Converting and Reconciling

Sally McCallum, Library of Congress

European BIBFRAME Workshop

Florence, Italy – September 2018

Outline

- MARC to BIBFRAME conversion
  - Conversion
  - Workflow
  - Works and Instances
- BIBFRAME to MARC conversion
  - Conversion issues
  - MARC title Authorities
  - Non-Latin and transliteration
  - URIs and omissions

MARC to BIBFRAME conversion

MARC to BIBFRAME conversion

- MARC Bibliographic Unit Record
  - Central is the description of the manifestation
  - Contains data associated with works and expressions like subjects, uniform titles, genre forms, etc. – must be split out
  - May contain information about several manifestations, such as print and electronic, or print and microform, vinyl and tape – must be split into multiple instances

MARC to BIBFRAME conversion

- MARC Authority title records
  - Similar to a part of a BIBFRAME Work record
  - Only made under certain circumstances. Fewer than 5% of books records in LC relate to a MARC Authority title record – must be matched with work data from bibliographic records
  - For some fields, these MARC Authority records use different tags from tags used for the same data in a MARC bibliographic record – must specify special conversion
    - E.g., MARC Authority 4XX transform like MARC Bibliographic 7XX

MARC to BIBFRAME conversion - Issues

- MARC changes over time
  - variety in data as cataloging rules changed – preprocess
  - extent of MARC – convert what you have
- Duplication of data in MARC
  - Coded, literal, controlled vocabularies
- Local data – local mappings
- Omissions
  - nac – no attempt to convert
  - ignore - not needed in BIBFRAME environment
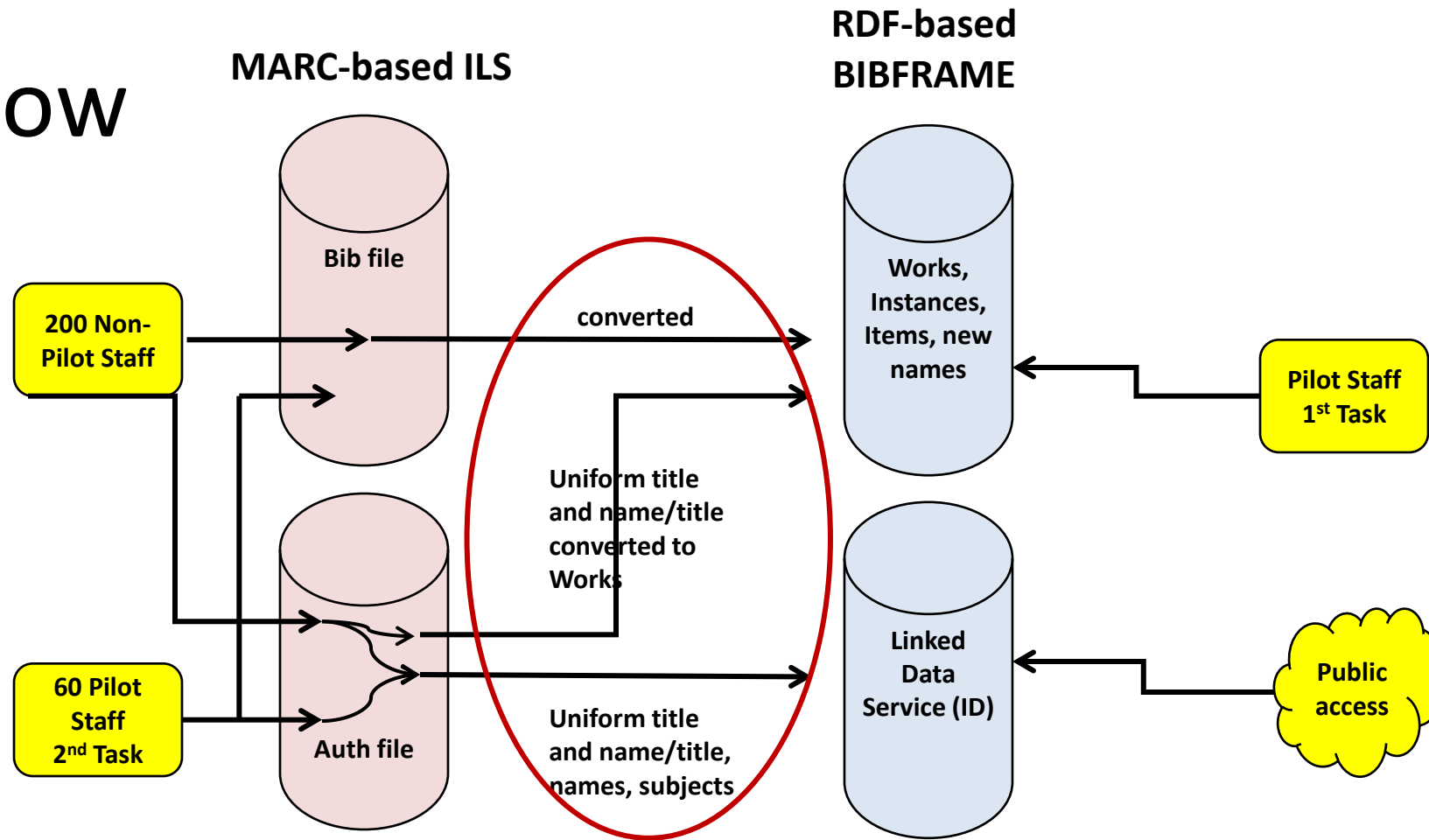 - All MARC content designation so others can supply conversion

