The following was delivered as the 2004 Phineas L. Windsor Lecture at the University of Illinois Graduate School of Library and Information Science, October 13, 2004

Cataloging for the Future

by

Barbara B. Tillett[1]

Abstract: Cataloging and catalogs are changing yet again to benefit from advances in technology. We have new ways of looking at the bibliographic universe to meet the needs of today's users. We must 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.

Objectives of the catalog

In the 1830's and 40's Anthony Panizzi, who was then the Keeper of the Books at the British Museum (translation "Librarian" at what is now the British Library) – Panizzi spoke eloquently about the purpose of library catalogs, about his 91 rules to create a catalog, and about the elements to include in bibliographic records and the ideas of **collocating** the editions of works under the name of the author – the principles of authorship and idea of the literary unit for organizing information. He justified the costs of creating a "full and accurate" catalog that would be more than a mere finding list or inventory of a collection, to help the user to find works and bring editions together. Cataloging costs were even a problem 170 years ago, and the cost of cataloging is still a major issue today, probably more than ever before. How can we justify cataloging and keeping some of the cataloging traditions? I hope to show you that some of the principles, objectives, and concepts that can be accomplished even better with new technology and that future systems will benefit from what catalogers and librarians have to offer.

The ideas of collocation and of enabling finding works under the name of an author were restated and expanded by Charles Ammi Cutter in 1876 as his "objects" of the catalog - enabling a user to find a book when he or she knows the author, the title, or the subject and to show the user what the library has by a given author, on a given subject, or in a given kind of literature, that is, collocating the bibliographic records, and to assist the user to choose a specific edition or to help with identifying the literary or topical character through notes.[2] We know the **finding and collocating objectives** are fundamental to specific user tasks and are worthy objectives of search systems for the future.

But are all of our cataloging traditions still necessary today? Let's look at main entry – an often-challenged concept.

In the 1960's we had the debates about **main entries** between Eva Verona and Seymour Lubetzky about literary units and bibliographic units - Eva Verona favored the bibliographical unit that achieved the first objective of finding a book, and Lubetzky favored the literary unit that Panizzi and Cutter advocated (using more uniform titles) to achieve the second objective of finding all editions, translations, etc. of a given work – the collocating objective.

---

[1] Dr. Barbara B. Tillett is Chief, Cataloging Policy & Support Office, Library of Congress and was one of the consultants to the IFLA Study Group on Functional Requirements for Bibliographic Records. Some portions of this paper are included in "FRBR and Cataloging for the Future," Barbara B. Tillett. *Cataloging & Classification Quarterly*, v. 39, no.3/4, (In press for 2005), p. 205-216.

[2] Cutter, Charles A. *Rules for a Printed Dictionary Catalogue*. Washington, Government Printing Office, 1876, p. 10.

Today, this debate is no longer relevant, because with computer systems we can achieve both objectives through the creation of information packages that can be displayed in various ways to meet all of the objectives equally. So we are starting to challenge some of the basic concepts of catalogs of the past, like main entry – realizing that both access by works and access to individual manifestations are important to some users. We don't need to prefer bibliographic unit or literary unit over the other, because our bibliographic and authority information can be rearranged by computer systems to display the elements that we have included in description and access points – showing the works (literary units) when that is useful and showing the individual manifestation (bibliographic unit) when that is needed. However to do this intelligently, the cataloger needs to clearly identify the elements: that is, specify the primary author, identify the work, and be clear about the elements that identify the work, expression, manifestation, and item he or she is cataloging. In fact, we have been doing this since the late 1960's with the MARC format, but we could do even better. I'll come back to this in a moment.

What about cataloging everything in the library? Actually libraries gave up trying to catalog everything a long time ago. At the end of the 1800's after a brief attempt to collaborate with publishers in creating catalog records for every article in selected journals, it was found to be uneconomical, and H.W. Wilson took up the challenge and started his own indexing company. Libraries catalog the materials in traditional collections (books, serials, maps, sound recordings, films, prints and photographs, manuscripts, archives), but also provide their users with access to resources through abstracting and indexing services and other reference tools that can be combined in searches through metasearch engines, portals, and gateways. Catalogs were once used side by side with other reference tools, but with the online access of all of these tools, they can now be searched together. Catalogs are going through a metamorphosis, merging with other tools to improve bibliographic control over a wider range of resources, and the resulting future tools will need to do a lot more. We not only want to retrieve the surrogate bibliographic records or citations that describe resources, we want the resources themselves, and increasingly we can get them in digital form or at least be told of a nearby library where we can obtain a copy or be told of an online vendor that will sell us a copy of what we want.

