

Authority Control: State of the Art and New Perspectives

Barbara B. Tillett PhD and MLS and BA

To cite this article: Barbara B. Tillett PhD and MLS and BA (2004) Authority Control: State of the Art and New Perspectives, Cataloging & Classification Quarterly, 38:3-4, 23-41, DOI: [10.1300/J104v38n03_04](https://doi.org/10.1300/J104v38n03_04)

To link to this article: https://doi.org/10.1300/J104v38n03_04



Published online: 23 Oct 2009.



Submit your article to this journal [↗](#)



Article views: 1277



View related articles [↗](#)



Citing articles: 2 View citing articles [↗](#)

STATE OF THE ART AND NEW THEORETICAL PERSPECTIVES

Authority Control: State of the Art and New Perspectives

Barbara B. Tillett

SUMMARY. Authority control is necessary for meeting the catalog's objectives of enabling users to find the works of an author and to collocate all works of a person or corporate body. This article looks at the current state of authority control as compared to the visions of the 1979 LITA (Library Information and Technology Association) Institutes and the 1984 Authority Control Interest Group. It explores a new view of IFLA's Universal Bibliographic Control (UBC) and a future vision of a virtual international authority file as a building block for the Semantic Web, and reinforces the importance of authority control to improve the precision of searches of large databases or the Internet.

KEYWORDS. Authority control, access control, universal bibliographic control, virtual international authority file, cataloging objectives, precision of searching

Barbara B. Tillett, PhD, MLS, BA, is Chief, Cataloging Policy and Support Office, Library of Congress, 101 Independence Avenue, SE, Washington, DC 20504-4305 (E-mail: btil@loc.gov).

[Haworth co-indexing entry note]: "Authority Control: State of the Art and New Perspectives." Tillett, Barbara B. Co-published simultaneously in *Cataloging & Classification Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 38, No. 3/4, 2004, pp. 23-41; and: *Authority Control in Organizing and Accessing Information: Definition and International Experience* (ed: Arlene G. Taylor, and Barbara B. Tillett) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2004, pp. 23-41.

<http://www.haworthpress.com/web/CCQ>
Digital Object Identifier: 10.1300/J104v38n03_04

The virtues of authority control have been debated and restated for decades. Catalogers for at least a century and a half have documented their decisions on how the single, authorized form of name for each entity should be represented in their catalog. They traced the various forms of names given to an entity to record the cross references they provided to users of their catalogs. They brought together their notes to further identify the entity for themselves and other catalogers building the catalog, documenting their research in the process of authority work. Some said it was unnecessary, most said it was essential to fulfill the objectives of the catalog to find and collocate the records for bibliographic resources. Still others said, stop debating and just get on with it, and we have, but ever mindful of the costs.

Since the 1970s people have claimed that authority work is the most expensive part of cataloging, and we still seek ways to automate and simplify the work to reduce costs. A giant step in that direction has been the move to share the work and share a resource authority file among many libraries. Examples have been the now famous cooperative program, NACO, the Name Authority Cooperative project involving the Library of Congress and other partners. There are also numerous examples of regionally and nationally shared authority files, like the Hong Kong Chinese Authority Name, known as HKCAN. Today's technology opens up opportunities for us to now link these many authority files and build on their strengths, to improve those resources, and to open new doors for service to users.

When we apply authority control in today's Web environment, we are reminded how authority control brings precision to searches, how the syndetic structure of references enables navigation and provides the end user with explanations for variations and inconsistencies, and how the controlled forms of names and titles and subjects help collocate (group together) works in displays. Even more today, we can envision using authority records to actually link to the authorized forms of names, titles, and subjects beyond the catalog for which they were originally intended to various online-accessible reference tools and resources, like directories, biographical dictionaries, abstracting and indexing services, and so on. Library catalogues can now be found in the mix of various tools that are available on the Web.

In 1979 LITA (Library and Information Technology Association) held a series of institutes entitled, "Authority Control: The Key to Tomorrow's Catalog" reminding us that the syndetic structure of catalogs in North America owes its origins to Charles Ammi Cutter. In his *Rules for a Printed Dictionary Catalog* in 1876, Cutter stated that controlling the forms of names was one of the "means" for meeting the objectives (he called them "objects") of the catalog. The objectives were "to enable a person to find a book of which either the author, title, or subject is known" (this is often called the finding objective),

“to show what the library has by a given author, on a given subject, in a given kind of literature” (this is called the collocating objective), and “to assist in the choice of a book as to its edition (bibliographically), as to its character (literary or topical)” (this is part of an identifying objective).¹ He stressed the importance of the syndetic structure of cross references in a catalog to get the user to the authorized form used to collocate the works of an author.

