

UCLA

UCLA Previously Published Works

Title

The single shared catalog revisited

Permalink

<https://escholarship.org/uc/item/9xw2n0zw>

Journal

Conversations with catalogers in the 21st century, 2011

Author

Yee, Martha M

Publication Date

2011

Peer reviewed

7 THE SINGLE SHARED CATALOG REVISITED

Martha M. Yee

Ten years ago, I wrote a short piece in the *ALCTS Newsletter* called “One Catalog or No Catalog?”¹ In it, I suggested that if we did not find a more efficient way of sharing the intellectual labor of authority control, we were in danger of losing our catalogs altogether. Ten years on, we can see two possible scenarios for creating a single shared catalog that we all cooperate in keeping under authority control. One is OCLC WorldCat; currently, OCLC is trying to persuade libraries to substitute OCLC access for local OPACs (Online Public Access Catalogs). The other is the nascent discussion in our field of the possibility of putting cataloging data onto the Semantic Web, connected with a new set of cataloging rules (RDA, Resource Description and Access) that claims to be RDF-ready (Resource Description Framework-ready).²

In this essay, I first review all the reasons why sharing a single catalog would be a good idea. Then I propose some specifications or “specs” for shared cataloging (and the resultant catalogs) in the future. Finally, I examine both OCLC WorldCat and the Semantic Web approach using RDA, enumerating the pros and cons of each approach for meeting the “specs.”

BENEFITS OF THE SINGLE CATALOG APPROACH

Cataloger Efficiencies

Catalogers would no longer have to put in the time necessary to keep thousands of completely separate catalogs under authority control, nor would they have to put in the time necessary to devise and manage complicated systems of record intake and outflow, record overlay, and the like, in thousands of

completely separate catalogs. It should be noted that computer algorithms for doing this kind of matching and overlaying tend to be seriously flawed and often undo the intellectual labor of catalogers by overlaying two records that are actually for different expressions, or failing to overlay two records that are actually the same manifestation.

Intellectual Work, Not Clerical Work

Catalogers, the best and the brightest of librarians, could focus their energies on intellectual work rather than clerical and managerial work. Cataloging work would become focused on entity identification and definition, i.e., on linking items to the manifestations, expressions, and works they contain; the persons and corporate bodies who created them; the subjects they discuss, illuminate, or depict; the genres or forms they exemplify; and the disciplines in which they were created. Each time such a link was made, it would be immediately available for the use of all other catalogers, thus providing further cataloger efficiencies.

More Time Available to Catalog Neglected Works

Time saved using the single shared catalog approach could allow catalogers to catalog special collections, cuts on sound recordings, poems in anthologies, journal articles, and other types of works that have been neglected or ignored in our catalogs.

Learning Efficiencies for Catalog Users

Catalog users worldwide would have to learn only one piece of catalog searching software. We could teach children in kindergarten how to search catalogs along with their ABCs, and they could use the knowledge for the rest of their lives.

More Works Linked and Available for Users

Catalog users would benefit immensely from the more efficient use of cataloger time and intelligence, since more of the world's cultural treasure would be linked into our hierarchies of works, creators, subjects, genre/forms, and disciplines.

SPECIFICATIONS (SPECS) FOR THE SHARED SINGLE CATALOG

Note that the most important prerequisite for any shared single catalog is an agreement on standards for entity definition among the catalog's constituent libraries. This is true with both methods of creating a shared single catalog: the OCLC WorldCat single catalog, or the Semantic Web approach. We need to reach global agreement on entity definition across language and cultural

boundaries as much as possible for entities such as person, corporate body, work, subject, genre/form, and discipline (classification). A list of key definitions follows.

Definitions

Entity: Entity is the general term used to encompass all of the following things that could be sought in catalogs by catalog users.

