Introduction to RDF

Resource Description Framework

Cataloging for the Future Series

Robin Fay, Cataloger & Trainer
LibraryScope | @georgiawebgurl

PLEASE INTRODUCE YOURSELF USING CHAT …

Tell me a little about yourself or your group

Area of the library?

Experience with web technologies, cataloging/ metadata?

RDF – Agenda

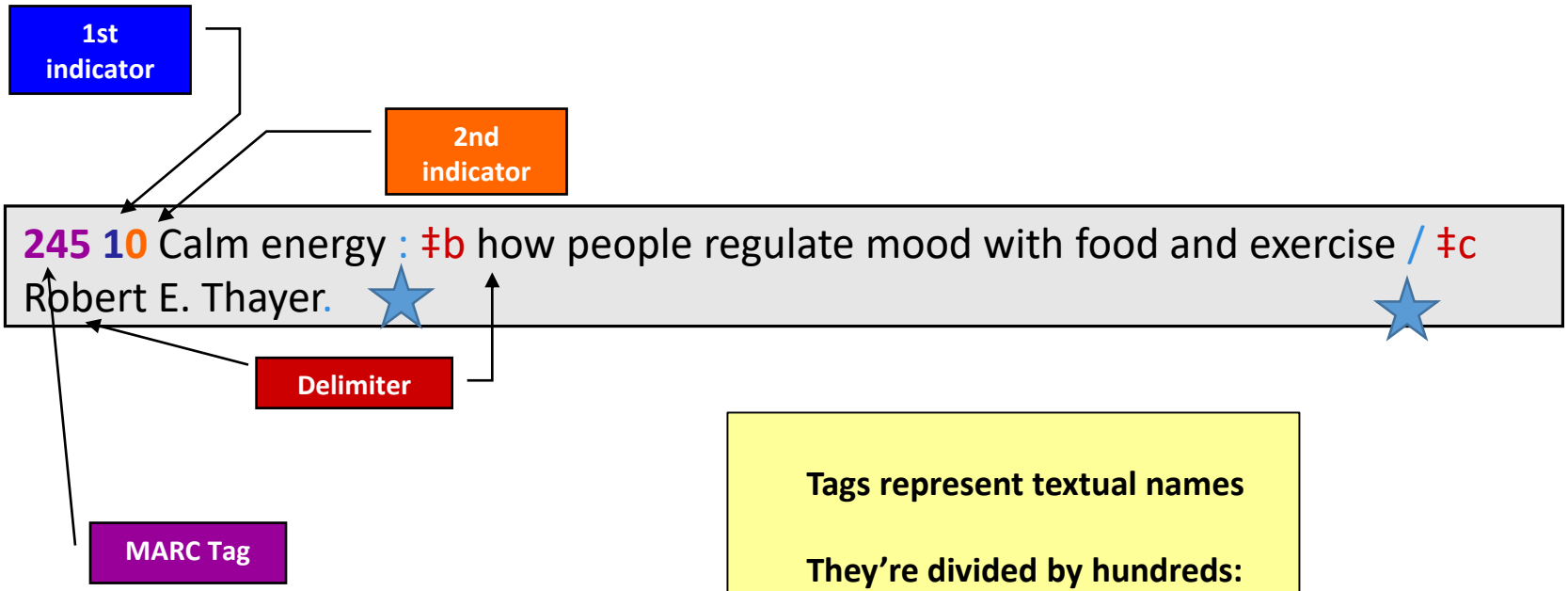
Where we most of us are (MARC)
XML as translator – MARCXML
RDA expressed as RDF
RDF basics
- overview & origins
- BIBFRAME

Resources & Quizzes

It's a MARC world… Our goal

- Describing the items we have
- Providing inventory/tracking control (barcodes, circulation records, item status, orders)
- Providing access via URLs, call numbers, etc.
- Helping our patrons find what they need (User Tasks) regardless of where they are (ILS, IR, physical stacks…)

A MARC field diagrammed



Tags represent textual names

They're divided by hundreds: e.g., 100, etc.

Subfields are represented by the double dagger ‡ , $, or |

START OVER | REGULAR DISPLAY | RETURN TO BROWSING | ANOTHER SEARCH | (Search History)

MARC TAG : FIELD

Indicators

ISBD

Subfields

```
LEADER 00000cam  2200000 a 4500a
001    50906019
003    OCoLC
005    20030715092633.0
008    021023s2003    ilu      b    001 0 eng
010    2002151683
015    GBA3-Y7095
020    0838908470
040    DLC|cDLC|dUKM|dC#P|dXFF|dKSU|dOCoLC
049    KSUU
050 00 Z666.5|b.C37
082 00 025.3|221
100 1  Caplan, Pris
245 10 Metadata fundamentals for all librarians /|cPriscilla
       Caplan
260    Chicago :|bAmerican Library Association,|c2003
300    ix, 192 p. ;|c28 cm
504    Includes bibliographical references and index
505 00 |tMetadata basics --|tSyntax, creation, and storage --
       |tVocabularies, classification, and identifiers --
       |tApproaches to interoperability --|tMetadata and the Web
       --|tLibrary cataloging --|gThe|tTEI header --|gThe|tDublin
       Core --|tArchival description and the EAD --|tMetadata for
       art and architecture --|tGILS and government information -
       -|tMetadata for education --|tONIX International --
       |tMetadata for geospatial and environmental resources --
       |gThe|tData Documentation Initiative --|tAdministrative
       metadata --|tStructural metadata --|tRights metadata
650  0 Metadata
650  0 Information organization
```

RDA – Resource, Description & Access

- Guidelines for how we describe items aka descriptive metadata.
- Provides more flexibility in describing content, especially digital content.
- It includes new MARC fields.
- RDA emphasizes relationships, transcription, and more data not less*. While we typically use it in MARC, RDA can be expressed in XML or RDF (a semantic web framework).
- Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier.

XML structure review

Conveys meaning by “marking up” other text and data with tags

Uses “tags” to demarcate elements that begin & end with angle brackets

Opening & closing tags

<title> opening tag

</title> closing tag

<title>The brown bear</title>

<element>value</element>

Does not specify what elements should be used, only how the elements should be tagged and allows for the creation of discreet, machine-actionable data elements and values.

Shares similar features with HTML, another markup language.

Bibliographic data in XML

Instead of transcribing data using characters to denote differences between types of data, clear & unique elements can be used to differentiate them (note MARCXML does this slightly differently)

264_1 … $c 1966

<publication date>1966</publication date>

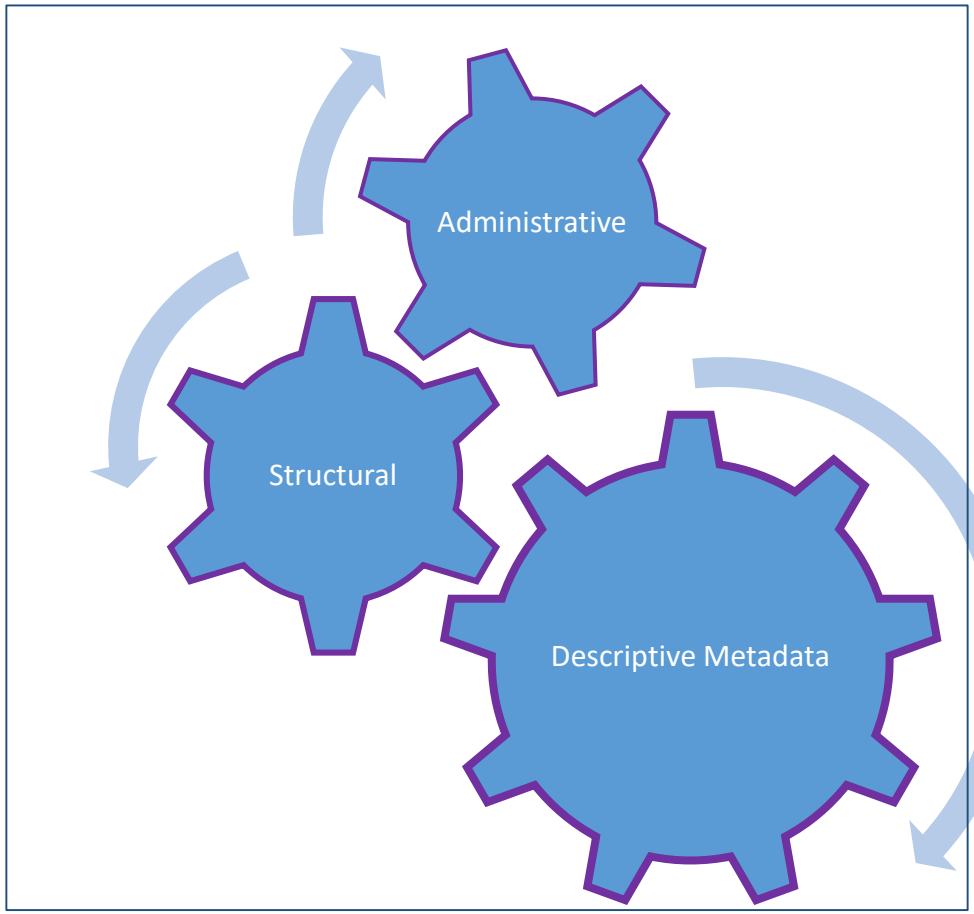
264_4 … $c ©1966

<copyright date>1966</copyright date>

264_1 … $c [1966?]

<approximate date>1966</approximate date>

Basic Types of Metadata

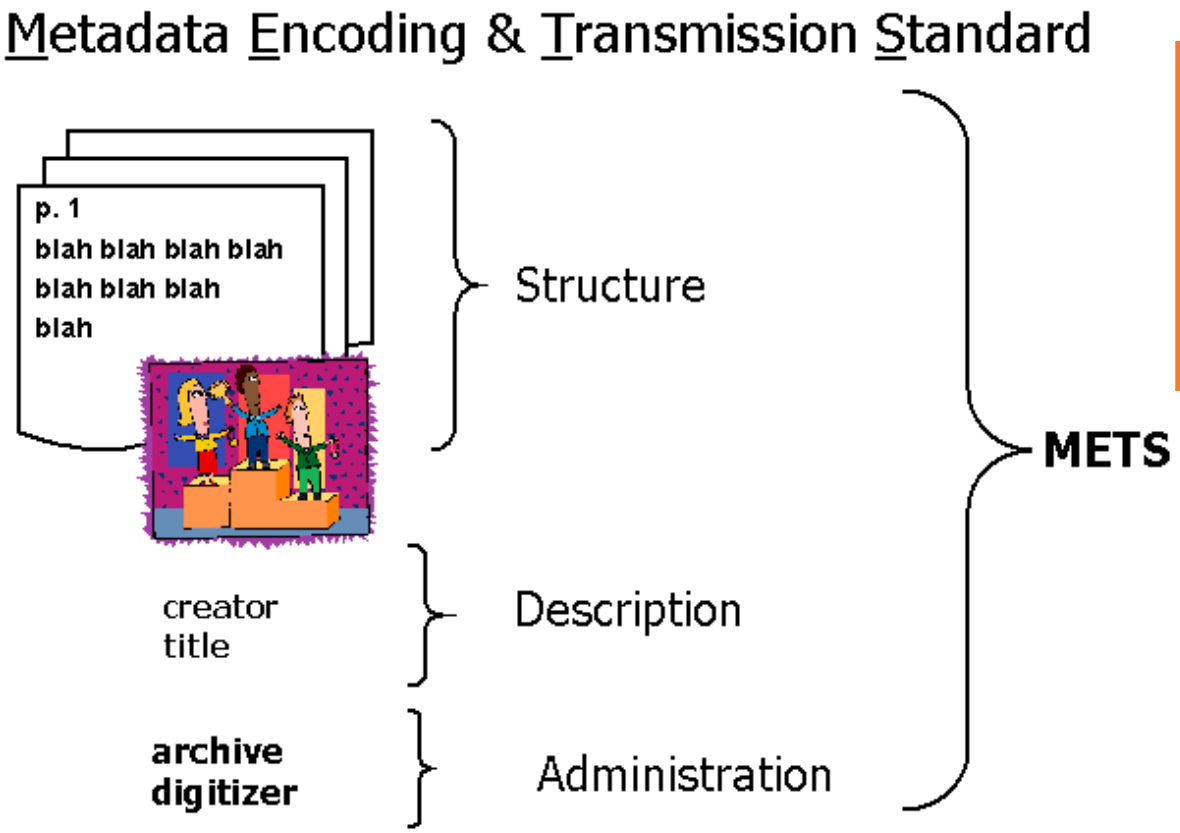


Administrative – Information to help manage the resource – file type, data created, location of creation, publication status, source - includes information to help the system identify and understand what to do with the file (such as knowing what tool to use to open it)

