Looking to the Future: Information Systems and Metadata

Presented by Dr. Barbara B. Tillett
Chief, Policy & Standards Division, Library of Congress
January 23, 2009

I'm here to share with you my own personal views on where we are going with RDA and my ideas about the sharing of metadata or information and the need to develop better information systems for the future.

RDA: Resource Description and Access is designed as a content standard for the digital environment, but will also function to create bibliographic descriptions and authority data for any cataloging scenario – whether for book or card catalogs or current integrated library systems or systems of the future. It's based on the FRBR (Functional Requirements for Bibliographic Records) and FRAD (Functional Requirements for Authority Data) conceptual models and on the IFLA International Cataloging principles.

I'm speaking today with the assumption that you already know something about FRBR and about the work going on with cataloging principles. If you'd like to know more in that area, see the links at the end of the handout.

You do not need to know the details of the FRBR conceptual model, just as you don't need to know the electronics behind using your cell phone, but it is helpful to know some basics and to understand the terminology so we as catalogers can work with system designers to build future systems to create and mine metadata to connect our users to our collections. FRBR and FRAD models describe the relationships and connections in our bibliographic universe that in turn will enable users to navigate through this universe to things they need or may like to know about.

You may also be interested in a recent presentation by Rob Walls from Libraries Australia on implementation scenarios for RDA – the URL is in the handout[1] and the FAQ from the National Library of Australia on changes to systems with RDA.

Internet

- **Catalogs are no longer the end points in isolation**
  - Global access to data
- **Integrate bibliographic data with wider Internet environment**
  - Share data beyond institutions

The evolution of technologies took a major turn with the creation of the Internet. Catalogs are no longer just stand-alone, end points in isolation, like book catalogs, card catalogs, or stand-alone OPACs of the past. Catalogs and especially bibliographic data from any source can now be integrated into the wider Internet environment. New kinds of links can be made, new displays can be generated for users from data packaged in new ways – all of it on a global scale in multiple languages and scripts. We now have the technology to provide global connection anywhere that computers can operate – that includes the digital connections of cell phones with Internet connections.

The computer systems can know where the user is located based on the GPS (global positioning system) and use that information to suggest nearby libraries that have the resources they select to use. OCLC's WorldCat already has a variation of this location specific feedback to users with Google's "Find in a Library" feature.

The new cataloging code, RDA, is being designed to prepare us for the technological capabilities of the Internet, today and into the future by having us identify the entities and relationships at the element level that machines can use better than they have been able to in the past in our MARC records. However, RDA will also work when we package the elements in MARC records as a transitional period. I have a slide that I hope will make this clearer in a moment.
RDA is not an encoding system or a presentation standard for displays, but instead specifies the elements and some values to be used to describe the things in our bibliographic universe - resources, persons, corporate bodies, etc.

The information systems and content in the future may all be freely accessible on the Web – I imagine it as something like the Internet cloud computing that we have today with Amazon, Google, and other systems – (this cloud computing image is based on one from Wikipedia) – where the elements that describe our resources are available to libraries and users everywhere in the world – the data may come from publishers, from the creators of the resources, from libraries and other institutions, or anywhere, and is accessible by any user anywhere at anytime.

Bibliographic data and digital resources are on the web now and we've started adding the controlled vocabularies to help identify resources – such as the controlled values for naming the types of content, types of carriers, and other elements in RDA that have controlled lists of values- they are already being registered on the Web and can be used to present displays and show pathways to related resources. Maybe Amazon or Google or OCLC will build the applications while libraries build and maintain the infrastructure of data packages, much as we are today.

3

I think we're at a crucial time for the development of new information systems, more global in nature, more Internet oriented, that can make cataloging easier and make the results of cataloging much more flexible and useful to our users.

RDA is taking the first steps in the right direction. We are building the infrastructure, preparing the building blocks to get there. But we can expect a gradual change over several years until there is consensus on this common purpose. Standing still with AACR2 is not an option if libraries are to remain viable.

Since the Toronto conference on the future directions for AACR in 1997, the JSC has been working to build the foundations needed for future resource discovery systems. <click> We have the FRBR and FRAD conceptual models with user tasks, <click> we have the International Cataloging Principles from IFLA giving us principles and objectives and some starting rules, and <click> we have RDA building on those foundations to give guidance on identifying entities and relating them; and some lists of controlled terms to use for some elements – and specific elements needed to identify entities. For now we have the MARC format, Dublin Core, MODS and MADS and some other schemas to package the data and we have crosswalks among the various communication schema, but we do not yet have agreed <click> data models or <click> new systems to help us reach the maximum benefits from these changes to <click> deliver information to our users in new ways.

