Tomorrow's metadata:

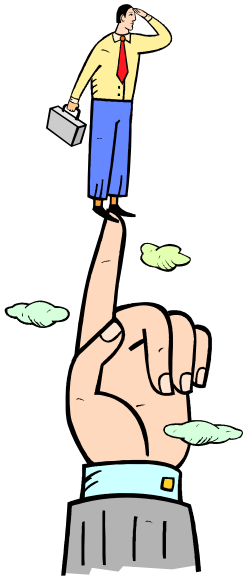
improving resource discovery for the user

RESOURCE DESCRIPTION & ACCESS

Chris Oliver

for ABQLA, May 7th, 2010

resource discovery



supported by

metadata

Improving resource discovery

To improve resource discovery:

- improve the metadata

plus

- technology to use the improved metadata

Improving metadata

- technological changes are happening
- we live in an online networked environment
- get library metadata ready to operate efficiently and effectively in current and newly developing environments

Tomorrow's metadata: RDA

- new metadata standard replaces AACR2
- changes the way we record metadata
- designed for now and the future
- benefits for our users and our libraries

Tomorrow's metadata: RDA

is our data incomprehensible?

- solving some problems in current online catalogues and databases

is our data inflexible?

- getting library data ready for future technological environments

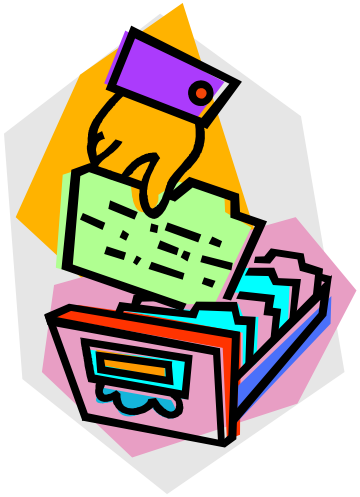
is our data invisible?

- making library data visible

1. Problems with online catalogues

- difficult to navigate and find
  - flat indexing
    - indexing order controlled by numeric value of characters, not by concepts
  - not intuitive to navigate
    - large sets of results with no meaningful groupings
- difficult to find, identify and select
  - ambiguous and embedded data

Advances



card

online

networked

Catalogue has expanded

- more resources available
- more functionality
- little change to metadata

What about the original purpose of the catalogue?

- find known item

and

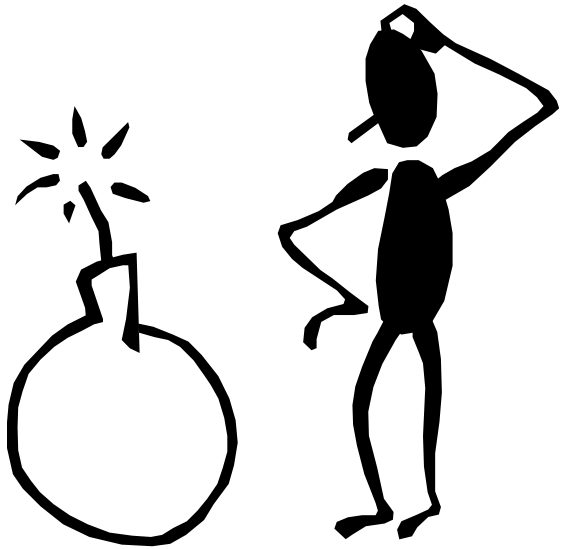
- enable resource discovery

Catalogue has expanded …

BUT THE PROBLEMS …

"online library catalogs … can befuddle users"

Marc Parry, reporter, Chronicle of Higher Education (Sept. 28th, 2009)



1. Problems with current catalogues

a. Navigating and finding

Importance of grouping

- organize by grouping -- based on a shared characteristic or shared relationship

Panizzi: - group together all works by the same author

1841 - group together a work and its translations

Cutter: - (1) find known item + (2) enable resource discovery

1876 - group by author, title, subject

Flat indexing/useless grouping

McGill's Aleph catalogue:

title words = Robinson Crusoe

197 hits — 1st hit = about a Japanese missionary

results

- unrelated (vague subject rel.)
- criticism
- audio-book
- text — 2007 book
- text — 2007 ebook
- adaptation as a movie
- criticism