Structural – The structure – similar to a TOC – how many pages, items, files, the order of information, part of or related to

Descriptive – Bib Data – Subject Headings, Title, Publisher, Date, Type

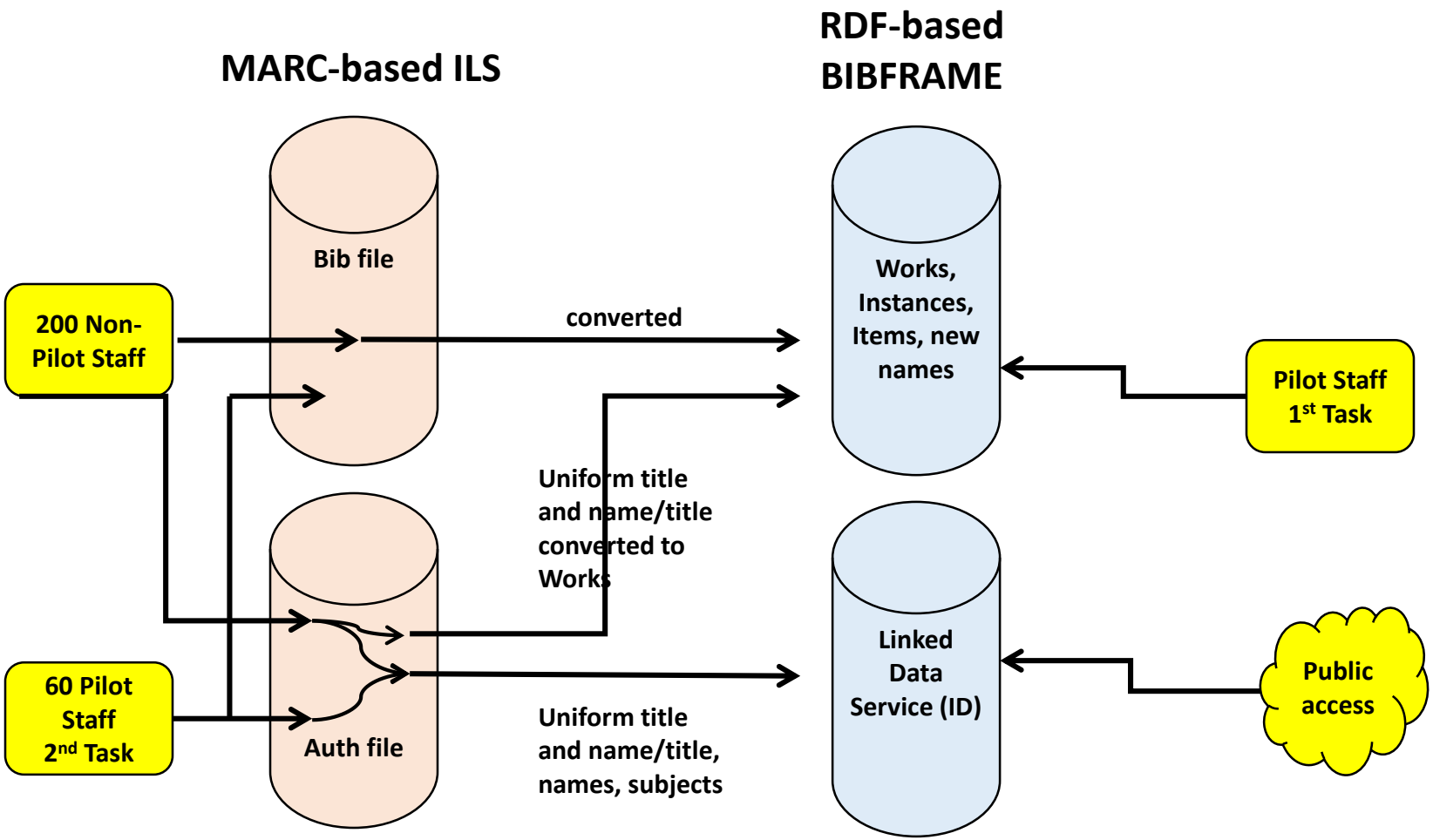
A great example from Karen Coyle!



METS is a standard which includes administrative, descriptive & structural metadata.

Karen Coyle 2004,
http://www.kcoyle.net/meta_purpose.html

LC's big project (started 2017)



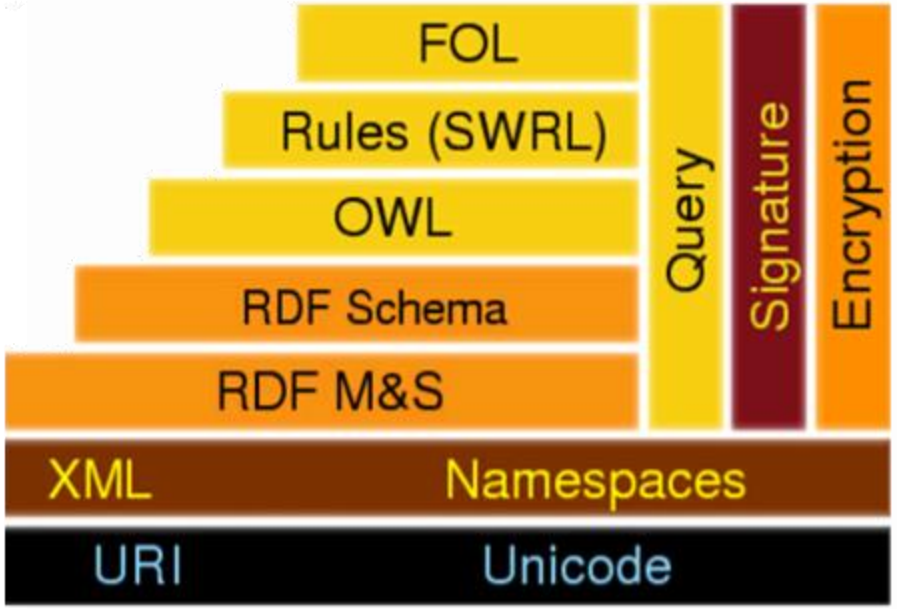
Since RDF and the concepts of the Semantic Web are from a different community, it has all new (to us!) terminology (the Semantic Web Glossary has much more)

RDF = Resource Description Framework

RDFS = Resource Description Framework Schema

OWL = Web Ontology Language

URI = Uniform Resource Identifier - think unique number – often a unique URL pointing to a specific item



Let's start URIs….

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards
4. Include links to other URIs, so they can discover more things

URIs = Uniform Resource Identifier

Tim Berners-Lee's Four Rules

Examples of URIs….

From Library of Congress Name Authority File

Details | Visualization

International Workshop on Semantic Web Services and Web Process Composition

URI(s)

- http://id.loc.gov/authorities/names/nb2004314894

Instance Of

- MADS/RDF ConferenceName
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)

- Library of Congress Name Authority File

RDF: Resource Description Framework is a general-purpose language for representing information in the Web (a metadata data model)

- is a W3C specification (WorldWideWeb Consortium)
- is a conceptual description
- is based upon making statements about web resources (triplets)
- **RDA can be expressed in RDF.**
- Often expressed in XML
- Think sentence structure :
  - subject – predicate(verb)-object
    - My dog eats dogfood.

So, we have the framework, but how do we apply it?

RDFS = Resource Description Framework Schema

A schema is

- **outline**: a schematic or preliminary plan
- A **structure** described in a **formal** language supported by the database management system ; in a relational database [such as MySQL), the schema defines the tables, the fields in each table, and the relationships between fields and tables.
- a description of the structure and **rules** a document must satisfy for an XML document type
- **BIBFRAME is a schema**

One more definition of interest

OWL = Web Ontology Language

- invented to link ontologies which are ***classification*** systems
- Attempts to define objects and their relationships
- Different “flavors”
- “interpreted as a set of "individuals" and a set of "property assertions" which relate these individuals to each other” (wikipedia 2009)
- Not a requirement
- Sounds familiar to catalogers, right?

Controlled vocabularies – short list

- VIAF – Virtual International Authority File
- LC NAF – Library of Congress Name Authority File
- ISNI – International Standard Name Identifier
- ORCiD – Researchers self-register for an ID
- DDC – Dewey Decimal
- AAT – Art & Architecture Thesaurus

Linked data the short version

- Linked data is about reusing data
- Instead of typing in a name, we just include a link that points to the record with the name. Saves time and keeps everything up to date. We will continue moving in this direction. So, records will include more links (actually URIs).
- Ontologies are a structure for linking together controlled vocabularies (like LCSH).
- We can use something called *vocabulary mapping* or *alignment* to create connections between existing vocabularies. (SKOS is one option)

Sharing Vocabularies: SKOS

SKOS is a standard using RDF:

- prefLabel
- broader
- narrower
- related
- closeMatch
- exactMatch

Sharing Vocabularies: SKOS

sh85121349 a skos : Concept ;

skos : inScheme http://id.loc.gov/authorities/subjects;

skos : prefLabel “Shepherds”@en;

skos : broader :sh85001441, :sh85007461, :sh85007805

skos : narrower :sh85036235 :sh85039437

skos : closeMatch http://d-nb.info/gnd/41685855-3

skos : exactMatch http://stitch.cs.vu.nl/vocabularies/Rameau/ark:/ 12148/cb119361753;

RDF Record snipped showing SKOS


<li rel="madsrdf:**hasBroaderAuthority** skos:broader"> **Rural population**

<li rel="madsrdf:**hasNarrowerAuthority** skos:narrower"> http://id.loc.gov/authorities/subjects/sh91003252" property=**Women peasants**

Our BIBFRAME/RDF example

```
        <bf:Source >
            <bf:code >lcsh</bf:code>
        </bf:Source>
    </bf:source>
  </bf:Topic>
</bf:subject>
<bf:subject >
  <bf:Topic rdf:about="http://bibframe.example.org/5226#Topic650-
  22" >
    <rdf:type
    rdf:resource="http://www.loc.gov/mads/rdf/v1#ComplexSubject" />
    <rdfs:label >Wheels--Fiction.</rdfs:label>
    <madsrdf:authoritativeLabel >Wheels--Fiction.
    </madsrdf:authoritativeLabel>
    <madsrdf:isMemberofMADSScheme
    rdf:resource="http://id.loc.gov/authorities/subjects" />
    <madsrdf:isMemberofMADSScheme
    rdf:resource="http://id.loc.gov/authorities/childrensSubjects" />
    <madsrdf:componentList rdf:parseType="Collection" >
      <madsrdf:Topic >
        <madsrdf:authoritativeLabel
        >Wheels</madsrdf:authoritativeLabel>
      </madsrdf:Topic>
```

While not SKOS, we can similarities in our BIBFRAME example.

