DCMI bridging two paradigms

- **“Records”**
  - Data structures as bounded sets of fields to be “filled in” with information
- **“Graphs”**
  - Data structures as unbounded “webs” of information connected by “statements”
- **DCMI Abstract Model (developed 2003-2007)**
  - Basis for a draft constraint language for application profiles
Filling in the blanks
Overview

- Bibliographic standards
  - International Federation of Library Associations and Institutions (IFLA)
  - Others
- Representation in Resource Description Framework (RDF)
- Creating triples from catalogue records
- Impact and implications for catalogues
IFLA standards

- RDF representations of standards for “universal” bibliographic control are being developed
- “FR” (Functional Requirements) family of models
  - For Bibliographic Records (FRBR)
  - For Authority Data (FRAD)
  - For Subject Authority Data (FRSAD)
- International Standard Bibliographic Description (ISBD)
  - Record structure and content standard for exchange of national metadata
- UNIMARC
  - Encoding for ISBD records (Bibliographic) and FRAD (Authorities)

Representation in RDF

- Entities => RDF classes
  - E.g. FRBR “Person”
- Attributes, tags, (sub)fields, relationships => RDF properties
  - E.g. ISBD “title proper”
  - E.g. UNIMARC “200 $a” (title proper)
  - E.g. FRBR “title of the manifestation”
- Controlled term values => SKOS vocabularies
  - E.g. ISBD Area 0 (content and media type)

FR family

- Each model has its own namespace
  - To reflect historical development
  - Each re-uses earlier RDF elements
- Consolidated model under development
  - Being informed by analysis of RDF representation
- FRBR RDF published
  - FRBRer (entity-relationship) ontology
    - Namespace elements plus OWL
    - FRBRoo (object-oriented)
      - Extension of CIDOC Conceptual Reference Model (for museums)
- FRAD and FRSAD imminent
  - Approved at IFLA 2011 conference

ISBD

- Element set and vocabularies for content and media types
- Namespace now published
- DC Application Profile in development
  - Models the ISBD record
    - What properties (fields)
    - Mandatory? Repeatable?
    - Aggregated statements
      - Sub-elements and punctuation
ISBD AP snippet

UNIMARC
- Proposal for RDF representation made at IFLA 2011
- Discussed with Permanent UNIMARC Committee
  - Decision taken to proceed

Other library standards in RDF (1)
- RDA: resource description and access
  - Content standard based on FR models
  - Refines the FR properties
  - Many more controlled vocabularies than AACR
    - Anglo-American Cataloguing Rules
- MODS/MADS (Metadata Object/Authority Description Schema)
  - Metadata structure based on MARC21
  - Library of Congress Name Authority File in MADS RDF
  - RDF representation of MODS just beginning ...

Other library standards in RDF (2)
- BIBO: Bibliographic Ontology
  - Classes and properties for citations and bibliographic references
- DCMI Metadata Terms (Dublin Core)
  - High-level common-denominator classes and properties for memory institution metadata
- Lots of controlled vocabularies
  - Library of Congress Subject Headings, Rameau (French subject headings), SWD (German subject headings), Dewey Decimal Classification, RDA vocabularies, etc.
From record to triples (in 9 stages)

- Very large numbers of records
  - Catalogue records, finding aids, etc.
  - 300 million; 1 billion?
- High quality metadata
  - In comparison with other communities
- Each record may generate many triples
  - 30 “raw” triples (no inferences) per MARC record?
- Very, very large numbers of triples
  - Billions? Trillions?

1. Take a record

<table>
<thead>
<tr>
<th>Field/attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record ID</td>
<td>54321</td>
</tr>
<tr>
<td>Title</td>
<td>Museum archives: an introduction</td>
</tr>
<tr>
<td>Author</td>
<td>Wythe, Deborah</td>
</tr>
<tr>
<td>Date</td>
<td>2004</td>
</tr>
<tr>
<td>LCSH</td>
<td>Museum archives</td>
</tr>
<tr>
<td>Media/GMD</td>
<td>Electronic</td>
</tr>
<tr>
<td>Content form</td>
<td>Text</td>
</tr>
</tbody>
</table>