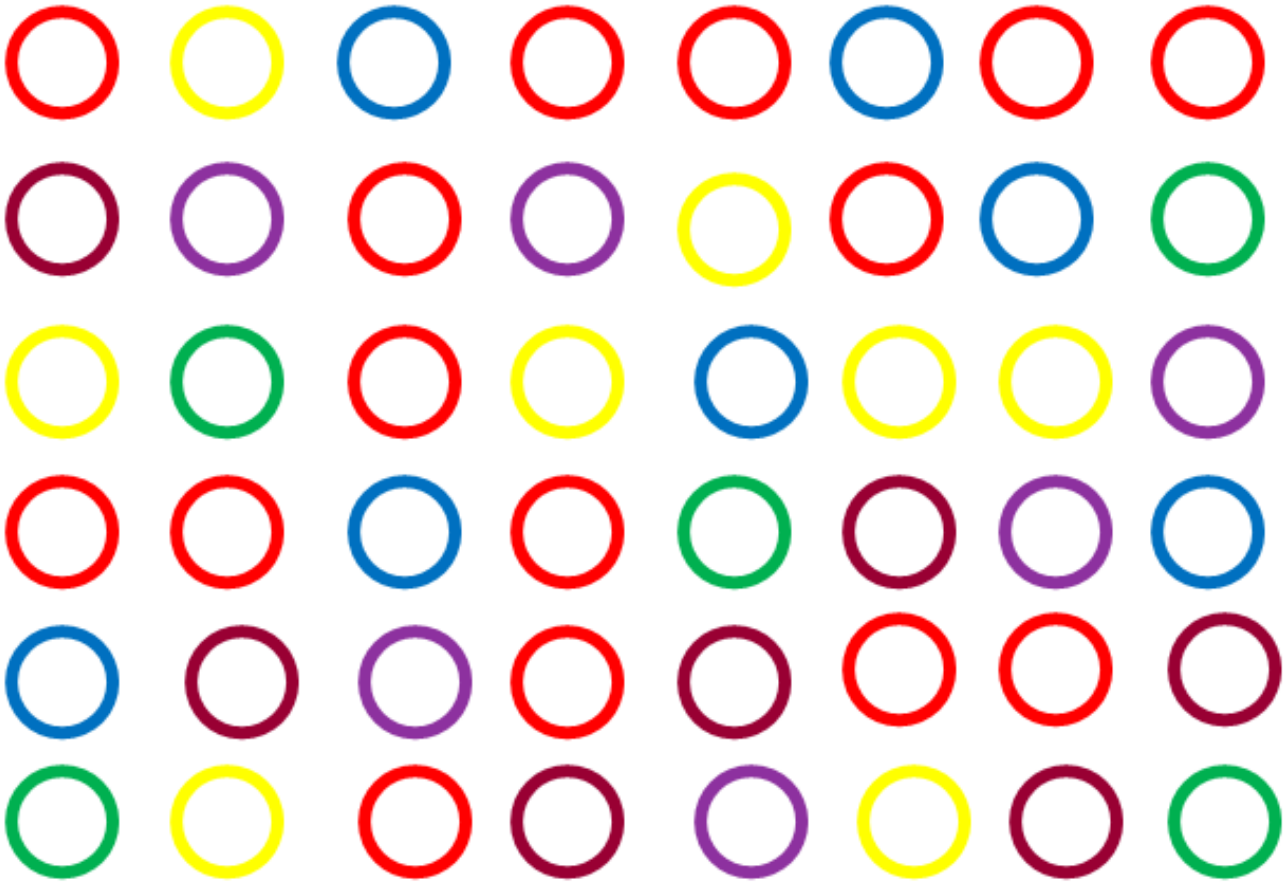
1			Oyabe, Zen'ichirō.	A Japanese Robinson Crusoe /	2009
2			Thomson, Shawn, 1966-	The fortress of American solitude : Robinson Crusoe and antebellum culture /	2009
3		<Recorded Sound> Find It	Defoe, Daniel, 1661?-1731.	Robinson Crusoe Read by Simon Vance.	2008
4			Defoe, Daniel, 1661?-1731.	Robinson Crusoe /	2007
5		<eBook> Find It	Defoe, Daniel, 1661?-1731.	Robinson Crusoe	2007
6		<DVD>		Robinson Crusoe Dan O'Herlihy, Felipe de Alba, Jaime Fernández, Chel López, José Chavez, Emilio Garibay.	2005

Results:

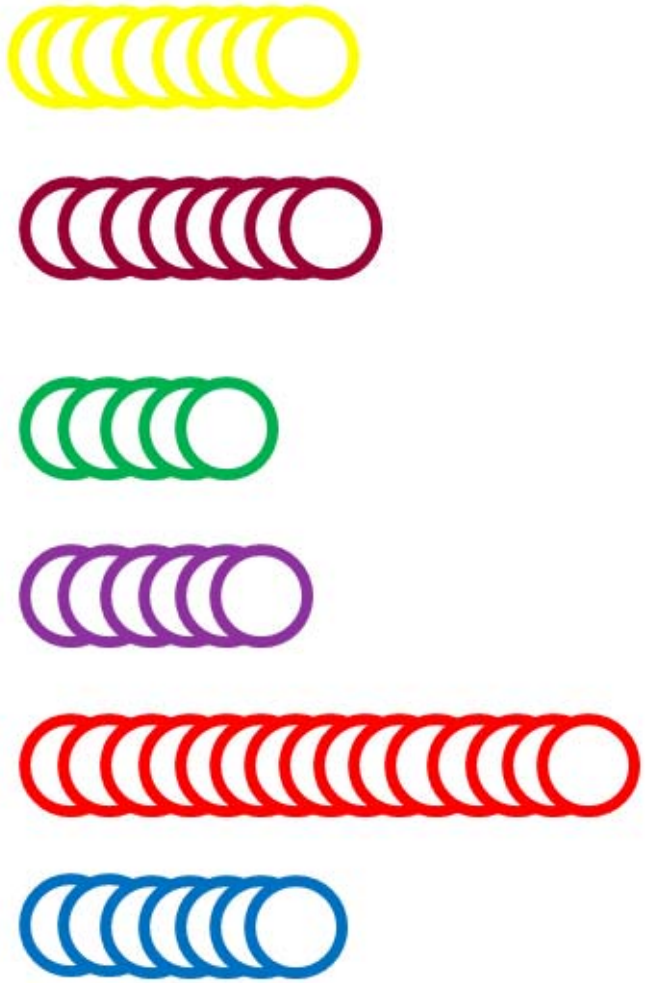
- criticism
- adaptation as a motion picture
- text
- dramatization
- audio-book
- translation
- adaptation as an opera
- basis for a libretto
- sequel

etc., etc.