We see our prefixLabel.
We can see where schema fits in
We do not have narrower, broad, or close match in this record.
MADS (Metadata Archival Description Schema) is used in both of these examples.

But what those RDF triplets?

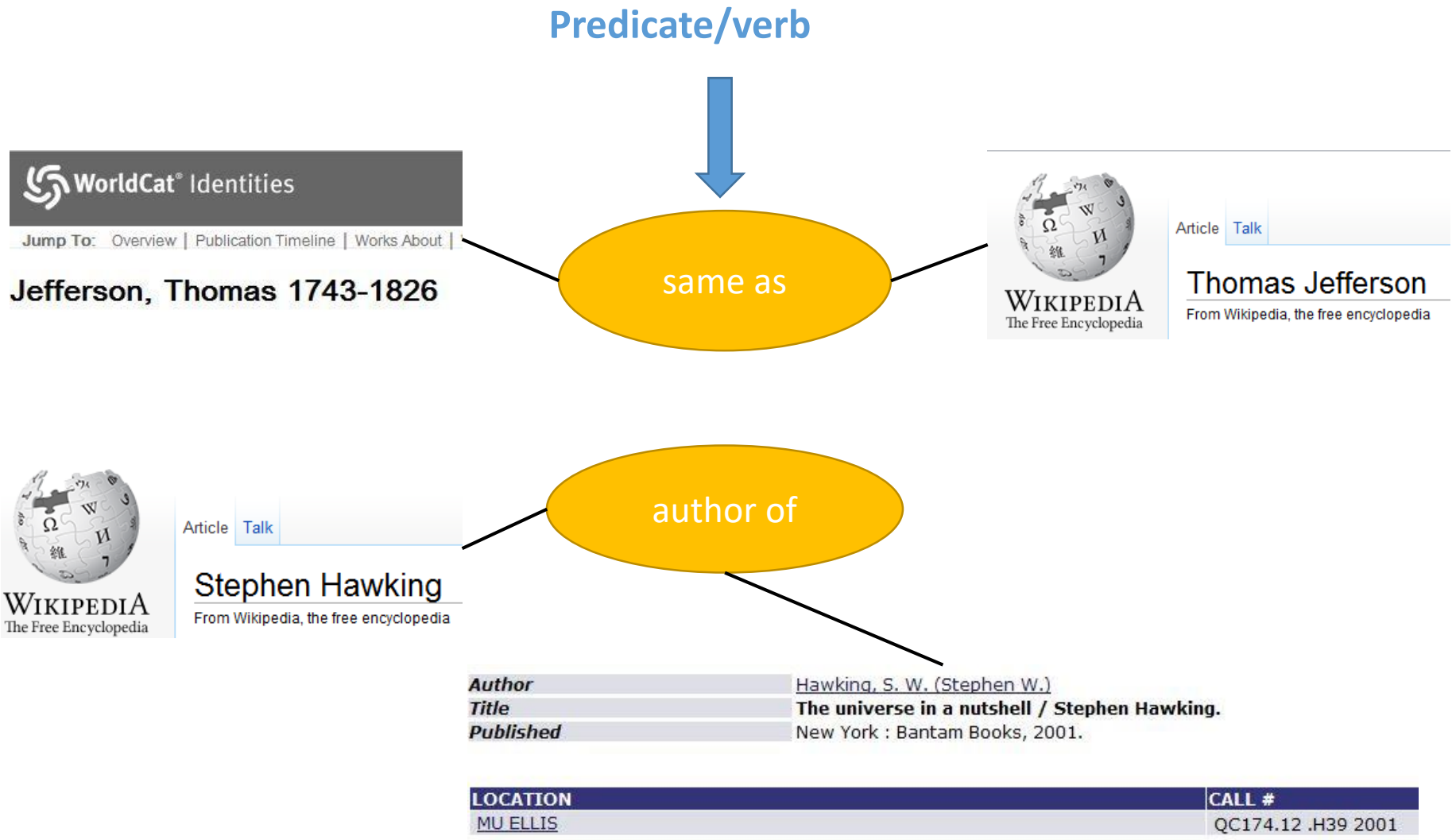
Let's take a look

- Remember my adorable puppy?

subject – predicate(verb)-object

- My dog eats dogfood.
- My dog likes me.

RDA Triplets – More practical examples



So, we can either input the information as we do in a MARC record using text or provided there are unique links to information we can use URIs. What could be the URI for a title?

Title	
Title proper	MARC21 for everyone
Other title information	a practical guide
Statement of responsibility	
Statement of responsibility relating to title proper	Deborah A. Fritz, Richard J. Fritz
Publication statement	
Place of publication	Chicago
Publisher's name	American Library Association
Date of publication	2003
Copyright date	©2003
Mode of issuance	single unit
Identifier for manifestation	083890842X
Media type	unmediated
Carrier type	volume
Number of units †	xvi, 188
Extent	pages
Dimensions	28 cm
Composite key °	MARC21 for everyone. American Library Association. 2003. Volume

Our data in a semantic viewpoint

Field/attribute	Value
Record ID	54321
Title	Museum archives: an introduction
Author	Wythe, Deborah
Date	2004
LCSH	Museum archives
Media/GMD	Electronic
Content form	Text

SOURCE: Getting triples from records: the role of ISBD
http://www.slideshare.net/scottishlibraries/isbd-record2triples

Our data in a semantic view

Field role and relationship

Can map to record such as viaf

"Bib" :Record id as subject

subject	predicate	object
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

riples

Ok, we have a good basic understanding of RDF and some of the semantic frameworks that apply to bibliographic data, so how do we create RDF?

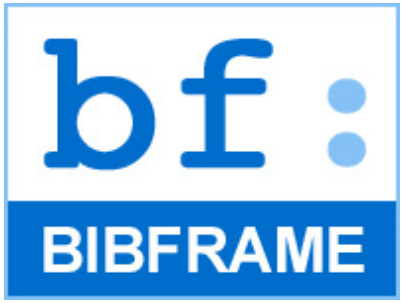
- You can write manually
- You can convert records to RDF
- There are tools for libraries but many tools for semantic web applications. See the resources list for more.
- Three tools to convert/create RDF records for libraries (RIMMF; MARCEdit and BIBFRAME. MARCEdit works with BIBFRAME.

BIBFRAME 2.0

“…aims to re-envision and, in the long run, implement a new bibliographic environment for libraries that makes "the network" central and makes interconnectedness commonplace.”

(read: is attempting to better position the library world for a linked data environment)

Primer for BIBFRAME: http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf



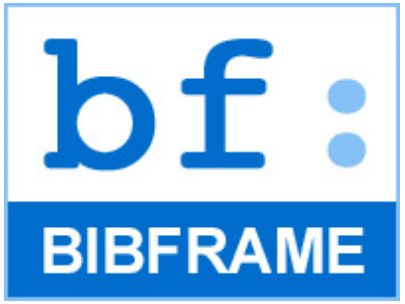
BIBFRAME 2.0

BIBFrame: **BIB**liographic **FRAME**work Initiative

Officially launched by the Library of Congress in 2011.
Now in version 2.0.

A new model for bibliographic data, that will be the basis for an new encoding standard that will replace MARC and will be XML-based.

Consists of the **BIBFRAME Model** is a conceptual/practical model that contains 4 high-level classes, or entities (Work, Instance, Authority, and Annotation) and the **BIBFRAME Vocabulary** which has a defined set of elements and attributes that describe resources and their properties.



BIBFRAME 2.0

Instead of bundling everything neatly as a “record” and potentially duplicating information across multiple records, the BIBFRAME Model relies heavily on relationships between resources (Work-to-Work relationships; Work-to-Instance relationships; Work-to-Authority relationships).

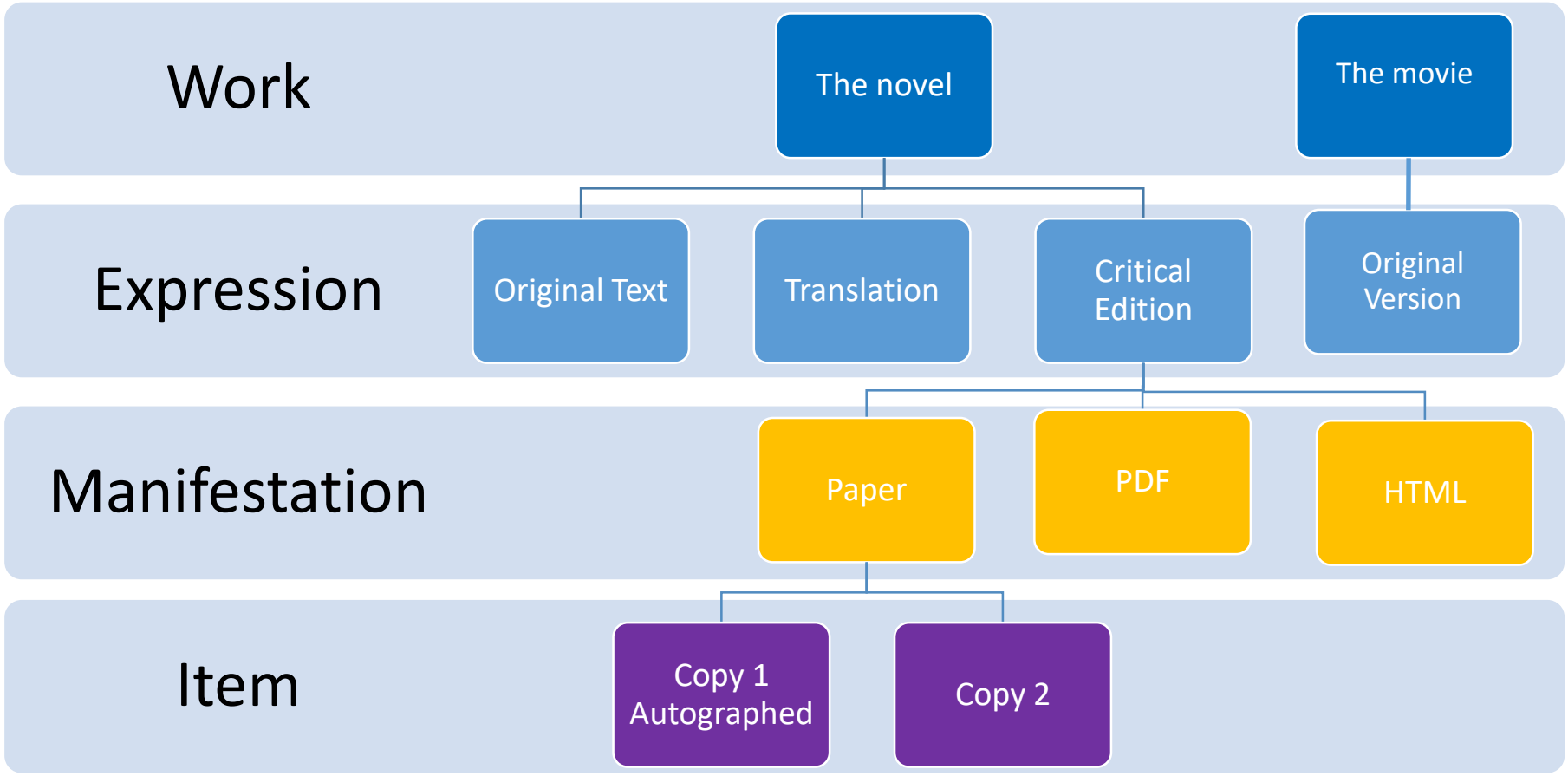
It manages this by using controlled identifiers for things (people, places, languages, etc). MARC employs some of these ideas already (geographic codes, language codes) but BIBFRAME seeks to make these aspects the norm rather than the exception.

