

For Festschrift International Librarianship – Today and Tomorrow (for the 65<sup>th</sup> birthday of Prof. Dr. Dr. h.c. Elmar Mittler)

K.G.Saur Verlag GmbH

Submitted November 2004

## **Change Cataloging, but Don't Throw the Baby Out with the Bath Water!**

By Dr. Barbara B. Tillett<sup>1</sup>

*Abstract: We will do cataloging differently in the future while retaining the best of basic cataloging principles and the benefits of authority control. Our tools not only will improve future catalogs but also information seeking systems of tomorrow's world.*

With metadata and Google and the Internet, increasingly we are hearing there is no need for libraries anymore; certainly no need for cataloging or cataloging rules or library catalogs or the MARC format. So, what's wrong with that point of view?

### We Don't Need Library Catalogs

User studies of the 1950's and 1960's told us about the information seeking behaviors of scientists and others<sup>2</sup>. People will go to the nearest at hand source to answer their information questions. That may be the tool at their desk (nowadays their PC with Internet connection) or if not found there, they'd try the person next to them – at work or at home or perhaps even call someone (or email them) for help. Way down the list would be going to a library or asking a reference librarian for help. So it's very natural that we'd expect people, especially the younger generation that has grown up with computers and Internet access to digital resources, to use an online tool first. Does that mean we don't need library catalogs or libraries anymore? No, but the tools we provide and the way we enable people to get the information they are looking for will change.

We can do better than today's library catalogs - the tool we use to enable users to discover what we have selected for them in our libraries. Panizzi persuasively and successfully defended a "full and accurate catalogue" when challenged by the Trustees of the British Museum in the 1840's.<sup>3</sup> He made the case that users would be better served with more than a finding list or inventory that listed the items in the library's collections. The collocation of works and listing all the materials by an author were worthy objectives for a library catalog that were also advocated by Charles Ammi Cutter in 1876 on the other side of the Atlantic Ocean<sup>4</sup>. Enabling a user to find known items or all resources on a given topic or by a given author, gathered together and arranged in a library catalog were among Cutter's "objects" of the catalog.

We also know from user studies such as those of the 1970's and 1980's and later that known item searching and searching by a subject are equally important and need to be provided in our online systems. We also know from researchers like Marcia Bates, Karen Drabentstott, and others that users do not usually know the controlled vocabularies used in

any particular field of study and would benefit by our leading them to the controlled terms.<sup>5</sup>

When we first moved to online public access catalogs (OPACs), we gained keyword searching at the expense of many other capabilities that card catalogs gave us – in particular the loss of non-roman script transcription in our bibliographic records. Most systems still today do not display or enable a user to search by every script in a library's collection, but with Unicode more systems should be providing that capability in the near future. We also lost a lot of the collocation of works, with many systems still today not grouping the author/title combinations as they should – just ask your music librarian friends. We can do much better and should be designing our future systems to build on the objectives of catalogs while adding new capabilities to guide users to the controlled terminology and related resources and eventually to the digital versions of the materials they are looking for. Systems like Endeca for guided searching and clustered displays have the capabilities that future systems should embody.

These future systems need to move beyond just what we traditionally included in library catalogs. We should be aiming for broader integrated systems – at the corporate level or beyond to assure the efficient re-use of bibliographic information for a wider variety of library operations and end-user tasks. At the very least we should connect our bibliographic and authority databases with the abstracting and indexing services to search and retrieve articles in journals and conference proceedings, but even go beyond that to connect with resources in biographical dictionaries, publisher and bookseller databases, and specialized databases for specific subjects and audiences beyond our own library collections. We need systems that go beyond even the FRBR<sup>6</sup> user tasks to find, identify, select, and obtain – to also meet the tasks of librarians as users and publishers as users, and copyright/right management agencies as users. We can and should do this.

We have the technology for combing the best of the relevance ranked retrieval systems (search engines like Google and Yahoo), with the guided search engines and clustered searching capabilities with user-centered features, like enabling the use of the bibliographic information to produce a bibliographic citation in the user's format of choice (as is done now with RedLightGreen) – those should be standard features of our future systems. Combine that again with Unicode so a user anywhere in the world can specify their preferred language and script (that they can read) and automatic transliteration to whatever scheme the user wants, and we take another leap to a better world.