Search the online catalog:



Grouping to guide the user

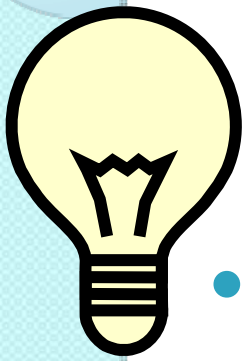


- texts
- translations of text
- audio-books
- dramatizations
- adaptations as a motion picture
- criticisms

Improving user experiences

- FRBR = *Functional Requirements for Bibliographic Records*
- conceptual e-r model
  - analyzes the bibliographic universe in terms of what is important to the user
  - gives a theoretical basis to improve the grouping or clustering of results

Why is the model useful?



Same bibliographic universe but look at it in a new light

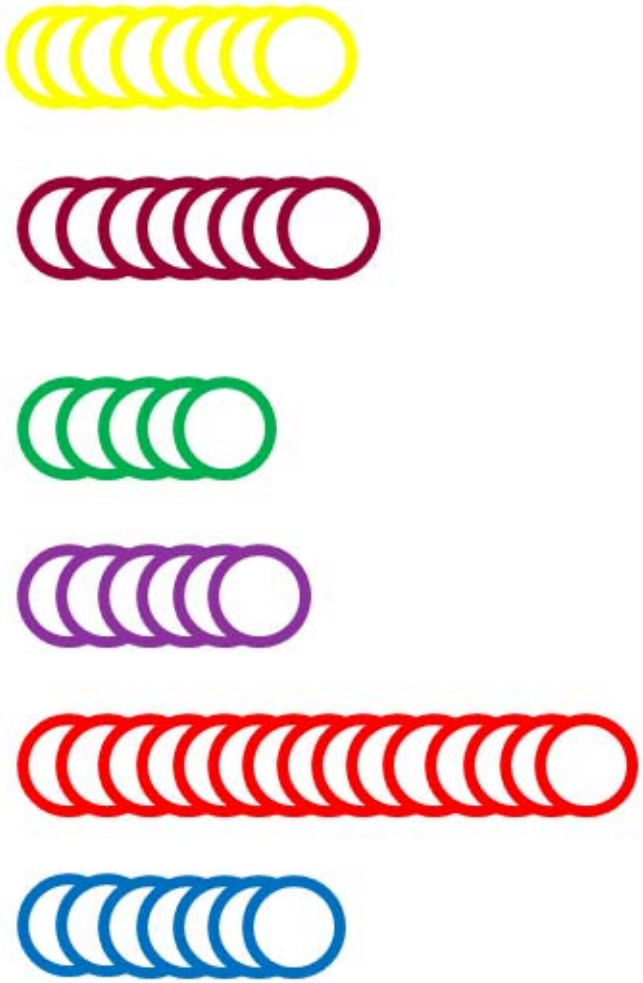
- enables a better understanding of the components of the bibliographic record and their function and value to the user
- looks at the bibliographic record within the context of large databases
- conceptually separates content and carrier

Meaningful grouping of results

- grouping together because they are related
- resources are related to each other when they …
  - share an attribute (characteristic)

    or
  - have a relationship to each other
- meaningful grouping
  - intuitive for user to understand
  - pathways for better navigation

Grouping to guide the user



- texts
- translations of text
- audio-books
- dramatizations
- adaptations as a motion picture
- criticisms

Need FRBR + metadata

McGill's WorldCat Local:

title words = Robinson Crusoe

7218 hits

1st hit = criticism

results

criticism

text 1975 ed.

related work

adaptation as a movie

1. **Robinson Crusoe**
by Pat Rogers
Book
Language: English
Publisher: London ; Boston : G. Allen & Unwin, 1979.
Database: WorldCat.org
Libraries that own this item: McGill University Library

2. **Robinson Crusoe**
by Daniel Defoe; Michael Shinagel
Book : Fiction
Language: English
Publisher: New York : Norton, [1975]
Database: WorldCat.org
Libraries that own this item: McGill University Library

View all editions and formats

3. 
Approaches to teaching Defoe's Robinson Crusoe
by Maximillian E Novak; Carl Fisher;
Book
Language: English
Publisher: New York : Modern Language Association of America, 2005.
Database: WorldCat.org
Libraries that own this item: McGill University Library

View all editions and formats

4. **Robinson Crusoe**
by Luis Buñuel; Oscar Dancigers; Luis Alcoriza; Hugo Butler; Dan O'Herlihy; Felipe de Alba; Jaime Fernández; Chel Lopez; José Chávez; Alex Phillips; Anthony Collins; Luis Hernández Bretón; Daniel Defoe; Olmec Productions.; Producciones Tepeyac.; Ultramar Films S.A.; VCI Entertainment (Firm);
DVD video
Language: English
Publisher: [Tulsa, OK] : Distributed by VCI Entertainment, [2004]
Database: WorldCat.org
Libraries that own this item: McGill University Library

Example of display of results:

- criticism
- text 1975 edition
- criticism
- adaptation as a motion picture
- criticism
- Japanese missionary
- criticism

etc.

