The Library of Congress BIBFRAME Initiative

History, overview, progression

October 21, 2016
Potomac Technical Processing Librarians
92nd Annual Meeting
Richmond, Virginia

Beacher Wiggins
Director for Acquisitions & Bibliographic Access
Library of Congress
To be covered—

- Where we were
- Where we went
- How we got there
- Where we are
- Where we are going
- Collegial approach
Linked Data and the Library of Congress

- LC engaged in linked data for several years
- First foray was sharing its authority data
- LC created its Linked Data Service (http://id.loc.gov) in 2009
- *Library of Congress Subject Headings* offered as first set of authority data
- Name authorities and various vocabularies followed
- Id.loc.gov played integral role in BIBFRAME Pilot
BIBFRAME’s beginnings were almost a decade ago

LC was pressured for years to develop a replacement for MARC

LC Working Group on the Future of Bibliographic Control’s *On the Record* was final push for LC to figure out a way replace MARC

The time was never quite right for a structure that was considered feasible

With introduction of linked data (RDF—Resource Description Framework) LC saw a viable structure
LC contracted with Zepheira to develop what became BIBFRAME model & vocabulary 1.0

Development of BIBFRAME 1.0 accomplished with input from community

Initially, LC had collaboration of early experimenters—British Library, Cornell, Deutsche Nationalbibliothek, George Washington, National Library of Medicine, and Princeton
This initial work and collaboration helped LC stabilize BIBFRAME model and vocabulary 1.0. This work continued for several years. By late 2014/early 2015, determination made that LC mount a pilot to test:
- efficacy of BIBFRAME
- ability of cataloging staff to create bibliographic data in BIBFRAME structure
Pilot preparation

- Some 45 staff identified for the Pilot

- Mix of catalogers and technicians that catalog
  - Materials in all languages, scripts and formats
  - Monographs, serials, cartographic materials, music (notated), sound recordings, moving image, and two–dimensional art (prints and photographs)

- Participants were to process materials they regularly received
Because LC must continue to distribute MARC records

Participants required to catalog in both the MARC 21 format and BIBFRAME
  ◦ Dual data creation affected the participants’ normal production
  ◦ No attempt to address the impact of BIBFRAME on production
Pilot participants were viewed as pioneers
Worked in a system still under development
Attended 16 hours of instruction on Semantic Web, Linked Data, and use of the BIBFRAME Editor
COIN—Cooperative & Instructional Programs Division staff members provided the training
Training materials available from the Cataloger’s Learning Workshop website
http://www.loc.gov/catworkshop/bibframe/
Module 1: Introduction to the Semantic Web and Linked Data (four and a half hours)

Module 2: Introduction to BIBFRAME Tools (two and half hours)

Training included using PowerPoint slides, quizzes, and exercises
Module 3 consisted of two Units:

- Unit 1—recap of major concepts of the Semantic Web and Linked Data
  - considered necessary because of time gap since participants first exposed to these concepts, and because some found the concepts themselves difficult to understand
- Unit 2—review of process
  - primary goal to provide hands-on training on use of BIBFRAME Editor to create BIBFRAME “descriptions”
  - secondary goals to explain Pilot ‘ground rules’ and to prepare participants to be effective testers and provide helpful feedback.
Pilot preparation—training (continued)

- Module 3—
  - Unit 1—
    - 40–slide PowerPoint presentation.
  - Unit 2—
    - 51–page manual, with plentiful screen captures to show participants what they should see at the various stages of working in the Editor.
Participants began using the BIBFRAME Editor immediately after being trained in its use.

Entered data into both the LC ILS (Voyager) and the BIBFRAME Editor.
  ◦ Created MARC records in LC ILS first.

Weekly ‘de-briefings’ held to help the participants, instructors, and developers.

Midway through Pilot, participants instructed to switch process:
  ◦ enter data into BIBFRAME Editor and then create MARC record in LC ILS.
Searching was available to primary datasets on LC Linked Data Service Authorities and Vocabularies web site, id.loc.gov
- Initially LC/NACO Authority File and *Library of Congress Subject Headings* (LCSH)
- Later, additional datasets from id.loc.gov were made searchable from the Editor

More datasets were searchable via the Editor, as well
- including some controlled lists from *Resource Description & Access* (RDA)
Later in the Pilot, ability to access previously input BIBFRAME descriptions was possible.

Some 2,500 BIBFRAME descriptions created.

Descriptions could not be edited.

Descriptions created in BIBFRAME did not constitute a database of record.

Descriptions not distributed as part of the Library’s cataloging distribution service.