In short, the BIBFRAME Model is the library community’s formal entry point for becoming part of a much larger web of data, where the links between things are paramount.

(from BIBFRAME FAQs: http://www.loc.gov/bibframe/faqs/)

Maps to FRBR WEMI

Based on a graphic in Tillett, Barbara "AACR2's Strategic Plan and IFLA Work towards an International Cataloguing Code‟ (2002)



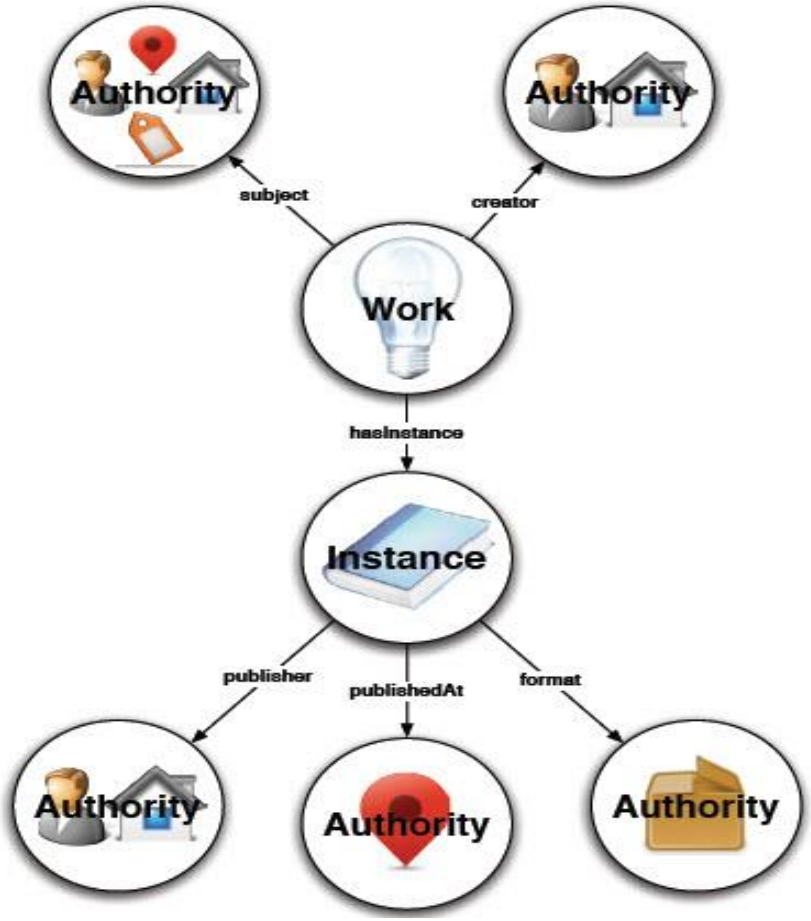
BIBFRAME Model

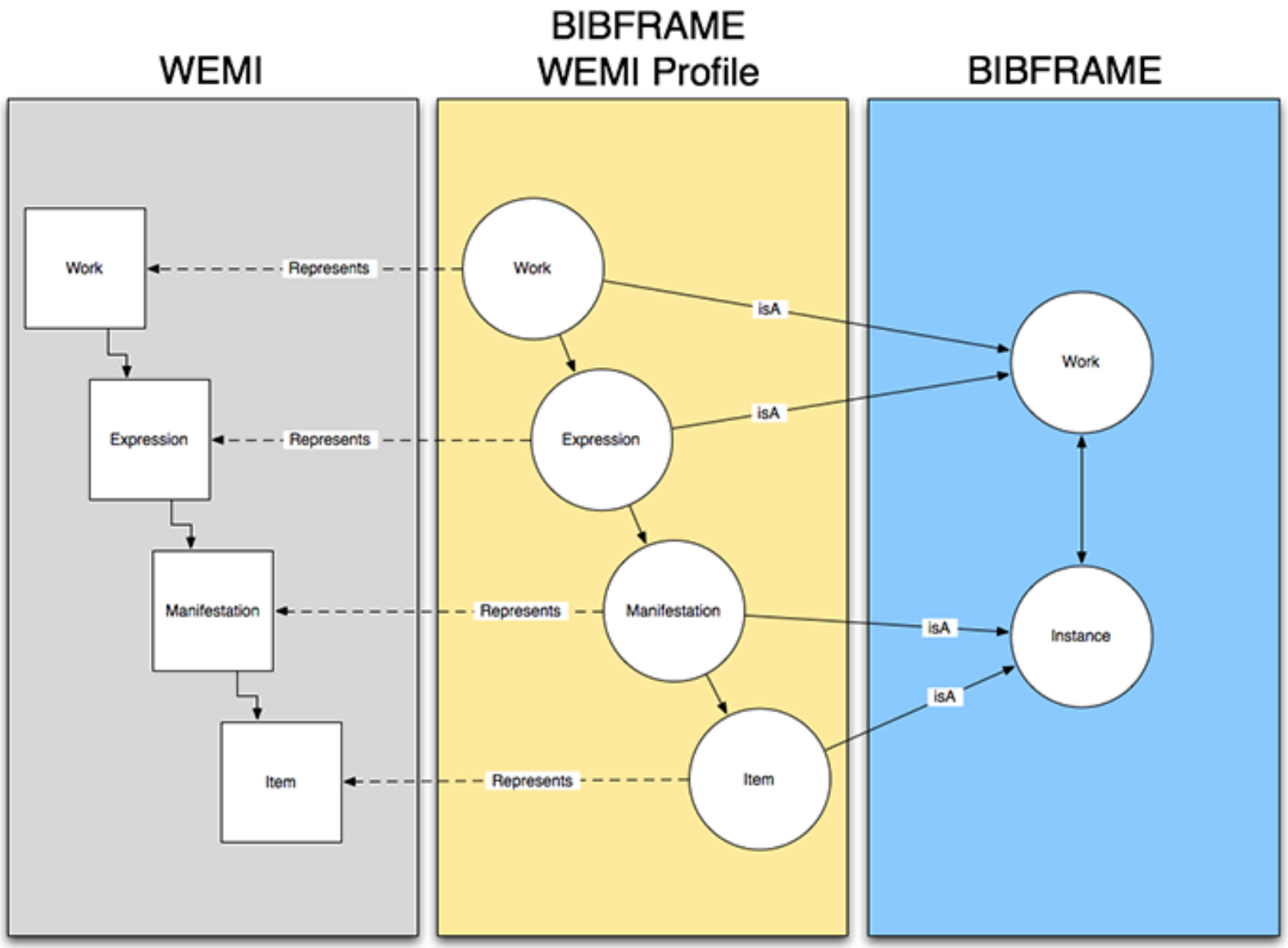
Core elements of the BIBFRAME model – similar but not exactly the same as FRBR

Work resource reflecting the conceptual essence of the cataloged item

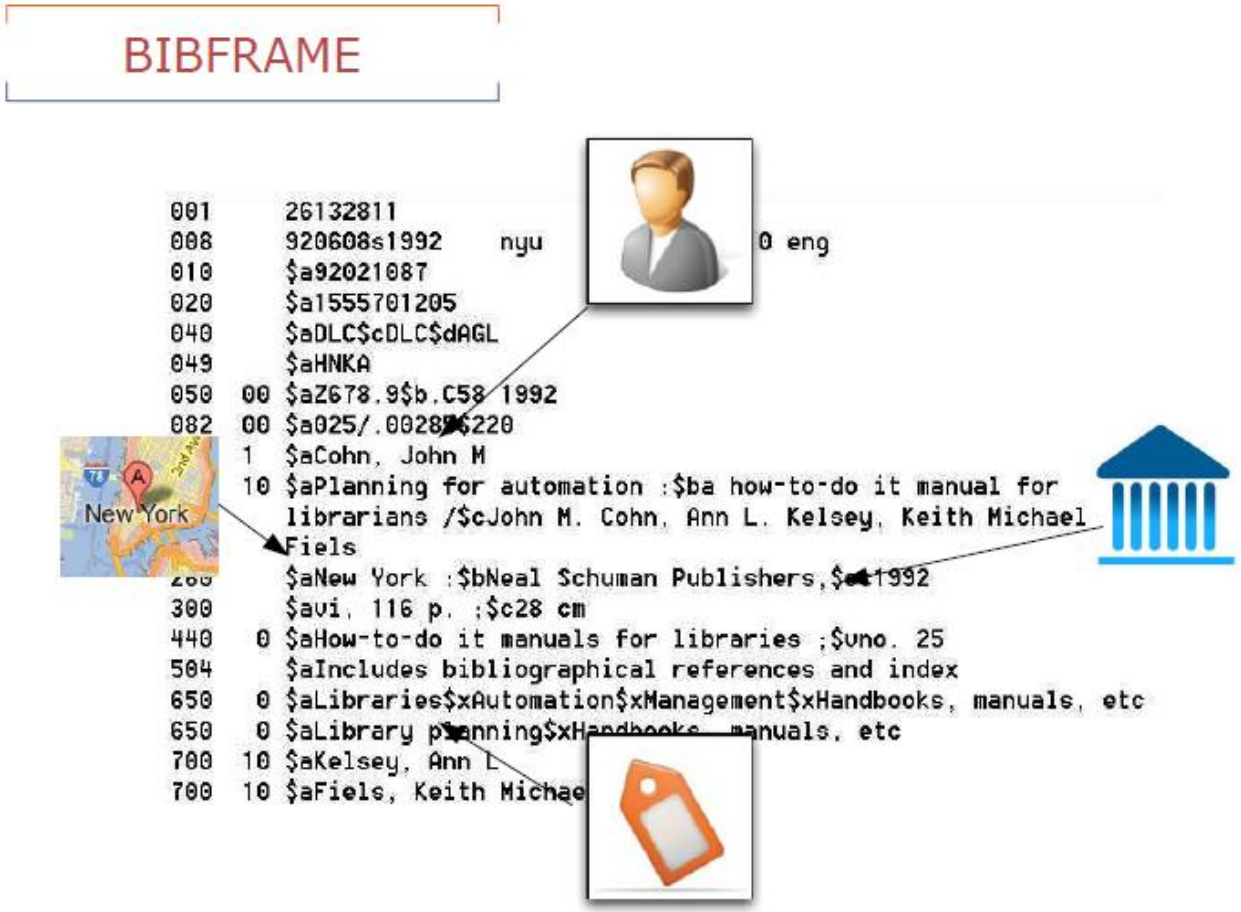
Instance Resource reflecting a material embodiment of a BIBFRAME work