Person: For person, the FRBR (Functional Requirements for Bibliographic Records³) definition is used, as follows:

(3.2.5): [A]n individual. . . Encompasses individuals that are deceased as well as those that are living.⁴

Corporate Body: For corporate body, the FRBR definition is used, as follows:

(3.2.6): [A]n organization or group of individuals and/or organizations acting as a unit. . . Encompasses organizations and groups of individuals and/or organizations that are identified by a particular name, including occasional groups and groups that are constituted as meetings, conferences, congresses, expeditions, exhibitions, festivals, fairs, etc. . . Also encompasses organizations that act as territorial authorities, exercising or claiming to exercise government functions over a certain territory, such as a federation, a state, a region, a local municipality, etc. . . Encompasses organizations and groups that are defunct as well as those that continue to operate.⁵

Work: For work, the FRBR definition is used, as follows:

(3.2.1): [A] distinct intellectual or artistic creation. . . Variant texts incorporating revisions or updates to an earlier text are viewed simply as *expressions* of the same work. . . Similarly, abridgements or enlargements of an existing text, or the addition of parts or an accompaniment to a musical composition are considered to be different *expressions* of the same work. Translations from one language to another, musical transcriptions and arrangements, and dubbed or subtitled versions of a film are also considered simply as different *expressions* of the same original work. . .

By contrast, when the modification of a *work* involves a significant degree of independent intellectual or artistic effort, the result is viewed, for the purpose of this study, as a new *work*. Thus paraphrases, rewritings, adaptations for children, parodies, musical variations on a theme and free transcriptions of a musical composition are considered to represent new *works*. Similarly, adaptations of a *work* from one literary or art form to another (e.g., dramatizations, adaptations from one medium of the graphic arts to another, etc.) are considered to represent new *works*. Abstracts, digests and summaries are also considered to represent new *works*.⁶

Expression: For expression, the FRBR definition is used, as follows:

(3.2.2): [T]he intellectual or artistic realization of a *work* in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms. . . . The boundaries of the entity *expression* are defined . . . so as to exclude aspects of physical form, such as typeface and page layout, that are not integral to the intellectual or artistic realization of the *work* as such. When an *expression* is accompanied by augmentations, such as illustrations, notes, glosses, etc., that are not integral to the intellectual or artistic realization of the *work*, such augmentations are considered to be separate *expressions* of their own separate *work(s)*. Such augmentations may, or may not, be considered significant enough to warrant distinct bibliographic identification. Inasmuch as the form of *expression* is an inherent characteristic of the *expression*, any change in form (e.g., from alpha-numeric notation to spoken word) results in a new *expression*. Similarly, changes in the intellectual conventions or instruments that are employed to express a *work* (e.g., translation from one language to another) result in the production of a new *expression*. If a text is revised or modified, the resulting expression is considered to be a new *expression*. Minor changes, such as corrections of spelling and punctuation, etc., may be considered as variations within the same *expression*.⁷

Manifestation: For manifestation, the FRBR definition is used, as follows:

(3.2.3): [T]he physical embodiment of an *expression* of a work. . . . As an entity, *manifestation* represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form. . . . Whether the scope of production is broad (e.g., in the case of publication, etc.) or limited (e.g., in the case of copies made for private study, etc.), the set of copies produced in each case constitutes a *manifestation*. All copies produced that form part of the same set are considered to be copies of the same *manifestation*. The boundaries between one *manifestation* and another are drawn on the basis of both intellectual content and physical form. When the production process involves changes in physical form the resulting product is considered a new *manifestation*. Changes in physical form include changes affecting display characteristics (e.g., a change in typeface, size of font, page layout, etc.), changes in physical medium (e.g., a change from paper to microfilm as the medium of conveyance), and changes in the container (e.g., a change from cassette to cartridge as the container for a tape). Where the production process involves a publisher, producer, distributor, etc., and there are changes signaled in the product that are related to publication, marketing, etc. (e.g., a change in publisher, repackaging, etc.), the resulting product may be considered a new *manifestation*. Whenever the production process

involves modifications, additions, deletions, etc., that affect the intellectual or artistic content, the result is a new *manifestation* embodying a new *expression* of the *work*.⁸

Subject: In this paper, subject is used to mean what a work is about. Thus it could potentially encompass any of the entities above, as well as concept, object, event, and place as defined in FRBR 3.2.7 to 3.2.10.

Genre/Form: In this paper, genre/form is used to mean any category that a work falls into. Examples would include novels, poetry, gangster films, or dictionaries.