Authority Record

```
000 01553cz a2200241n 450
001 4407626
005 2008080108465
008 891129nl acann
010 __ la n 8822807
035 __ la (DLC)n 8
040 __ la DLC la D
100 10 la Miller, Mi
400 10 la Miller, Mi
400 10 la Miller, Mi
400 10 la Miller, Mi
670 __ la His Ventu
670 __ la Ph. call to
670 __ la His 10 mi
670 __ la His 10 mi
670 __ la Your first
```

Bibliographic Record

```
000 01272cam a2200313 a 450
001 15346181
005 20081031140931.0
008 080627t20082009inua 001 0 eng
906 __ la 7 lb cbc lc orignew ld 1 le ecip lf 2
925 0_ la acquire lb 2 shelf copies lx policy
010 __ la 2008028824
015 __ la GBA894004 l2 bnh
016 7_ la 014680487 l2 Uk
020 __ la 9780789738035
020 __ la 0789738031
035 __ la (OCoLC)223934946
050 00 la TK5105.88813 lb .M56 2009
082 00 la 004/.36 l2 22
100 1_ la Miller, Michael, ld 1958-
245 10 la Cloud computing : lb Web-based
260 __ la Indianapolis, Ind. : lb Que, lc 200
300 __ la xiii, 293 p. : lb ill. ; lc 24 cm.
500 __ la Includes index.
650 _0 la Web services.
650 _0 la Cloud computing.
650 _0 la Electronic data processing lx Distr
035 __ la (OCoLC)ocn223934946
040 __ la DLC le DLC ld BTCTA ld BAKE
```

```
001    14919759
030    DLC
005    20090121141236.0
014    2008028824
561 0# $aFrom the collection of S. Smith.
```

Holdings Record

For the transition period, we can continue to use MARC authority records as a place to put the attributes for entities that are typically related to many other thins, like the set of names and identifying information for a person that can play different roles with respect to works, expressions, manifestations, an items.

For the transition period we have <click> MARC bibliographic records that have identified various attributes – RDA's elements – of a given resource, that is the manifestation and its embodied works and expressions.

For the transition period we have <click> MARC holdings records that primarily identify specific inventory of items and the physical pieces within a particular collection that exemplify the manifestation being described in the bibliographic record.

For the most part, MARC does a good job of labeling the elements, although in some cases, like names of persons,<click> the level of granularity perhaps is not what may be needed in the future – particularly for machines to help us optimally parse the data (for example MARC does not separately provide a subfield for the surname from the given name and it cannot always be assumed by the presence or absence of a comma), and there is a MARC-RDA task group helping identify where it would be most helpful to adjust MARC to enable future evolution.

"Bridge" Period

- **Mapping tables for RDA and MARC, Dublin Core, MODS/MADS, and ISBD**
- **Decisions in PCC and local choices for alternatives and options, if needed – documented in *RDA Online***
- **Development of RDA Workflows and other training materials (e.g., changes from AACR2)**

6

We already have some tools in RDA to help during this transition period: <click> There are tables that show the crosswalks or mapping between the RDA elements needed for identifying each entity and the corresponding tags or elements in MARC, Dublin Core, MODS and MADS, and elements in ISBD.

And we expect this year to document the decisions for the Program for Cooperative Cataloging regarding their choices for RDA's alternative instructions and options if PCC decides they wish to require consistency among that program's members – those decisions can be documented in RDA Online and can be used to show a PCC view of RDA.

We also expect more "RDA Workflows" <click>to be available to provide a simple step by step approach for catalogers to build records. There also will be more training materials – for example we already have the 5JSC/Secretary/7 document on the JSC Web site that shows the changes in RDA from AACR2 – where the rules differ.

This bridge period still uses MARC records and is just one scenario where RDA will work.

I mentioned RDA includes some elements for which there are controlled lists of terms or values - the JSC considers all such lists to be open. In RDA we document the valid terms as they are established – or an individual bibliographic agency may choose to document their own terms for their user community – for example in a different language.

The terms used for the categories of entities established <click> in FRBR like work, expression, manifestation, and item – these terms will be registered through IFLA (Gordon Dunsire is working on that). The values or terms to be used for media types, content types and carrier types were developed in collaboration between the RDA editor and the publishing community that developed ONIX and are being registered on the web – most are also included in the DCMI/RDA Task Group's registry of RDA elements with Dublin Core (the URL is at the end of your handout). <click> Similarly we expect there to be more freely accessible registries on the web of controlled vocabularies, such as the LCSH in SKOS format (at <id.loc.gov/authorities> in the next few weeks).

RDA Cataloging Scenarios

- **Scenario 1:**
  - Relational/object oriented database structure
  - Alternative/ future: linked data, semantic web structures
- **Scenario 2**
  - Linked bibliographic and authority records
- **Scenario 3**
  - 'Flat file' (no machine-actionable links)

8

In 2006, Tom Delsey worked with the JSC and prepared a document describing 3 examples for cataloging scenarios to give an idea of how RDA could be implemented in different cataloging environments (That document was updated in January 2007 and is available as one of the working documents on the JSC Web site as 5JSC/Editor/2): <click>

Tom Delsey's Scenario 1 is an application of a relationship or object oriented database structure with linked records. However, I like to think of Scenario 1 as going beyond databases to linked data structures on the semantic web. To me, the linked data is the future that the JSC has kept in mind in structuring and developing the new content standard. <click>

Scenario 2 is what many libraries have today with MARC records in an integrated library system – with linked bibliographic and authority records. <click>

Scenario 3 is the packaging of bibliographic and authority data separately in stand alone records or flat files with no machine-actionable links, like book catalogs and card catalogs.

RDA data can readily be mapped to any one of the scenarios or variations – these were just chosen to be illustrative. Each scenario supports the objectives that RDA is designed to fulfil but the different data structures have a bearing on the efficiency of data creation and maintenance, and on the ease and effectiveness with which users are able to access the data and navigate the system, which I will show you in just a moment.