#### We Don't Need the MARC Format

When we first moved to the MARC format in the late 1960's, the creators had great foresight to clearly label the elements of bibliographic description and access as well as adding some extra information to make it easier to process and communicate the records. But we can do better. The rage now is XML, and certainly that will evolve in the future to the next generation. For now, we can move through this transition period by turning the MARC records into MARC XML records so they can be harvested with Open

Archive Initiative protocols (OAI) or SRU/SRW (the next generation after Z39.50) and be found by the major search engines.

We do need something like the MARC format in the future to clearly label the elements of our information packages that identify the resources we wish to describe and provide controlled and uncontrolled access points at a “level of granularity” that makes them as flexible as possible for creative re-use in the future. We want to be able to repackage the information we now provide for different purposes and offer different displays. That is only possible, if we clearly label what we provide. The MARC format does that now, but can always be improved upon.

People now building digital libraries quickly realize the importance of following standards and consistently including a basic set of elements to enable optimal retrieval of relevant material. They see the lowest common denominator driving the capabilities of national digital library projects and limiting the effectiveness of the systems being built today.<sup>7</sup> Deciding on the use of a particular metadata scheme, like Dublin Core, is not enough. The collaborating institutions in creating digital libraries and future systems need to agree on basic standards beyond just the metadata scheme. In the National Science Digital Library project, some of the institutions use only a few elements and without agreed standards some mis-use the elements creating chaos when a system tries to search the resulting data. The metadata schema needs to be rich enough to enable a machine to recognize the components and to index and present them in a useful way.

#### We Don't Need Cataloging Rules

In 1674, Sir Thomas Hyde wrote in his preface to the Bodleian Library catalog that inexperienced people who make indexes for their private collections of books think it only takes writing down the titles from the title pages. So why should we need cataloging rules? He pointed out that to create an alphabetical catalog for a large collection of a multitude of books from all over the world brings up “intricate and difficult problems that torture the mind.”<sup>8</sup> Panizzi described similar misunderstandings on the part of those inexperienced in cataloging. Once you have a collection of over say 2,000 items, a human being can no longer remember every item and needs a system to help find things. When you add very large collections of many types of material in many languages and scripts, there must be some sort of organization or there will be chaos. We can do better than our current cataloging rules and rule interpretations, but rules are definitely needed. Just as we are seeing with the building of digital library systems, having easy to follow rules based on sound principles for knowledge organization, and keeping the user's needs first is very important to creating a truly useful database to describe and access information resources.

The current move to revise the Anglo-American Cataloguing Rules (AACR) reflects an ongoing need to improve the rules and to make them principle-based and easy to understand and follow. The plans for the next edition of AACR are to explain the principles behind the rules and the concepts of bibliographic control and authority control, then to give clear, easy to follow general rules that apply to the description of all

types of material, followed by rules of special types of content and special physical media. Then the rules would move on to choosing access points to enable a user to find that description and then rules about controlled forms of names given to persons, corporate bodies, and names of works/expressions (in the FRBR terminology). It is hoped this new set of rules would be more universally applicable – worldwide as well as by a wide variety of people that would use it as the content standard for any metadata schema they chose to employ.

We have too many exceptions that have evolved over the years, once again creating the case law approach that Andrew Osborn warned about in the early 1940's in his "Crisis in Cataloguing" article<sup>9</sup>. We need to step back to see where consistency is important, to assure the provision of basic elements of description and access (even Panizzi identified the basic elements for his descriptions and the ISBDs defined the elements and the order of those elements basic to any bibliographic description). We need to keep those basics, and get rid of the special case law for every situation. We need to eliminate special streams for cataloging certain kinds of materials to keep workflows simple. Catalogers should be able to know when something is important to bring out and what is not needed for their target user group. Training catalogers to know what is needed to meet the basic user tasks is essential. We cannot just have the inexperienced person giving information they think might be important. Without guidance and rules the results will run the gamut from extreme detail to nothing more than a keyword or two and the resulting retrieval systems will be working with the lowest common denominator resulting in unhappy users. Panizzi also wrote about this result in the 1840's, but it still holds today.

The current use of AACR2 with Library of Congress Rule Interpretations is the result of going to the other extreme to try to address the questions of how to apply the rules for the approximately 300 catalogers at the Library of Congress and the hundreds of other catalogers in libraries that follow the same practices in cooperative programs and bibliographic utilities. Keeping the goal of consistency at the forefront, these rule interpretations give catalogers who desire specific instructions for every situation, the guidance they want. We would benefit by developing the next rule interpretations to help the cataloger understand which elements are important to provide in a strictly consistent way (in order to consistently identify the resource and help find it to begin with) and which should just faithfully and accurately describe the resource that is the object of the cataloging record (to help with the other user tasks of selecting and obtaining).