Even though libraries don't catalog everything in their collections, today they are going beyond their own collections to inform their users of relevant resources anywhere in the world. Web resources are being selected and added to catalogs and other Internet resources are being linked for our users to access on the tool that is also their online catalog. Not only do they get the search tool, they may also get to see the digital object itself.

Is there then still value in having future tools keep the objectives of catalogs of the past, namely to enable finding and collocating resources and also to incorporate the benefits of authority control? Of course! And can we do it better and less expensively than we do now? We must!

Trained Catalogers

Reducing costs can happen through better use of people and technology and through a cataloging process that returns (once again) to the basics for providing bibliographic control over a large volume of diverse material. It requires skilled people and smart systems to make it easier for the user. It is a fallacy to think an untrained person can produce a "good enough" product. Training is necessary so the product is more easily indexed and usable. Sir Thomas Hyde wrote in his preface to the Bodleian Library catalog of 1674 that inexperienced people who make indexes for their collection of private books think all it takes is writing down the titles from the title pages. He goes on to say 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."[3] (He said it in Latin, but that's the gist of it.)

We see this today in trying to use metasearch engines to retrieve materials across a wide variety of source databases that are built by people of varying degrees of training, based on different metadata standards. One example is the National Science Digital Library. The NSDL uses software to massage the metadata they get from the various contributors and typically have to "dumb-down" the transformed data. Some of it is simply not usable, because it is not structured in a clear way that lets a machine recognize the components. For example, Dublin Core metadata elements are used, but the element "creator" may contain a single name in any order or multiple names – with no indication for the machine to know what's going on.[4] Some contributors to the NSDL include full information, others provide very minimal information, so trying to make sense of the resulting searches is challenging.

It's interesting to note that non-librarian designers of digital libraries discovered they needed authority control once they got beyond their prototype systems of a few hundred records. Names and titles can be presented in a wide variety of ways and in many scripts, so librarians follow standards to enable gathering like things together. Controlled vocabularies for names and subjects are proven components of good retrieval systems. So here's another cataloging tradition that has value into the future – controlled vocabularies. And we know authority control is expensive. To build and maintain controlled vocabularies and authority files requires trained experts, but the value becomes clear for the precision of searching and enabling the collocation objective of catalogs – especially when you want to obtain resources in a very large, multilingual, diverse collection. Yes, it may be expensive, but we'll look in a moment how to help reduce that cost.

Creating a subject heading string requires knowledge of the syntax rules, so there have been proposals to get rid of the need for that by using faceted terms that can be post-coordinated.[5] Yet we know that post-coordinated searching reduces the precision of searches and increases the recall, making it harder for users to find what they need in very large databases.

Doing authority work takes skill and time, so there have been decisions to eliminate access points by using minimal level or core records, such as the "core records" in the Program for Cooperative Cataloging. This results in the abandonment of some of the basic objectives of catalogs to enable a user to find all the works of an author and to collocate the various manifestations of a work and to find all the works under a given subject. Cost-cutting decisions to eliminate subject access run against studies that tell us subject access is so vital, especially for digital materials and Web resources. Cutting costs by reducing or eliminating name and subject access and the corresponding authority control result in serious disservice to users – and remember that users

---

[3] *Catalogus impressorum librorum Bibliothecae Bodlejanae in Academia Oxoniensi*. Curä & operä Thomas Hyde. 1674.

[4] Hillmann, Diane, Naomi Dushar, Jon Phipps. *Improving Metadata Quality: Augmentation and Recombination*, available at http://metamanagement.comm.nsdl.org/Metadata_Augmentation--DC2004.html

[5] Dykstra, Mary. LC Subject Headings Disguised as a Thesaurus. *Library Journal,* 113 (March 1, 1988), p. 42-46.

Cochrane, Pauline A. *Improving LCSH for Use in Online Catalogs: Exercises for Self-Help with a Selection of Background Readings*. Littleton, CO: Libraries Unlimited, 1986.

Chan, Lois Mai, Eric Childress, Rebecca Dean, Edward T. O'Neill, and Diane Vizine-Goetz. "A Faceted Approach to Subject Data in the Dublin Core Metadata Record," *Journal of Internet Cataloging,* 2000.

include library staff who need some of that information to do their jobs of selection, acquisitions, cataloging, and reference.

Beyond access points, descriptive elements have also been reviewed and reduced or eliminated. Panizzi argued for a base set of elements, justifying them by the usefulness of each element to the user, and cataloging rules since his time have refined his choices. We now have the ISBDs (International Standards for Bibliographic Description) that give us prescribed elements and their order in a bibliographic record. However, we have also seen how shortsighted decisions to eliminate some of the basic elements have resulted in inconsistent records that are lost in retrievals and collocations – again the user is the loser.

