Focus for BIBFRAME

Sally McCallum

Library of Congress

2017 European BIBFRAME Workshop

Focus – "more" and "keep"

- Increase visibility on the Internet for library resources
- Increase retrieval power
- Increase cultural community interactions
  - Keep data interchange
  - Keep resource sharing

- Since 2000 momentum for change has been growing

Outline

- Linked Data startup
- BIBFRAME startup
- BIBFRAME 2.0
- Recap

2006 – Linked data startup

- 2006 - Tim Berners-Lee's 4 principles
  - URIs, http URIs, RDF+SPARQL, rich links to more

- 2007 – LC's *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*
  - Use technology to get broader use of library curated vocabularies
    - e.g., names, subjects, etc.
  - Replace the MARC format with a data interchange framework that makes library data more readily available on the web.

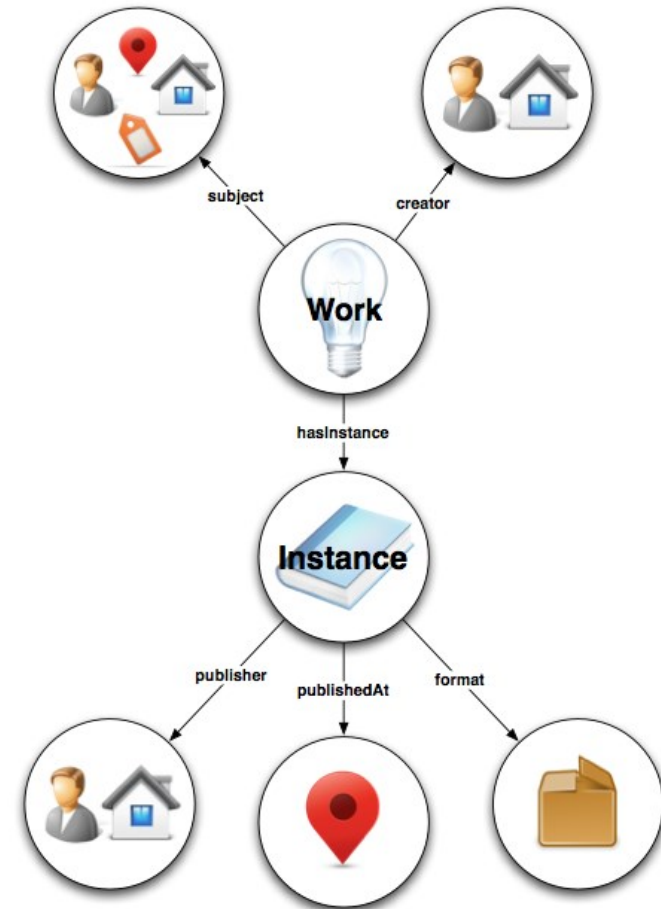# 2009 – LC Linked Data Service (id.loc.gov)

- "Use technology to get broader use of library curated vocabularies"
  - First "curated vocabulary" made available: LCSH
  - Then added --
    - Name authorities (9 million records)
    - Other LC thesauri, such as Thesaurus for Graphic Materials, AFS Ethnographic Thesaurus, **…**
    - Description terms: languages, countries, geographic areas, **…**
    - Cultural organizations codes
    - Preservation (PREMIS) vocabularies, such as event type, cryptographic hash functions, **…**
  - Focused on <u>machine</u> search and retrieval, giving responses in RDF
    - Uses MADSRDF and SKOS
  - Today over 800,000 hits a day on average

2012 – BIBFRAME Startup

- "Replace the MARC format with a data interchange framework that makes library data more readily available on the web"
  - Why replace MARC
    - Age and structure – e.g., length and linking
    - Cataloging rule change (RDA)
    - Modeling activities by museums, archivists, etc.
    - Resource shifts to electronic
  - Challenges
    - LC has 18 million existing MARC records, OCLC has over 330 million
    - Multiplicity of "curated vocabularies" – how to interlink them?
    - Enhancement of description of non-book material, esp. AV
    - Enabling better differentiation of carriers – print, electronic, tape, vinyl, CD, …
    - The huge MARC-based infrastructure – national and international – systems, vendors, services

2012/14 – Discussion & modeling

- Discussion - "Early experimenters", listserv, etc.