MARC to BIBFRAME conversion specifications

521 - TARGET AUDIENCE NOTE (R)	I - intendedAudience - IntendedAudience
Indicators	
First - Display constant controller	
# - Audience	ignore
0 - Reading grade level	## - note - Note - rdfs:label "reading grade level"
1 - Interest age level	## - note - Note - rdfs:label "interest age level"
2 - Interest grade level	## - note - Note - rdfs:label - Interest grade level"
3 - Special audience characteristics	nac
4 - Motivation interest level	nac
8 - No display constant generated	nac
Subfield Codes	
$a - Target audience note (R)	## - rdfs:label
$b - Source (NR)	## - source - Source rdfs:label
$3 - Materials specified (NR)	See $3 spec

MARC to BIBFRAME conversion programs

- Excel specs plus xslt conversion

- With Indexdata's YAZ or Metaproxy search tools, can configure conversion with look ups

- Github page for converter has instructions on how to configure and use: https://github.com/lcnetdev/marc2bibframe

Pilot workflow

MARC-based ILS

RDF-based BIBFRAME

Bib file

Works, Instances, Items, new names

200 Non-Pilot Staff

converted

Pilot Staff 1st Task

Uniform title and name/title converted to Works

60 Pilot Staff 2nd Task

Auth file

Uniform title and name/title, names, subjects

Linked Data Service (ID)

Public access

MARC to BIBFRAME workflow - Works

- Works from MARC Bibs and Authorities
  - MARC Bibs with uniform titles (240) → derive BIBFRAME Work description using uniform title information for Work title
  - MARC title Authorities → convert to BIBFRAME Works
  - MARC Bibs without uniform titles → Create BIBFRAME Work description using 245 for Work title
  - Merge Work descriptions for same Work

MARC to BIBFRAME workflow - Works

- BIBFRAME Work consolidation
  - Consolidate subjects, genre, other data in BIBFRAME Works that were developed from merger of several Work descriptions
  - Make BIBFRAME Work descriptions from MARC linking entries and MARC 7XX, 6XX, and 8XX
  - Merge Work descriptions generated from linking entries into existing Works or create new Work

MARC to BIBFRAME workflow – Instance/Item

- Creating BIBFRAME Instances and Items

  - Create BIBFRAME Instances from MARC Bib

  - Separate BIBFRAME Instance descriptions for different carriers

  - Create BIBFRAME items from MARC bib

Work/Instance conversion challenges

- Tried to keep matching simple using title, author/title, and LCCN where possible

- Special conversion for Authority title records required because 1XX, 4XX, 5XX treated differently in MARC Bib and Authority

- MARC Authority title records are not made for all MARC Bib 240 strings

- Titles for serials difficult to disambiguate

- For some media many MARC Bib records have the title: "untitled"

- Nonsorting differences create difficulty matching MARC Bib 7XX and MARC Authority 4XX and 5XX with 240/245

Works/Instances from MARC - Stats

- Converted approximately
  - 19 million MARC Bibliographic records
  - 1.2 million MARC Authority title records
- Created
  - Over 19.2 million BIBFRAME Work descriptions
  - Over 23.7 million BIBFRAME Instance descriptions
  - RDF file of over 4 billion triples

Completing the circle - BIBFRAME to MARC conversion

BIBFRAME to MARC conversion

- Motivations
  - <u>Double keying of bibliographic data in Pilot</u>
  - Need for transformation fluidity
    - complexity of transformation to BIBFRAME and RDA data models
    - MARC records will be converted multiple times, back and forth
  - Need to supply both full MARC records and full BIBFRAME descriptions into foreseeable future
  - Many systems will probably need MARC for some subsystems that may still require it