Discipline (Classification): In this paper, discipline is used to refer to the perspective taken toward a particular subject in a given work. For example, the subject "water" could be discussed from the perspective of a hydrological engineer, or it could be described from the perspective of a poet.

The "specs" for a single shared catalog would include the following:

Spec 1

When a cataloger changes either an entity definition or a preferred heading for a particular entity or adds variant name access for a particular entity, it should be possible to make this change in one place rather than multiple places. In our current environment, another way to say this would be that it should be possible to change a single authority record and have the change automatically reflected in all other records that cite that particular entity, without any additional cataloger labor being necessary.

Spec 2

For any given entity, it should be possible to designate different preferred forms of heading for different languages and for different cultural settings (e.g., a public library versus a technical special library) so that computer programs can automatically supply the correct preferred forms in each different environment.

Spec 3

It should be possible to provide subject access to entities beyond works-expressions-manifestations. For example, the ideal single, shared catalog should be able to provide subject access to persons in order to allow a researcher to assemble a list of, for example, women authors of the nineteenth century who published in England, and then go from there to a list of all of the works by such authors found in the catalog. It should also be able to provide subject access to corporate bodies, to allow a researcher to assemble a list of government agencies around the world that regulate immigration, for example. Similarly, it should be able to provide subject access to proper names of geographic entities, to allow a researcher to assemble a list of lakes found in the state of Wisconsin, for

example, and then go from there to a list of all works about those lakes found in the catalog.

Spec 4

When a user (or a cataloger) searches for an entity in such a way that there are multiple potential matches, it must be possible to display an alphabetical list of results consisting of the entities matched, with the list containing a single heading for each entity matched. In other words, the same entity should never appear in the list twice under two different forms of name. Ideally, this heading would be the “name commonly known” for each matched entity in the language, script, and/or transliteration desired by the user (or cataloger).

Spec 5

When a user (or a cataloger) searches for an entity using one or more keywords, all entities described by those keywords should be produced in the search results, even when the keywords appear in variant name fields for the entity, such as cross-references in authority records or variant title fields in bibliographic records. This is particularly critical for work searching. A user (or a cataloger) should be able to search for a work combining a variant name for the author and a variant title and still succeed in retrieving the work. This does not happen in any current systems.

Spec 6

All hierarchies should be made available to users (and catalogers) for exploration. These include the following:

- Classification hierarchies
- Broader, narrower, and related term hierarchies in subject headings
- Cross-references to main headings (corporate names, personal names, subject headings) should be shown to any user whose search matches the variant form of the main heading plus a corporate subdivision, a subject subdivision or work title appended to a personal name, or a subject subdivision
- Work-expression-manifestation-item hierarchies for any work, expression, manifestation or item desired by a user

Spec 7

Once a user (or a cataloger) has chosen a particular work of interest and wishes to survey the available expressions, the user or cataloger should be able to request an arrangement of the available expressions in any of the following orders or suborders:

- Alphabetical by language
- Alphabetical by surname of editor
- Alphabetical by surname of translator
- Alphabetical or numerical by edition statement
- Alphabetical by publisher
- By initial publication date in ascending or descending order

Spec 8

If the user (or the cataloger) interested in a particular work wishes to survey works related to that work, it should be possible for the user to request an arrangement of the available related works subgrouped by media, e.g., motion pictures based on the work, television programs based on the work, plays based on the work, etc.

For more detailed specs, please see my earlier work, *Principles for the Display of Cataloger-Created Metadata*.⁹

OCLC WORLDCAT

Unfortunately, in many ways, the choice of OCLC WorldCat as presently constituted would amount to the “no catalog” rather than the “one catalog” choice presented above, for the following reasons:

1. OCLC WorldCat is not under authority control. Many entities are represented by more than one form of heading in OCLC WorldCat.
2. Users (non-catalogers) are not even allowed access to the cross-references and scope notes found in authority records, let alone given default (or any) searches that match on authority records.
3. Users (non-catalogers) are not allowed to browse headings; only a keyword search of bibliographic records (i.e., manifestation records) is allowed.
4. It is not possible to search for a particular work using variants of both the author’s name and the title, with matching being done on authority records; and then to display works matching the search query in order by work identifier (main entry).
5. The default display for large search results is, from my point of view, in “no discernible order,” which OCLC inexplicably calls “relevance” order. Whatever the order is, it has no relevance to me! There is limited sorting capability as an option. However, the user must choose to sort search results either by author or by title; a work sort (main entry, author-title sort) is not an option.
6. Uniform titles are not used to organized displays.