The current cataloging rules are based on principles that are now being reviewed and re-evaluated. The Paris Principles of 1961 were focused on choice and form of entry, but today we have the potential to allow more flexible information packages than the static catalog card, to enable us to achieve the collocation of literary units (Seymour Lubetzky<sup>10</sup> – author/title uniform titles for works and expressions) as well as the bibliographic units (Eve Verona<sup>11</sup> – title proper of the manifestation). There is no need to argue about main entries in those terms, as access can be given and citations formed, as long as we identify the primary creator of the work and clearly indicate the relationship between other persons or corporate bodies and the work, expression, manifestation, and item being described. Unfortunately following an administrative decision to cut cataloging costs in

the 1980's, the use of relator terms and codes was dropped except for special situations of illustrators of children's books and some special uses for musicians and legal defendants, etc. That simple coding enables clear relationships to be defined and assists in collocating displays and filtering searches. We need to re-evaluate that decision to take a longer view of the value of such information to the information packages we provide.

We might even take more advantage of current MARC structures to move more to using authority records for serials and other works/expressions. There we could put the subject cataloging information – once, instead of redundantly with each manifestation of that work. The subject headings and classification numbers that fit the work/expression (the content) would then be inherited by the linked bibliographic records for the various manifestations of that work/expression. We can do it better.

Users are looking for articles, not the serial titles as such, so it would be far better to spend time using an authority record for the serial linked to our inventory control (holdings) records and acquisitions/subscription/order records. Current serials cataloging is being done for other catalogers and should be re-evaluated from an end-user's perspective.

#### We Don't Need Authority Control

We had over a decade of debate about whether or not authority control was necessary and finally decided it is necessary and we should get on with it.

Studies like the Cranfield Projects<sup>12</sup> (Cyril Cleverdon's studies) looked at precision and recall, demonstrating that the two are inversely related – greater recall means poorer precision and greater precision means poorer recall – high recall being the ability to retrieve everything that relates to a search request from the database searched, while precision is retrieving only those relevant to a user. With authority control, where one has the linking of synonyms or variant forms of names that are linked to an authorized form (or one chosen as a default display form) we get greater precision. This saves the user's time and effort first in trying to think up all the variant forms to search (as with spelling variations – British versus US spellings) and secondly avoids having to sort through retrieved records with what at first glance may appear to be the same name or term but in fact may be a different entity or topic (such as homonyms or undifferentiated names).

Keywords give us limited recall and terrible precision. Yet we are often told these days that users don't care, they would just as soon have a keyword ability, even though they realize it's not retrieving all it could. So why not combine that keyword with more guided searching based on the controlled vocabularies (personal names, corporate names, uniform titles, and subject terms) and increase their precision of searching with little or no increase in their effort?

Also, we don't have to decide on one term being the authorized term, if we build clusters of the variant forms (in our authority records for now). We can let the user decide which

form to display or use in their use of that bibliographic information (as for creating a bibliography in a certain citation style), reserving one as the default display form. That can even extend, as has been suggested in IFLA descriptions of the “Virtual International Authority File” concept, to allowing a user to select the language and script preferred for the names and subject terms they search by and see displayed in the retrieved search results. By linking the authority files of the world, one could envision a switching capability that would allow a user in China to see the Chinese form of the name for Confucius, while a user in the United States could see “Confucius” as the well-known English form of that philosopher’s name.<sup>13</sup> We can do it better than we do it now.

Authority control or the clustering of the related variant forms of names or subject also enable us to fulfill the collocation function of a catalog – to bring together all the works/expressions of an author, the performances of an artist, the motion pictures of a director, the prints of a photographer, and so on. The syndetic structure of references also leads the user from variant forms of names and terms to those used in the controlled vocabularies. This can be augmented as noted before through guided search engines or clustered search systems that can present the user with terms from controlled vocabularies that are connected to the keyword they entered.

We also may find, if our integrated system automatically build basic authority records for us and automatically check the local authority file, that we don’t need to do as much authority work as now, particularly for uniform titles of works and expressions. OCLC studies have shown us that less than 20% of their WorldCat database represents works with more than a single manifestation – over 80% have only one work and one manifestation – no need to create a uniform title authority record – or one could just be created automatically as a placeholder until we get another manifestation.<sup>14</sup> We could have our systems alert us when they detect a second or third repeat of the same title or author/title combination. This would augment what the cataloger may know about related works/expressions – the human factor is still really important as the machines cannot always recognize the related works under different forms of title or in different languages.