Another effort to reduce costs has been the development of special processing for special types of materials. The result has been a proliferation of processing streams to meet special needs and the resulting special case law approaches have led to complex documentation to record the myriad conditions and pathways. We have justified this by explaining that with a large workforce you need to document all these special conditions in order to assure consistent products. This wasn't just the warning of Andrew Osborn in 1941 about the "crisis in cataloging,"[6] and the proliferation of case law rules, but it's again a major concern today. We need to step back again to see where consistency is important and to assure there are general rules to cover those situations and also to recognize other areas of description where consistency doesn't matter, provided that the resulting information in the cataloging record accurately reflects the item being cataloged and is clearly understood by the user. Back to basic principles.

In 1932 Keyes Metcalf in expressing concern about the cost of cataloging suggested giving up cataloging as we do it now; finding some way of increasing production; or through centralized or cooperative cataloging reducing the amount of cataloging that must be done.[7] Today we are trying all three approaches and still need to do more.

We know we cannot continue to catalog as we always have – certainly not as we did in the days of handwriting cataloging information on slips and cards or even typing card sets or buying them from the Library of Congress or a vendor. We know we can't even catalog as we have for the past nearly 40 years of online cataloging through our still very primitive library systems and online catalogs (OPACs). The online catalogs actually lost some of the benefits of earlier catalogs. We lost the use of the full array of original scripts – the Japanese, Chinese, Hebrew, Arabic and many other beautiful scripts of languages that we used to transcribe when creating descriptive information – we've been limited to those languages that the MARC format would accept, and even those could not be displayed in most integrated library systems – this remains a limitation of most integrated library systems today. We lost the collocation of many works and expressions in OPACs that, to this day, typically do not properly index uniform titles. Just talk to your music librarian colleagues to hear their dismay about current online catalogs displaying musical works. Our tools are not as good as they should be or as they can be.

Yet, those OPACs brought us the ability to filter and limit searches and to bring together records in ways we could not do in book or card catalogs. They gave us the mixed blessing of keyword searching for great recall at the expense of precision. And now we see the Internet capabilities of Google and Yahoo and Amazon for even faster keyword access to information. We have the

---

[6] Osborn, Andrew. "The Crisis in Cataloging," *Library Quarterly*, v. 11 (Oct. 1941), p. 393-411. Also published by American Library Institute, 1941. 19 p.

[7] Keyes D. Metcalf, "Cooperative cataloging: activities in 1932 and plans for 1933," *Library Journal*, 58 (Feb. 1, 1933), p. 107.

promise of even newer search engines than Google that are built on graph theory to retrieve the richness of the data we've been providing in bibliographic and authority records; systems to help guide users by enabling them to see the wealth of terms and names associated with the topic they are looking for. Still the cataloging remains expensive.

Cooperative Cataloging

For over a century we have tried to reduce costs through cooperative or centralized approaches to cataloging. The business model for cooperative cataloging is based on the notion that resources are created in multiple copies and that a bibliographic record can be re-used to describe a copy held anywhere, that is, cost savings by sharing the work of cataloging. The records created by one library can be re-used by many libraries worldwide. Those costs can be lessened through cooperative cataloging initiatives to catalog more with fewer people, by using shared standards that in turn enable the re-use of the resulting bibliographic and authority records on a global scale. "Catalog it once for all!"[8] to save the cumulative cost of redundantly cataloging the same manifestations in each library that acquired a copy. One cataloger would catalog the manifestation once for all catalogers and for all time.

This is also the idea behind IFLA's "Universal Bibliographic Control" or UBC – where each national bibliographic agency is responsible for creating bibliographic records for its nation's published products and for creating authority records for all of its authors. (IFLA is the International Federation of Library Associations and Institutions.) The idea is that those records would then be used worldwide by anyone else acquiring those materials or needing the authority record for the person or corporate body represented by the authority record. The problem with that concept is that not all libraries use the same language or script or cataloging rules, so records created in France, for example, might be of minimal use in Russia or China where they might prefer to transliterate to a script their users can read. IFLA is taking a new view of Universal Bibliographic Control to link records for the same entity and take more advantage of today's technology to enable greater sharing of bibliographic and authority information. We can retain the diversity of cultural differences and citation practices with cataloging rules to meet the local user needs, while agreeing on basic elements and clear identification of the elements of description and access for increased interoperability worldwide.

The economies of scale make sense for published materials that are typically printed or produced in multiple copies. That does not hold up so well for unique materials, like objects in museums or archives and many special collections of rare materials or other resources that we may want to make available to a targeted group of users, like in-house technical reports and unpublished reports of research or courseware or instructional materials. So we don't get the economies from cooperative cataloging for these special materials – but we could reduce costs for those through better tools that support the cataloger – systems that help suggest controlled subject terms and classification numbers and help with authority work for names. Such systems can help all types of cataloging.