Seymour Lubetzky in his *Principles of Cataloging* in 1969 noted that one could accomplish the collocation objective by enabling an inquiry under any variant form to retrieve the works of an author under any of the names he/she used—even a citation from a bibliography, and thus eliminate the problem of choice of name by which to identify an author in the catalog.²

During the 1979 LITA Institutes, Ritvars Bregzis also made the case that we don’t need the extra work of devising an authorized form of name or uniform title, as long as we can make the associations, the relationships among related works. He wrote that “the computer technology has given us an opportunity to return to the record syndetic structure of the catalog, the structure in which the authentic form of the identificatory information describing the publication, being also the most frequently cited form, is given its own identity as a component of the catalog.”³ We do this using references to direct the user to where the collection of records are filed that collocate the works of an author. This collection of records can be accomplished on a screen on a computer terminal, which the computer can manage without actually storing the records in any particular order in the database—but instead indexing them for displays as needed. As Ritvars Bregzis noted, the inclusion of an authorized form of heading is not a requirement for the bibliographic record, as long as we can link it in the computer to the related works of the author. Although he did not foresee the use of ISADN’s or work/expression level citations, such tools are available to use today to meet the same objectives. Michael Gorman also spoke at those institutes. He and others have suggested how to accomplish the linking,⁴ using a structure where there were unique records for each physical item linked to records for the related persons, corporate bodies, works, subject, and other records for physical items. There would be an authority record for each work with more than one bibliographic description and identified by more than one title. Gorman also said there would be no composite author/title authority records. This view would fit very well in today’s FRBR conceptual model of the bibliographic universe.

When I started the Authority Control Interest Group (ACIG) in ALA in 1984, I conducted an opinion poll to get discussion started. At that time 18% of the respondents had online catalogs and 90.4% were using some bibliographic utility for shared cataloging. I compared the results of the ACIG opinion poll with two other surveys conducted earlier, and found that somewhere

between 13 to 28% of the libraries had no authority files, but instead relied on the records from other libraries. However, of the at least 72% that did maintain their own in-house authority files, we know it was at great expense when aggregated nationally. We also recognized the value of the NACO project in helping reduce the overall cost to the nation's libraries. Some of the desired capabilities were a comprehensive, internationally shared resource authority file with the ability for any library to add to the file. Catalogers wanted keyword access to the files in addition to the search keys and direct text string searches. The respondents wanted browsable files, especially for author/title uniform titles and links to the bibliographic file, as well as ways to easily navigate among parts of hierarchies and earlier/later names. As for maintenance, it was recognized that names change and the respondents wanted a way to automatically identify those changes that should be reflected in their local catalogs, with a fast, easy method to update and resolve conflicts. More identifying information was wanted—more dates, and there were pleas to bring back the history and scope notes. It was also clear that local libraries felt it was important for their catalogs to be customized to meet the needs of their target user groups.⁵

SO ARE WE THERE YET?

The dreams of the 1979 LITA Institutes are still with us, and the hopes of ALA's Authority Control Interest Group of the mid-1980s remain unfulfilled.⁶ Some have blamed our MARC formatted records and the online library systems built around these records. Is that what is holding us back? Or can we use the MARC Format and make it do what we want? Do we expand MARC to encode the links and relationships or come up with a new mechanism? In the Anglo-American cataloging world using the MARC Format, we used to specify the relationships or roles of persons and corporate bodies by adding a "relator" term to access points in bibliographic records (as prescribed in AACR2 rules) and sometimes the corresponding MARC code (and this is still done for some music and rare book areas), but generally the practice was abandoned as being too expensive to continue in our shortsighted efforts to cut near-term costs. We find that now inhibits our fully implementing FRBR, where such roles are essential to clear identification, so it may be time to revisit that administrative policy decision.

We've recently seen a renewed focus on user needs and the tasks users perform when using a catalog, stated again in the IFLA *Functional Requirements for Bibliographic Records*⁷ (FRBR). FRBR lists four "user tasks"—things we feel a user wants to do relative to the bibliographic universe:

Find an entity or entities in a database using attributes or relationships (Elaine Svenonius has suggested this should actually be in two parts—to locate and to collocate entities.⁸)

Identify—to confirm that the entity found corresponds to the entity sought

Select—to choose an entity meeting the user requirements for content, physical format, etc.

Obtain—to acquire an entity or to access an item
and we could add a task (as Elaine Svenonius does):

Navigate—that is, the catalog should enable a user to navigate through related materials that may be in the collection or indeed in the entire bibliographic universe.

FRBR is an evolving model and is currently being extended in the realm of authority control through the work of IFLA's Working Group on the Functional Requirements for Authority Numbers and Records (FRANAR), chaired by Glenn Patton, who will be speaking at this conference. We may find that this FRBR conceptual model enables us to meet the objectives of a catalog in new ways.