Our catalogs and practices reflect the technology available to us at any point in history – when we had book catalogs we designed rules that helped us best present data in that linear form. When we had card catalogs and could take advantage of duplicating a main entry card to provide full access in many places in an alphabetical linear arrangement, our rules reflected that. But we haven't moved much beyond that for online catalogs. However, we now have the ability to repackage the bibliographic data for the Internet to give needed information collocated appropriately to respond to the user's query or to offer the user pathways to discover related resources.

③ **Bibliographic entry in book catalogs**

③ **Main entry/added entries in card catalog**

② **Bibliographic, Authority, and Holdings records – MARC format**
  - **Stand alone or linked**

① **Information packets/Description sets = clusters of attributes to identify each entity and indicate relationships**
  - **Linked**   9

In Scenario 3 we have book and card catalogs organized around "**entries**" <click>

In Scenario 2 with the MARC format we shifted to a focus on individual **records** that typically were intended to stand alone so they could be sent to other systems and shared by making copies and distributing them. The bibliographic record data, authority data, and holdings data sometimes was connected through record control numbers or physically linked inside local integrated library systems.

<click>

With future Scenario 1, now we are thinking in terms of "records" could probably be better understood as future information packets or "**Description sets**" as they have been called by the Dublin Core community. These are the sets or clusterings of data – the information needed to identify each entity and to indicate the relationships. In an email conversation I recently had with Corey Harper ,he noted that these "clusterings of more atomic "statement" level data can be built in a variety of ways, and that the concept of the "record" itself is losing its centrality in this context."

Let's look at some examples of these scenarios.

Scenario 3 describes using RDA to build card or book catalogs or flat file databases, where bibliographic, authority, and holdings information is packaged separately.

This is an entry from the British Museum's printed book catalog of 1841 that collocated works of an author. Let's look at that entry in the book catalog from an FRBR perspective, wearing our FRBR glasses.

FRBR gives us a new perspective for seeing what's been there al along. FRBR let's us view the bibliographic universe in a new way, let's us describe the things in this universe with a new vocabulary that also helps us talk to designers of systems and application programs to work together to build better resource discovery systems for the future.

British Museum 1841
("full and accurate" book catalog)

ACOSTA(CHRISTOVAL).
Tractado de las drogas, y medicinas de las Indias
Orientales, con sus plantas.   *Burgos*, 1578. 4º
Another copy.
The same. *Ital.*   *Venetia*, 1585. 4º
Another copy.
Tractado en loor de las mugeres. *Venetia*, 1592. 4º

Person Work Expression Manifestation Item

11

We have the FRBR entities of person (purple), work (blue), expression (green), manifestation (orange), and item (red). They are all there by virtue of the presence of the identifying data elements or attributes for each of those entities.

For book catalogs of the past, we had collocation under the name of a person (in purple) for the various works (in blue) in all of their expressions (in green), manifestations (in orange), and items (in red), reflecting what we can find at the particular library – in this case the British Museum. All the FRBR entities are reflected in these entries by their identifying information.
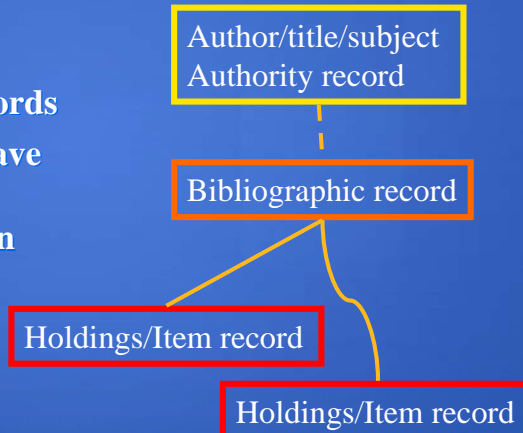
However, this information was in their book catalog, which you either had to use at the British Museum, purchase for yourself, borrow from someone, or refer to in your local library!

For now most of us are in an environment like this Scenario 2, using the MARC format in an integrated library system of self-contained records.

These records may or may not have any connection to each other, but I personally hope that RDA will inspire systems designers to develop much better systems for the future that are linked and FRBR-based. There are some current experiments with FRBR, like OCLC's WorldCat, VTLS's Virtua system, and experiments at the National Library of Australia and National Library of Sweden, as well as the music catalog project at the Indiana University and Kent State University research about FRBR - just to name a few, but we still need more and we hope RDA will give the impetus to move forward.

Relational / object-oriented database structure
Scenario 1
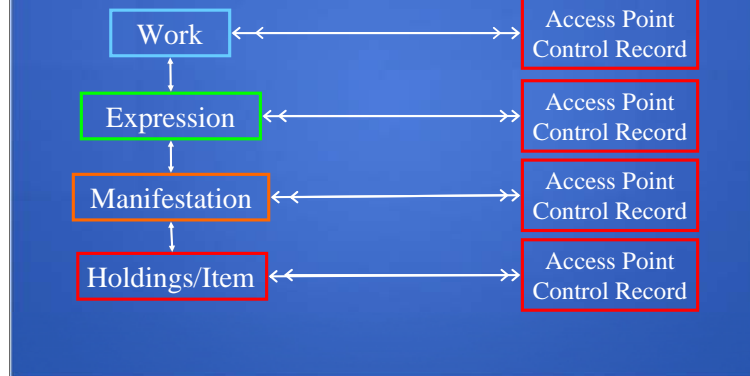
This scenario shows records for controlled access points and indicates that the work could be linked to the access point control record for the creator, the expression could be linked for example to the access point control record for the name of a performer or a translator, and even the manifestation record could be linked for example to the access point control record for the name of a publisher or the item record could be linked to the access point control record for the name of the holding institution or owner.