Better ...

Some grouping –

2nd hit = display for the 1975 edition of the text

but also shows:

View all editions and formats

"*frbr-ization*"

OCLC groups together texts

translations

audio-books

Better ...

1. click View all editions and formats

   cluster of 91 hits

   Defoe's work
   - book
   - ebook
   - audiobook
   - translations

2. facets = use existing MARC 21 coding

   allow search to be refined quickly

Problems

- results of View all editions and formats
  = display shows a jumble of formats, languages, etc.
- completely hides the fact that there are **two** different works:

Robinson Crusoe

Robinson Crusoe Pt. 2

- an abridgment also buried in the results

Problems with grouping

- imperfect grouping because of inadequate metadata

for example: clusters of Defoe's work

hit no. 2

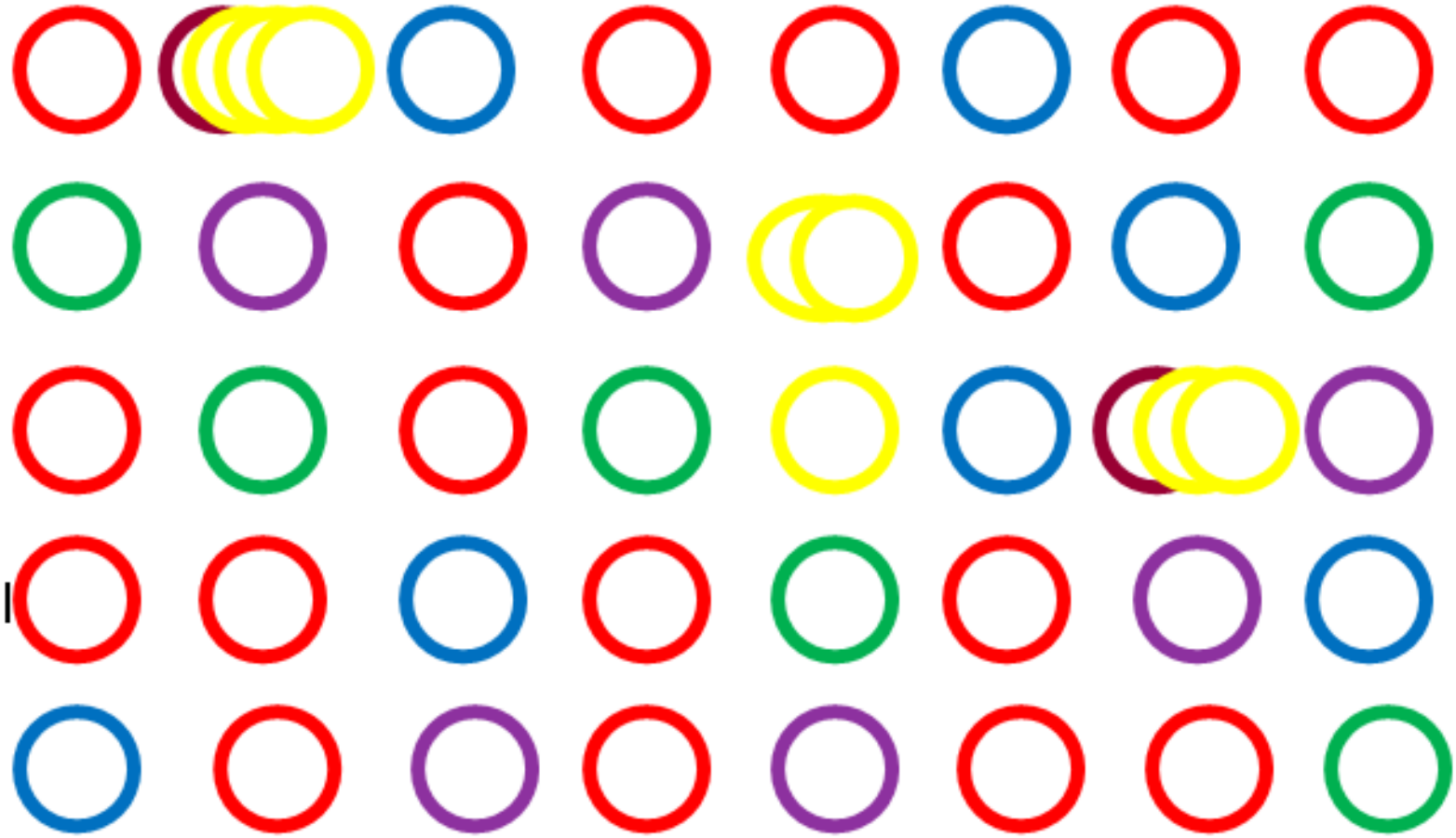
hit no. 23

hit no. 38

hit no. 54

for example: some relationships are unclear adaptations mixed the original

Slightly “FRBR-ized” catalogue



Truly “FRBR-ized”