OBJECTIVES OF AUTHORITY WORK

And what about the objectives of authority work? Through authority records, catalogers in the days of book and card catalogs maintained a record of their decisions for the authorized form of a heading and the variant forms for which cross reference entries (in book catalogs) or cards (in card catalogs) were made. These records were mainly needed for larger catalogs and cataloging units to maintain consistency among multiple catalogers. The use of a consistent form of heading enabled libraries to help avoid the costly unintentional ordering or cataloging of materials already held in the library's collection. It enabled a user to save time and effort by showing the user the references to the authorized forms for headings and collocating works under a single form with references to related entities when appropriate.

The authority record documented the references made to the authorized heading and that enabled maintenance of the catalog. When a heading changed or was to be deleted (for example, when material was withdrawn from the collection and the heading no longer was needed), the "tracings" of the references were used to pull the associated reference cards or remove the references from book catalogs.

The authority records also documented the cataloger's authority work. Notes about sources that were checked to establish the authorized form were added to the authority record, sometimes both the sources where information

was found as well as sources that were checked but no information was found. Other notes for catalogers were added as needed to further explain the identity of the entity covered by that heading, distinguishing it from others with similar names.

OBJECTIVES FOR AUTHORITY RECORDS

The Web environment opens up new uses for authority records and new objectives to augment the traditional objectives. For example, as we have found through sharing the authority records from the Library of Congress online and particularly now on the Web, the sharing of the workload reduces cataloguing costs.

Our community has expanded, especially in Europe these days, where libraries are viewed with archives, museums, and rights management agencies as cultural “memory institutions.” We are moving from the stand-alone authority files of a single institution, or even from the shared online files towards a goal of sharing authority files among all communities. Shared authority information has the added benefit of reducing the global costs of doing authority work while enabling controlled access and better precision of searching.

Other objectives for authority control are:

- to simplify the creation and maintenance of authority records internationally when we can all access the authority records
- to enable users to access information in the language, scripts, and form they prefer or that their local library provides for them.

The existence of authority records for an entity also opens up new possibilities for links to other resources, like the home page for the entity described and links to digital resources such as biographical dictionaries, abstracting and indexing services, telephone directories, and other references sources on the Web. More about these objectives later.

As integrated library systems were created in the 1970s and 80s and new generations were developed in the late 1990s, we realized some of the promises of automated authority control. Even some of the early systems provided direct links between the bibliographic records and the authority files. This structure often placed a code in the bibliographic record for the authorized name found in the associated authority record and the system could pull up the authorized form for displays of the full bibliographic record. This structure also made maintenance of the headings and references much easier, as the correction needed to be made only once in the authority record and all associated

bibliographic records would display properly. Other systems without linked bibliographic and authority records also developed global update capabilities, and some still do not have it, so there is a wide variety on the market today.

Most systems display reference information to direct the user to the authorized forms of headings, but some systems do not use the authority records at all.

Many systems also offer validation of the form of the heading, matching the form entered by the cataloger in a bibliographic record to the forms in authority records and reporting back whether an authority record exists or not. These capabilities are a great help in automating authority work, but still have not gone far enough.

Web catalogs and associated integrated library systems provide the traditional authority control functions of creating and updating authority records and displaying the cross references but have primarily been seen as a tool for catalogers. As we open our authority files for access through the Internet, we find the authority file becoming a useful tool for other librarians and information professionals and even for end-users.

Catalogers and others can use the authority file as another reference tool for name variations and information to identify entities, as well as a channel for reaching bibliographic records, and from there, reaching directly to digitized resources. The records in these automated files also enable navigation to related entities.

The concept of record may disappear, but it is not yet clear what will emerge in its place. It's still easier for us to use the record construct than to make the larger mental leap to totally new structures; but they will come. We can think of a small step where the records would serve to control the various forms of names for an entity rather than having any single heading be the only authorized form (back to the 1979 LITA Institutes). We've talked about this for decades, and the Getty does it in several of its controlled vocabulary tools. For example, in the *Union List of Artist Names*⁹ there is a listing of the forms of name for an entity that have been found in various resources, brought together to use for searching and displays. The entry indicates the reference source where that form of name is used. In their search and retrieval systems, the system uses all the variant forms when generating a Web search, but there is a downside to this. The users are not yet told why they are getting all the variant forms that are retrieved—they do not realize it is the same entity.

When we control all the possible variations for the names of an entity, and we associate them with the bibliographic records for the bibliographic resources that they have some role in creating, producing, or owning, we need to explain that to the user.

Let's say I searched under Lewis Carroll, the author of *Alice's Adventures in Wonderland*. Why am I getting back information about this mathematician, Charles Lutwidge Dodgson, 1832-1898? In fact, it's the same person with two separate bibliographic identities and not a case of variant names.