2. Disaggregate to single statements

<table>
<thead>
<tr>
<th>Record</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>54321</td>
<td>(has) title</td>
<td>Museum archives: an introduction</td>
</tr>
<tr>
<td>54321</td>
<td>(has) author</td>
<td>Wythe, Deborah</td>
</tr>
<tr>
<td>54321</td>
<td>(has) date</td>
<td>2004</td>
</tr>
<tr>
<td>54321</td>
<td>(has) LCSH</td>
<td>Museum archives</td>
</tr>
<tr>
<td>54321</td>
<td>(has) media type</td>
<td>Electronic</td>
</tr>
<tr>
<td>54321</td>
<td>(has) content form</td>
<td>Text</td>
</tr>
</tbody>
</table>
3. Create URI for record
- Must be unique, so 54321 no good on its own
- http URIs are a good thing (W3C)
- So add record ID to a unique http domain
  - E.g. http://MyLibraryX.com (unique to the library)
  - http://MyLibraryX.com/54321
  - (or http://MyLibraryX.com#54321)
- This is not a URL!

4. Replace record ID with URI

<table>
<thead>
<tr>
<th>URI</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mlx:54321</td>
<td>(has) title</td>
<td>Museum archives: an introduction</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>(has) author</td>
<td>Wythe, Deborah</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>(has) date</td>
<td>2004</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>(has) LCSH</td>
<td>Museum archives</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>(has) media type</td>
<td>Electronic</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>(has) content form</td>
<td>Text</td>
</tr>
</tbody>
</table>

“mlx” = qname (xmlns) = shorthand for “http://MyLibraryX.com/”

5. Find URIs for attributes
- Attributes are modelled as RDF properties (predicates) in “element set” namespaces
  - E.g. Dublin Core terms (dct); ISBD (isbd); FRBR (frbrer); RDA (rdaxxx); Bibliographic Ontology (bibo); etc.
- Choose a namespace, find property with same (or closest) “meaning” (e.g. definition) as attribute
  - Nearest property minimises loss of information
- Get URI for property
- If no suitable property, choose another namespace
  - Properties do not have to come from single namespace
- Match and mix!

5 (cont). Find URI for title
- http://iflastandards.info/ns/isbd/elements/P1014 (isbd:P1014)
  - hasTitleProper
- http://RDVocab.info/Elements/titleProper (rdaGR1:titleProper)
5 (cont). Find URI for author
- dct:creator
- rda:role:author
- (isbd does not cover “headings”)

5 (cont). Find URI for date
- dct:date
- isbd:P1018
  - hasDateOfPublication
  - production
  - distribution
- rdaGr1:dateOfPublication

5 (cont). Find URI for LCSH
- LCSH is a subject vocabulary
  - Controlled terms
- So attribute is really “subject”
  - And the term itself is the value
- dct:subject

5 (cont). Find URI for media type
- Assuming record uses new ISBD Area 0 ...
- isbd:P1003
  - hasMediaType
5 (cont). Find URI for content form

- Assuming record uses new ISBD Area 0 ...
- isbn: P1001
  - hasContentForm

6. Replace attributes with URIs

<table>
<thead>
<tr>
<th>URI</th>
<th>URI</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mlx:54321</td>
<td>isbn:P1014</td>
<td>Museum archives: an introduction</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>rdarole:author</td>
<td>Wythe, Deborah</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>isbn:P1018</td>
<td>2004</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>dct:subject</td>
<td>Museum archives</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>isbn:P1003</td>
<td>Electronic</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>isbn:P1001</td>
<td>Text</td>
</tr>
</tbody>
</table>

7. Find URIs for values

- If object of a triple is a URI, it can link to the subject of another triple with the same URI
- Linked data!
- Values from controlled vocabularies may have URIs
  - Possible vocabularies: author, subject, ISBD Area 0
  - NOT: title, date
- For author: Virtual International Authority File (VIAF)
- For LCSH: Library of Congress Authorities & Vocabularies
- For ISBD Area 0: Open Metadata Registry

7 (cont). Find URI for author

- Author: Wythe, Deborah
- VIAF: http://www.viaf.org/
  - viaf:31899419/#Wythe,+Deborah
7 (cont). Find URI for subject (LCSH)

- LCSH: Museum archives
- LoC: http://id.loc.gov/authorities/
  - lcsh://sh85088707#concept

8. Replace values with URIs

<table>
<thead>
<tr>
<th>subject</th>
<th>predicate</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>mlx:54321</td>
<td>isbd:P1014</td>
<td>“Museum archives: an introduction”</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>rdarole:author</td>
<td>viaf:31899419/#Wythe, +Deborah</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>isbd:P1018</td>
<td>“2004”</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>dct:subject</td>
<td>lcsh://sh85088707#concept</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>isbd:P1003</td>
<td>isbdmt:T1002</td>
</tr>
<tr>
<td>mlx:54321</td>
<td>isbd:P1001</td>
<td>isbdcf:T1009</td>
</tr>
</tbody>
</table>