BIBFRAME to MARC conversion specifications

intendedAudience/IntendedAudience	**521**	**R**	**[a bf:Instance] bf:intendedAudience _:b1 .** **_:b1 a bf:IntendedAudience .**		**Each intendedAudience/IntendedAudience will create a MARC 521.**
intendedAudience/IntendedAudience/note/Note/rdfs:label	521 I1	NR	[a bf:Instance] bf:intendedAudience _:b1 . _:b1 a bf:IntendedAudience . _:b1 bf:note _:b2 . _:b2 a bf:Note . _:b2 rdfs:label "reading grade level" .	If (value == "reading grade level") {"1"} else if (value == "interest age level") {"2"} else if (value == "interest grade level") {"3"} else {"8"}	
intendedAudience/IntendedAudience	521 I2	NR	[a bf:Instance] bf:intendedAudience _:b1 . _:b1 a bf:IntendedAudience .	" "	
intendedAudience/IntendedAudience/rdfs:label	521 $a	R	[a bf:Instance] bf:intendedAudience _:b1 . _:b1 a bf:IntendedAudience . _:b1 rdfs:label "Moderately motivated." .		
intendedAudience/IntendedAudience/source/Source/rdfs:label	521 $b	NR	[a bf:Instance] bf:intendedAudience _:b1 . _:b1 a bf:IntendedAudience . _:b1 bf:source _:b2 . _:b2 a bf:Source . _:b2 rdfs:label "Follett Library Book Co." .		
intendedAudience/IntendedAudience/bflc:appliesTo/bflc:AppliesTo/rdfs:label	521 $3	NR	[a bf:Instance] bf:intendedAudience _:b1 . _:b1 a bf:IntendedAudience . _:b1 bflc:appliesTo _:b2 . _:b2 a bflc:AppliesTo . _:b2 rdfs:label "first movement" .		
	521 $6	NR			See specs for linkage