You need to tell the user about variant names used for the same person or corporate body or work, or somehow convey the relationship of the variant names to the entity and related entities. Yes, we want to collocate the works of the person or corporate body, but we also want the user to understand what is going on.

STANDARD NUMBERS

Another way to enable collocation when there are several variant forms of name used by an entity, and this is a method that has been suggested many times over the years,¹⁰ is the use of standard numbers. Why not just store the language-neutral number for the entity in the bibliographic record and link that to the authority record where the display form would reside? Or, as some authors have suggested, let the user choose the form he/she wants to see. Several early integrated library systems offered and still use this technique.

In 1980 IFLA also proposed using an ISADN, International Standard Authority Data Number. There have also been suggestions for an ISAN, International Standard Authority Number, as well as the ISO International Standard Text Code (ISTC to identify works and expressions). Still another suggestion has been to just use the authority record control numbers, such as Library of Congress Control Numbers, as this unique, persistent identifier.¹¹ I personally would like to test using the unique, persistent record control numbers and to see if that works.¹² Or we might use the number assigned to an information package (i.e., a future version of what we now call a "record") for an entity under OAI (Open Archive Initiative) protocols that I will mention in a moment. That would avoid having to set up an expensive international organization to manage the distribution and maintenance of such numbers. The organizational overhead and costs are partially why the original IFLA suggestion never came to fruition.

The future prospects are tempting us to believe that we might reach some of those futuristic goals of the 1979 LITA Institutes and beyond—to allow the user to choose the displayed form, to automatically generate many of the variant forms (permuted, direct order, abbreviated, etc.—in fact some systems now offer this capability), and to link to other Internet resources, including digital objects (some systems can do this now, too), reference tools, as well as other tools in a future semantic Web.

CHALLENGES AT THE GLOBAL LEVEL

As we know there are many challenges to accomplishing authority control or even sharing authority records on a global scale. There are different cataloging rules that rightly focus on the needs of their users. There are systems we might wish to link to that have no rules at all. Clearly there is a great challenge with different languages and scripts, and there is the technical challenge of accessing and displaying records that are encapsulated in different communication formats, particularly the various “MARCs”—MARC 21, UNIMARC, RUSMARC, etc., and XML.

In the digital library world and libraries in general, there is increased recent focus on the need for interoperability. This is being proposed in many ways, including the fact that we can now map different communication formats with Z39.50 protocols (in fact the LEAF Project explores this model).

We have developed crosswalks to the “MARCs,” including crosswalks from MARC 21 to and from ONIX, and others. We have also mapped MARC 21 into an XML format. These crosswalks and mapping strategies can help us to search and retrieve library resources effectively with publishers’ databases, abstracting and indexing services, and other resources on the Web.

Over the past few years there have been several projects that help us get closer to providing authority control on a global scale. There are several sponsored by the European Union, such as the AUTHOR Project that converted a sampling of authority records from the seven participating countries to the same communication format, UNIMARC.¹³ The LEAF project that I just mentioned is looking at linking authority files for archival purposes using Z39.50 protocols and OAI (Open Archive Initiative) protocols. The <indecs> and INTERPARTY projects were looking for cooperative work among libraries, museums, archives, and rights management communities in sharing authority information. HKCAN is the Hong Kong Chinese Authority for names that provides a successful shared authority file among the libraries in the consortium, enabling romanized forms of headings and Chinese traditional and simplified character forms.

Within the International Federation of Library Associations and Institutions the *Guidelines for Authority Records and References*¹⁴ (GARR) was issued in 2001. In 1998, the IFLA MLAR (Minimal Level Authority Records) Working Group identified essential data elements needed in authority records (today we’d call these metadata).¹⁵ This work continues through the IFLA Working Group on FRANAR (Functional Requirements for Authority Numbers and Records). They are reviewing and updating the MLAR findings and recently enlisted the help of Tom Delsey in extending the FRBR model to authority records.

Within the digital metadata community, there is a Dublin Core “Agents” working group that continues to explore recommendations for dealing with authority information in the digital environment, as does the DELOS/NSF Working Group on “Actors/Roles.” The Archival community also is developing an Encoded Archival Context for authority metadata using XML.

At OCLC, discussions continue about authority records in CORC (now known as Connexion). This is a potential OCLC project that looks towards global expansion to build an authority file. Connexion now provides simultaneous creation of both MARC 21 and Dublin Core bibliographic records.

Another development over the past few years has been the acceptance of Unicode within Microsoft tools, such as the Windows operating system, that facilitates more global compatibility with multi-script capabilities. And the worldwide expansion of NACO and SACO to users of the *Anglo-American Cataloguing Rules* and *Library of Congress Subject Headings* also promotes authority control on a global scale. We will hear more about these various projects and initiatives later in this conference.

