* itself (e.g. for a person: sex, nationality, language, dates of birth and death, bio-bibliographical data), together with elements of a more technical and housekeeping relevance (institution creating the record, adopted cataloguing rules, used sources),
- the tendency to consider equivalent the original forms of names and the translated or adapted forms, preferring the latter ones – when they exist – in the various linguistic and national contexts.

This second trend implies, I think, both immediate and future drawbacks on which we should reflect more, but upon which I cannot touch now. I deem the first one, instead, very interesting and stimulating, although not free from risks. It is easy but perhaps necessary to reassert that basically libraries have responsibilities of a bibliographic nature, related to the control and availability of publications, while, for example, the census of persons and corporate bodies, except for their being authors of publications and within the limits of elements functional to a bibliographic search, is not one of their specific tasks. Today, authority control is carried on in a deeply changed context, in which other considerations are against the above reassertion. On the one hand, activities for access point control in more and more comprehensive and wide cooperative databases involve an enormous amount of work and imply vast information and competence: considering this investment only as a "housekeeping" function, strictly confined to cataloguing technicalities, seems reductive, while its potential for a wider information function is clear.

We must not forget that, while the traditional library catalogue covered a modest fraction of the bibliographic universe and usually played its role only for on-site users, today's great bibliographic databases are even bigger than the major bibliographic reference works of the past and tend to surpass them also in exhaustivity. At the same time they are within reach of a very vast on-site and distant public, much more so than the large reference (or biographic, encyclopaedic, etc,) works actually to be perused – in most cases – only in the reference rooms of great libraries. Even more, today's large databases are within reach of a public not limited to the librarys' users and, with greater reasons, to strictly bibliographic search objectives.

Therefore the wider information objectives that authority files may take on must be carefully – bravely too – defined and assessed, starting from the realization that these objectives exceed the proper and exclusive field of libraries – the one of bibliographic control – and must therefore be pursued in cooperation with a wider range of partners, starting with cultural and research institutions. Libraries can bring into these new forms of cooperation a considerable wealth of competence and resources, a wealth of concepts and methods developed in their long cataloguing and indexing experience and applied in huge, structurally complex information databases.

In order to plan effective, wide-ranging information initiatives, shared with other partners, the clearing prompted by the entities/relationships analysis is greatly valuable; it shows how to avoid duplicating information, and conversely, how to share it, for different applications too, provided it had been isolated (i.e., from our viewpoint, separated from the bibliographic records).

Very interesting incentives come, for example, from comparison with the field of archives that shows us a formulation opposite, under certain aspects, to the one typical in a library environment. In the archival field, description of documentary material is often brief and "light", while information on creators, in other words, according to FRBR terms, the one related to entities of the second group rather then to the ones of the first group, tends to be rich and elaborate. The information function of this "other half" of archival treatment (for the history of institutions, families, etc.) is not marginal: on the contrary, it is obviously a primary, basic function.

I do not want in the least to sustain that these differences reflect simple habits, indeed I am convinced that basic functional differences exist (and persist) between archives and libraries. But, in spite of differences, it is useful to look carefully at analogies and stimuli aroused by these same differences, not in order to homologate or level, but to positively cross-fertilize diverse methodologies and experiences.

In order to try and outline possible future perspectives it is always useful to look back, to measure and assess the road already covered. The last quarter of the 20th century, from the cataloguing point of view, was characterized by the standardization of bibliographic records, particularly in their descriptive elements, with the development of MARC formats and ISBDs. The modern, better organized and more structured bibliographic record has become the basic element, the *building block*, for the application of computer systems in libraries and then, with the development of cooperation and webs, for today's huge bibliographic databases.

These databases are basically made up of proper bibliographic records, records of the features of specific documents, while other elements, if not absent, are definitely on a secondary level.

On the contrary, the new signals of interest and attention that we perceive in recent years urge us to look "upstream" and "downstream" of the bibliographic record in the strictest sense: on the one hand, starting from the already classical authority control functions, towards the development of information systems and relationship networks related to entities which are not strictly bibliographic (works and authors, organizations, events, concepts, etc.), on the other hand, towards the enrichment of the traditional function of document description and identification with further information on contents or digital reproductions of documents parts, down to the access to the full text. The bibliographic record, then, might be relieved of information not in its province (being related to entities different from publications) and of all the functions more suitably worked out upstream or downstream (in access systems or, when needed, but links to the images and/or the texts of the publications themselves).

A "light" bibliographic record would no longer be, as it is the case today, the paramount component of library information systems, instead, it would keep its central role – probably permanently linked to the library function – rather as "joint", nimble, swift passage or interconnection point between intelligent systems for access and content organization and intelligent systems for remote use of digital resources.