- original expressions

- expressions in different languages

- expressions as spoken word

- derivative relationships

- subject relationships

- whole-part relationships

Need good metadata

- apply FRBR concepts with current metadata

  → some improvement

- improve the way metadata is recorded

  foundation for →

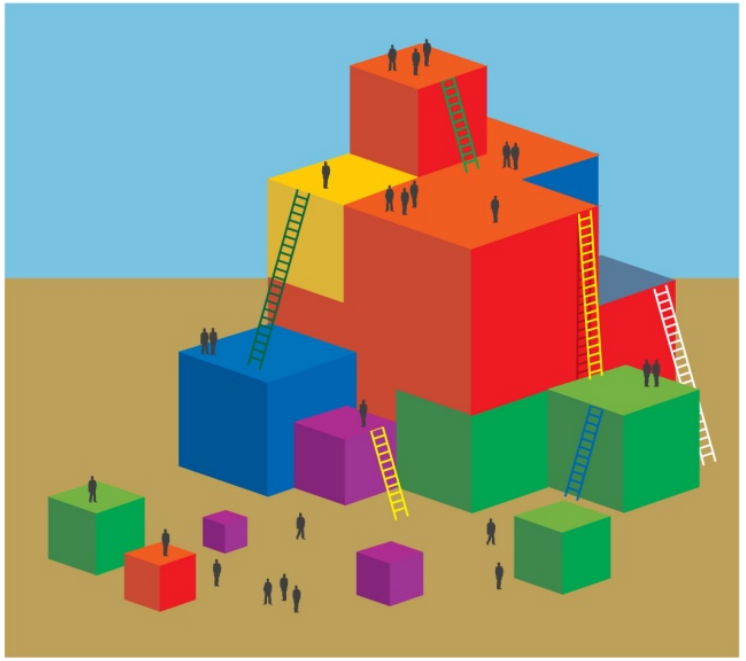
  - improved displays
  - improved navigation

RDA = Metadata for navigating

- add “expression” entity to improve grouping of identical versus similar content
- clear distinction between content and carrier
- enhanced descriptions
- emphasis on relationships
  - recording relationships
  - specifying the nature of the relationship

RDA = Metadata for navigating

RDA metadata = building blocks to support better grouping of results



1. Problems with current catalogues

b. Finding, identifying and selecting

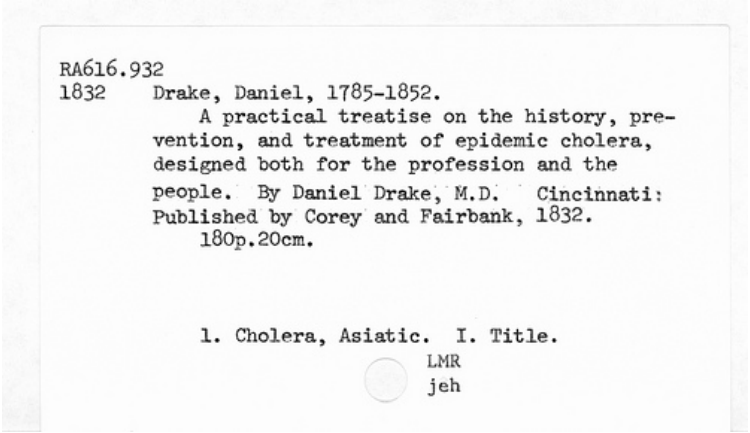
Limited ways to refine search

- “limits” in traditional catalogue
  - usually based on existing indexes

    e.g. year language keyword etc.
- “facets” in next generation catalogs
  - options based on current MARC 21 coding
- only a few characteristics of a resource can be used reliably when searching
- keyword falls short

Limitations of AACR2

- developed for the card catalogue environment



RA616.932
1832 Drake, Daniel, 1785-1852.
A practical treatise on the history, prevention, and treatment of epidemic cholera, designed both for the profession and the people. By Daniel Drake, M.D. Cincinnati: Published by Corey and Fairbank, 1832.
180p.20cm.

1. Cholera, Asiatic. I. Title.

LMR
jeh

  - very succinct descriptions
  - intended to be read and interpreted by humans
  - recorded as paragraphs

Ambiguous information

- ambiguous information

e.g. date of “publication”coded in 008
or 260 subfield c

what is that date?

publication ***OR*** copyright
creation
distribution
production … etc.

Ambiguous information

- index shows:

  name of a person ------- book

  e.g. Hume, David, 1711-1776

  1612 edition of Theophrastus' *Characters*

what's the relationship between person and book?

author of the work ***OR*** editor

translator

former owner

Embedded information

- relationship information:

  name of a person ------- book

→ requires that a human read and interpret information in the record

- information embedded in a non-specific note

  e. g. info about video format characteristics, font, base and applied materials

→ requires that a human read and interpret information in the record

RDA Data Elements

- distinct and precise elements for each kind of data
- recommended controlled vocabulary for content of many elements
- each element has the potential to be usable:
  - to index
  - to search
  - to build meaningful displays of data
- data in element can be used by humans and machines

RDA = Usable Metadata

- precisely defined elements and controlled vocabulary in many elements
- sufficient data to present comprehensible results for the user
- precise data for automated processing to create better displays of data so the user can find, identify and select
- stop: paragraph style
  data embedded in long character strings
  data stored in ambiguous elements