Authority Resource reflecting key authority concepts that have defined relationships to works and instances

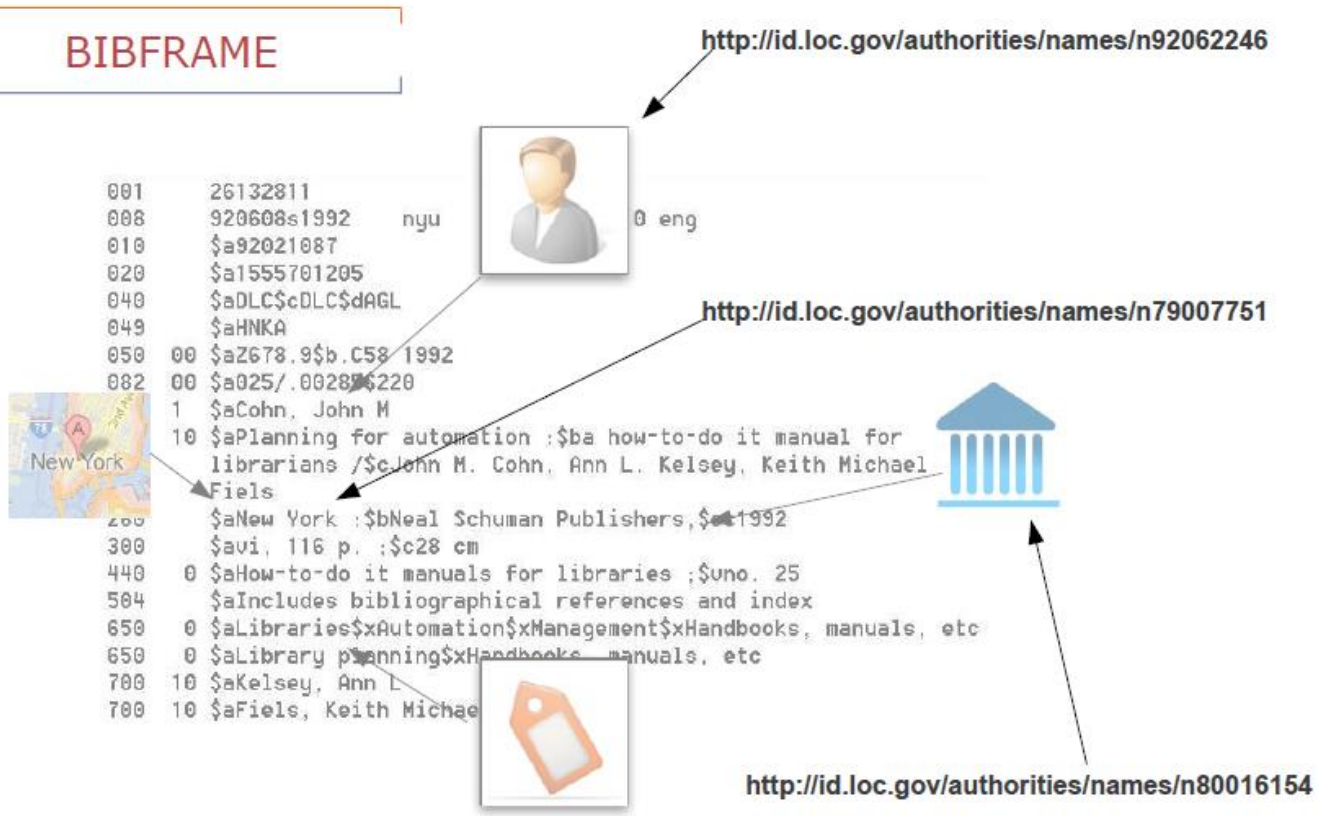



WEMI
BIBFRAME
WEMI Profile
BIBFRAME
Work
Represents
Work
isA
Work
Expression
Represents
Expression
isA
Manifestation
Represents
Manifestation
isA
Instance
isA
Item
Represents
Item

Remember our linked data discussions? BIBFRAME can make that happen. Instead of typing in the name, we can use a URI which displays as the current form.



BIBFRAME



Kevin Ford
NDMSO, Library of Congress
Email: kefo@loc.gov / Twitter: @3windmills

The BIBFRAME Vocabulary is comprised of the RDF properties, classes, and relationships between and among them.

1) Datatype and Object Property

Any given BIBFRAME property is either a datatype property or an object property.

A datatype property is one whose object is always a literal. An example is bf:version.

2)URIs and Labels

When referencing a resource, provide the URI, label, or both.

3)URIs and blank nodes BIBFRAME takes no position on the issue of URI vs. blank node.

While it is recognized that URIs are linked-data friendly and blank nodes are not, both are acceptable in BIBFRAME and the choice is an implementation decision.

4) Classes and Types

Classes are generally used to indicate type.

5) Reciprocal Properties

For any given BIBFRAME property, a reciprocal property should be defined, if appropriate.

http://www.loc.gov/bibframe/docs/bibframe2-rdf-conventions.html

The BIBFRAME Vocabulary is comprised of the RDF properties, classes, and relationships between and among them.

6) Metadata about the Description

Do not represent metadata about a description of a resource as a property of the resource itself.

7) Proliferation of Properties

Avoid proliferation of properties by defining a single general property when multiple potential properties have the same meaning.

8) rdfs: and rdf: Properties

Use rdf:value and rdfs:label as appropriate when supplying the value of a resource.

9) Formal constraints

Explicit domains and ranges for a property are generally not specified.

10) Naming Properties and Classes

Class names are nouns and property names suggest verbs.

http://www.loc.gov/bibframe/docs/bibframe2-rdf-conventions.html

List of classes (types) from Bibframe.org

Complete List of Classes

Select a class name below and jump to a panel describing that class and its usage below:

Agent	Event	MovingImage	Serial
Annotation	Family	Multimedia	StillImage
Archival	HeldItem	MultipartMonograph	Summary
Arrangement	HeldMaterial	NotatedMovement	TableOfContents
Audio	Identifier	NotatedMusic	Tactile
Authority	Instance	Organization	Temporal
Cartography	Integrating	Person	Text
Category	IntendedAudience	Place	ThreeDimensionalObject
Classification	Jurisdiction	Print	Title
Collection	Language	Provider	Topic
CoverArt	Manuscript	Related	Work
Dataset	Meeting	Relator	
DescriptionAdminInfo	MixedMaterial	Resource	
Electronic	Monograph	Review	

Each is further defined with a description and usage. http://bibframe.org/vocab-list/

Agent

bf:Agent

Entity having a role in a resource (Person, Organization, etc.).

Label: Agent

SubClass Of: bf:Authority

Full Class List | Top of Page

Annotation

bf:Annotation

Resource that asserts additional information about other BIBFRAME resource.

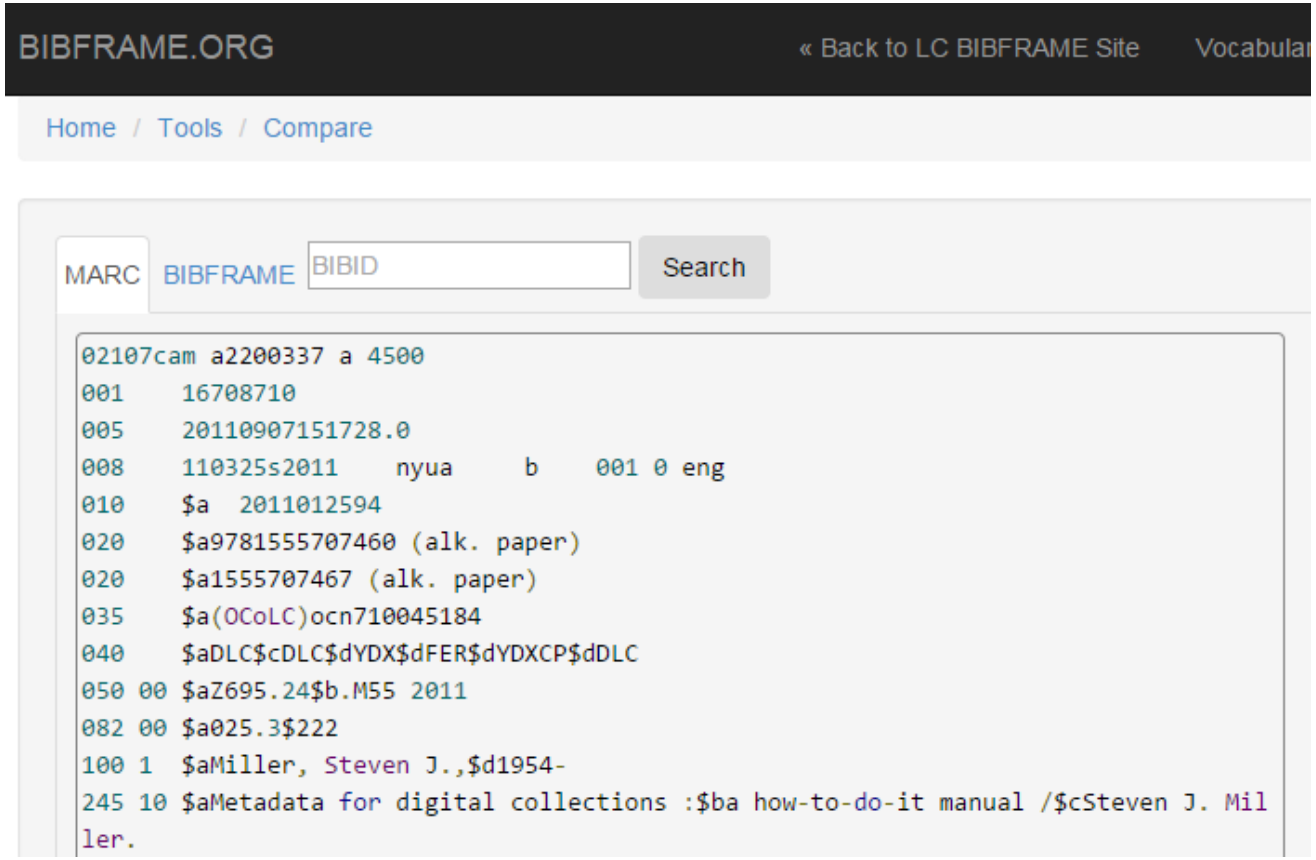
Label: Annotation

SubClass Of: bf:Resource

Full Class List | Top of Page

Similar to DublinCore / DCMI's structure on the web.