The availability of millions of authority records worldwide, multiple automated national and regional authority files, and the technological capabilities of the Internet and protocols are all coming together now, and we are really at the brink of making a virtual international authority file a reality.

NEW VIEW OF UBC

We’re also making an historic change to how we view Universal Bibliographic Control (UBC). The IFLA UBC principles for authority control are parallel to those for bibliographic control, namely that:

- each country is responsible for the authorized headings for its own personal and corporate authors (they didn’t mention uniform titles, series, or subjects), and
- the authority records created by each national bibliographic agency would be available to all other countries needing authority records for those same authors. Even more, that the same headings would be used worldwide.

In the 1960s and 1970s when this was really catching on, technology had not yet advanced to make such sharing practical on an international level. Plus, the lack of funding for an international center to manage such a program prevented that visionary concept from becoming reality. As for the same form being acceptable worldwide, the IFLA developers at that time were primarily

from North America and Europe and apparently did not acknowledge the necessity for multiple scripts when dealing with users worldwide.

For the past couple of years a *new* view of Universal Bibliographic Control has been emerging from several working groups within IFLA. This new perspective reinforces the importance of authority control, yet puts the user first. It's a practical approach that recognizes users in China may not want to see the heading for Confucius in a Latinized form, but in their own script. Similarly users in Japan or Korea would want to see the heading in *their* own script and language. National bibliographic agencies still need to have their own authority records for their own bibliographic control, but we can link them globally to create a virtual international authority file that will enable sharing of authority information and enable future displays that show a user's preferred form.

We can link the authorized forms of names, titles, and even subjects from the authority files of national bibliographic agencies and other regional agencies through a virtual international authority file. There are several models for how this might work, and we need to do more pilot projects of prototypes of these models to test which would be best to pursue.

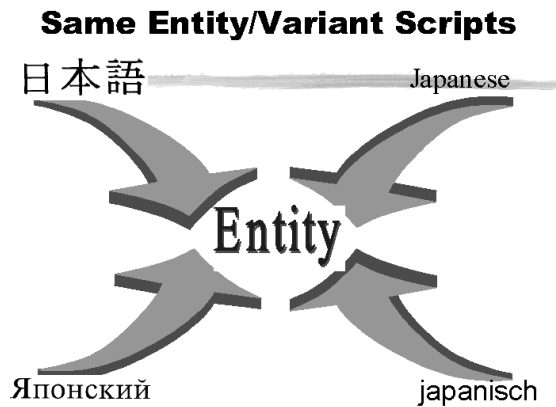
In order to be of most use to the library users in each country, the scripts should be the scripts they can read! Figure 1 shows that the names we give to an entity can be expressed in many languages and in many scripts. For example, we could write a name in English or German with a roman script, in Russian in Cyrillic scripts, or in Japanese (in any of three scripts!) and in many other languages and scripts.

Transliteration may serve as a way for some users to be able to decipher records, but much better is the accuracy of using original scripts. In fact, we should eventually be able to display the script and form of a heading that the user expects and wants.

I believe that many catalogers within IFLA realize the value of preserving parallel authority records for the same entity. This allows us to reflect the national and cultural needs of our individual users, and at the same time allows us to set up the syndetic structure of cross references and authorized forms of headings to be used in our catalogues intended for a specific audience following our own cataloging rules. It also allows us to include variants in alternate scripts, at least as cross references for now.

As we look at linking we must recognize that different cataloguing rules have differences in what they consider entities—AACR2's choices are not universal. For example, German rules (*Regeln für die alphabetische Katalogisierung*—RAK) do not recognize that a ship's logs can be under an entry for the name of the ship, so they would not have an authority record for the name of the ship. Similarly for events, for example, the meetings of corporate bodies, AACR2 creates a hierarchically subordinate heading for a meeting

FIGURE 1. Names in Many Languages and Scripts



under the name of the corporate body. The German rules would not create a heading for the “meeting.” There are also different practices for undifferentiated names—the Germans recently changed their rules to differentiate more names—they more commonly used undifferentiated forms for personal names using just initials for forenames.

However, even under the same cataloguing rules, say AACR2, when we get more information to differentiate a person, we can make a new authority record to differentiate that person from others grouped together under an undifferentiated form of name. As a result, the record for the undifferentiated name can reflect different associated entities over time.

If we agree that sharing authority information on a global scale is worthwhile, how do we get there? Several major authority files exist, built according to their own cataloguing rules and rule interpretations. We need a one-time project to link the existing records for the same entity—a retrospective matching project. One suggestion has been to use matching algorithms, such as those developed by Ed O’Neill and others at OCLC, building on bibliographic clues for machine matching at a fairly high level of accuracy. A “proof of concept” project to test this approach is underway between OCLC, the Library of Congress, and the Deutsche Bibliothek (German National Library) in Frankfurt, Germany. We would still have manual matching and checking to do, but we expect machine matching will be a great help. We could also have the computer add linking text strings and record control numbers or an entity identification number to facilitate later links and pathways to preferred forms for displays. Or we may find we do not need to specifically record these links, if our future systems are smart enough to make the links for us.