In some current library systems, records are stored behind the scenes in a relational or object-oriented database structure that mirrors the FRBR and FRAD conceptual models. They are already FRBR-ized to a degree. There typically are separate records containing the identifying data elements for the FRBR entities such as person, corporate body, concept – each record containing the necessary identifying data for that entity. Here's where I prefer to use the term description sets rather than records, because we often think of a MARC record, when we say record.

But here we have a description set for each resource that may either be in one record (like our MARC records) or in linked records, where the various records may be connected through the text of the authorized access points or by persistent identifiers, or the description set may not be in actual records at all, but in packets of information that could be stored in any number of ways within a system.

There is no need to redundantly record data that is inherited from the related entity record or description set – for example there is no need to repeat a title at the expression level when it is recorded in the work description set to which it is linked – we'll see more about this in a moment.

Scenario 1

- Linked "records" for entities
  - works, expressions, manifestations, items, persons, corporate bodies, families, concepts, etc.

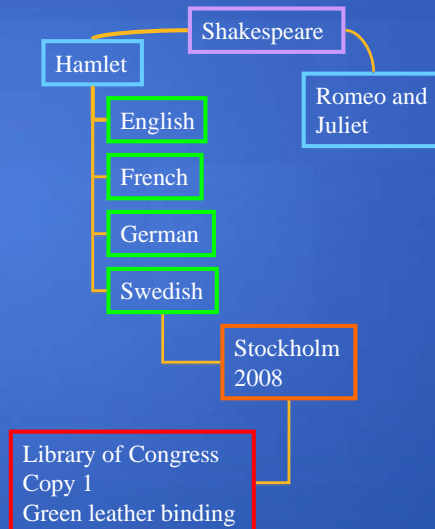Concept — Work — Expression — Manifestation — Item — Person — Corporate body

14

This picture shows a scenario that links clusters of data describing each of the FRBR entities, making the relationships explicit. All of this data can be mined and displayed in different ways depending on the user task.

Scenario 1

- **Display**
  - **All the works associated with a person, etc.**
  - **All the expressions of the same work**
  - **All the manifestations of the same expression**
  - **All items/copies of the same manifestation**

Shakespeare

Hamlet

Romeo and Juliet

English

French

German

Swedish

Stockholm 2008

Library of Congress
Copy 1
Green leather binding

15

We hope future systems will be developed to take full advantage of mining the metadata catalogers provide and have been providing. It should be easier to fulfill the functions of a catalog to display all the <click> works associated with a person, all the <click> expressions of the same work, all the <click> manifestations of the same expression, and all the <click> items and their special characteristics,

Scenario 1 – Related Works
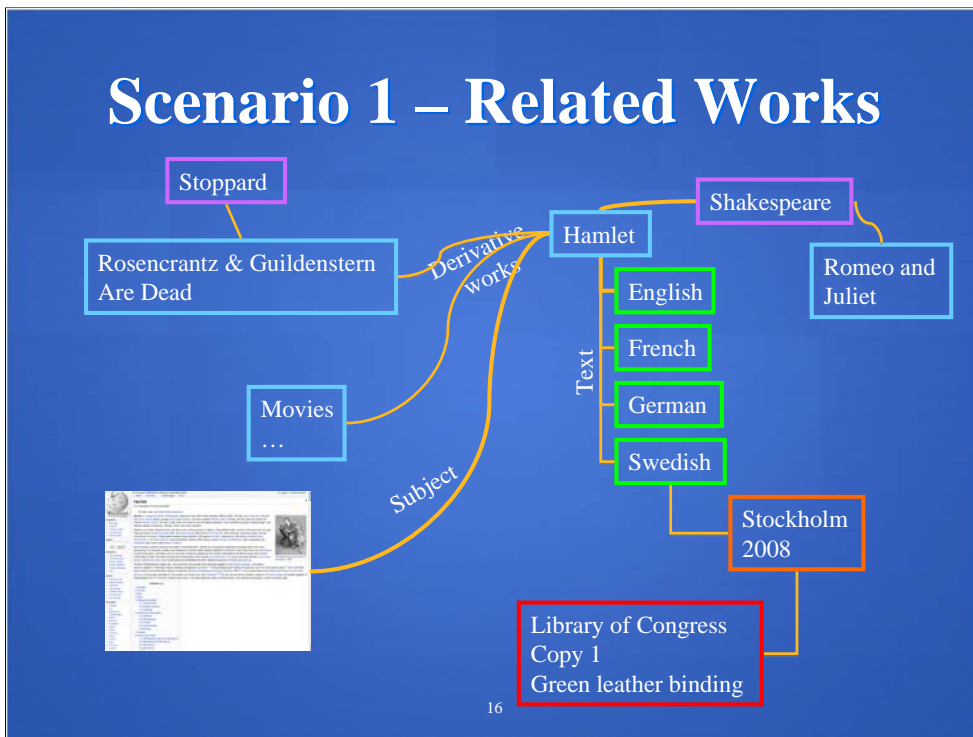
plus all related works <click> to movies or plays based on Hamlet – all of this to guide a user through our rich collections and beyond – this shows the connection to the Wikipedia article about Hamlet.
And once we are able to share this linked data on the Internet, we can offer resource discovery systems that will make cataloging much easier by describing once the works/expressions and their relationships to other works/.expression and subject and linking for new manifestation – new resources we add to our collections.
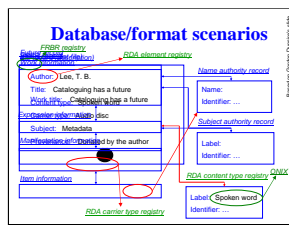
Here's another way of looking at these scenarios and the way RDA intends to use FRBR terminology and concepts to help improve the way we structure records or future information packages for future systems. This is an animated view based on the slide from Gordon Dunsire at the University of Strathclyde. Let's go through it.