Examples – BibFrame in MARC



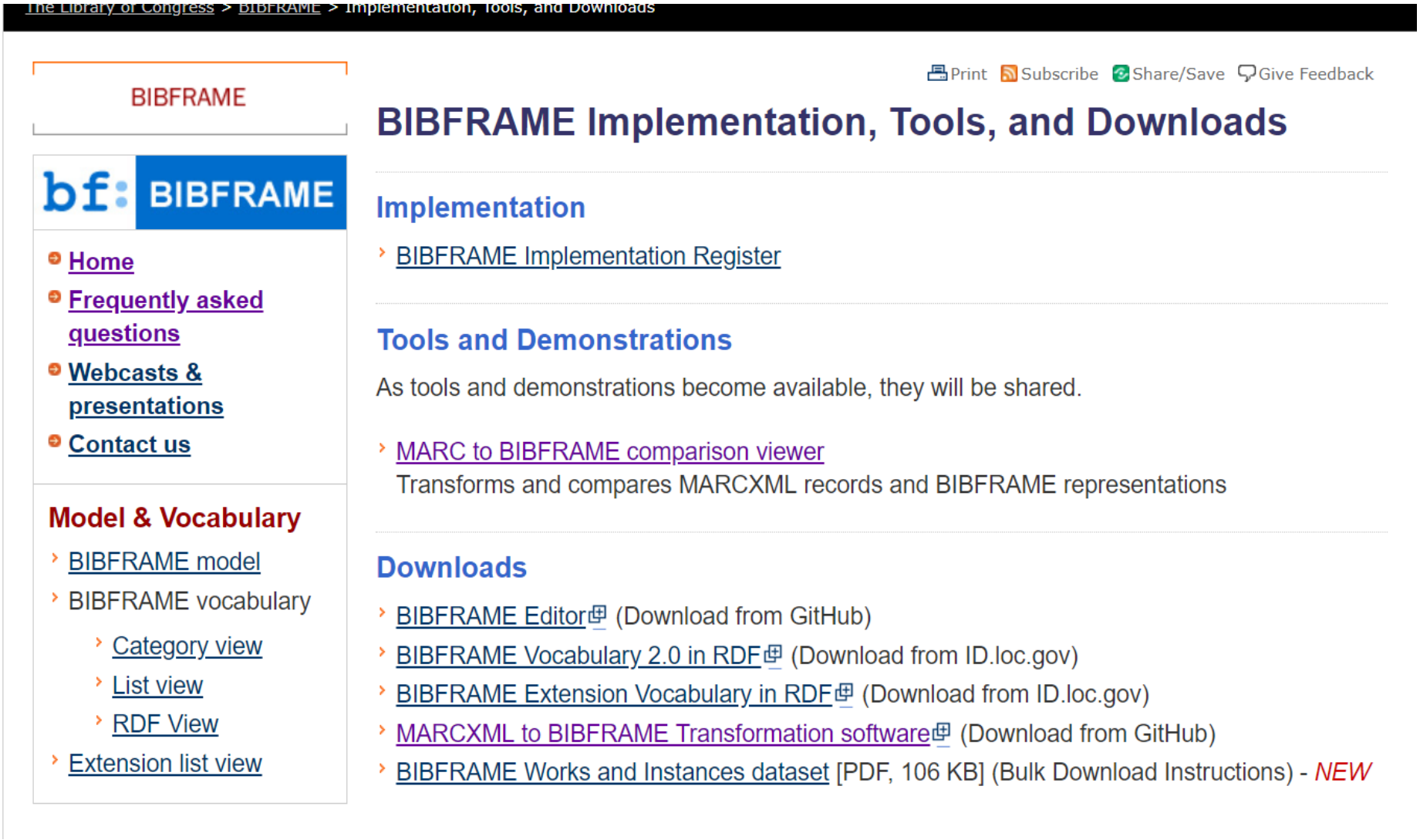
Tools you can use from LC & the BIBFRAME 2.0 Project

MARC 21 to BIBFRAME 2.0 Conversion Tools

- Conversion Specifications
  Library of Congress MARC to BIBFRAME 2.0 conversion specifications
- Conversion Programs
  XSLT conversion programs that apply the Library of Congress conversion specifications
- MARC to BIBFRAME comparison viewer
  For comparison of MARC bibliographic records, the BIB ID or the LCCN may be used
  For comparison of MARC authority records only the LCCN can be used

https://www.loc.gov/bibframe/

Tools you can use



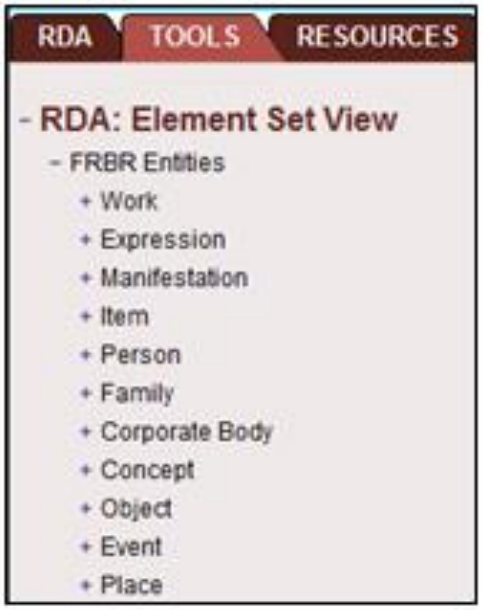
RIMMF for testing, training, and learning

Free software to create records for training

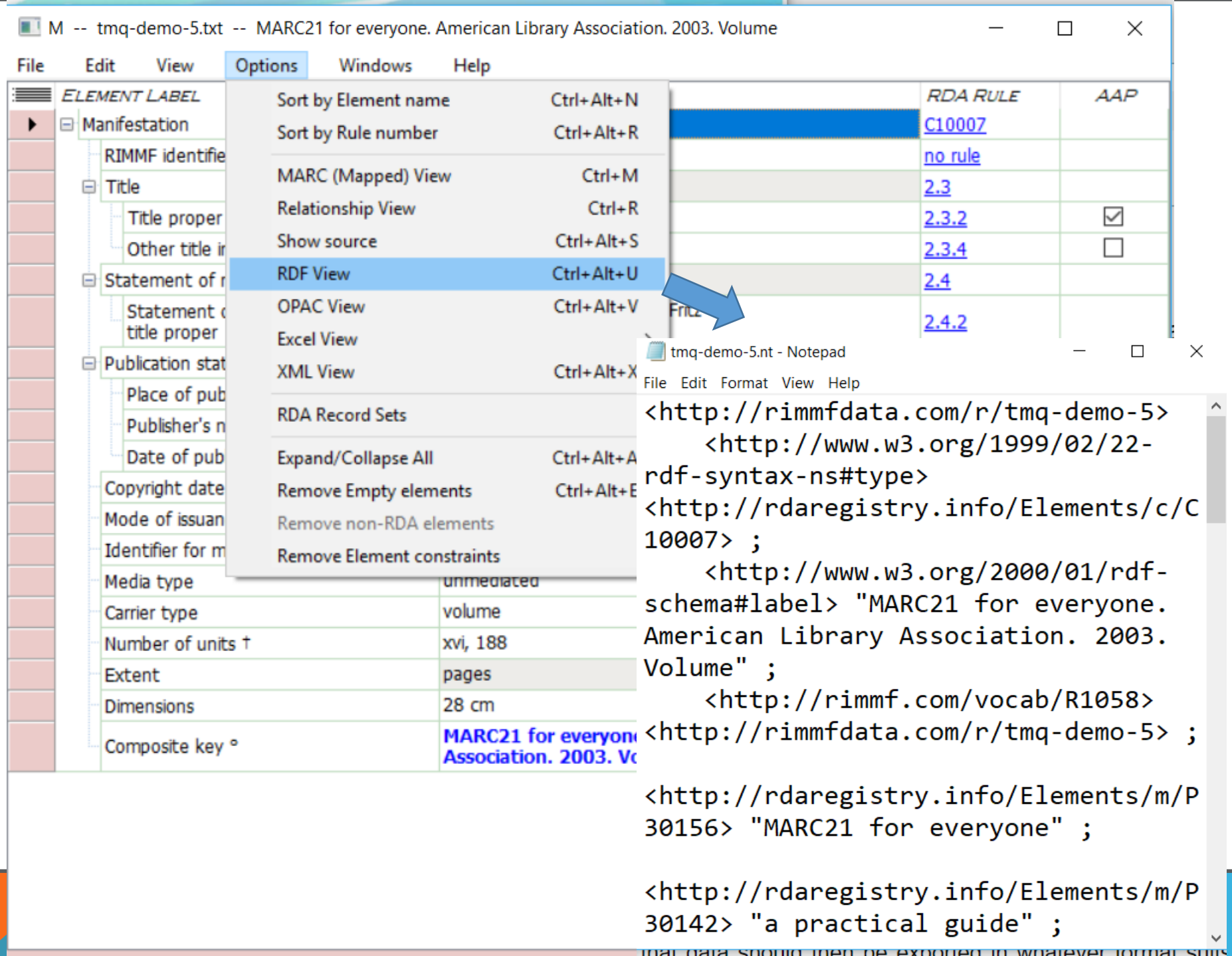
(RDA in Many Metadata formats) using a form

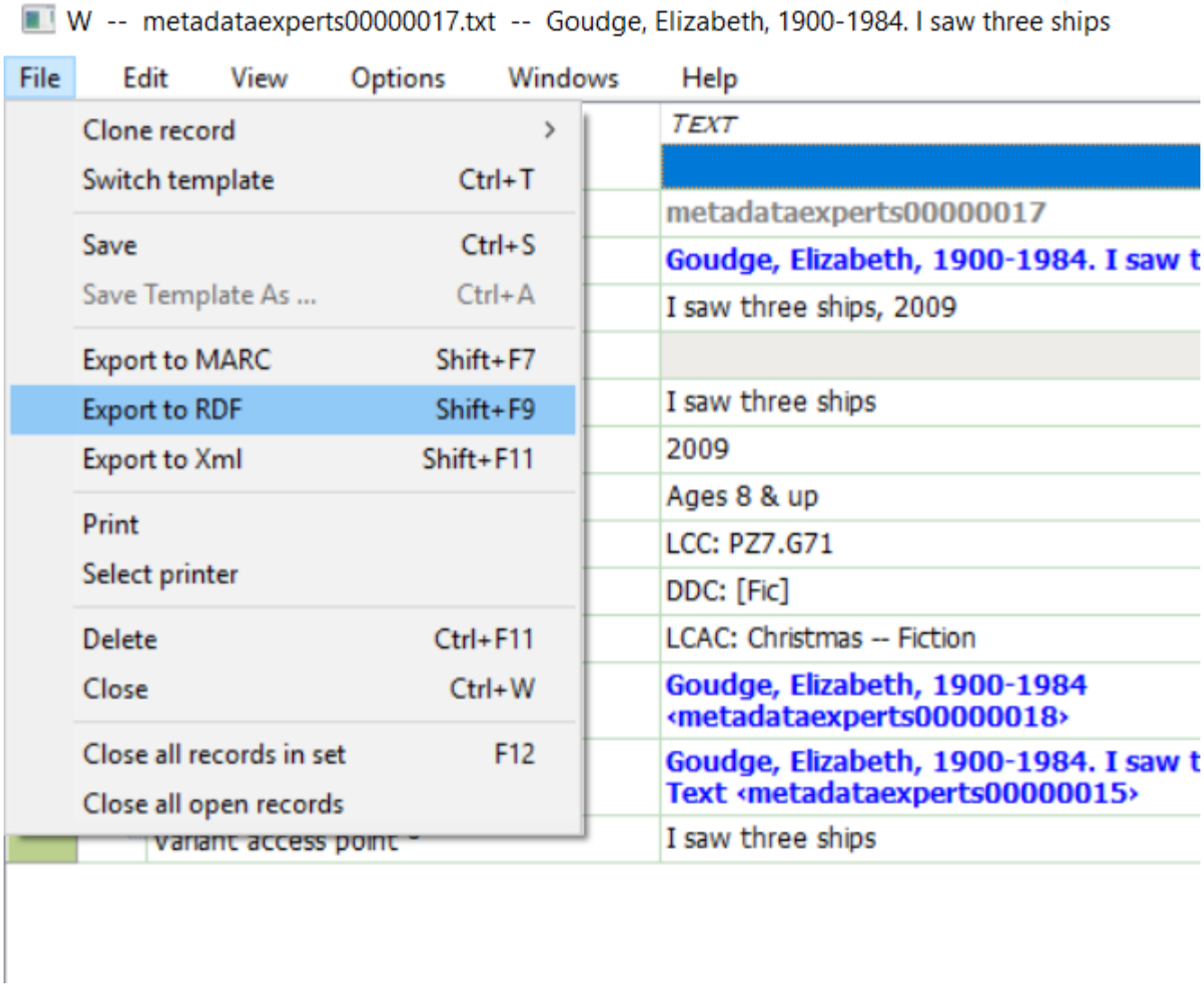
RDA to RDF, RDA to XML and even MARC !

