BIBFRAME Introduction

Sally McCallum

Library of Congress

European BIBFRAME Workshop

September 2018

Genesis - briefly

- 2007 – LC's *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*
  - Replace the MARC format with a data interchange framework that makes library data more readily available on the web
  - Radical changes coming in the cataloging rules
    - Questioning data model, transcription
  - More variation in resources – electronic, internet, audio and visual media
- 2009 – Began experimenting with linked data via LC Linked Data Service (ID.LOC.GOV) – LCSH, Name Authority file
- 2011 – Began working toward BIBFRAME

Outline

- Goals and Aspirations
- BIBFRAME Data Model
- BIBFRAME Vocabulary
- RDF
- Pilot 1
- Pilot 2
- Time for questions and discussion

BIBFRAME Goals

- Supply search engines with description in a form they can exploit – more mainstream format for data

- Extensibility to new media

- New views of different types of metadata
  - descriptive, authority, holdings
  - coded data, classification data, subject data
  - preservation, rights, technical, archival

- Reconsideration of the data-related activities
  - exchange, internal storage, input interfaces and techniques

Goals

- Use/exploit linking
  - Traditionally with text and identifiers
  - semantic technology with URIs
- Accommodate needs for different types of libraries
  - large, small, research, public, specialized, …
- MARC transition
  - enable reuse of data from MARC
  - provision of transformations to new models
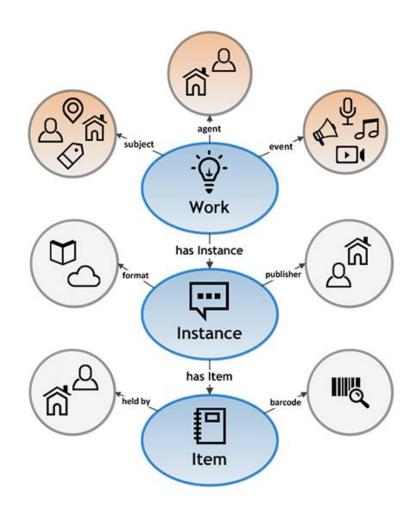- Scalable

Potentials created

- Promise of inferencing

- Efficient creation of descriptions

- New ways to share

- Expansion to resources outside our community – museums, archives, etc.

BIBFRAME Data Model

- Reviewed FRBR, RDA, CIDOC CRM, INDECS, FIAF
- Opted for a _modified_ and flexible RDA/FRBR with
  - Work (RDA/FRBR work and RDA/FRBR expression)
  - Instance (RDA/FRBR manifestation)
  - Item (RDA/FRBR item)

Data Model components - Works

- bf:Work – A conceptual view of a resource
  - Titles
  - Agents (Contributors)
    - Person, Family, Organization, Jurisdiction, Meeting
  - Subjects
    - Topics, Temporal, Places
    - Event as subject
    - Titles, Agents
    - Classification
  - Event as content
  - Form/genre
  - Audience
  - Cartographic coordinates
  - Etc.

Works

- Relationships between Works
  - Expression
  - Translation
  - Adaptation
  - Edition
  - Series
  - Arrangement
  - Succeeding/preceding
  - Etc.
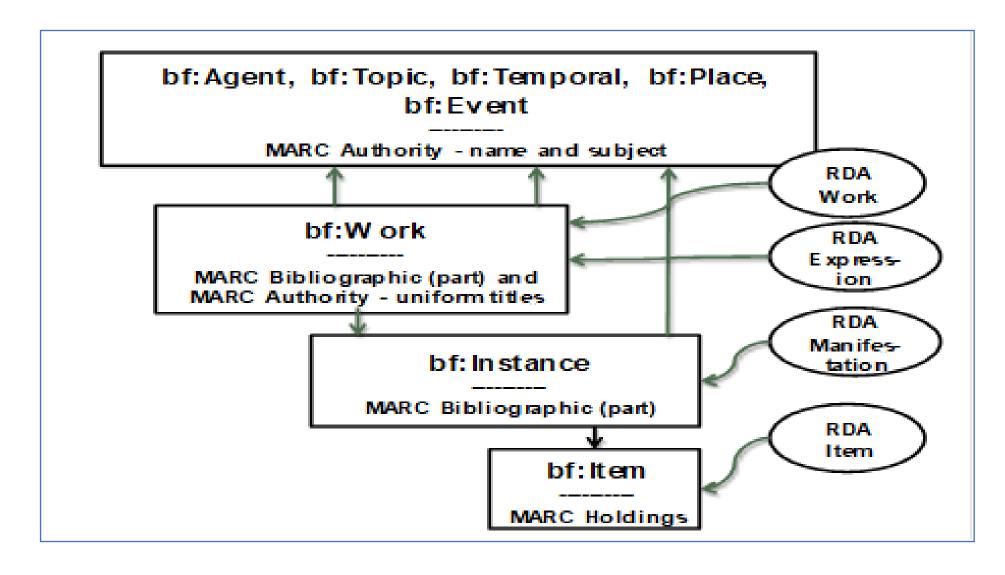- How specific? Depends …

Data model - Instances and Items

- bf:Instance – A physical embodiment of a bf:Work
  - Title found on resource
  - Statements of responsibility, edition, publication, series as found on the resource
  - Provision activities – publication, distribution, production, etc.
  - Dates associated with the physical resource
  - Format of resource
  - Extent of resource
  - Identifiers
  - Normalized data for carriers, resource composition, etc.
- bf:Item – A single example of a bf:Instance
  - Call numbers
  - Holding institutions
  - Barcodes