7. In search result displays for a particular work, expressions of the work are not differentiated from works that are about that work, or works that are related to that work.
8. Users are not shown important expression information in bibliographic record displays. Part of this is OCLC's fault for truncating displays of individual records, but part of it is the fault of OCLC-contributing libraries. Many OCLC-contributing libraries (especially public libraries) operate under the theory that users just want works and aren't interested in expressions, so they leave expression-identifying information out of their cataloging records. Examples of expression-related data that is often missing from the bibliographic records themselves include illustrator, translator, and editor statements of responsibility, series, and edition statements.
9. OCLC requires that a new record be made for every manifestation change (format change or change in distribution information); the problem with this is that there is no way for a computer to tell when a new record has been made for a new expression and when it has been made for a new manifestation that is an exact copy of another manifestation in the catalog. As a further complication, OCLC now even encourages making a new record for the same manifestation of the same expression of the same work for each different language of cataloging. Paradoxically, despite this worship of the manifestation, the record-merging algorithms that are applied to the batchloading of records ignore expression and work identification data in bibliographic records to such an extent that two different expressions of the same work, or even two different works, can easily be merged.

It is a real shame that the developers of OCLC have so far failed to mine the treasure to be found in our records, particularly in our authority records, but also the expression information to be found in bibliographic records. There are signs that this might be changing. The WorldCat Identities project begins to demonstrate what OCLC could accomplish if it figured out better ways of providing users with access to authority records first (instead of the current emphasis on bibliographic records). After all, the entities of most interest to users (works, persons, corporate bodies, subjects, genre/forms, and disciplines [classification]) are all represented by authority records, not by manifestation-based bibliographic records. In this connection, it should be noted that in the list of the OCLC Top 1000¹⁰ (the works held by the most libraries), none are single edition works by single work authors.

RDA/SEMANTIC WEB

The Semantic Web with RDA cataloging rules is the second potential method that might eventually be used to create a single shared catalog. All assertions below concerning RDA should be accompanied by the following caveat.

The last draft of RDA released for the review of the cataloging constituencies in November 2008 was subsequently modified by the JSC (Joint Steering Committee for Development of RDA) before being submitted to the publishers, and the true final draft as submitted to the publishers is currently not accessible to anyone other than the RDA developers themselves. My assertions below are based on the November 2008 draft¹¹ and may be inaccurate concerning the draft finally published.

The premise of the Semantic Web seems to be that we would assemble records—or, more accurately, perhaps, displays and indexes—from bits of data scattered throughout the Internet. Can this work? Is the Internet fast enough to do it? RDA seems to take it on faith that a huge increase in granularity is a good thing without any prior experimentation to demonstrate how these tinier and tinier bits of data will be reassembled into coherent displays and indexes.

It appears that RDF resists hierarchy. Hierarchy is one of our main tools for allowing users to navigate vast amounts of information efficiently. RDF-based RDA, for example, expects relationships to be demonstrated by means of one-to-one links (no hierarchy) rather than one-to-many (hierarchical) links. With our current cataloging rules (AACR2), when we catalog a parody, we link it up to the work it parodies, but we don't make a link in the opposite direction. Currently, when a person searches for the work, it is possible to display all the parodies of that work by means of this one-to-many link. In RDA, not only will we link a parody to the work parodied, we will also be required to make a link to the parody from every manifestation and expression of the work parodied that we have already cataloged. Thereafter, every subsequent manifestation and expression of that work must have a link inserted that leads to the record for the parody. This hierarchy-resistant approach is simply not efficient for either the cataloger or the catalog user.