http://www.marcofquality.com/wiki/rimmf3/

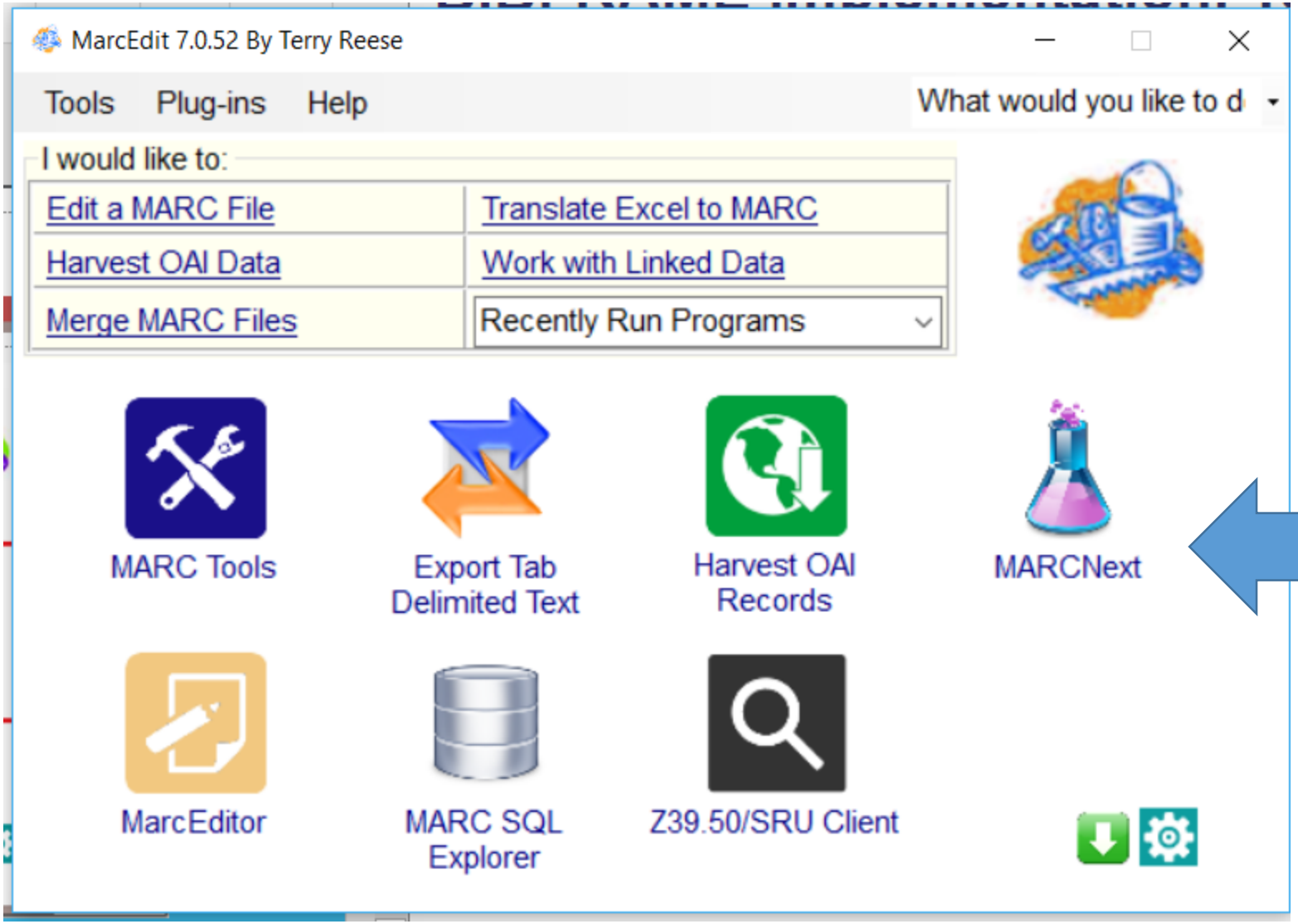


"Screen image from the RDA Toolkit (www.rdatoolkit.org) used by permission of the Co-Publishers for RDA (American Library Association, Canadian Library Association, and CILIP: Chartered Institute of Library and Information Professionals)"

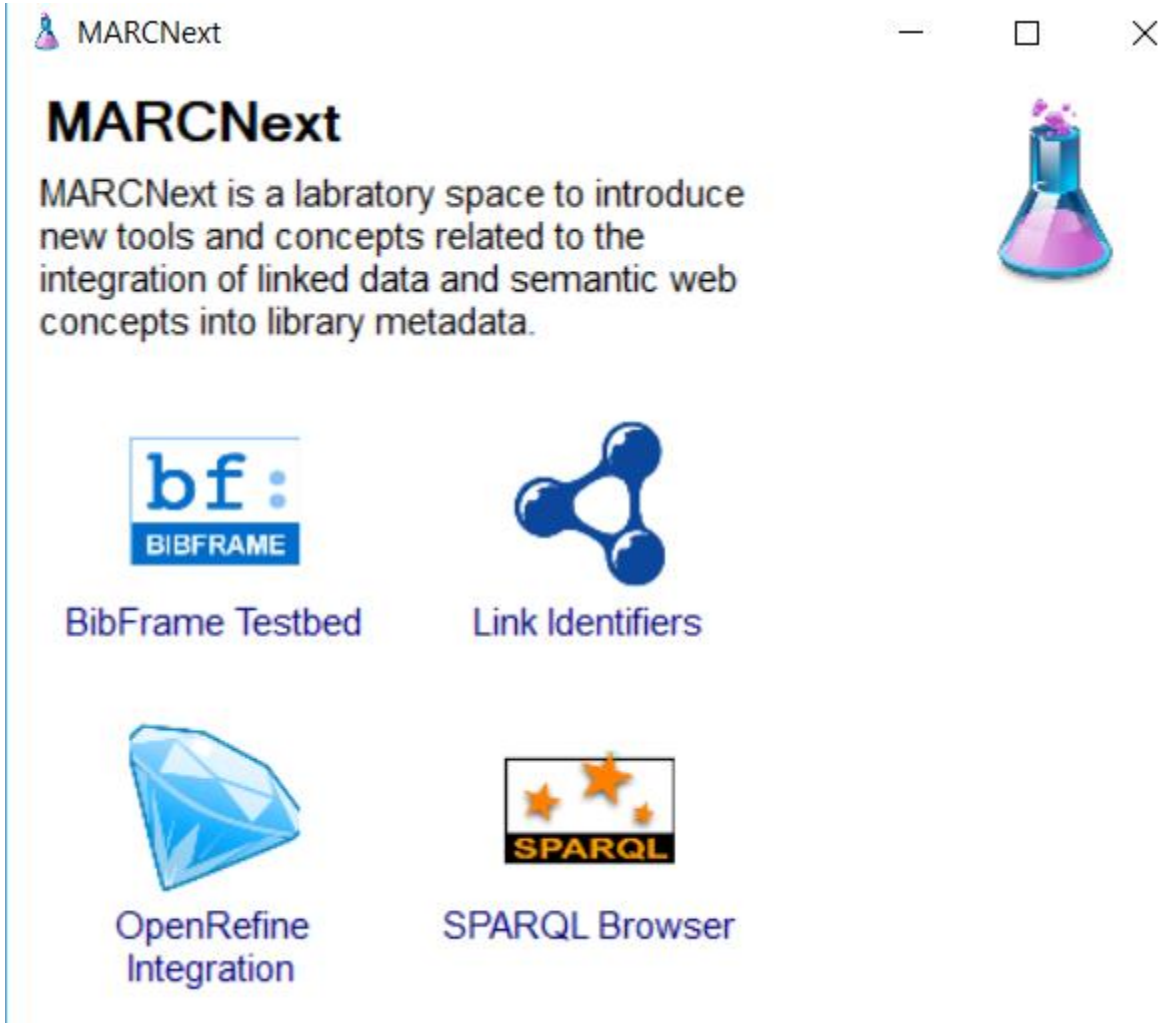

M -- tmq-demo-5.txt -- MARC21 for everyone. American Library Association. 2003. Volume
File Edit View Options Windows Help
Sort by Element name Ctrl+Alt+N
Sort by Rule number Ctrl+Alt+R
MARC (Mapped) View Ctrl+M
Relationship View Ctrl+R
Show source Ctrl+Alt+S
RDF View Ctrl+Alt+U
OPAC View Ctrl+Alt+V
Excel View
XML View Ctrl+Alt+X
RDA Record Sets
Expand/Collapse All Ctrl+Alt+A
Remove Empty elements Ctrl+Alt+E
Remove non-RDA elements
Remove Element constraints
ELEMENT LABEL
RDA RULE
AAP
Manifestation
Title
Statement of r
Publication stat
Copyright date
Mode of issuan
Identifier for m
Media type
Carrier type
Number of units †
Extent
Dimensions
Composite key °
C10007
no rule
2.3
2.3.2
2.3.4
2.4
2.4.2
unmediated
volume
xvi, 188
pages
28 cm
MARC21 for everyone
Association. 2003. Vo
tmq-demo-5.nt - Notepad
File Edit Format View Help
<http://rimmfdata.com/r/tmq-demo-5>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://rdaregistry.info/Elements/c/C10007> ;
<http://www.w3.org/2000/01/rdf-schema#label> "MARC21 for everyone. American Library Association. 2003. Volume" ;
<http://rimmf.com/vocab/R1058>
<http://rimmfdata.com/r/tmq-demo-5> ;
<http://rdaregistry.info/Elements/m/P30156> "MARC21 for everyone" ;
<http://rdaregistry.info/Elements/m/P30142> "a practical guide" ;