This example is missing many of the data elements, like the publication statement, simply because trying to show it all on a PowerPoint slide gets too complicated otherwise…so just imagine the se elements as examples.<click>

In the past we had bibliographic data on a catalog card - we transcribed data from the item and typed or wrote it on the catalog card. The recording of the metadata was then used for displaying that data to our users in the card catalogs. We included such metadata as the author, title, content type, carrier type, subject terms, even possibly the provenance data as we see here. Later we re-used such metadata in a new package – the MARC record. <click>

With some online systems, the author data is actually stored in an authority record and subject terms <click> are in subject authority records <click> with links between the bibliographic and authority records. We record the data in MARC format but it is displayed in a different way through our online catalogs to our users. <click> In a FRBR-based system, we separately identify <click> item level data, such as provenance of that particular item; <click> manifestation level data such as title proper and carrier type and the publication information such as the imprint – place of publication, publisher's name, and date; <click> expression level data, such as content type; and <click> work level data, such as the subject headings and the name of the creator of the work – and the work's title, which in many cases is based on the title proper of the 1st manifestation. <click>

All of this data is linked and is used to identify each resource. It may be contained in a single package or through linked packages of data, depending on the system design. I hope future systems will make this structure invisible to users and easy for catalogers to create and maintain. <click>
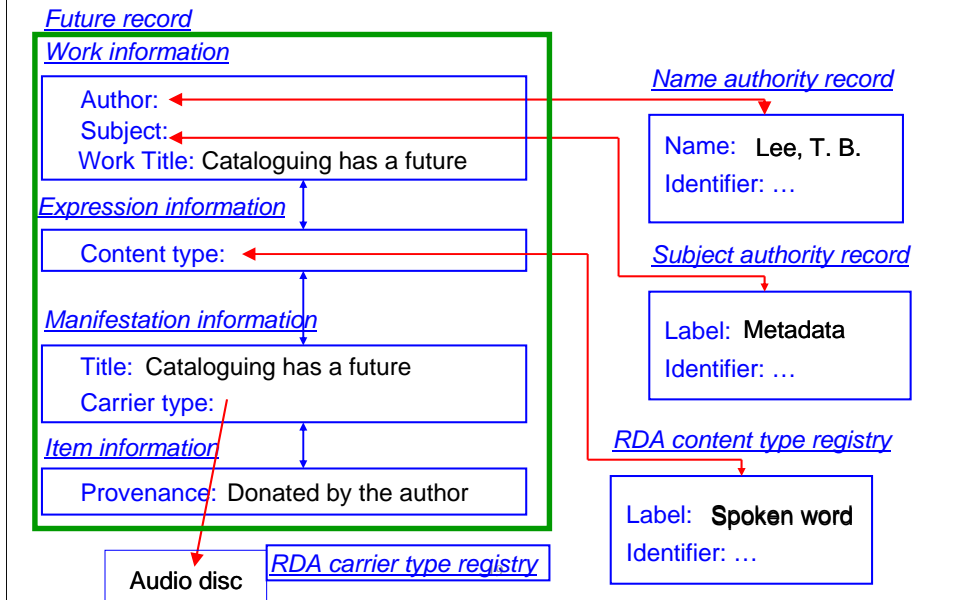
We also have started to build machine-actionable registries for the controlled vocabularies we are using for <click> content type and <click> carrier type. In the future we may wish to link all names to authority records or registries – even for names now in notes <click> or publisher's statements. The terms we use for the roles, <click> like "author", "composer", "artist" – are also controlled vocabularies that we are making available as Web accessible registries. In fact, all of the RDA elements and sub-element terms are being put into registries on the Web. Likewise the <click> ONIX terms for content types are in a registry on the Web and <click>IFLA is working to put all of the FRBR elements on the Web.
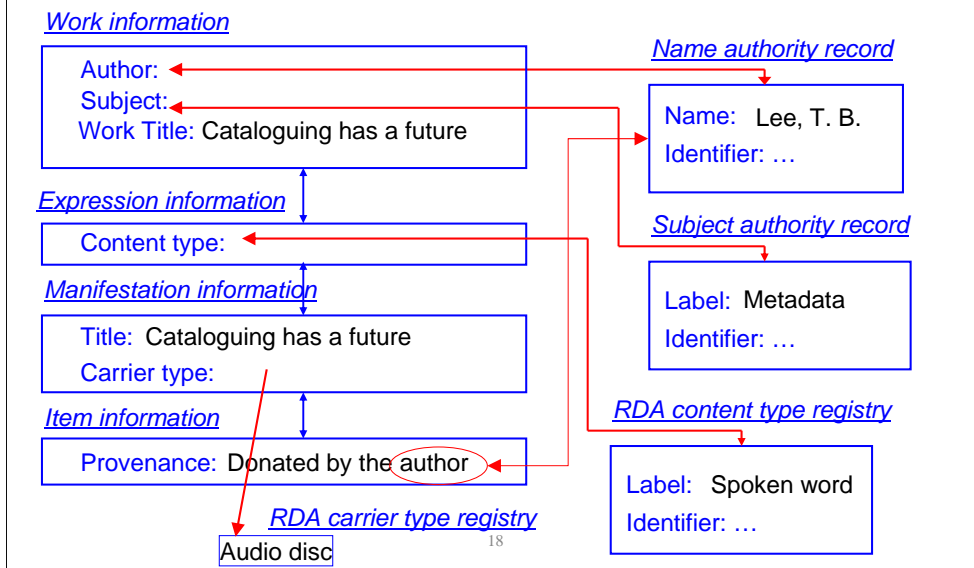
So our future vision is that this linked data would be available for re-use on the Web using the registries and repositories of description sets – done once and shared by all.