Technology

Just as we have many times in the past, we can take advantage of the technology available to us at the time. For example, librarians made the most of printed card technology to mass-produce cataloging records that could be purchased and added to local catalogs. With the development of the MARC format in the late 1960's we saw online systems that helped catalogers enter

[8] Tillett, Barbara B. "Catalog It Once for All: A History of Cooperative Cataloging in the United States Prior to 1967 (before MARC)," *Cataloging & Classification Quarterly*, v. 17, no. 3/4 (1993), p. 3-38.

bibliographic information and enable wider sharing of that information eventually worldwide. Bibliographic utilities since the 1970's have provided a rich resource of centralized sharing of bibliographic and authority data through networked systems. With integrated library systems, we could not only bring in those shared records from the bibliographic utilities, but we could re-use bibliographic information at various stages of acquiring and cataloging materials, and on through the life cycle of resources to their binding, storage, circulation, interlibrary loan, and preservation. Now with Web OPACS and catalogs merging with other tools in gateways and portals, our users can be anyone in the world and our collections can be global as well. Library systems are still in their infancy, just as the card catalog was a century ago, and our systems have yet to realize the full potential of enabling access to the world's information resources. Information in libraries will be one of the prime resources for future search and retrieval systems. Google has already recognized that and is courting libraries to digitize their collections. Our bibliographic and authority records are a treasure that can be shared for the benefit of users worldwide in many communities – even beyond libraries.

FRBR Applications for Cataloging

Today we are working towards greater global standardization in order to more easily share bibliographic and authority information and thereby help to reduce the costs and also increase the ability to meet changing user needs. IFLA has led the way for generations and recently offered a conceptual model of the bibliographic universe to provide a framework for re-thinking cataloging and catalogs.

This conceptual model is part of the "Functional Requirements for Bibliographic Records" also known as FRBR (sometimes pronounced "ferber"). Very importantly, FRBR describes the entities in this universe and their attributes and relationships. It reminds us of the importance of bibliographic relationships to collocate the works and expressions associated with a person or corporate body – something lost in many online catalogs of today. It also reminds us the user comes first, that we catalog in order to meet specific user tasks to: 'find,' 'identify,' 'select,' 'obtain,' and I add 'relate.'

I have written before that "FRBR describes fundamental concepts that have been stated in many ways in the past and now puts those concepts in new terms. It reinforces those concepts just at this point in time when librarians are working closely with other knowledge managers and information providers (from publishers to museums, archives, rights management organizations, and the Internet and computer system developers) to organize and provide access to both the traditional materials and the growing corpus of digital objects. These new digital forms of material and new packages for information introduce an opportunity for tremendous change in how we provide bibliographic control. Understanding the same concepts and sharing a vocabulary among all information providers are essential steps in enabling change."[9]

Cataloging Principles

In 1961 IFLA sponsored a meeting of cataloging experts to promote international cooperation and standardization of cataloging practices to facilitate the sharing of bibliographic information. The product of that meeting was the "Paris Principles" that are the foundation of nearly every cataloging code used throughout the world today. This was a giant step towards global bibliographic control.

[9] Tillett, Barbara B. "FRBR and Cataloging for the Future," *Cataloging & Classification Quarterly*, v. 39, no. 3/4 (in press for 2005), p. 205-216.

In 2003 IFLA began a multi-year series of meetings of the world's cataloging rule makers and national cataloging experts to re-examine the Paris Principles, to update them for today's environment, and to expand them to cover both descriptive and subject cataloging now reflected in online catalogs and planned for future systems. The 2003 meeting, held in Frankfurt, Germany, brought together 54 experts from 32 European countries to compare their current cataloging codes. We just held the 2004 meeting in Buenos Aires, Argentina for the Latin American and Caribbean countries. One goal of these meetings is to see if the existing cataloging codes could be harmonized and perhaps to recommend rules for an international cataloguing code that would still allow for cultural and national variations where that was important to users. The resulting draft "Statement of International Cataloguing Principles"[10] (Frankfurt Draft Principles) is being refined over the course of further annual regional meetings worldwide through 2007 and a subsequent "worldwide" review beyond the cataloging experts and rule-makers themselves.