7 (cont). Find URIs for ISBD Area 0

- Media type: Electronic
- ISBD media type
  - isbdmt:T1002
- Content form: Text
- ISBD Content form
  - isbdcf:T1009

9. Publish triples (linked data)

- mlx:54321 | isbd:P1014 | “Museum archives: an introduction”
- mlx:54321 | rdarole:author | viaf:31899419/#Wythe, +Deborah
- mlx:54321 | isbd:P1018 | “2004”
- mlx:54321 | dct:subject | lcsh://sh85088707#concept
- mlx:54321 | isbd:P1003 | isbdmt:T1002
- mlx:54321 | isbd:P1001 | isbdcf:T1009
Linked data chains

mlx:54321 | dct:subject | lcsh:/sh85088707#concept  

lcsh:/sh85088707#concept | skos:related | rameau:XXX  

rameau:XXX | frbr:isSubjectOf | mly:98765  

mly:98765 | rda:titleOfTheWork | “Managing archives in museums”  

rameau:XXX | skos:prefLabel | “archives du musée”

Linked data cluster = “record”

mlx:54321 | isbd:P1014 | “Museum archives: an introduction”  

mlx:54321 | rdarole:author | viaf:31899419/#Wythe,+Deborah  

mlx:54321 | isbd:P1018 | “2004”  

mlx:54321 | dct:subject | lcsh:/sh85088707#concept  

mlx:54321 | isbd:P1003 | isbdmt:T1002  

mlx:54321 | isbd:P1001 | isbdcf:T1009

Duplication and legacy records

- Many copies of legacy records
  - Copied and amended for local use
  - Danger of minting multiple URIs for the same resource
  - National bibliographic agencies have significant role to play
    - As memory/cultural institutions
    - The linked-data memory/culture of a nation
    - Proposal from IFLA Namespaces Task Group to IFLA Bibliography Section

FRBRization

- FRBR splits record into four functional parts
  - User-centred functions
  - Subject of a FRBR triple is one of the parts, not the resource as a whole
  - But subject of ISBD triple is the resource as a whole
  - Class collisions can be avoided by using unbounded (no domain or range) versions of properties
A short history of the evolution of the library catalogue record

In the beginning ...
Lee, T. B.
Cataloguing has a future. - Audio disc (Spoken word). - Donated by the author.
1. Metadata
... the catalogue card

From flat-file record ...

From flat-file description ...

... to relational record
... to FRBR record
From FRBR record ...

| Work          | Manifestation | Item
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Author:</td>
<td>Title:</td>
<td>Provenance:</td>
</tr>
<tr>
<td>Subject:</td>
<td>Cataloguing has a future</td>
<td>Donated by the author</td>
</tr>
<tr>
<td>Content type:</td>
<td>Carrier type:</td>
<td></td>
</tr>
<tr>
<td>Spoken word</td>
<td>Audio disc</td>
<td></td>
</tr>
</tbody>
</table>

Name authority
- Name: Lee, T. B.

Subject authority
- Term: Metadata

RDA content type
- Term: 

RDA carrier type
- Term: 

Amazon/Publisher
- Title: ... to extinction!

Where is the record?

- Implicit, not explicit
  - Everywhere and nowhere
- A semantic Web will allow machines to create the record just-in-time
  - We will not have to maintain records just-in-case
- The user will have control over the presentation
  - I want to see an archive or library or museum or Amazon or Google or Flickr or ? display
  - And by avoiding duplication, we can all get on with describing new stuff ...

The hyperdimensional (Tardis) card

W3C Library
Audio shop
Spoken word archive
Lee Museum

"TARDIS four port USB hub, for office-bound Time Lords: Open a time vortex on your desk" – Pocket-lint

Metadata focus

Shift of focus of metadata creation, maintenance, storage, preservation (by professionals, amateurs, machines)

From Record

To Statement(s) = triple(s)

But metadata display ...
... aggregates triples (from multiple sources) to create records on the fly
Thank you!

- gordon@gordondunsire.com
- http://metadataregistry.org/
  - Open Metadata Registry
  - IFLA Namespaces Task Group (needs updated)
- http://dublincore.org/dcmirdataskgroup/
  - DCMI/RDA Task Group (in revision)