2. Designed for now and the future

“Well-formed” Metadata

- instructions on how to record values of elements
- controlled vocabularies where appropriate
- overall structure governed by a formal model

RDA = “well-formed” metadata

- data recorded in precisely defined data elements
- each data element contains one type of data
- controlled vocabulary is used as the value recorded in many elements
- underlying model for the data = FRBR/FRAD model

Designed for now and the future

Now

- designed to work in the current environment
- compatible with AACR2 records
- co-exist with AACR2 records in the same database

Future

- positioned to take advantage of new database structures based on linked data
- function in the semantic web
- visible in the web alongside other types of metadata

RDA

RDA = content standard

not an encoding standard

not a presentation standard

RDA data can be encoded using:

- MARC 21
- other encoding schema such as Dublin Core, MODS, etc

RDA data can be presented using :

- ISBD conventions
- other display conventions or standards

RDA = content standard

- not locked into MARC 21 encoding
- not locked into MARC 21 record structure

- can be used with web-friendly encoding schema based on XML
- can be used by other metadata communities

Designed for now and the future

Now

- encode in MARC 21
- record data according to simplified transcription instructions

Future

- encode in XML using schema such as Dublin Core, MODS, etc
- transcription instructions allow for automated data capture from other sources, e.g. publishers, digital repositories