The Frankfurt Draft statement declares that it is "built on the great cataloguing traditions of the world, and also on the conceptual models of the IFLA documents *Functional Requirements for Bibliographic Records* (FRBR) and *Functional Requirements and Numbering of Authority Records* (FRANAR)." The introductory text goes on to say, "It is hoped these principles will increase the international sharing of bibliographic and authority data and guide cataloguing rule makers in their efforts to develop an international cataloguing code." An international code would be rules and principles to guide rule makers with rules that have worldwide acceptance while also allowing for options when needed to meet cultural differences. There will continue to be the need for codes that reflect the needs of different cultures, sometimes due to different scripts, different citation traditions, different conventions of naming, and so on.

Anglo-American Cataloguing Rules (AACR)

This also implies there will be a continuing need for *Anglo-American Cataloguing Rules*. AACR is the cataloging code now used in all the Anglo-American countries as well as in many other places throughout the world. Following the 1997 International Conference on the Future of AACR held by the Joint Steering Committee for Revision of the Anglo-American Cataloguing Rules (JSC), the JSC recommended the analysis of AACR and reviewed the logical structure of AACR2 with the desire to restructure the rules so they would be easier to apply and to articulate the underlying principles to build cataloger's judgment. Plans have been underway for several years now to create a new edition of the rules. One of the goals for restructuring the rules is to increase the consistency in application of the underlying principles across all types of materials. The Introduction for AACR3 will explain the FRBR terminology and basic cataloging concepts, including the FRBR user tasks and objectives for library catalogs, and eventually will include the future IFLA Statement of International Cataloguing Principles.[11] The timing with the IFLA initiative is quite good, as IFLA will probably have a nearly final version of the Statement of International Cataloguing Principles by late 2006, which is when AACR3 will be sent to the publishers.

[10] The draft "Statement of International Cataloguing Principles" is available on the Web at: http://www.ddb.de/news/ifla_conf_index.htm the Web site for the 1st IFLA Meeting of Experts on an International Cataloguing Code, Frankfurt, 2003.

[11] The plans for AACR3 began with the general sense that the rules are basically sound but could be improved, for example, the format of the code itself could be more helpful to catalogers in day to day application, and it was felt that a principle-based code would help train new catalogers and help build catalogers judgment. The JSC's strategic plan outlined general goals and tasks to reach the next edition of the rules. The JSC Strategic Plan is available on the Web at http:///www.nlc-bnc.ca/jsc.

One of the beauties of including the FRBR concepts in the new edition of the rules is to remind us of the basic objectives to enable finding and collocating bibliographic records. This has been the objective of library catalogs, but should also apply equally to search engines of the future. FRBR describes the model to facilitate the collocation of related entities in the vast bibliographic universe. FRBR specifies the basic attributes or data elements that need to be present in national bibliographic records. The FRBR model also identifies important relationships in the entity-relationship model and those are already in the rules but can be made clearer. Records built on the FRBR concepts will clearly identify the bibliographic entities, so displays in online catalogs or retrieval systems can show the families of works and related works, as well as their expressions and various manifestations in multiple physical formats, even down to specific distinctive items and where they are located or accessible. Hopefully this will result in clearer displays for users to get a better picture of what is available related to what they are looking for.

In the past few years there have been several test systems and prototypes based on the FRBR concepts. One test of the FRBR relationship information is OCLC's research into "XISBN" whereby a searcher of an online search engine or of an online bookseller's Web site can invoke an application that would capture the ISBN (International Standard Book Number) of the retrieved item and find the FRBR related expressions and manifestations. It grabs their ISBNs and displays the resulting set of bibliographic records – to suggest to the user that there are other manifestations beyond the single one searched for. This ISBN information can also be used to provide information on where specific items for each manifestation are located in the user's geographic area (based on the user's zip code linked to libraries in that location). This capability builds on the objective to collocate the various manifestations of a work.

The ISBN in FRBR terms is an identifier for the manifestation, one of its attributes. The provision of the ISBN in bibliographic records comes through cataloging rules, and in turn a system can use that attribute to suggest other manifestations to the user that may equally meet his or her needs – and equally to a requestor for interlibrary loan or to a patron at a reference or information desk. The user may not care which edition, as long as it is one in a language he or she can read and contains the content he or she wants. The format of being a paperback or a CD or a cassette also may not matter, as long as the user has the means for using that format to get to the content. So using the connections to the literary unit, to the work/expression, a system can help a user fulfill the task of obtaining desired information.

FRBR, MARC Records, and Future Cataloging Systems

The FRBR entities of work, expression, manifestation, and item are often confused with the current MARC record structures by some catalogers. They want to know how FRBR maps to what we are doing now. In fact, the bibliographic records we create in MARC format contain attributes of all of the FRBR entities. You do not have to have separate records for each entity, but it is helpful to separately identify or code the specific attributes – as we do now with MARC tagging or XML DTDs (data type definitions). In that way we can more easily manipulate that data for various displays, and many of the attributes can be used for collocating or clustering together the manifestations of the same expression and the expressions of the same work and even the family of related works. If we clearly identify or label the attributes, they can be manipulated by machines for searches and displays.