Package for data sharing

Future record

Work information
Author:
Subject:
Work Title: Cataloguing has a future

Expression information
Content type:

Manifestation information
Title: Cataloguing has a future
Carrier type:

Item information
Provenance: Donated by the author

Audio disc | *RDA carrier type registry*

Name authority record
Name: Lee, T. B.
Identifier: …

Subject authority record
Label: Metadata
Identifier: …

RDA content type registry
Label: Spoken word
Identifier: …

If we find we need to share a description set for a resource-share a "record" as we do today among libraries, <click> we could construct one to package as a MARC record or some other type of record, depending on the need just as we do now to export a MARC record from an ILS.

Packaging a record is dynamic and dependent on the purpose of the sharing or transmission. Some systems are already doing this, like the ExLibris infrastructure that can bring in cross references from an authority record to accompany bibliographic data as needed to ingest into Primo. Our we might want to bring in biographical authority records from archives for a special collection catalog.
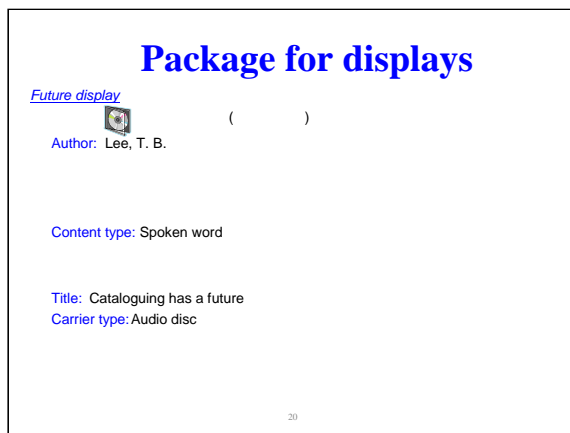
We could even have the systems provide icons or other interesting devices to help our users quickly see the options of what's available. (click through)

We have become very accustomed over the past 40 years with the MARC format to think in terms of packaging the data about our resources and the associated entities as bib or authority records. Those bib and authority records have included identifying information as well as clues about significant bibliographic relationships. But that view of how we package the attributes will evolve as our information discovery tools mature.

The National Library of Sweden has been experimenting and applying FRBR in systems with linked data for a semantic web environment that seems to epitomize the direction we are heading. We want people to find things that are available to them, much like Amazon, Google, or any business trying to put their customers in touch with the products and services they have to offer. We have an inventory of resources that we need to describe and show how they inter-relate so our customers, our users, can find or learn about resources we have that will meet their information needs.

These things in our inventory – in our bibliographic universe – are described in the FRBR conceptual model as entities that have relationships and can be described by their attributes. RDA calls those attributes elements, to be in line with the semantic web, RDF (Rich Data Framework) structures and schemas.

The data elements we use to identify or describe the entities and relationships are clearly labeled and can be packaged and re-used to meet various needs – different displays can be created, much as we do now with different displays when we search our OPACs for a subject and see the resources that are available on that topic versus a display when we want to see all the description of a particular manifestation – such as in a full display, a short display, a Dublin Core display, or a full MARC record display.

The future FRBR "record" will actually be linked data packets or describing sets of data.

Wish List

- **Templates or RDA Online workflow "wizards"**
- **ILS links from specific elements in input screens to RDA instructions**

We've put a lot of effort into building the infrastructure and we still have a lot of work ahead of us, but what do we still need to develop to reap the optimal benefits from RDA?

The JSC hopes RDA Online will include workflow wizards at some point and be linked to templates probably in integrated library systems to help catalogers needing guidance. These workflow wizards will give simple step-by-step instructions with links to the relevant RDA guidelines.

We also hoped that ILS Vendors and bibliographic utilities like OCLC would build in RDA so their input screens could link directly from the specific data elements to the RDA instructions.

Wish List

- **Import descriptive metadata**
  - **Publisher/author supplied (e.g., ONIX)**
  - **Third-party supplied**
    - **Book vendors, contractors**
- **Validation of required "core" elements linked to mode of issuance**

22

As for my own personal wish list, I want systems that will import the descriptive metadata that comes from publishers (as OCLC and LC are now doing with ONIX data) or for any metadata that accompanies digital objects either literally or is intended to go with the object.

I want to see our systems provide validation of the RDA core elements appropriate to the mode of issuance of the resource being described.