I fear that RDF not only resists hierarchy, but may also resist the provision of ordering of elements. It is clear that RDA completely removes display from the rules. Since, from the catalog user's point of view, cataloging is display design, RDA seems successfully to have removed cataloging from the cataloging rules. Everything we call cataloging (effective indexing and effective displays) is pushed out of RDA and into "application" or "implementation."

I also fear that RDF resists linking related data elements, such as all of the data elements (place, publisher, date) pertaining to the publication and/or distribution of a particular manifestation, or such as the data elements (number, type of instrument) pertaining to instrumentation of musical works or arrangements. For example, we may need to associate one place of publication with the publisher and another place of publication with the distributor. In order to deal with this, RDA has become more granular, defining place of publication separately from place of distribution and place of manufacture. It now also requires that all resources be given information about publication, even those that are not published, such as motion pictures. Even with this solution, there are still problems when a piece of data is repeatable and another piece of information needs

to be associated with it, as when there are two publishers, each with its own place of publication.

As another example, to express a complex instrumentation for a particular musical work, we may need to associate more than one number with more than one instrument, e.g., 2 two organs and 5 accordions (not 2 accordions and 5 organs). If number cannot be linked to type of instrument, the user interested only in music for 2 accordions might have to retrieve and look through music for some other number of accordions and 2 organs, pianos, violins, etc. RDA development constantly ran into problems caused by the inability of the RDA model to allow linking of related data elements such as these. This cannot be seen as simply a “display” concern, when failure to solve it will prevent the design of effective display. Rather, this is a concern with the model itself, which does not reflect the reality of what a catalog is, what kind of intellectual work it does, and what function it serves.

I have been assured that RDF does not prevent collecting the same piece of data in two or more different forms, e.g., a transcribed form for the publisher, a controlled form for the publisher, and a supplied note explaining something about the publisher. I am still skeptical about whether this is possible, though, partly based on the fact that RDA seems to avoid collecting the same piece of data several different ways. In fact, RDA is completely ambiguous about the function of a particular piece of data. Will this data function as a note, or will it function as an access point (i.e., a heading in an index)? We frequently cannot tell from the rules themselves. RDA’s excuse is always that this is a “display” concern, and RDA does not deal with display. Because RDA does not deal with display, examples initially provided in the first drafts of RDA were unreadable and had to be converted to ISBD before any of the constituent reviewers could make any sense of them.

Can all bibliographic data be reduced to either a class or a property with a finite set of values? Can everything that catalogers do be reduced to a set of pull-down menus? Cataloging is a discursive art, a kind of descriptive writing. It is not simply the coding of data.

Skepticism is critical; we *cannot* adopt Semantic Web standards such as RDF until there have been demonstration projects that show it is possible to take RDF data and produce a catalog, with effective indexing and effective displays, according to our specs (above).

That said, the idea of converting our current shared creation of manifestation-based bibliographic records into shared creation of entity definitions (URIs, or Uniform Resource Identifiers) is very appealing. The URI for a particular entity such as a work, a person, a corporate body, a subject, a genre/form, or a discipline (classification) could then be used as a node for all variant names for that particular entity in every known language and script. This, in turn, could enable designers of indexes and displays to build in features that would allow users to select a particular language, script, even educational level or disciplinary perspective (e.g., kindergartner versus physicist), which could then determine what forms of name are preferred for that user in displays of

multiple matched entities. Users could be allowed to search directly for the entities they seek, such as a particular work or a particular creator, unlike the current situation in which catalog systems force users to search for particular manifestations (no matter what their actual entity interest); and unlike the current situation, in which users' searches fail when they do not choose to search under the variants that happen to appear on a particular title page for a particular manifestation that is in the system being searched.

Perhaps we need either to adapt RDF or to spearhead a more sophisticated way to encode data on the Internet in order to create record-like structures, each with a URI, each representing an entity (works, persons, corporate bodies, subjects, genre/forms, and disciplines), and each clustering together all of the information we want to collect about that entity, including variant names. Admittedly, I'm feeling my way in the dark here, though, in an effort to provoke true RDF/Semantic Web experts to suggest solutions to the problems I'm identifying.