Our systems for creating bibliographic descriptions should keep MARC or other communication formats invisible to the cataloger and the user. For example, catalogers could be assisted in the creation of bibliographic descriptions through the use of style sheets that capture as much as possible in terms of transcribed information from the object to be cataloged. We could provide automatic transliteration when needed. The systems of the future could suggest possible related

resources. They could suggest possible subject headings and classification numbers based on information already in the database of bibliographic and authority records for similar material.

We can also envision putting subject headings and classification numbers at the work and expression levels that would then link to data for the various manifestations that embody that work and expression. For now, we could enter that subject information in an expanded authority record for the citations (or we now call them uniform title authority records) for the works and expressions. If we assign the classification numbers and subject headings to the work and expression, that information could be linked to or re-used for the bibliographic records describing the various manifestations without needing to redundantly assign subject headings and class numbers to each manifestations separately. We'd just need to add to the call number to provide a unique location for each item or copy. When we have systems that can use data from authority and bibliographic and holdings records (with today's MARC record constructs), there are many exciting opportunities for data entry and displays. It liberates the data within MARC records themselves to be applied to displays that fulfill the objectives of the catalog and fulfill user tasks.

Future Systems

Future systems could also suggest controlled vocabulary terms and names that would be verified by the cataloger and would offer the possibility to connect to a virtual international authority file to check for controlled names created by other authoritative sources worldwide. That virtual international authority system would also be a means for future switching of display forms for names and terms, again to keep the user first. The user would specify their preferred language and script, and the system for controlled vocabularies would enable switching to their preferred choice.[12]

Applying cataloging concepts and fulfilling objectives for catalogs would help future systems to gain the benefits for end-users of improved precision of searches and better clustering of related entities. Those future systems will help the user search all potential sources of information through intelligent systems, like the Semantic Web or similar global networked environments.

We have other prototypes of what systems could do in applications like RedLightGreen from RLG. It incorporates some of the FRBR concepts in a Web-based search engine. It uses a Google-like search to retrieve the works and expressions in its database and alerts the user to available translations and editions (which are a type of expression in FRBR terms). The user focus of RedLightGreen is on undergraduate students, so the system lets the student know the local libraries where he or she can get the desired item, and it enables that student to re-use the retrieved bibliographic information (the FRBR attributes) repackaged in their choice of standard citation format to use in bibliographies and footnotes in papers they are writing.

---

[12] For more information on the vision for a virtual international authority file, see "Authority Control on the Web," Barbara B. Tillett. *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.
"A Virtual International Authority File," Barbara B. Tillett. *International Cataloguing and Bibliographic Control,* v. 30, 2001.
"A Virtual International Authority File," Barbara B. Tillett. *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)

There are also systems that enable end-users to go to online sellers, like Amazon, Barnes and Nobel, Borders, etc. to purchase items they have found. This bridges libraries with bookstores and vendor systems to give users the option to borrow or purchase. Our library concepts and research can benefit the development of those commercial products as well.

There are also new front-end systems, like Endeca, for guided searching of databases. Endeca has built on library concepts of giving users suggested terminology to refine their search and to point out related resources not only through the traditional reference structures and links from authority records, but through suggesting related search terms and concepts found anywhere in associated bibliographic and authority records. It re-purposes the traditional elements of bibliographic and authority records to assist the user in seeing the range of possibilities for searching. The user enters any term or name or piece of information they want to search and the system responds with a screen of clustered terms used in the database that are associated with the information the user keyed in. From the research of Marcia Bates[13] and Karen Markey Drabenstott[14], we know users very likely don't know the vocabulary or controlled forms used by the library system; and with systems like Endeca, they don't need to know, because the system knows and can offer up an array of suggested terms, names, places, times found anywhere in the bibliographic and authority records that are connected to the term used for the search. The system makes full use of the attributes and terms we have provided through cataloging and presents the information in a creative new way to give the user a better picture of the universe they are exploring.

Combining Bibliographic Description with Items

Our bibliographic records more and more include links to the digital objects themselves or to related electronic finding aids and other resources. Some systems now embed our bibliographic descriptions with the digital information objects themselves – either copying the MARC record or mapping it to a METS[15] record with MARC XML or just by cutting and pasting the contents of the MARC records to use as the descriptive metadata for the digital objects.