Some local systems already provide us with computer-assisted mechanisms for automatic checking of headings against an existing authority file, and we could see this expanded to launch a search against a virtual international authority file, if no match was found locally. We can also envision the capability of displaying the found matches from the virtual file for a cataloguer to edit or to merge information, if desired, into the local authority record, including capturing the information for future linking.

We can also envision extending authority control to users through display of public notes and references (as most systems do today) and through links to related resources, such as official Web sites for the entity, authoritative biographical dictionaries, and other identifying resources.

We could soon realize future switching capabilities to display forms the user wants. Some systems now provide community specific retrievals to concentrate on the subject needs of a community in selecting resources for online searches, and other systems like “my library” or “my opac” even go beyond that to specific retrievals customized for individual users. Those systems could build in the authority preferences for user preferred scripts and displays for controlled vocabularies.

We want to have the authorized form preferred by a library as the default offered to most users, but we can also envision offering user-selected preferences through client software, or “cookies” that let the users specify once what their preferred language, script, or cultural preference is—for example, spelling preferences when cultures have variations, like American English and spelling preferences in the United Kingdom (e.g., labor and labour).

Also, for example, when a Russian-speaking user comes along, the local system or the “cookies” on the user’s system, could specify he/she wants to see the Cyrillic form of headings and we could display it for them. You can also imagine displaying any script or a Braille keyboard output, or we could provide voice recognition response, built on users’ profiles or their “cookies.” This might be accomplished by putting variant forms in variant scripts all in one authority record, or it may be better to link parallel authority records that each reflect the needed syndetic structure of the cataloging rules upon which they are each based.

Within a single authority system, we might incorporate the references appropriate to those cataloging rules governing the catalog for which the authority record was intended. Let me show you how this might look applied to a Library of Congress authority record for Confucius.

Figure 2 is an example of what a Library of Congress authority record might look like with Unicode capability to include original scripts as cross references in a library’s catalog. Actually with Unicode the roman script

FIGURE 2. Authority Record for Confucius

Tag	I1	I2	Subfield Data
010			#a n 80050515
035			#a (DLC)n 80050515
040			#a DLC #c DLC #d DLC #d NIC
100	0		#a Confucius
400	0		#a Konfuzius
400	0		#a K'ung Fu-tzu
400	0		#a Kongzi
400	1		#a Kong, Qiu
400	0		#a K'ung-tzu
400	1		#a K'ung, Ch'iu
400	0		#a K'oshi
400	0		#a Konfu't'si'i
400	0		#a Kongja
400	0		#a Kung Fu
400	1		#a K'ung, Fu-tzu
400	0		#a Confucio
400	0		#a Конфуций
400	0		#a 孔夫子
400	0		#a 孔子
400	0		#a 孔丘
400	0		#a こうし
400	0		#a コウシ
400	0		#a 공자
670			#a Jakobs, P. M. Kritik an Lin Piao und Konfuzius, c1983: #b t.p. (Konfuzius)
670			#a Konfu't'si'i, 1993: #b t.p. verso (551-479 B.C.)
670			#a His Gespräche (Lun y'ü), 1910: #b t.p. (Kungfutse)
670			#a Web connection #u http://www.friesian.com/confuci.htm
700	0		#a 孔夫子 #5 HKCAN

diacritics would appear after the letter rather than before the letter shown here, but this just gives you an idea of what it would be like.

There is no particular order to the arrangement of the references, except to place the non-roman scripts following the roman scripts, but even that ordering is not necessary for the computer—it just makes the record easier for the cataloger to follow. This model shows English, Italian, German, Chinese, Japanese, Korean, Russian, and transliterations (including Wade-Giles and pinyin for the Chinese, since the Library of Congress just switched to use pinyin).

Notice also the new MARC 21 capability to include the URL for a Web page in the last 670 note field. This also shows the use of a linking 700 field to show that an authority record was located at the National Library of China and shows the form of authorized heading according to their rules. In Hong Kong

they have a regional name authority file, known as HKCAN, that uses the 7XX fields for the authorized form in the traditional Chinese script. They use Innovative Interfaces INNOPAC system and are able to use this information in OPAC displays to direct users to additional material cataloged under that alternate form. That enables bibliographic control for collocation under the name of the person or corporate body.

VIAF MODELS

So, what models might we explore for this international authority file? We currently have a *distributed model*, where a searcher would use a standard protocol like Z39.50 or soon “ZING,” the next generation of Z39.50. A recently agreed-upon extension to the Bath profile for Z39.50 will enable searching and retrieval of authority records. Through this protocol we can search the *independent authority files* of participating National Bibliographic Agencies and regional authorities.