Wish List

- **Import controlled metadata**
  - Registries for RDA/ONIX terms
  - VIAF (language/script appropriate to user)
- **Drop down menus for controlled vocabularies**
  - Media, content, carrier types
  - Names of persons, families, corporate bodies
  - Subject headings

23

I want APIs (Application programming interfaces), that is, computer routines that provide means to easily import controlled metadata from registries on the Web or international authority data systems, like VIAF – the Virtual International Authority File – that can also enable the display of languages and scripts appropriate to the user.

I also want systems to provide drop down menus for the controlled vocabularies, such as the particular terms for specific RDA elements that have controlled lists of terms (like the ISO list of scripts, the forms of musical notation (like staff notation, letter notation, etc.) forms of tactile notation (like braille, Moon type, etc.), like the media types (like audio, computer, video, etc.), content types (like cartographic image, computer program, performed music, still image, text, three-dimensional moving image, etc.), and carrier types (like computer discs, microfiche, volumes, videocassettes, etc.), and to suggest the appropriate links to names of related persons, families, or corporate bodies identified in the attributes and relationships; as well as suggest subject headings.

I want to see the automatic suggestion of classification numbers for LCC, Dewey, and other systems that wish to register their schemes <click> based on what is included in the description sets and linked digital tables of contents, summaries, abstracts or text. I'd like to see suggestions <click> made based, as we now can do in Classification Web on matches with existing similar works and correlations <click> between classification and subject heading systems.

Wish List

- Automatic generation of work/expression data and links to "creator"
  - Based on identifying elements for the "first" manifestation
- Automatic prompting and validation of work/expression data
  - Suggests possible matches for new cataloging

We will not need to have catalogers manually keying in separate records for all the works and expressions, because the description set of metadata will mostly already be there or be generated based on the first manifestation received – saving a lot of time and effort. <click>

We know from OCLC studies of WorldCat that over 80% of the works in their database are embodied in just one manifestation.[2] So when we build the description and access points for a resource, the elements that identify the manifestation can be used by our future computer systems to automatically build the attributes for the name of the work and the relationship to the creator of the work, as well as provide a placeholder date of the work/expression. When we build a MARC record now, we add the language, which is expression data, and the future systems should know to label it as expression identifying data.

Our future systems should automatically prompt us of existing <click>work/expression data. The data from the new manifestation should be matched, and exact or fuzzy matches should be displayed to prompt the cataloger to verify it's the same work or expression. The computer systems should <click> suggest relationships and make it easy to designate a relationship, such as with a touch screen, where you point to link; and default to general types of relationships with the option to select more specific ones as needed or add new ones to a controlled list of designation type terms.

---

[2] Bennett, Rick, Brian F. Lavoie, and Edward T. O'Neill. 2003. "The Concept of a Work in WorldCat: An Application of FRBR." *Library Collections, Acquisitions, and Technical Services* 27,1 (Spring). E-print available at http://www.oclc.org/research/publications/archive/2003/lavoie_frbr.pdf. (PDF:354K/32pp.)

Wish List

- **Easy way to share "maintained" data worldwide**

And I want to see catalogers have an easy way to share data that is maintained. This was a goal that was also mentioned in "On the Record" by the Working Group on the Future of Bibliographic Control". When the data package or description set is fixed, it's there for everyone to use in displays -- without the need for further human intervention.

Or there can be automatic updates to local systems that choose to keep their own data.

Wish List

- **FRBR collocating (expand and collapse elements for displays)**
- **Simple displays of pathways to related resources and information about related entities**
  - **Open options if user wishes to explore – don't overwhelm with all possible relationships**

For our end users including catalogers, I want to see more FRBR- based systems that make it simple and easy to expand or collapse the set of elements that are displayed for collocated entities and relationships to help them easily move about the surrogate bibliographic universe – but to do it in a way that does not overwhelm them with graphics but let's them be in control of how much or how little or which direction they take to explore the available resources. And I'm sure you can think of lots more for your own wish list!

Resource Discovery System

- **User-focus**
- **Builds on existing descriptive metadata clearly labeled**
- **Identifies all names and other identifying information for an entity**
- **Identifies significant relationships to enable collocation and navigation of the bibliographic universe**
- **Re-uses data globally for more efficient operations**

So what is that vision of the future? A user- focused resource discovery system that uses existing descriptive metadata from the resource itself as much as possible, that identifies variant names for an entity to increase the recall of relevant matches while clustering the identifying information about an entity to distinguish between entities to increase the precision of matches, and re-use of the identifying data about entities and their relationships for more efficient operations and to offer pathways to related resources in our bibliographic universe.

We get there by providing a clearly labeled set of data elements for the entities and relationships that matter to us in the bibliographic universe – the things in our collections and the associated people, corporate bodies, and families, and the various relationships, including subject relationships that help us connect our users to information we have for them – whether they are looking for something specific (a known item search) or we can inform them about related resources or information that they may find relevant to their information needs.

And to do this in a way that builds on data that is readily available.

RDA is helping us move in that direction.

Considerations to get there

- **Bridge**
- **What incentives are needed to reach Scenario 1?**
- **How would staff of different institutions, programs, publishers, systems, national, etc., function in and support Scenario 1?**
- **What are the economic/legal obstacles to overcome if some controlled vocabularies are proprietary, restricted, or less easily available on the Web?**
- **Does someone need to start the ball rolling for Scenario 1? Who? How?**

So how do we get to the future Scenario 1. I did mention that we have some things for the bridge period where we are still using MARC records.