- Developed Initial Model (1.0)
  - Innovative and Simple
    - **Works** (FRBR/RDA Works and Expressions)
    - **Instances** (FRBR/RDA Manifestations)
    - **Annotations** for holdings, reviews, cover art, …

2012/14 – Tools & Experimentation

- Tools to support exploration
  - "Format": BIBFRAME RDF Vocabulary 1.0 developed
  - Data: MARC to BIBFRAME conversion tools, BIBFRAME output from LC's system via Metaproxy enabled
  - Input: editor tool, profile editor
  - Support: expansion of Linked Data Service – additional controlled lists, several LC Classification schedules

- Enabled lots of discussion and experimentation by the community

2015/16 – First LC Pilot

- Exploration of the cataloger experience inputting BIBFRAME descriptions
  - Is the Work/Instance dichotomy clear and useful for catalogers?
  - Does search support finding of information needed by catalogers?
  - Is the MARC data transform adequate for cataloger use?

  - Are type-ahead and drop-downs efficient for their tasks
  - Is the labeling of the editor entities appropriate?
  - Are links to cataloging rules useful?

2015/16 – First LC Pilot

- 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
  - 40 catalogers, 12+ languages, 9 different media
  - Training of catalogers – linked data, BIBFRAME, editor, a little RDF
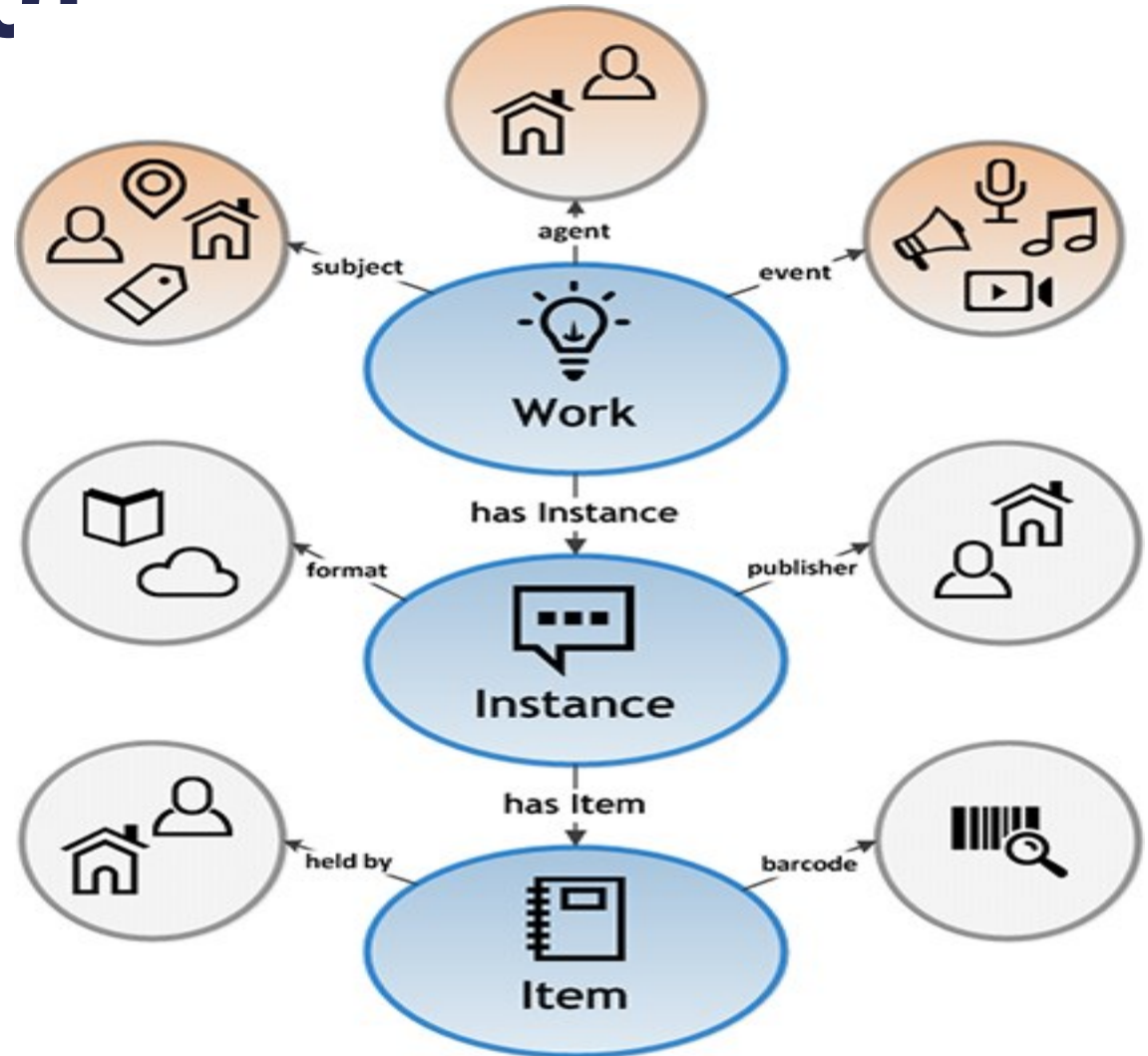
2015/16 – First Pilot