ADDITIONAL RDA PROBLEMS

There is one way in which RDA (and FRBR before it) imposes a rigid hierarchy that does not correspond to bibliographic reality and that, contrary to what Tom Delsey claims, I don't believe is required by RDF. Both FRBR and RDA make the assumption that each element of the bibliographic description should correspond to one and only one level of the FRBR Group 1 (work-expression-manifestation-item) hierarchy. Subsequently, adjustments had to be made to this assumption. For example, it is patently clear that a manifestation can have a different title from its expression, and an expression can have a different title from its work. The model underlying RDA is still riddled with problems due to this flawed reasoning. One example lies in the rules on relationships that assert that a film work can have a director, but only a film expression can have an editor. As another example, any data that is transcribed from a title page is linked only to manifestation, despite the fact that most of the elements of the bibliographic description were put there originally because of their value in differentiating expressions. Examples of transcribed statements and other manifestation-linked data that should be linked to expression include statements of subsidiary responsibility, such as illustrator and editor statements; edition statements, such as 2nd rev. ed.; and statements of extent, which have proven in past experiments to be one of the most reliable indicators of content change between expressions. These are all rigidly linked to manifestation, rather than expression in RDA (following FRBR's tables, which do not correspond to FRBR's own definitions of the group 1 entities). As an aside, in the latest RDA-FRBR mapping table,¹² there does seem to be some wavering with regard to extent, which is linked to "manifestation/expression," with a question mark, but in the structure of RDA itself,¹³ extent is still covered in the "describing carriers" chapter.

RDA has completely turned its back on the opportunity to call for identifying entities by the name commonly known in the community of the catalog, e.g., English in the United States, Russian in Cyrillic in Russia, and Chinese in

the vernacular in China. In the past, we thought we had to impose foreign names on English-speaking users in order to share our cataloging internationally, even though that practice violated the principle of using the “name commonly known.” Now, however, we are closer and closer to having a linked international authority file that would allow the sharing of cataloging internationally without requiring the violation of cataloging principles. VIAF¹⁴ (Virtual International Authority File) gathers together the authority records from many different national libraries, and links together the authority records that pertain to the same entity. Currently, its function is primarily to serve authority-control librarians, but it could potentially support a system to serve catalog users directly.

Because of this missed opportunity in RDA, U.S. users will not only have to find works by and about the KGB under the Russian name, as they do now, they may now have to find it under the Russian name in Cyrillic (depending on decisions made at the Library of Congress about “the preferred script of the agency creating the data” (RDA 9.2.2.5.3¹⁵ and 11.2.2.12¹⁶)). They may now have to find works by and about Mao Tse-Tung under his name in Chinese characters.

RDA does have room for rules governing subject headings, genre-form headings, and classification, but these have not yet been developed, so we can’t see yet how well RDA will be able to satisfy our “specs” in this regard.

One of the main recommendations made in Toronto at the launch of the project to transform AACR2 into RDA was that RDA be designed to solve the multiple-version problem by providing at least the option of creating an expression-based record from which manifestation records could be hung. At the end of all these years of development, we can see that RDA looks exactly like AACR2 in this regard, with a manifestation-based bibliographic record containing elements that describe both manifestation and expression, and with no attempt to tease these two very different types of data apart so that computers could compose hierarchical displays of all of the expressions of a work, and all of the manifestations of an expression.

It appears that the designers of RDA have completely forgotten about the necessity for a catalog to provide the catalog searcher with lists of matched entities to browse through. The cavalier statement that the scenario 1 (relational, object-oriented database) implementation of RDA can be done with no access points whatsoever¹⁷ implies that the problem of designing displays of multiple matched entities is being completely ignored by RDA. This kind of approach assumes that every user can design a search that produces one and only one entity. It totally ignores the reality on the ground of users with incomplete and inaccurate information, faulty memories, and vague descriptions. It completely ignores the problems we are all having now in systems like Google that produce millions of hits and then cannot organize these hits into any useful order. For this reason, the specs in this paper prominently require the ability to create ordered displays of multiple entities using names commonly known. For both OCLC and RDA to ignore this user need is scandalous.

CHANGE OF NAME IS CHANGE OF IDENTITY?