Combining the description of an object with the object itself isn't new. We could say we've been doing that for many years with Cataloging-in-Publication (CIP) information, that is, bibliographic description in the items themselves. That CIP data or bibliographic description is printed on the verso of the title page, so it's not directly searchable by a computer system. The analog content of the item also is not searchable, but once we provide the MARC record for the CIP data to computer systems, particularly those searchable through the Web, we increase the access to that information. And once we digitize the content of the item, both the description and the full content of the digital object can be searchable by computer systems right on the user's own PC, fulfilling the user task to obtain the item they want.

The cost of cataloging digital or electronic materials might also be reduced through the availability of more tagged or labeled descriptive metadata built in to the objects themselves as they are created. Microsoft and other companies could design systems to automatically provide basic metadata as electronic resources are being created, going beyond what they do now in suggesting a title and providing the name of the creator, and providing the date of creation. Those systems could help the creator to access easy to use keywords from controlled vocabularies and

[13] Bates, Marcia J. (October 1989). Rethinking Subject Cataloging in the Online Environment. *Library Resources & Technical Services.* 33(4):400-412.

[14] 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.

[15] METS is the Metadata Encoding and Transmission Standard.

authority files to include with their creations for future indexing and retrieval. OCLC has already experimented with such a tool using the latest XML-enabled version of Word. Catalogers, who get copies of the resulting works, could enhance the basic descriptive metadata, adding other controlled subject terms and classification numbers as needed and providing relationships to names of persons, corporate bodies, works, etc. to enhance the Web discovery of both the descriptions and the items themselves.

For images, we have cameras now creating their own metadata when you take a picture with them – certainly the date the picture was taken, and some add the photographers' name (or rather the name of the owner of the camera who entered the information at the start), and it captures information about the settings used. It could even include the satellite geospatial positioning information to indicate where the picture was taken – and all of that could be used later for retrieval. For documents, when we use software like Microsoft Word, the software automatically indicates the date the document was created or modified, the author (based on the registered owner of the software), the document's file size, the software used to create it (type of file), and it suggests a title for the document when you save it – using the first words of the first line of text – ironically that is how the Babylon clay tables, scrolls, and early manuscripts were identified – the first lines of text; and music- by the first few notes of the work. All of that is metadata that can be used to retrieve that document later.

Some bibliographic and authority records already include hotlinks to digital items for online access, to "obtain" in FRBR terms. This brings with it the need to alert the user to access restrictions for obtaining some materials that require licenses or other payment for use or those that are restricted, confidential, or top secret materials. There are several experimental systems for digital rights management, and some integrated library systems include such alerts as well.

I mentioned before that we have some adjustments to make in bibliographic descriptions to be more precise about what we are describing – a work, an expression, a manifestation, an item – we already do that to some degree in our MARC records, but the MARC format itself has benefits and drawbacks for enabling future manipulation of the data elements for different displays. Some systems have experimented with "FRBR-izing" legacy data, like VTLS and Innovative Interfaces. The Library of Congress and OCLC each have different kinds of FRBR algorithms that are available for experimentation and map MARC data into FRBR concepts.

As we re-examine cataloging practices in light of FRBR, we see the usefulness of explicitly identifying the roles of persons and corporate bodies with respect to the bibliographic entities, more than the MARC format does through the tags of main or added entry, to use what the MARC format already includes as relator terms and codes. Through another cost-cutting decision many libraries in the United States agreed to stop providing such role information for nearly all types of roles (exceptions being for some illustrators and some roles in music – like composers and arrangers). These are things that could be built into the creation of descriptive metadata in future cataloging systems in ways that are easy for the cataloger to use. This "role" information is not part of the heading for the person or corporate body, but instead is an identification of the relationship between that person or corporate body and the bibliographic entity being described. It is useful information for filtering large search results.

We also see the value of standardized citations for works and expressions (what we now call uniform titles for works and expressions) to collocate the small but important percent of materials

that exist in multiple manifestations. Some of OCLC's research has shown less than 20% of their WorldCat database of over 56 million records have more than a single manifestation per work.[16] We can see the potential for decreasing cataloging work and costs by linking subject headings and classification numbers to those citations that in turn can be linked to records for the manifestations that embody those works and expressions – or they could be linked to or embedded with the digital object they describe, increasing the precision of retrieval for that object.

What if everything was digitized?

But what if all the world's information – print, graphic, sound, moving image, everything -- was available in digital form? We could access by words or sounds or images found in the digital objects themselves, but retrieval would be more precise if we included controlled vocabularies (names, titles of works, and subjects).