Another model is to have one *central authority file with links to all others*. This model requires the central agency to match entities and make the links while the other participants continue to maintain their own file. A cataloger would then get access to all of the authority records for that entity worldwide by a single search of the central file.

Yet another model is to have a centralized agency coordinate the work of many participants with the centralized agency maintaining a centralized *union authority file* and libraries could contribute to it as they wished. NACO uses this model in a controlled way. It might also be a model for a more open system where any library could contribute authority records, such as has frequently been proposed for OCLC. However, such an open model also lends itself to less consistent information, unless the contributors adhere to mutually agreed upon standards and there are checks to avoid unintentional duplication of records for the same entity. In this model the user or the users’ local system would just search the single file.

A variation on this centralized union authority file model, instead of linking to one authority file, would be to link them all to a centralized server or “*virtual*” *union authority file*. It is “virtual” because the full authority records remain in the national and regional authority files with only minimal data harvested by the server. We may find that this model is the best approach in terms of record maintenance. It might employ the Open Archives Initiative (OAI) protocols with a central server that harvests metadata from the national or regional authority files. The records for the same entity would be linked at the central server. That information would be refreshed in the server whenever

there are changes in the national files. This means the day-to-day record maintenance activities continue to be managed as they are now by the National Bibliographic Agency (or regional authority).

There are many other models we could imagine. I am sure you can think of others, and we need to try them out to see which will work best in today's Internet environment.

The German National Library (Die Deutsche Bibliothek) and the Library of Congress together with OCLC have started a proof of concept project to test the centralized union authority file model using OAI protocols. We envision this project in at least four stages.

The first stage of this project began in 2002 to link our existing authority records for personal names. OCLC has matching algorithms they are testing to compare the LCNAF (Library of Congress Name Authority Files—about five million records) and the DDB's Personal Name Authority File (PND—about one million records). They use the bibliographic records and information in authority records to do this matching. We want to see how much the machine can match and how much human work will be needed. It is hoped that if this proves successful, it can be the basis for a true Virtual International Authority File.

In Stage 2, as we make the links we will be building one or more servers with this "metadata"—one will be housed at OCLC, probably another at the OCLC European office (PICA), and another at the DDB. We were not planning on having a separate server at the Library of Congress for this project.

As we continue to populate the OAI server with matched records, the user (at this stage the user would be a cataloger) would be able to check the system (probably using SiteSearch or a similar software) to see if the authority record already exists for the entity the cataloger is trying to establish.

We would hope, later on, that vendors would build in software to automatically launch a search of the VIAF if the entity was *not* found in the local authority file.

For Stage 3 of the proof of concept project, we also want to test using the OAI protocols to do the ongoing maintenance of updating the information in the server by harvesting metadata for new and updated or deleted information in the home authority files.

A possible last stage, Stage 4, would be to test the end-user-display capabilities to switch the preferred form of language displayed on his/her machine. This stage is in the future, but it reminds us of the opportunities that libraries have now to contribute to the infrastructure of the future Internet environment.

We can envision a shared international authority file being an integral part of a future "Semantic Web." You may have heard about this in the *Scientific American* article by Tim Berners-Lee, founder of the Internet.¹⁶ The idea is to

make the Internet more intelligent for machine navigation rather than human navigation of the Web. It involves creating an infrastructure of linked resources and the use of controlled vocabularies called “ontologies.” These ontologies could be used to enable displays in the user’s own language and script.

Here’s where libraries have an opportunity to contribute to the infrastructure of the future Web—we already have controlled vocabularies in our various authority files. Those would be linked with other controlled vocabularies of abstracting and indexing services, of biographical dictionaries, of telephone directories, and many other reference tools and resources, to help users navigate and to improve the precision of searches, so users could find what they’re looking for.

All of these tools would also link to their respective databases for bibliographic and other resources. For example, Library of Congress authority files would link to the bibliographic and holdings databases of the Library of Congress, and even to our digital repositories for the linked digital objects themselves.

You can see that we would also build in the search engines and future tools that as a collective resource would connect us to the entire digital world.

All of this, of course, would have built-in, appropriate security and privacy assurances, and ways to identify and acknowledge resources that we can trust and rely on, and somehow, miraculously, all of the copyright issues will be resolved. It’s great to think about the possibilities and opportunities for testing this out and to think about how we can improve upon our dreams. The Internet has brought us a new way to convey information and has opened up possibilities and opportunities that we never dreamt of even a few years ago. Catalogers can build authority records using the Web and all communities (publishers, rights management agencies, archives, museums, and other libraries) can use this information and reduce costs worldwide. Authority control will help users of the Web to benefit from collocation and search precision that authority control enables. And, very importantly, it also means we can do it in ways that are meaningful to users in their preferred language and script.