I'd also like to raise the following question: Our current principle is that "change of name is change of identity." Recent changes in pseudonym rules (via LCRI (Library of Congress rule interpretation) 22.2 and RDA itself) are fragmenting the person entity further and further. There is no research to indicate that this principle corresponds to the way library users perceive the entities they seek, and much anecdotal evidence suggests that it runs counter to their perceptions. Show me a biography or corporate history that is about a "bibliographic identity" rather than being about a person or corporate body across any name changes that might have occurred (e.g., marriage, change in stage name, trial-and-error attempt to create a more distinctive pseudonym, change in corporate name over time, etc.).

Of equal importance, this principle surely makes it more difficult to share entity definition across language and other cultural boundaries. There is no guarantee that a change of name in one language will be mirrored in every other language. If we want to let users choose language and script, surely it would be more efficient and more logical to group all variant names for an entity together and then characterize them as to language, script, user community, etc.

The reason behind the principle was never user need (other than the temporary convenience of public library users who don't want to be bothered by having to know about the writer behind a pseudonym on a work of pulp fiction), but was always for the convenience of the cataloger. Ironically, online systems already automatically provided users with the ability to search under the forms of name on title pages, since the default search in most online systems is a keyword search of bibliographic records. Instead of using authority records and uniform headings in bibliographic records to provide an alternative kind of access that collocated works regardless of forms of name on the title page, we foolishly decided to duplicate the effect of title page keyword searching in our authority records. The excuse was that users interested in collocation could use the "see also" references in authority records to carry out collocation. However, most current systems never provide users with access to authority records or to the "see also" references at all. Thus, collocation of all of the works of a given author is removed as an option from most current systems.

In order to have a future, it is crucial that we libraries differentiate ourselves from Google, Amazon.com, and similar commercial outfits. We offer intellectual lumping (gathering together and ordering) of everything having to do with a given entity, as opposed to Google's and Amazon's destructive splitting of everything under the language used in the documents indexed.¹⁸ This gathering and ordering function is one of the primary things we have to offer society. If we start splitting in imitation of Google, we are offering no added value, and we are much more expensive.

If we maintain the transcription principle, it ought always to be possible to use free text searching to search for forms of name on items, leaving entity records free to group everything by or about an entity together regardless of form of name on particular items. Perhaps it is time to reexamine this principle?

WHAT WILL WE NEED TO AGREE ON IN ORDER TO SHARE A SINGLE VIRTUAL CATALOG?

I would say that much of the current chaos evident in OCLC is due to our inability to agree on common standards for naming and identifying works, expressions, and manifestations. Although we do have standards for naming works that are either anonymous classics or works of single personal authorship, other works are not consistently named. The object of a bibliographic record, which in the nineteenth century would have been an expression, in the twentieth century became either a manifestation or an expression, without much theoretical consideration on our part.

Some of those in our field who welcome the Semantic Web seem to think that it will be enough just to throw bits of data onto the Internet. I don't see how this can work to create the hierarchical displays our users need unless the bits of data are rigorously tied to FRBR entity levels. In order to share a single virtual catalog, surely we will need to agree with each other on entity definition. In the past, our *de facto* entity definition could be deduced by our practices regarding the object of a record. Most of the FRBR entities are represented by authority records (works, persons, corporate bodies, concepts, objects, events, places). The object of a bibliographic record, which used to be an expression, became willy-nilly the manifestation in the twentieth century, when various reproductive techniques became common. Semantic Web designers insist that records will not exist on the Semantic Web. If that is the case, I fail to see what the bits of data on the Semantic Web can cluster around if not the FRBR entities. What are catalog users seeking in the Semantic Web if not bibliographic entities such as authors, works, and works on a subject? What will we display to them if not clusters of data that represent a sought entity? If the Semantic Web allows us to do that more effectively than we have been doing heretofore, I doubt we will regret the loss of the "record." It may even have the salutary effect of refocusing catalogers' and system designers' attention on the entities users are most interested in (works, authors, and subjects) and away from the entities that have dominated our attention heretofore (physical manifestations acquired, paid for, and shelved in libraries). In effect, the display to the user of the data pertaining to the entity of interest will become the record. If we cannot agree on entity definitions that match the entity definitions of our users, however, it is hard to see how we can provide any kind of coherent display of these bits of data to our users.