(Kirk and Paul will address in their presentation later, today).
Workflow

- No changes were made in workflow
- Participants were still creating MARC records in the LC ILS
- Not operating in production mode
- BIBFRAME descriptions created will eventually be discarded, since these were in version 1.0
Lessons learned

- Good understanding of RDA needed for working in the BIBFRAME Editor
- Need to focus on using RDA terminology rather than MARC coding
- Participants wanted to see and analyze BIBFRAME RDF—Resource Description Framework serializations created during Pilot
- Reinforced training objectives of including Semantic Web and Linked Data presented in Modules 1 and 2
Pilot System

- Network Development and MARC Standards Office—NDMSO created technical components that supported the Pilot
- Included most of LC’s MARC bibliographic records transformed into
  - BIBFRAME descriptions
  - controlled authority and term lists with URIs
  - BIBFRAME input editor for the participants to use
    (Kirk and Paul will delve deeper later)
Pilot System (continued)

- Pilot’s focus was input of data and impact on catalogers
- End user (researchers) access was not studied
- System did not support
  - holdings
  - acquisitions processes
  - distribution of BIBFRAME descriptions
- 2,500 records created in the Pilot made available in a bulk download file
Can catalogers input BIBFRAME descriptions into a BIBFRAME oriented system?

- Pilot participants submitted some 2,500 descriptions to the system
- Eight profiles for different resource types established to assist with input:
  - monographs, serials
  - notated music
  - Cartographic materials
  - BluRay DVD, Audio CD
  - 35mm Feature Film
  - prints/photographs
Is the Work/Instance dichotomy clear and useful for catalogers?

- Modeling of Works and Instances was clear
- Participants generally just looked for the RDA rule and viewed it or put in the value
- How it was packaged by the BIBFRAME model was not that important to know
- Underscored the dichotomy between the FRBR/RDA and BIBFRAME models
Do type ahead and drop downs make work easier?

- Dropdowns and lookups were popular features
- They improved
  - accuracy of data strings
  - provided the data linking URIs without keying them
  - made input more efficient
Is the labeling on the editor clear and useful?

- BIBFRAME editor used labels
  - closely synchronized with RDA
  - linked to key RDA rules for an element
- Participants found the labels and RDA rule links very helpful
- Treatment of Expressions in BIBFRAME model required additional explanation
- BIBFRAME model considers an Expression a Work with links between the RDA Work and RDA Expression
Can adequate searching be implemented?

- Searching as implemented was adequate but could be improved
- Look ahead fields were very useful for known item searching
- Some “what do you have like this” searching was helpful
- Known item searching usually sufficed
Can the MARC records be transformed adequately for cataloger use?

- Decision made to simulate BIBFRAME environment
- Required conversion of LC file of 18 million MARC bibliographic records to provide BIBFRAME file against which to catalog
- 13.5 million records converted
  - split into Work and Instance records
    - 13.4 million Work records
    - 13.85 Instance records
- Transformation was credible, but still a work in progress
Can the MARC records be transformed adequately for cataloger use? (continued)

- Good enough to illustrate Work/Instance separation, although not thoroughly tested in the Pilot
- MARC Authority records needed by the catalogers were loaded into the LC Linked Data Service
- For Pilot, name authorities were changed from weekly load to daily load to provide up-to-date authority lookup
- Providing input of newly created authority descriptions into the BIBFRAME system was desirable but could not be met in the timeframe
Conclusion

- Pilot (October 1, 2015 – March 31, 2016) achieved its aim and considered a success
- Input from catalogers participating in testing the system enabled those developing BIBFRAME to make considerable strides in its development
- BIBFRAME 2.0 model and vocabulary
  - Released in April 2016
  - will form the basis of the next phase of a pilot in early 2017
Conclusion (continued)

- LC will continue to refine BIBFRAME model and vocabulary 2.0
- Participants will continue to create BIBFRAME descriptions one day per week, so as not to lose their skills
- Work will continue on refining BIBFRAME tools
- Specifications for transformation of MARC data to BIBFRAME being developed
- These will lead to conversion programs for use and testing by community
Conclusion (continued)

- Preparation of files, e.g., to convert LC MARC records for new pilot participants
- Preparation of a more robust infrastructure to accommodate a more rigorous pilot
- Review and augment LC’s Linked Data Service to be more interactive for pilot participants
- Prepare/update needed documentation
LC, as member of LD4P—Linked Data for Production, will work with five institutions funded by a Mellon grant to test BIBFRAME 2.0

- Stanford
- Cornell
- Columbia
- Harvard
- Princeton

Each of these libraries is conducting pilots processing materials/collections that complement the materials LC Pilot encompass...
LC will continue to collaborate with other organizations
- OCLC
- Zepheira
- NLM (Tina Schrader will present later, today)
Thank you!

- For more information on BIBFRAME—
  [http://www.loc.gov/bibframe/](http://www.loc.gov/bibframe/)

- Beacher Wiggins  [bwig@loc.gov](mailto:bwig@loc.gov)
- Director for Acquisitions & Bibliographic Access
- Library of Congress
- 101 Independence Avenue, SE
- Washington, DC  20540
- (202) 707–5137  FAX—(202) 707–6269