- 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
  - www.loc.gov/bibframe/docs/pdf/bibframe-pilot-phase1-analysis.pdf

2016 – Redevelop BIBFRAME (2.0)

- Based on Pilot One, began planning Pilot Two
- Influences for adjusting model and vocabulary
  - Community wide comments - Listserv, GitHub, …
  - Expert advice
  - Pilot experience, PCC comments
  - Audio Visual media study
    - Report: ***BIBFRAME AV Modeling Study***
  - Proposal papers for key areas, e.g.,
    - Titles, agents and roles, items, identifiers, notes, …
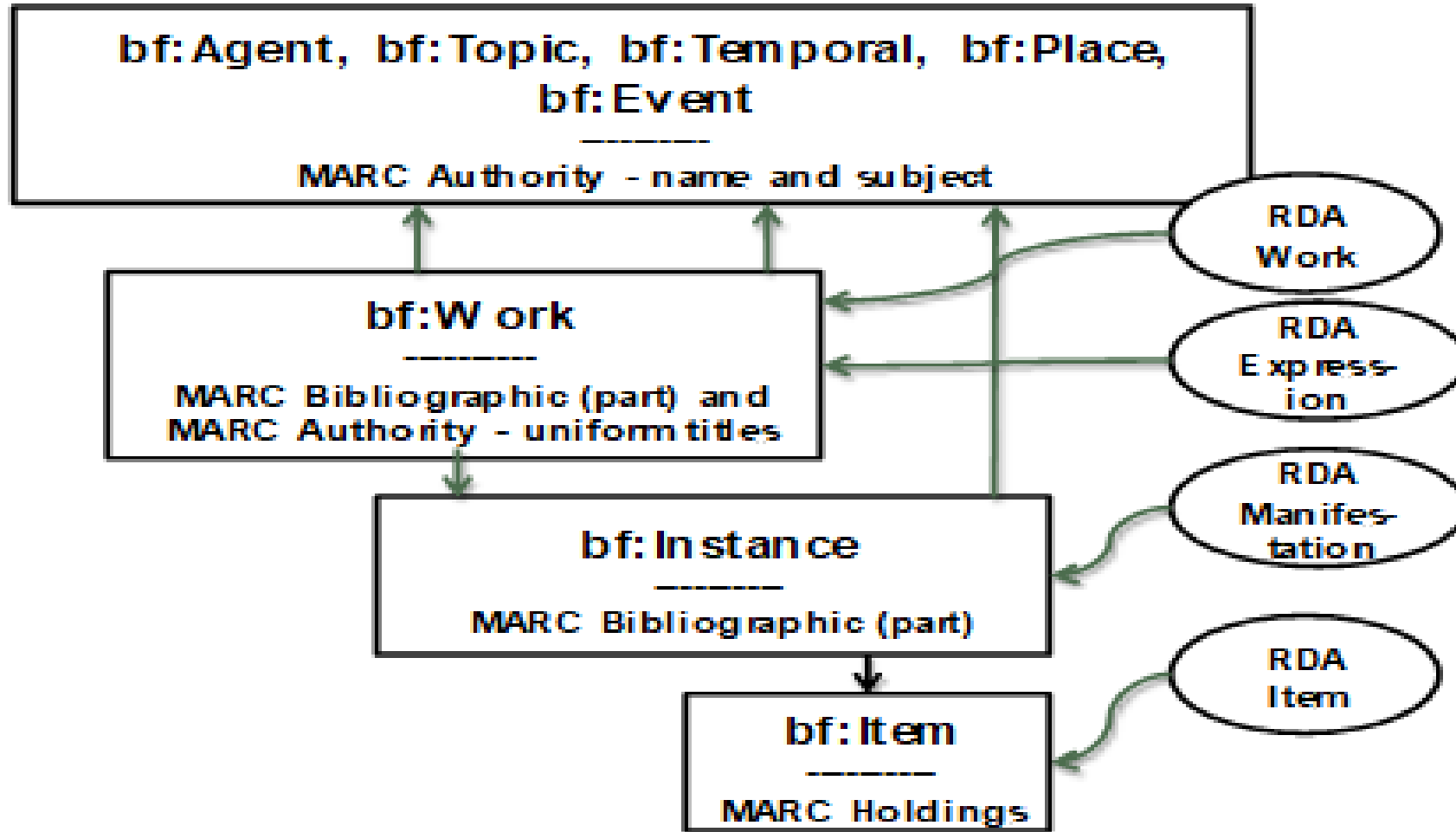
Adjusted model (2.0)