We can open up the valuable information within our authority records to users worldwide and use the authority records as tools to connect, not only to bibliographic data, but to biographical dictionaries, telephone directories, abstracting and indexing services, official Web sites for the entity, and more. The authority records can be a key part, a building block for the infrastructure of the Semantic Web and beyond.

We still need more research and testing, but we also have a lot to offer the world, and this wonderful resource, created and maintained by libraries worldwide, can offer us a fresh perspective for talking with other communities and moving together to the future.

NOTES

1. Cutter, Charles A. *Rules for a Printed Dictionary Catalogue*. [Washington, D.C., Government Printing Office, 1876] p, 10.
2. Lubetzky, Seymour. *Principles of Cataloging*. Los Angeles, Calif.: Institute of Library Research, University of California, 1969, p. 94.
3. Bregzis, Ritvars. "The Syndetic Structure of the Catalog," *Authority Control: The Key to Tomorrow's Catalog*; proceedings of the 1979 Library and Information Technology Association Institutes, edited by Mary W. Ghikas. [Phoenix, AZ], 1982, p. 24.
4. Gorman, Michael. "Authority Control in the Prospective Catalog," *Authority Control: The Key to Tomorrow's Catalog*; proceedings of the 1979 Library and Information Technology Association Institutes, edited by Mary W. Ghikas. [Phoenix, AZ], 1982, p. 166-177.
5. Tillett, Barbara B. "1984 Automated Authority Control Opinion Poll," *Information Technology and Libraries*, v. 4, no. 2 (June 1985), p. 171-178.
6. Arlene Taylor also recently discussed this point: Taylor, Arlene G. "Authority Control: Where It's Been and Where It's Going," presentation at the NELINET sponsored conference *Authority Control: Why It Matters*. Worcester, MA, 1999. Also available online at: <http://www.nelinet.net/edserv/conf/cataloging/cts/1999/taylor.htm>.
7. *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 or as html: www.ifla.org/VII/s13/frbr/frbr.htm.
8. Svenonius, Elaine. *The Intellectual Foundation of Information Organization*. Cambridge, Mass.: MIT Press, 2000, p. 17-18.
9. From its web site, the Getty says, "The Union List of Artist Names (ULAN)—is a structured vocabulary containing more than 220,000 names and biographical and bibliographic information about artists and architects, including a wealth of variant names, pseudonyms, and language variants." See online at: http://www.getty.edu/research/conducting_research/vocabularies/ulan/ (formerly: <http://www.getty.edu/research/tools/vocabulary/ulan/>).
10. Poncet, J. "Authority Files in Machine Systems," in *The Interchange of Bibliographic Information in Machine Readable Form*: papers given at the Western European Seminar on the Interchange of Bibliographic Information in Machine Readable Form held at Banbury, England, 12-16 May 1974, sponsored by the British Council, the British Library and the Library Association, R. E. Coward and M. Yelland, eds.—London: Library Association, 1975, p. 96-98. Malinconico, S. Michael. "Bibliographic Data Base Organization and Authority File Control," *Authority Control: The Key to Tomorrow's Catalog*; proceedings of the 1979 Library and Information Technology Association Institutes, edited by Mary W. Ghikas. [Phoenix, AZ], 1982, p. 1-15. Also in *Wilson Library Bulletin* (Sept. 1979).
11. For example, in the recommendations of the IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN. *Mandatory Data Elements for Internationally Shared Resource Authority Records*. IFLA, UBCIM, 1998, p. 1. Also available online at: <http://ifla.org/VI/3/p1996-2/mlar.htm>.
12. I realize that Library of Congress Control Numbers occasionally do change for the same entity, but by and large they can be considered "persistent."

13. For example, see Danskin, Alan. 1998. "International Initiatives in Authority Control." *Library Review* 47, no. 4: 200-205.

14. *Guidelines for Authority Records and References*. 2nd ed., revised by the IFLA Working Group on GARE Revision. München: K.G. Saur, 2001. 46 p.

15. *Mandatory Data Elements for Internationally Shared Resource Authority Records*: Report of the IFLA UBCIM Working Group on Minimal Level Authority Records and ISADN, chair Barbara B. Tillett, Françoise Bourdon, Alan Danskin, Andrew MacEwan, Eeva Murtomaa, Mirna Willer. International Federation of Library Associations and Institutions, Universal Bibliographic Control and International MARC Programme, 1998. 95 p. Also available online at: <http://ifla.org/VI/3/p1996-2/mlar.htm>.

16. Berners-Lee, Tim, James Hendler and Ora Lassila. "The Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities," *Scientific American*, May 17, 2001. Also available online at: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.