In addition, in order to provide users with browsable displays of all of the entities that match their search queries, surely we will also need to agree on names commonly known in the different language and cultural contexts from which catalog users come. Otherwise, we will not be able to provide a name for each entity matched in the form of a browsable list.

CONCLUSION

I believe society will benefit immeasurably from having one catalog to search, as opposed to having no catalog to search. The benefit will be considerably weakened if that one catalog is not a good catalog, however. We library catalogers have more than a century of experience with cooperating. Surely we could cooperate to devise electronic solutions to the problem of building a single user-friendly virtual catalog over the Internet, the maintenance of which we could then share.

NOTES

1. Martha M. Yee, "Viewpoints: One Catalog or No Catalog?" *ALCTS Newsletter* 10, no. 4 (1999):13–17, <http://repositories.cdlib.org/postprints/3084/> (accessed May 11, 2010).

2. Martha M. Yee, "Can Bibliographic Data Be Put Directly Onto the Semantic Web?" *Information Technology and Libraries* 28, no. 2 (2009): 55–80, <http://repositories.cdlib.org/postprints/3369> (accessed May 11, 2010).

3. International Federation of Library Associations and Institutions (IFLA), "Functional Requirements for Bibliographic Records: Final Report, as Amended and Corrected through February 2009" (n.p.: IFLA, 2009), http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf (accessed May 10, 2010).

4. *Ibid.*, 25.

5. *Ibid.*, 25–26.

6. *Ibid.*, 17–18.

7. *Ibid.*, 19–20.

8. *Ibid.*, 21–22.

9. Martha M. Yee, "Principles for the Display of Cataloger-Created Metadata, February 15, 2002 Draft," <http://www.slc.bc.ca/yee.pdf> <http://slc.bc.ca/yee.pdf>

10. "OCLC Top 1000," www.oclc.org/research/top1000/default.htm (accessed May 10, 2010).

11. "RDA: Full Draft, 11/24/08," RDA: Resource Description & Access Toolkit, RDA Constituency Review, <http://www.rdatoolkit.org/constituencyreview> (accessed May 5, 2010).

12. "RDA to FRBR Mapping, 5JSC/RDA/RDA to FRBR Mapping/ Rev/3, 1 July, 2009," <http://www.rda-jsc.org/docs/5rda-rdafrbrmappingrev3.pdf> (accessed May 10, 2010).

13. "RDA Prospectus, 5JSC/RDA/Prospectus/Rev/7, 1 July, 2009," <http://www.rda-jsc.org/docs/5rda-prospectusrev7.pdf> (accessed May 10, 2010).

14. "VIAF: the Virtual International Authority File," <http://www.viaf.org> (accessed May 10, 2010). See also Janifer Gatenby, "Identifiers and GLIMIR. OCLC Symposium for Publishers and Librarians, 18 March 2009," http://www.oclc.org/multimedia/2009/files/GatenbyJ-Identifiers_GLIMIR.pdf (accessed November 10, 2009).

15. "RDA: Full Draft, 11/24/08," "9. Identifying Persons. 9.2. Name of Person. 9.2.2. Preferred Name for the Person. 9.2.2.5. Different Forms of the Same Name. 9.2.2.5.3. Names Written in a Non-Preferred Script," 10.

16. *Ibid.*, "11. Identifying Corporate Bodies. 11.2. Name of the Corporate Body. 11.2.2. Preferred Name for the Corporate Body. 11.2.2.12. Transliteration," 21.

17. Association for Library Collections and Technical Services, Cataloging and Classification Section, Committee on Cataloging: Description and Access, "Minutes of the Meeting held at the 2008 ALA Midwinter Meeting in Philadelphia, Pennsylvania, January 11, 12, and 14, 2008," <http://www.libraries.psu.edu/tas/jca/ccda/docs/min0801.pdf> (accessed May 10, 2010).

18. "Lumpers and Splitters," Wikipedia, http://en.wikipedia.org/wiki/Lumpers_and_splitters (accessed May 10, 2010).