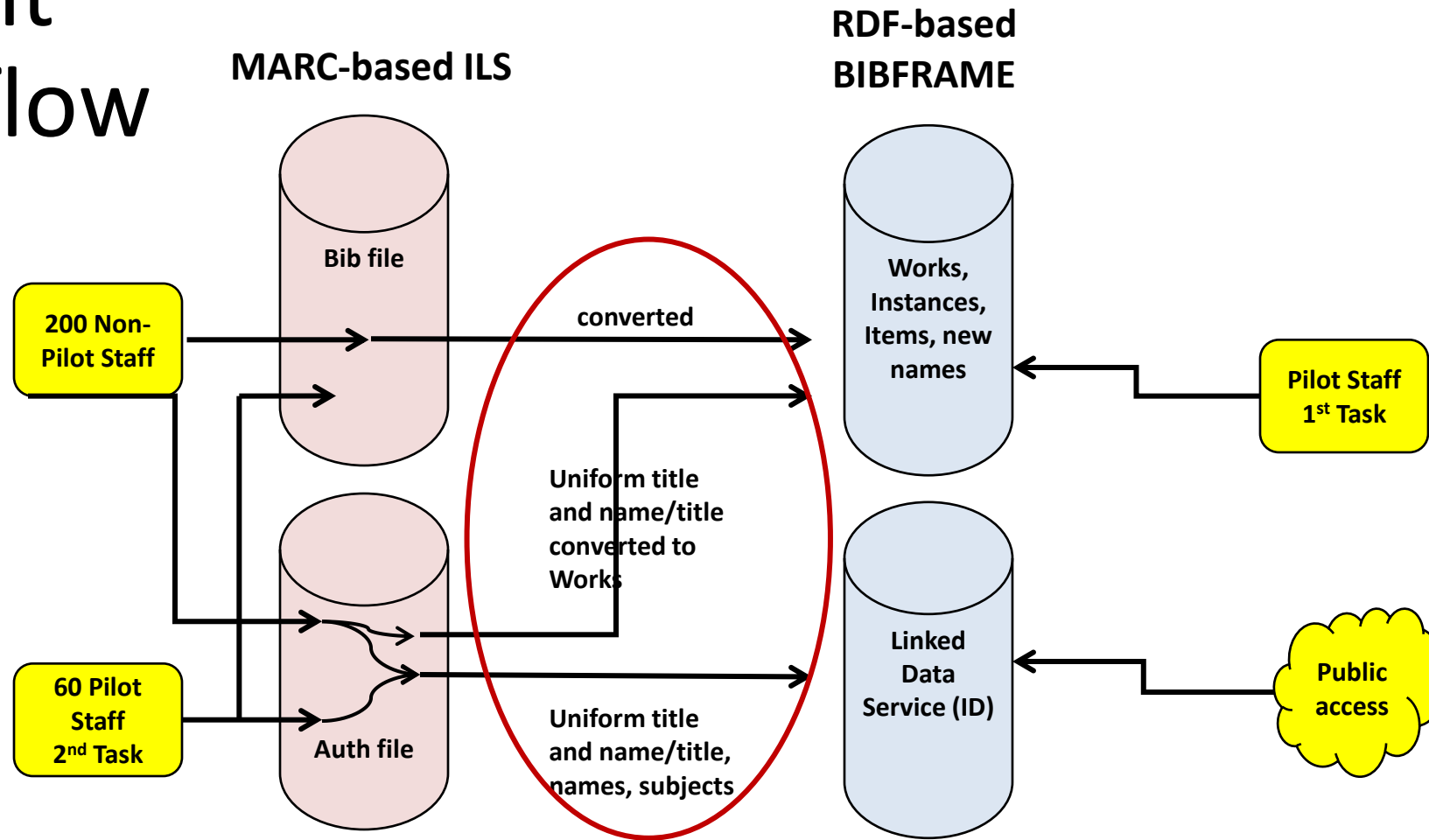# BIBFRAME to MARC conversion issues

- MARC title Authority records – complicated to convert

- Non-Latin script descriptions – different path taken by BIBFRAME

- Preservation of URIs - need to be able to keep anywhere

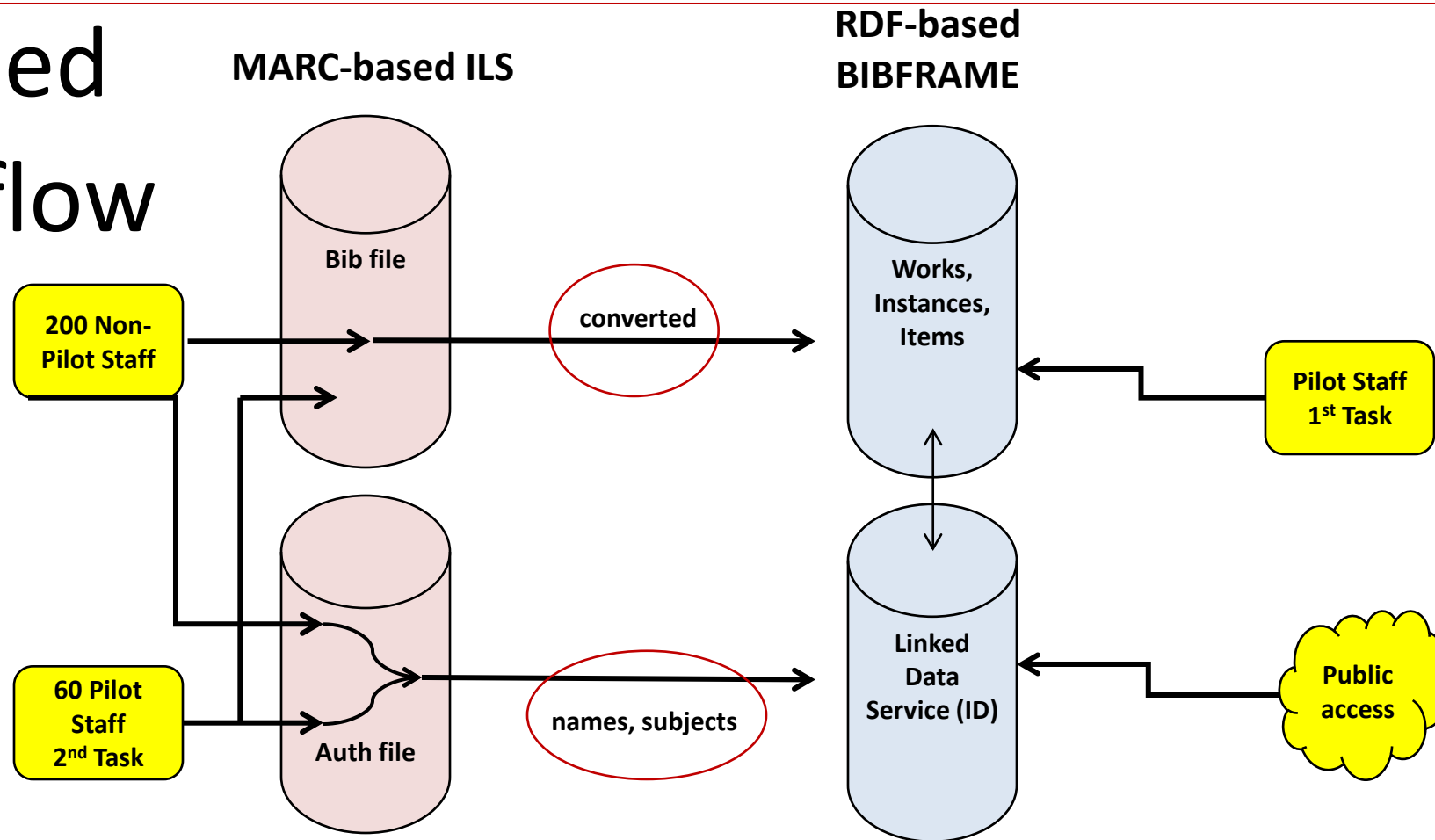- Omissions – many not possible to regenerate