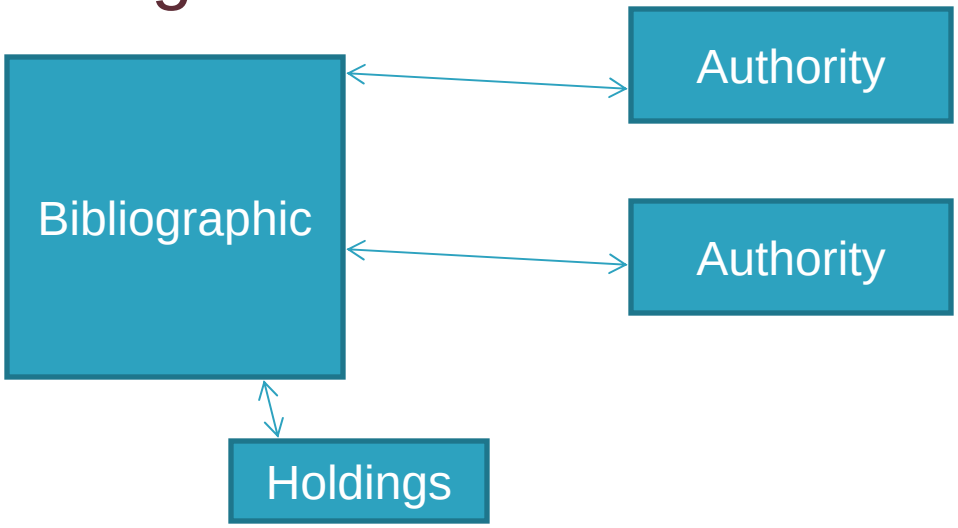
Flexibility of RDA data

- in existing database structures

bibliographic records = description**+** access points

+ authority records – linked to access points

+ holdings records – linked to bib records



Flexibility of RDA data

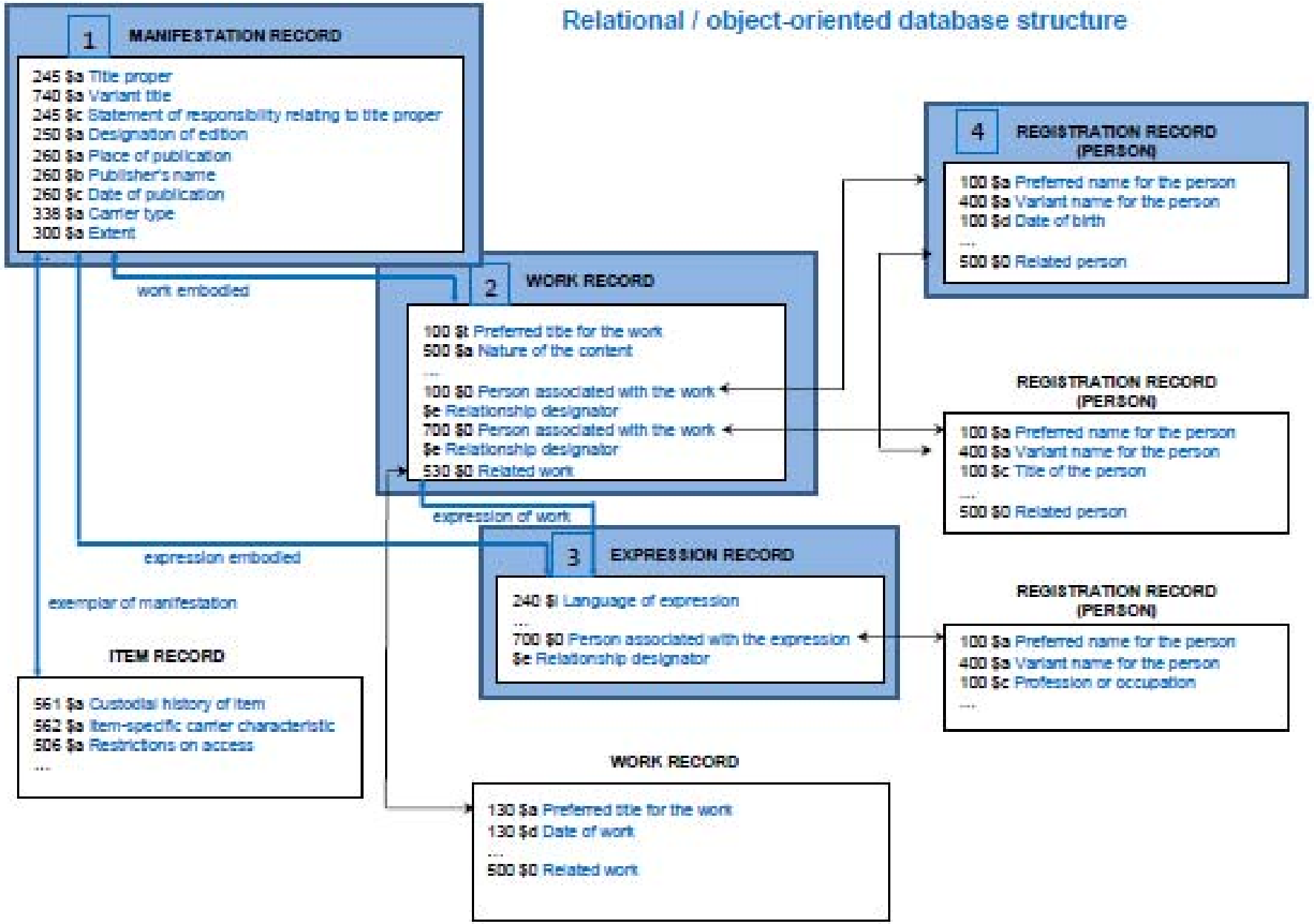
- newly emerging database structures

e.g. a database mirroring FRBR/FRAD model

manifestation record **+** item record
+ work record **+** expression record
+ record for person, family, corporate body
+ relationships = links between the above entities

illustration: from Tom Delsey's presentation to the Deutsche Nationalbibliothek, Frankfurt, Germany, June 2nd, 2009:
http://www.rda-jsc.org/docs/td20090602.pdf

Relational / object-oriented database structure



AACR2 + MARC 21

for example

name of a person ------- title of book

AACR2 + MARC 21

- type of relationship embedded in text of bibliographic description
- bibliographic record contains name of person and title
- may have an authority record that also ties together name of person and title of work

AACR2 + MARC 21

bibliographic record

245 00 $a Alice in Wonderland, or, What's a nice kid like you doing in a place like this? /$cHanna-Barbera Productions.

700 1# $a Carroll, Lewis, $d 1832-1898. $t Alice's adventures in Wonderland.

authority record

100 1# $a Stoppard, Tom. $t Rosencrantz and Guildenstern are dead

(for illustration of following example)

RDA + MARC 21

for example

name of a person ------- title of book

RDA + MARC 21

- type of relationship embedded in text of bibliographic description
- bibliographic record contains name of person and title
- may have an authority record that also ties together name of person and title of work
- relationship designators in bib and auth records ($e, 4, i)

RDA + MARC 21

bibliographic record

245 00 $a Alice in Wonderland, or, What's a nice kid like you doing in a place like this? /$cHanna-Barbera Productions.

700 1# $i parody of (work) $a Carroll, Lewis, $d 1832-1898. $t Alice's adventures in Wonderland.

authority record

100 1# $a Stoppard, Tom. $t Rosencrantz and Guildenstern are dead

500 1# $w r$i based on (work) $a Shakespeare, William, $d 1564-1616. $t Hamlet

RDA + post MARC 21

for example

name of a person ------- title of book

RDA + post MARC 21 record environment

- “record” for person entity, work entity, expression entity (if needed), manifestation entity, item entity
- defined relationships between entities
- defined relationships means that MARC record structure is no longer required