Brewster Kahle of the Internet Archive was cited recently in the *BBC News* World Edition as follows: "Using a robotic scanner, Mr. Kahle said the job of scanning the 26 million volumes in the US Library of Congress, the world's biggest library, would cost only $260m…He estimated that the scanned images would take up about a terabyte of space and cost about $60,000…to store. Instead of needing a huge building to hold them, the entire library could fit on a single shelf."[17]

Besides the point that the Library of Congress has over 116 million items – and not just the 26 million book volumes in its collections and that we continue to add thousands of new items in analog form to the collections each day, the point Mr. Kahle was making is that digitizing it all is

---

[16] 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

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].

[17] "Visionaries outline web's future," article from *BBC News*, World edition (last updated Friday, 8 October, 2004, 09:39 GMT 10:39 UK at http://news.bbc.co.uk/2/hi/technology/3725884.stm

probably do-able with today's technology, and someday those scanned digitized texts, images, and sounds will also become searchable with advances in optical character recognition and sound and image recognition. It's not good enough just to digitize it all – we must be able to retrieve individual items and groups of items that meet user needs – to find, identify, select, and obtain information. (Libraries are also working on how to preserve digitized information and assure its existence hundreds of years from now.[18])

In the future, for the born digital and the digitized items, all the "chief source" information will be there, and we will find ways to give more weight to information found on a chief source (that's the title page for books, remember), just as Google can now give more weight to metadata in titles and from highly ranked sources in addition to frequency of occurrence of searched terms or frequently accessed sources. But we know from history that searching is improved when we used controlled vocabularies. Libraries have been in the business of selecting the most relevant resources for its users, and we could continue to augment the machine-generated metadata for those selected resources with added access information using our controlled vocabularies. And those controlled vocabularies will be augmented by more sophisticated search systems, suggesting terms to users, like Endeca and other systems I mentioned. We could also make our tools available to other communities worldwide to use to augment the digital information linked to the digital object.

We need application software to enable retrievals and displays such as topic maps and other cluster display capabilities and guided search engines, utilizing information we provide in bibliographic and authority records that augments information we could use from digitized objects themselves. Tomorrow's catalogers will continue to develop and maintain the most useful controlled vocabularies and authority databases worldwide.

We need the virtual international authority file as a functioning building block along with other name and topic lists and mappings of vocabularies to enable users to search and see names and terms displayed in the language and script they can read while enhancing the precision of Web searching. The virtual international authority file would help catalogers reduce the costs and labor of authority work through more international sharing of the work and automatic incorporation of found authority information into local systems. It will speed up their work and avoid redundant effort. In the future, a virtual international authority file system would enable creators and realizers of intellectual and artistic content to automatically include (or link to) controlled forms of names and topics at the point of recording the content they create.

We need international cataloging rules (or compatible national rules) to make it easier to re-use the products of cataloging and to help reduce global costs for bibliographic control. Future rules would provide guidance for those serious about description and access, while providing general principles anyone could use. The description and access points provided by creators and publishers will be enhanced by catalogers, just as we do now, but we can expect to get more usable information from publishers, who have trained catalogers on their staff. Sharing bibliographic and authority information worldwide will help reduce costs.

We will see new information retrieval systems, that is, corporate integrated systems, which enable the creation, maintenance, and use of XML-based data packages. Those data packages would build on the FRBR entities, attributes, and relationships for elements we now place in

[18] One example is the National Digital Information Infrastructure Preservation Program led by the Library of Congress with millions of dollars to research and develop viable preservation strategies for digital materials.

bibliographic and authority records but in future with links to the objects themselves (whether through "call numbers" for a given library location, a URL or other address of the item, or embedding the description in an "information package" with the digital object itself).

We will still need trained catalogers – whether they work for publishers or libraries or giant searching conglomerates or online vendors, like Amazon. Cataloging requires skilled people to save users time and effort in locating information they need. We want to continue to let users know about related material and some of those relationships will be provided by machine-generated links, but others will need human intelligence to identify. Cataloging follows some basic principles and concepts and objectives that put the user first, providing collocation and differentiation of entities that are meaningful to users. Those principles should be enhanced by future search engines and global networks. The skills and knowledge of catalogers will help future generations develop even better systems and methods for organizing the ever-increasing universe of information.

So all of this describes my vision for the future:

I am confident that we will build future systems for machine-generating or capturing bibliographic information and there will be better systems for manipulating that information to meet user needs. But I also believe all of the machine-generated information will be greatly improved by human intervention -- by trained librarians, who select materials for their users from among the world's creations -- by trained librarians who add appropriate terms from controlled vocabularies and who build and maintain those controlled vocabularies and international authority files. Those tools will be part of future global networked systems helping to improve the precision of searching and meeting the users' needs in whatever language or script they prefer. The expertise of librarians is of enormous value, and we have a lot to contribute to the design of better ways to organize information and help users find information they need.

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