BIBFRAME to MARC conversion issues:
MARC title Authorities as Bibliographic Works

Current workflow

MARC-based ILS

RDF-based BIBFRAME

Bib file

Works, Instances, Items, new names

200 Non-Pilot Staff

converted

Pilot Staff 1st Task

Uniform title and name/title converted to Works

60 Pilot Staff 2nd Task

Auth file

Uniform title and name/title, names, subjects

Linked Data Service (ID)

Public access

Simplified workflow

MARC-based ILS

RDF-based BIBFRAME

Bib file

Works, Instances, Items

converted

200 Non-Pilot Staff

Pilot Staff 1st Task

Auth file

names, subjects

Linked Data Service (ID)

60 Pilot Staff 2nd Task

Public access

Making MARC title Authorities into MARC Bib Works

- Make MARC Bibliographic title "Works"?
- What is missing from the Bib format?
  - Indication that MARC record is a Work description
  - Variant title fields (MARC Authority 4XX)
    - Bib format has variant titles in field 246 or use 7XX with new indicator
  - Related title fields (MARC Authority 5XX)
    - Bib format has related titles in fields 7XX
  - Related agent names (MARC Authority 5XX)
    - Bib format has related agents in fields 700-720
  - Series control fields, MARC Authority 640-646
    - E.g., Numbering examples, analysis, tracing, classification practice
    - Create a special series of fields if still needed

Making MARC title Authorities into MARC Bib Works

- What is in the Bibliographic format that is missing from Authorities?
  - Richer related resource linking
    - MARC Bib 7XX, added entries and linking entries
  - Genre/form-type information
    - MARC Bib 655, various 008
  - Notes that apply to all Instances, e.g.,
    - MARC Bib 524, Preferred citation
    - MARC Bib 520, Summary
    - MARC Bib 586, Awards note
  - Subject Information
    - MARC 6XX, Subject access fields

BIBFRAME to MARC conversion issues:
Non-Latin in MARC

Non-Latin in MARC Bibliographic

- MARC <u>Bibliographic</u> records – current practice
  - Latin only in regular (non-880) fields
  - All non-Latin in 880 fields
  - Fields 880 linked to corresponding regular fields
    - 880 10 $6 245-02/$1 $a 樊巖 集

  - Transliterated data: Main entry, titles, imprints, notes, etc.

Non-Latin in BIBFRAME

- Instance descriptions
  - Transcribed data is only in the script of the resource, no duplication for transliteration
    - title statement and statement of responsibility, edition statement, publication statement, notes, etc.
  - Instance description is linked to the Work where there is transliterated Latin script for the title, contributors, etc.
  - All Instances are linked to Works

Reconciliation for non-Latin

- Cease using MARC 880 fields and transliterating description fields
- Continue transliteration for MARC Authority records and Work records.

BIBFRAME to MARC conversion issues:
URIs

URIs

- Preservation of URIs at the subfield level and control fields
  - URIs in records for components of subjects, e.g., in BIBFRAME:

```
<bf:subject>
  <bf:Topic>
    <madsrdf:componentList rdf:parseType="Collection">
      <madsrdf:Geographic rdf:about="http://id.loc.gov/authorities/subjects/sh00005709"/></madsrdf:componentList>
    <madsrdf:componentList rdf:parseType="Collection">
      <madsrdf:Topic rdf:about="http://id.loc.gov/authorities/subjects/sh99001795"/></madsrdf:componentList>
    <madsrdf:componentList rdf:parseType="Collection">
      <madsrdf:GenreForm rdf:about="http://id.loc.gov/authorities/subjects/sh99001964"/></madsrdf:componentList>
  </bf:Topic>
</bf:subject>
```

  - URIs for coded data in fixed length control fields, e.g., 008

- Thanks!

- Questions and discussion?