### Future Systems

We can do better with our library catalogs and integrated library systems by taking a broader view of those tools that were the library catalogs and combining them with databases from publishers, abstracting and indexing services, biographical dictionaries and other reference tools, and links with specialized databases for specific subjects and audiences worldwide. To do that we need to be sure our bibliographic and authority information packages are clearly labeled to an appropriate level of granularity to assure flexibility in repackaging that information – for collocation of works and expressions, for switching the display forms of names and subject terms, and for repackaging for citations or records for other systems worldwide that follow the same or different cataloging practices. Our future systems need to help with automatic tagging of data, but our catalogers also need to understand the tagging and the indexing capabilities in order to know the important elements to include for optimal retrieval of the records they create.

Systems should automatically create base authority records for us from information in the bibliographic descriptions, letting the cataloger augment that basic record as needed. This also takes a trained person to recognize when an entity is the same of different from others already established.

Technology can help us and librarians have always made full use of the technology available at the time- whether it was the move from clay tablets to book catalogs, or on to card catalogs where we could easily re-use the cataloging work of others to the even greater capabilities of sharing machine-readable bibliographic and authority records. We can always do better and our systems will continue to evolve and get better with the skill of librarians and catalogers contributing to the mix of ever more sophisticated tools to organize information and search and retrieve information important to users. Libraries need to work closely with the developers of our future systems, to share our wealth of knowledge about organizing information. We have a lot to offer such collaborative endeavors. We know a lot about providing bibliographic control and authority control, but we can always do better, assisted by tomorrow's systems. Our knowledge, our principles, and basic cataloging rules and concepts will still provide the foundation for tomorrow's systems that ever better meet user needs.

We will do cataloging differently in the future while retaining the best of basic cataloging principles and the benefits of authority control. Our tools not only will improve future catalogs but also information seeking systems of tomorrow's world.

## **References**

---

<sup>1</sup> Dr. Tillett is the Chief of the Cataloging Policy and Support Office, Library of Congress, Washington, D.C. Thanks to Sue Vita for the subtitle. The reader is also directed to see a related publication by Dr. Tillett: "Cataloging for the Future" presented October 13, 2004 as part of the Windsor Lecture at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign (in press).

<sup>2</sup> See for example the studies of Mel Voigt on the *Scientists' Approaches to Information*, 1959. (ACRL publication)

<sup>3</sup> "Mr. Panizzi to the Right Hon. The Earl of Ellesmere. – British Museum, January 29, 1848." In *Foundations of Cataloging: a Sourcebook*, edited by Michael Carpenter and Elaine Svenonius. – Littleton, Colo. : Libraries Unlimited, 1985, p. 18-47.

<sup>4</sup> Cutter, Charles A. *Rules for a Printed Dictionary Catalogue*. Washington, Government Printing Office, 1876, p. 10.

<sup>5</sup> Bates, Marcia J. (October 1989). Rethinking Subject Cataloging in the Online Environment. *Library Resources & Technical Services*. 33(4):400-412.

Drabenstott, Karen M. (2000). Web Search Strategies. In *Saving the User's Time through Subject Access Innovation*, edited by William J. Wheeler. Champaign, IL: Graduate School of Library and Information Science, University of Illinois.

<sup>6</sup> *Functional Requirements for Bibliographic Records : Final Report*, IFLA Study Group on the Functional Requirements for Bibliographic Records. (Consultants: Tom Delsey, Elizabeth Dulabahn, Elaine Svenonius, Barbara Tillett). München : K.G. Saur, 1998.

---

Available as: [www.ifla.org/VII/s13/frbr/frbr.pdf](http://www.ifla.org/VII/s13/frbr/frbr.pdf) or as html:

[www.ifla.org/VII/s13/frbr/frbr.htm](http://www.ifla.org/VII/s13/frbr/frbr.htm)

<sup>7</sup> Hillmann, Diane, Naomi Dushay, Jon Phipps. "Improving Metadata Quality: Augmentation and Recombination." Available at

[http://metamanagement.comm.nsdlib.org/Metadata\\_Augmentation--DC2004.html](http://metamanagement.comm.nsdlib.org/Metadata_Augmentation--DC2004.html)

<sup>8</sup> *Catalogus impressorum librorum Bibliothecae Bodlejanae in Academia Oxoniensi.* Curä & operä Thomas Hyde. 1674.