But here are some of the further questions that we'll need to think through together: <click> what are the incentives for institutions to reach Scenario 1. There will be cost savings for the cataloging operation and service providers through shared linked data, especially when there is no need to 'exchange' records, but we need to leave open the option for those institutions that want to continue our current models. We need to share this story with library directors to help them better understand and build a common purpose.

We need to imagine how staff from different institutions including libraries, archives, publishers and distributors would function to support Scenario 1, what would be the role of national bibliographic agencies beyond maintaining controlled vocabularies? How many of those institutions would be able to make their vocabularies freely available and what mechanisms need to be in place for any models that would require payment for data. I just saw the Dublin Core 2009 conference announcement and it looks like many of these issues will be addressed there!

Who should start the ball rolling and how?

FRBR-Based Collocation
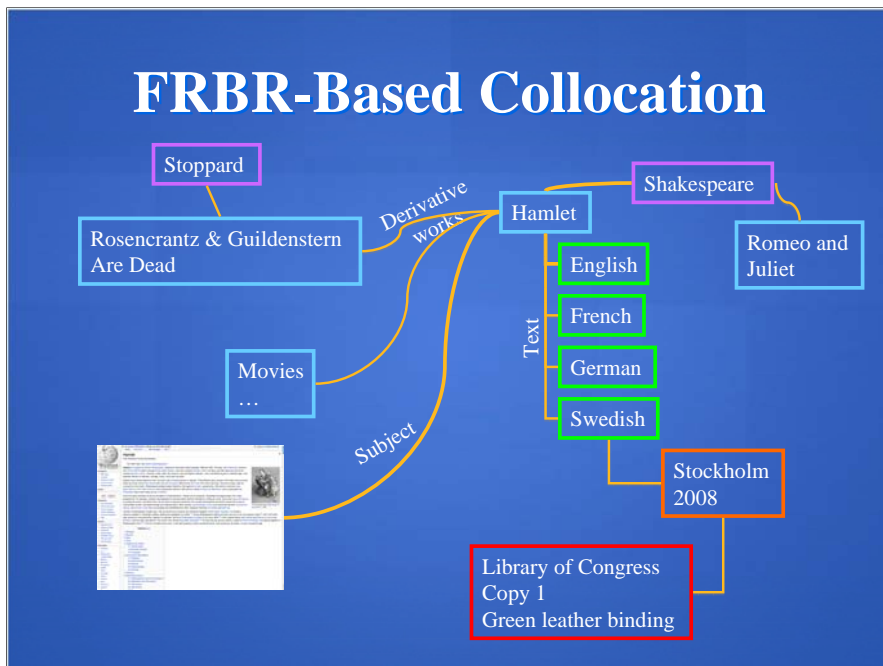
I look forward to discussions today and in the near future that will help us move closer to FRBR-based systems to help users – to collocate related resources and to offer the pathways for users to explore our collections and the resources available to them worldwide so they can find what they need.

We may need to begin with small steps to make the best use of existing MARC record structures, but hopefully that will position us for future structures and systems that take full advantage of the model for streamlined cataloging operations and more effective user service.

The first release of RDA reflects many compromises that carry over traditions of case law from AACR2, but that is intentional so there will not be too much change, because we are being told by administrators to avoid the trauma of moving to AACR (first the Miami compromise for such things as corporate entry under place names – better known as "superimposition") and then the even greater trauma of de-superimposition when we moved to AACR2 that caused the splitting and closing of card catalogs.

However, as we are able to engage various communities, we can begin work towards a more principle-based standard and one that results in metadata that is re-usable globally.

Thank you for your attention.

Links

- **Rob Walls' presentation on Implementation of RDA in Australia**
  http://www.nla.gov.au/lis/stndrds/grps/acoc/documents/Walls2008.ppt
- **RDA Database Implementation Scenarios**
  http://www.collectionscanada.gc.ca/jsc/docs/5editor2.pdf
- **Encoding RDA data**
  http://www.collectionscanada.gc.ca/jsc/docs/5editor2.pdf
- **RDA, FRBR/FRAD, and Implementation Scenarios**
  http://www.collectionscanada.gc.ca/jsc/docs/5editor4.pdf
- **MARC development web site**
  http://www.loc.gov/marc/development.html
- **DCMI/RDA Task Group wiki**
  http://dublincore.org/dcmirdataskgroup
- **XC Project – recent description by Jennifer Bower**
  http://www.extensiblecatalog.org/MetadataReports

Also note the FAQ on the Libraries Australia site.

More Links

- **LC Webcasts**
  - **RDA Overview Background**
  http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=4320
  - **Cataloguing Principles**
  http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=4327

- **IFLA**
  - **FRBR**
  http://www.ifla.org/VII/s13/frbr/
  - **FRAD**
  http://www.ifla.org/VII/d4/wg-franar.htm
  - **ICP**
  http://www.ifla.org/VII/s13/icc/

- **JSC**
  http://www.collectionscanada.gc.ca/jsc/

[1] Refer you to a presentation on the implementation of RDA in library systems that was presented this past year by Rob Walls, Director, Database Services, Libraries Australia entitled, "Implementation scenarios, encoding structures and display". As he said, his presentation: "Starting at a conceptual level with a description of several database implementation scenarios; then moving through the nuts and bolts of data encoding with MARC21 and Dublin Core; and finally looking at specific implementation issues for library systems and Libraries Australia."
<http://www.nla.gov.au/lis/stndrds/grps/acoc/documents/Walls2008.ppt >