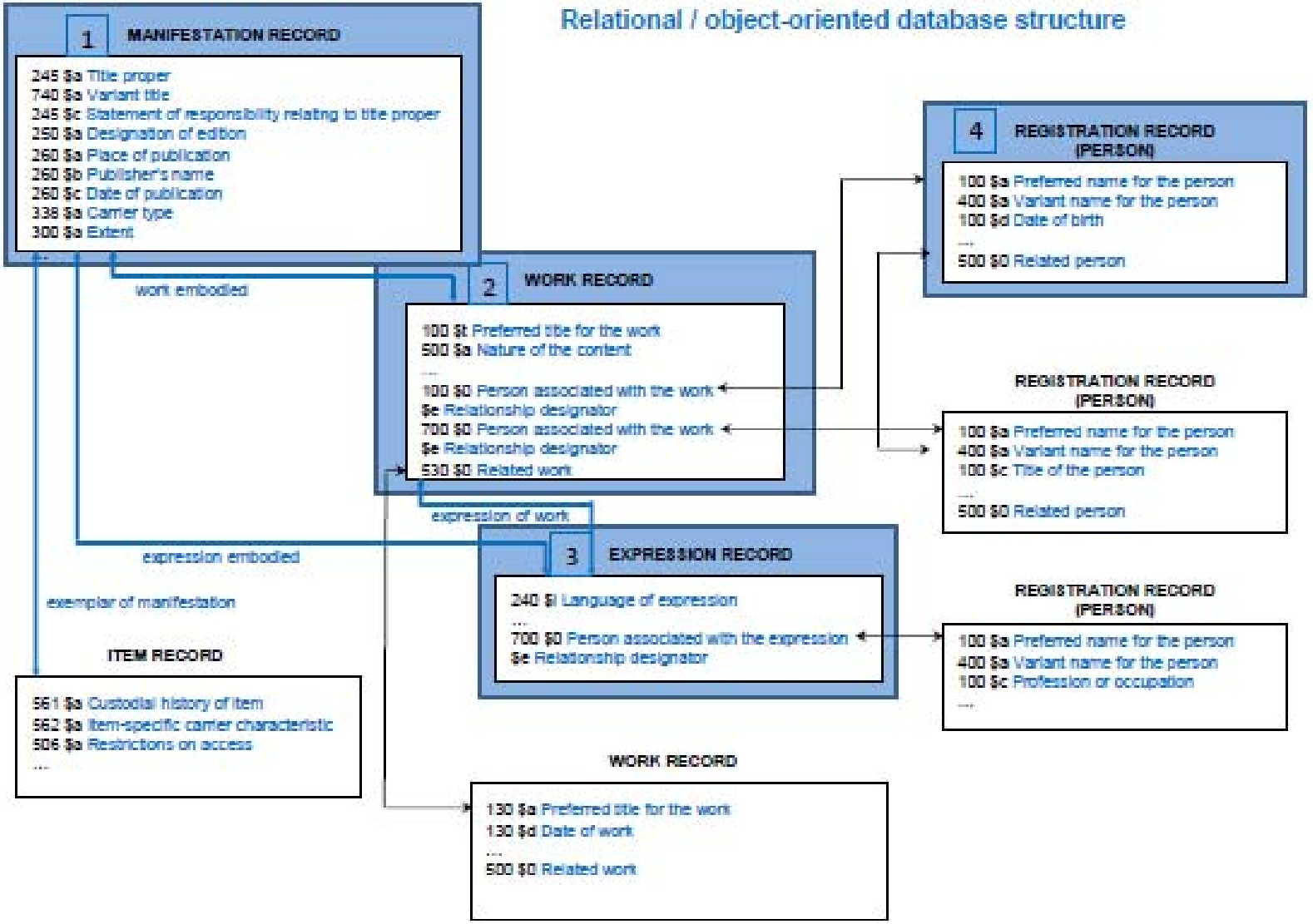


illustration: from Tom Delsey's presentation to the Deutsche Nationalbibliothek, Frankfurt, Germany, June 2nd, 2009: http://www.rda-jsc.org/docs/td20090602.pdf

RDA = Flexible Metadata

- metadata to support machine-actionable processing of data
- metadata to support research discovery on the web
- metadata that can be stored and used in existing and newly emerging database structures

3. Making library data visible

Hidden from the web

- online catalog
  - abundance of metadata
  - invisible to web search engines“dark data”
- MARC 21 - not designed for the web
  - MARC originally automated the printing of cards
  - library specific record format
  - used in closed databases
  - web cannot access and use MARC data

Making library data visible

No user expects information silos:

- users expect that all metadata is on the web
  - ➢ library data needs to be visible on the web

- users do not ask whether the data they need comes from a library or a digital repository or an archive
  - ➢ library data should interact and co-exist with metadata of other cultural memory communities

    e.g. museums, archives, digital repositories, etc.

Making library data visible

- release library data from MARC 21 record structure
- library data available on the web
- library data that can link to related resources in public web spaces
- open the door to using bibliographic data in new ways

RDA = content standard

- not locked into MARC 21 encoding
- not locked into MARC 21 record structure

- can be used with web-friendly encoding schema based on XML
- can be used by other metadata communities

Not just for libraries

- possibility for other communities to adopt/adapt
- instructions designed to describe a wide variety of resources
- connecting with other cultural heritage communities

  e.g. additions and changes for archives

Making connections

- developed with an awareness of practices in other metadata encoding communities

  e.g. abstracting and indexing community

- compatible with metadata standards of other resource description communities

  e.g. development of controlled vocabulary for content and carrier types with members of ONIX (standard for the publishing community)

More international

- beyond “Anglo-American”
- options for use of
  - other languages
  - other scripts
  - other calendars
  - other numeric systems
- increasing interest from countries that never used AACR

RDA = Visible Metadata

- libraries produce valuable metadata
- enable library data to be visible and usable on the web
- connect with other metadata communities
- internationalize

Tomorrow's metadata: RDA

- supports resource discovery
- changes the way we record metadata
- designed for now and the future
- takes us out of the library silo
- connect us with other metadata communities
- positions us to take advantage of tomorrow

RDA moves us forward

RDA …

takes us from where we are

moves us to a new track

stops us from disappearing into fog of obsolescence

Questions: chris.oliver@mcgill.ca

Photos from Flickr:

Catalog card by Public Library of Cincinnati and Hamilton County

http://www.flickr.com/photos/cincinnatipubliclibrary/3392293647/in/set-72157616028126172/

Switch: snow and fog by Luke S.
http://www.flickr.com/photos/varocker07/70700316/

clip art from Microsoft Office 2007