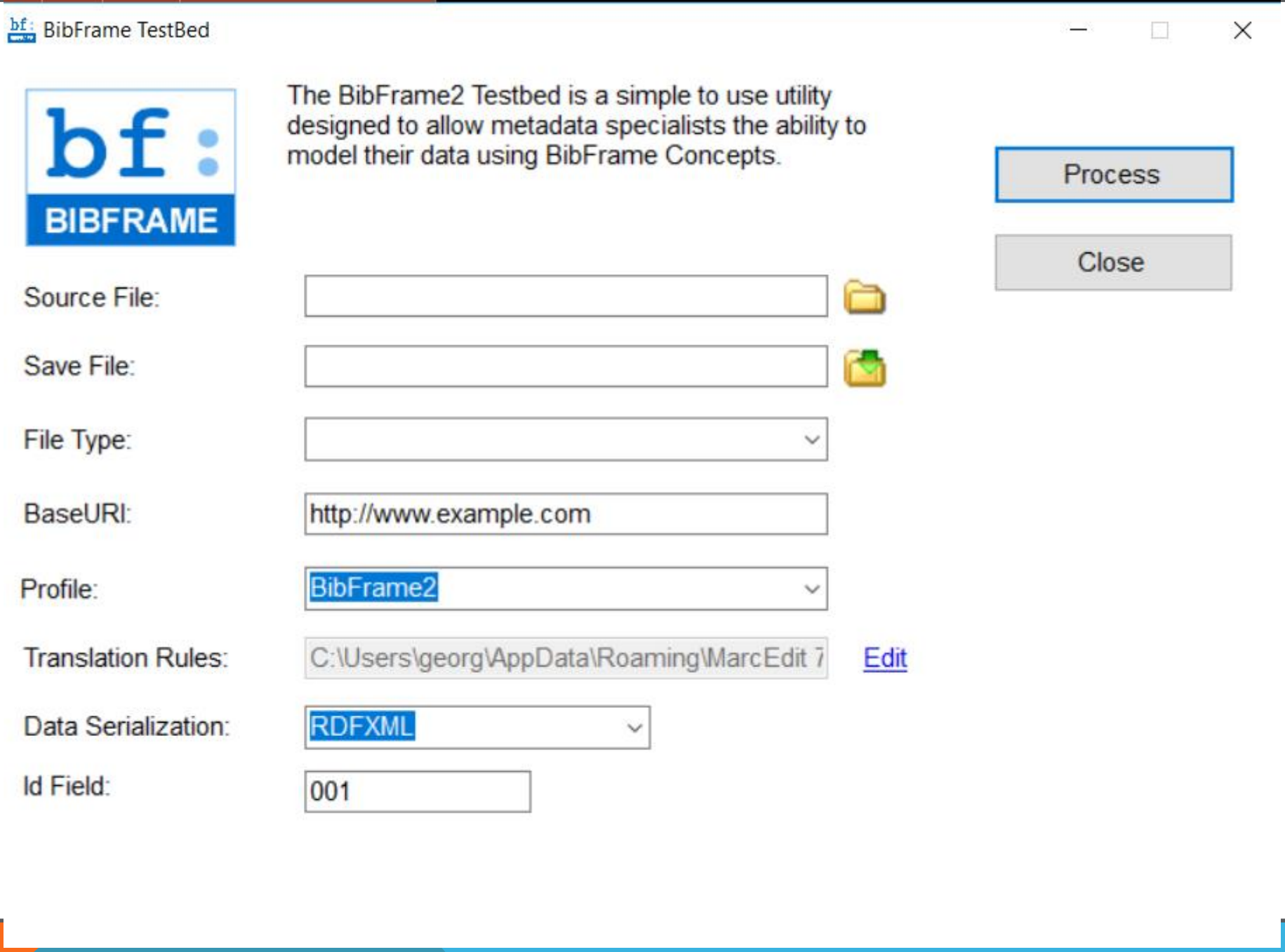

W -- metadataexperts00000017.txt -- Goudge, Elizabeth, 1900-1984. I saw three ships
File
Edit
View
Options
Windows
Help
Clone record
Switch template Ctrl+T
Save Ctrl+S
Save Template As ... Ctrl+A
Export to MARC Shift+F7
Export to RDF Shift+F9
Export to Xml Shift+F11
Print
Select printer
Delete Ctrl+F11
Close Ctrl+W
Close all records in set F12
Close all open records
TEXT
metadataexperts00000017
Goudge, Elizabeth, 1900-1984. I saw t
I saw three ships, 2009
I saw three ships
2009
Ages 8 & up
LCC: PZ7.G71
DDC: [Fic]
LCAC: Christmas -- Fiction
Goudge, Elizabeth, 1900-1984
‹metadataexperts00000018›
Goudge, Elizabeth, 1900-1984. I saw t
Text ‹metadataexperts00000015›
I saw three ships

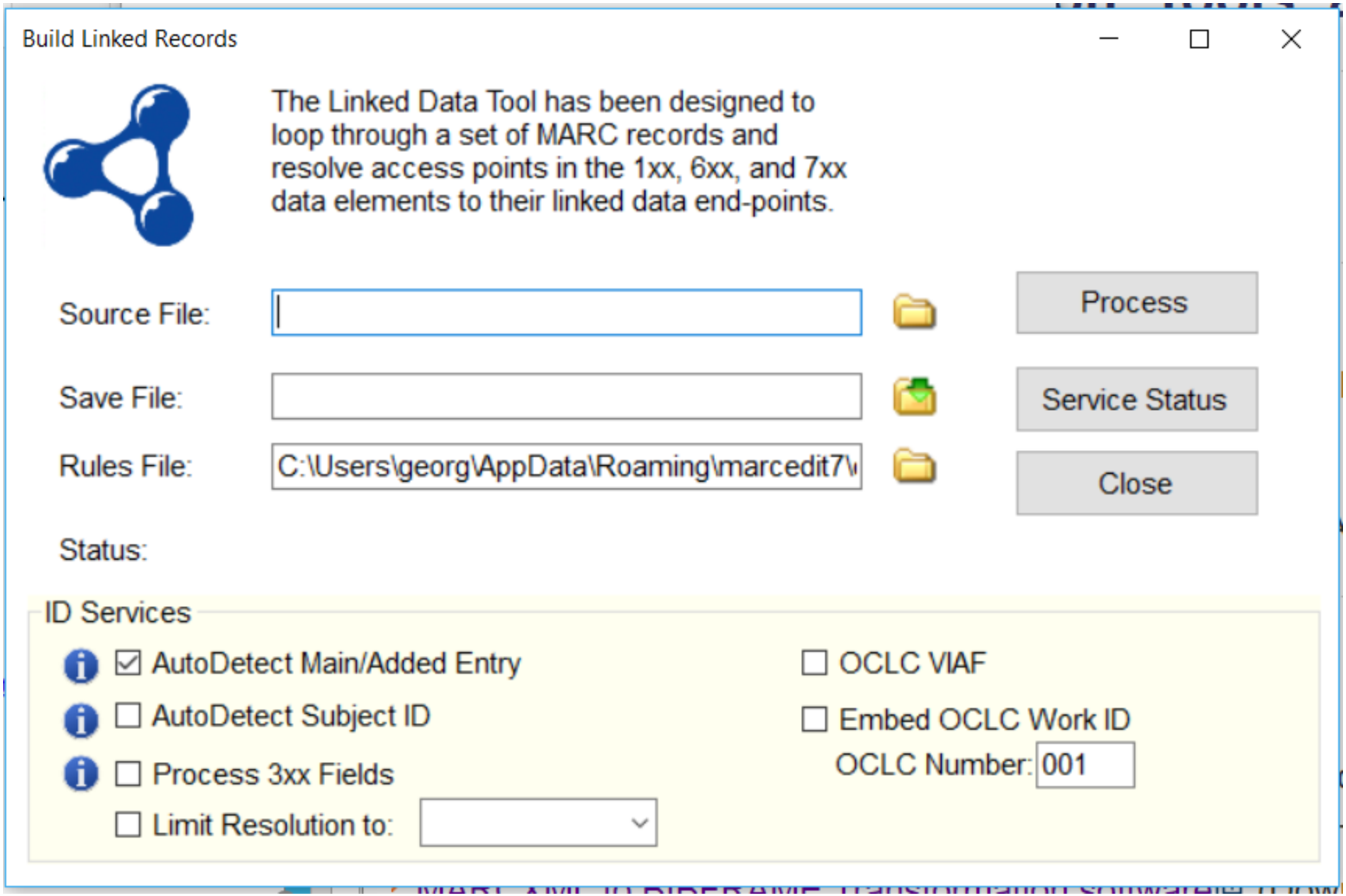
Tools you can use : MARCEDIT



MARCEDIT/MARCNEXT




BibFrame TestBed
bf:
BIBFRAME
The BibFrame2 Testbed is a simple to use utility designed to allow metadata specialists the ability to model their data using BibFrame Concepts.
Process
Close
Source File:
Save File:
File Type:
BaseURI:
http://www.example.com
Profile:
BibFrame2
Translation Rules:
C:\Users\georg\AppData\Roaming\MarcEdit 7
Edit
Data Serialization:
RDFXML
Id Field:
001


Build Linked Records
The Linked Data Tool has been designed to loop through a set of MARC records and resolve access points in the 1xx, 6xx, and 7xx data elements to their linked data end-points.
Source File:
Save File:
Rules File:
C:\Users\georg\AppData\Roaming\marcedit7\
Process
Service Status
Close
Status:
ID Services
AutoDetect Main/Added Entry
AutoDetect Subject ID
Process 3xx Fields
Limit Resolution to:
OCLC VIAF
Embed OCLC Work ID
OCLC Number:
001

Results

- Publications
- Ontologies
- Technologies and Models
- Old technologies
- Services
- Methodologies
- Linked data
- Material used in papers
- Benchmarks

MARiMbA

MARiMbA is a command-line tool, designed with librarians in mind, to transform MARC (MAchine-Readable Cataloging) records to RDF, following Linked Data best practices [1][2][3].

The tool supports the whole mapping and transformation process from MARC metadata to RDFS/OWL vocabularies. It is a tool aimed at facilitating the Linked Data generation process and at allowing librarians to carry out the RDF generation without any technical support. In order to achieve this, MARiMbA has the following features:

- The tool works with MARC authority and bibliographic formats.
- All work is done using spreadsheets. There is no need to learn any additional mapping or transformation language (e.g. XSLT).
- The tool analyses MARC input records in order to generate easy-to-use mapping templates. These templates are focused on facilitating the decision-making task, errors discovery and the evaluation of the whole transformation process.
- It allows the user to use any vocabulary formalized as RDFS/OWL.
- It includes a minimal configuration file that allows the user to adjust some features of the process . However, the tool is preconfigured to be used out of the box, following the FRBR model (Functional Requirements for Bibliographic Records).
- It includes a lightweight SPARQL server (Fuseki) that allows the user to perform queries against the generated data with no extra configuration or data loading.

MARiMbA has been successfully used to transform around 7 million MARC 21 records from the Spanish National Library, which produced around 60 million RDF triples. The resulting data are available via SPARQL at http://datos.bne.es/sparql. Additionally, an RDF resource example can be found at http://datos.bne.es/resource/XX1718747.

How to use it?

MARCXML

- MARCEditor
- MARCXML Toolkit
https://www.loc.gov/standards/marcxml/marcxml.zip
- MARC Conversion Stylesheets
https://www.loc.gov/standards/marcxml/
- MARC / Java project http://projects.freelibrary.info/freelib-marc4j/tutorial.html
- MARC Tools
https://www.loc.gov/marc/marctools.html#toolslist
- Conversion Tools for Libraries
- https://sourceforge.net/projects/dnb-conv-tools/
- XML2MARC (command line tool)
https://metacpan.org/pod/distribution/MARC-XML/bin/xml2marc

RDF Tools

- **Resource Guide**

  https://docs.google.com/document/d/19k1uhopm3PiFxShqDXEShQd1wgGHDTmyOgj7aplWNhM/edit?usp=sharing

  Continues Semantic Web Glossary for Libraries and Quizzes: RDF and XML.

thank you!