<sup>9</sup> Osborn, Andrew. "The Crisis in Cataloging," *Library Quarterly*, v. 11 (Oct. 1941), p. 393-411.

<sup>10</sup> Lubetzky, Seymour. "The function of the main entry in the alphabetical catalogue – one approach," In: International Conference on Cataloguing Principles, Paris, 9<sup>th</sup>-18<sup>th</sup> October, 1961. *Report.* – London : International Federation of Library Associations, 1963, p. 139- 143.

<sup>11</sup> Verona, Eva. "The function of the main entry in the alphabetical catalogue – a second approach," In: International Conference on Cataloguing Principles, Paris, 9<sup>th</sup>-18<sup>th</sup> October, 1961. *Report.* – London : International Federation of Library Associations, 1963, p. 145- 157.

<sup>12</sup> Cleverdon, Cyril, Jack Mills, and Michael Keen. *Factors determining the performance of indexing systems. Vol. 1: Design. Part 1: Text.* ASLIB Cranfield Research Project, Cranfield, Bedford, England, 1966, p. 121-377.

Cleverdon, Cyril and Michael Keen. *Factors determining the performance of indexing systems. Vol. 2: Test results.* ASLIB Cranfield Research Project, Cranfield, Bedford, England, 1966. 299 p.

<sup>13</sup> Tillett, Barbara B. "Authority Control on the Web," *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium : Confronting the Challenges of Networked Resources and the Web*, Washington, D.C., Nov. 15-17, 2000, sponsored by the Library of Congress Cataloging Directorate, edited by Ann M. Sandberg-Fox. Washington, D.C.: Library of Congress, Cataloging Distribution Service, 2001, p. 207-220. <http://www.lcweb.loc.gov/catdir/bibcontr/tillett.html>

Tillett, Barbara B. "A Virtual International Authority File," *Record of Workshop on Authority Control among Chinese, Korean and Japanese Languages (CJK Authority 3)*, March 14-18, 2002, held at National Institute of Informatics (NII) in cooperation with National Diet Library. National Institute of Informatics, 2002, p. 117-139 (Also in Japanese, p. 140-153)

<sup>14</sup> Hickey, Thomas; & Vizine-Goetz, Diane. *Implementing FRBR on large databases* [online]. [Dublin, Ohio]: [OCLC], 2002 [cited 31 December 2002]. Available from [http://staff.oclc.org/~vizine/CNI/OCLCFRBR\\_files/frame.htm](http://staff.oclc.org/~vizine/CNI/OCLCFRBR_files/frame.htm)

Hickey, Thomas B., O'Neill, Edward T., & Toves, Jenny. Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR). In: *D-Lib Magazine* [online], Sept. 2002, v. 8, no. 9. Available from <http://www.dlib.org/dlib/september02/hickey/09hickey.html> (ISSN 1082-9873)

O'Neill, Edward. *FRBR: application of the entity-relationship model to Humphry Clinker*: ALCTS/CCS/ Cataloging and Classification Research Discussion Group,



---

Saturday, June 15, 2002...Atlanta... [online]. [Buffalo, NY: Judith Hopkins] June 2002? cited 27 August 2002]. Available from

<http://www.acsu.buffalo.edu/~ulcjh/FRBRoneill.html>

O'Neill, Edward. FRBR (Functional Requirements for Bibliographic Records): application of the entity-relationship model to Humphry Clinker. In: *Library Resources and Technical Services* (2002) v. 46, no. 4, p.150-159. (ISSN 0024-2527)

OCLC. *OCLC research activities and IFLA's Functional requirements for bibliographic records* [online]. Dublin, Ohio: OCLC, cop. 2002 [cited 16 July 2002]. Available from <http://www.oclc.org/research/projects/frbr/index.shtm> With links to OCLC's four projects: *Case study: the FRBRization of Humphry Clinker* <http://www.oclc.org/research/projects/frbr/clinker/index.shtm>, *Extending the case of Clinker* <http://www.oclc.org/research/projects/frbr/works.htm>, *Algorithm development and testing* <http://www.oclc.org/research/projects/frbr/algorithm.htm>, *Fiction Finder* <http://www.oclc.org/research/projects/frbr/fictionfinder.htm> [cited 31 December 2002].