- Fundamentally the same
- Adjustments
  - Add Items
  - Clarify Events
  - Replace Authority class with Agent and Concept classes

Adjusted vocabulary (2.0)

- Continue using the Resource Description Framework (RDF)
  - But consider practicality and rules
- Examples of vocabulary improvements
  - Make a clear distinction between Datatype (literal) and Object (resource, URI) properties
  - Enable supplying URI, label (literal), or both
  - Distinguish types by <u>class</u> where practical
  - Define reciprocal properties, if appropriate

MARC, BIBFRAME, and RDA

2017 – LC Pilot Two

- More complete
  - Realistic cataloging environment
  - Converted whole MARC catalog to BIBFRAME to catalog against
    - 17 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
- 60 catalogers
  - BIBFRAME input first
- Linked Data Service (ID) a key part of the BIBFRAME "machine"

Data issues

- Variability of retrospective MARC data
  - Cataloging styles
  - Duplication – coded, text, controlled terms
  - Models – unit records + title authorities, FRBR
  - Merge with full partial and other types of records
- Transcription vs. access
- Subject analysis
- Aggregations

Explorations

- Scalability, robustness
- Validation of machine creation of BIBFRAME from MARC
- Experiment with taking in BIBFRAME RDF from a vendor
- Test bibliographic extensions
- Offer download of "tuned" BIBFRAME file for others to explore discovery
- Identifiers
- Mapping from BIBFRAME to MARC
- . . .

Meanwhile

- Examine maintenance models
  - E.g., MARC, SRU/Z39.50, W3C, MODS
- Carry out "agile" development of pilots to test and "prove out" some of the theories (and rhetoric) of linked data and RDF
- Share what we learn and do with the community
- Keep up with community discussion and ideas to inform and help mold
- Another pilot or invest in a production environment?

Recap - Input channels

- Current input channels
  - BIBFRAME listserv - BIBFRAME@LOC.GOV
  - GitHub issue trackers for resources posted there - github.com/lcnetdev/
  - Pilot projects – LC and others
  - Consultant and expert analyses

Recap - Exposure

- Extensive web sites for ontologies, vocabularies, proposals, etc.
  - www.loc.gov/bibframe
    - links plus models, conversion specs, notes, analyses, etc.
  - id.loc.gov
    - Vocabularies like LCSH, LCC, NAF,
    - Ontologies like bf:, bflc:
  - bibframe.org
    - Currently, BIBFRAME 1.0 material archive – redevelopment?
  - github.com/lcnetdev/ - for programs and downloads

Recap - Sharing

- Components available for community to explore
  - **BIBFRAME vocabulary**
    www.loc.gov/bibframe/docs/index.html
  - **MARC to BIBFRAME conversion specs**
    www.loc.gov/bibframe/mtbf
  - **MARC to BIBFRAME conversion programs**
    https://github.com/lcnetdev/marc2bibframe2
  - **MARC to BIBFRAME comparison viewer**
    http://id.loc.gov/tools/bibframe/compare
  - **Editor profiles**
    https://github.com/lcnetdev/bfe/tree/development/static/profiles/bibframe

Thanks for your attention!