Striking differences from MARC

- MARC uses the unit record, BIBFRAME Work and Instance descriptions are separate and linked

- MARC title and name-title Authorities are BIBFRAME Works
  - Subjects are associated with Works
  - Work descriptions made for <u>all</u> bibliographic entities

- Instances are separated by carrier, rather than combined into one description – print and microfilm in MARC are 2 BIBFRAME Instances

- Enables expression of most data as URI links

To recap: MARC, BIBFRAME, and RDA

Vocabulary (ontology) for BIBFRAME

- RDA-based property names and data elements included
  - E.g., MARC = target audience; BIFRAME = intended audience
  - E.g., sound, projection, video, and digital characteristics
- Elements not related to cataloging rule are generally MARC-based
  - E.g., shelfmark, barcode, generation process
- Uses RDF structures

RDF in BIBFRAME

- Resource Description Format (RDF)
  - Practical vs. theoretical (rules)?

- Practical principles
  - Make a clear distinction between Datatype (literal) and Object (resource) properties
  - Enable supplying URI, label (literal), or both
  - Generally distinguish types by class
  - Define reciprocal properties, if appropriate
  - Use rdfs: and rdf: properties if they fit: e.g., rdfs:label and rdf:value
  - Avoid proliferation of properties, abstract classes and properties, and over constraining
  - Do not represent metadata about the description as a property of the resource itself
  - Use of blank nodes is an implementation decision

RDF in BIBFRAME

- Naming properties and classes
  - Class names should be a noun
  - Property names should suggest a verb
  - Don't overthink
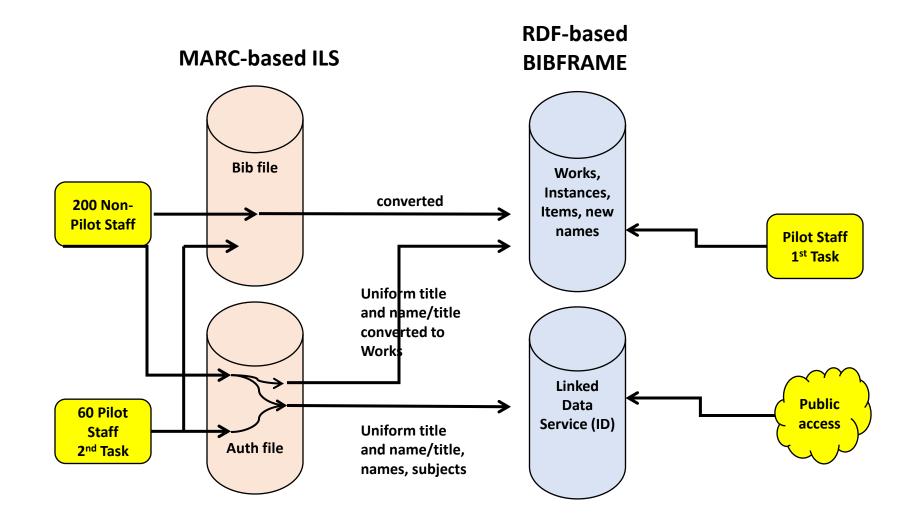
Pilot 1 – 2014/15

- Dimensions of the first pilot
  - Base files to catalog against converted to BIBFRAME
    - 14 million bibliographic records
    - Title authority records converted to BIBFRAME Work descriptions
  - Editor adapted for books, serials, music, maps, sound recordings, …
  - Linked Data Service enhanced slightly
  - 40 catalogers, 12+ languages, 9 different media
  - Training of catalogers – linked data, BIBFRAME, editor, a little RDF
  - BIBFRAME editor created descriptions sved but not integrated into the file

Pilot 1

- Report and assessment
    - Catalogers adapted!
    - Liked type-aheads and drop downs which improved accuracy
    - Liked links to RDA rules in element labels
    - Interested in RDF serializations
    - Work/Instance observed but often ignored
    - Often communicated using MARC tags instead of RDA labels
    - Search limited to known item inadequate – browse needed
    - Need for ability to also input name authorities

Pilot 2 – 2017-

- More complete
  - Realistic cataloging environment
  - Converted whole MARC catalog to BIBFRAME to catalog against
    - 19 million MARC bibliographic records converted to BIBFRAME Works, Instances, and Items
    - 1.2 million uniform title authority records converted to BIBFRAME Works
    - Merged and matched BIBFRAME Works
    - Continuing to refine …
  - BIBFRAME files kept up to date daily with data from NON-Pilot catalogers
- 60+ catalogers
  - BIBFRAME input first
- Linked Data Service (ID) was augmented

MARC-based ILS

RDF-based BIBFRAME

Bib file

Works, Instances, Items, new names

Auth file

Linked Data Service (ID)

200 Non-Pilot Staff

60 Pilot Staff 2nd Task

Pilot Staff 1st Task

Public access

converted

Uniform title and name/title converted to Works

Uniform title and name/title, names, subjects

Pilot 2 results so far

- Improved environment
  - BIBFRAME Database dynamically updated so in synch with MARC file
  - Cataloger save, update, clone
  - Catalogers provided with multifaceted search capability
  - LCSH lookups by segment rather than whole strings
  - Added term lists needed in descriptions to Linked Data Service (ID)
- Developments begun
  - Began work on importing BIBFRAME data from external site
  - Began developing programs for <u>BIBFRAME to MARC</u>
  - Enhancements to the Editor
  - Generated a first BIBFRAME database file of triples for download.

Time for questions

- Later, I will talk more about the details of the conversion and reconciliation (match and merge) processes
- Later, Jodi will give more detail about the editor characteristics

- Thanks!