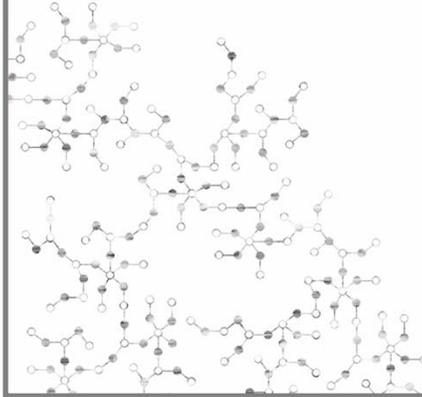


Semantic Web: Intersection With Libraries

F. Tim Knight
Osgoode Hall Law School Library
York University

A Conversation on the Semantic Web and Legal Information

Annual Meeting Pre-Conference Workshop
Canadian Association of Law Libraries/
L'Association canadienne des bibliothèques de droit
May 24th, 2014 in Winnipeg, Manitoba



Now I'd like to take a look at the connection between the Semantic Web and some recent developments in the library profession.

One of the biggest developments in the library world has been the introduction of the new cataloguing rules, Resource Description and Access (RDA).

What is RDA?

- Successor to the **Anglo-American Cataloguing Rules** (AACR2)
- Aligned with the **Statement of International Cataloguing Principles** (2009)

“Several principles direct the construction of cataloguing codes. The highest is the convenience of the user.”

http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf

2

You may have heard of RDA, but if you're not a cataloguer you may not have had any firsthand experience with it yet. I won't go into a lot of detail this morning, but I thought it might be useful to provide a little background.

RDA is the successor to AACR2. It is also aligned with the Statement of International Cataloguing Principles a document that “replaces and broadens the scope” of the original Paris Principles released in 1961. This is important because the primary purpose of those principles is to create cataloguing guidelines that will serve the convenience of catalogue users.

What is RDA?

- A 'content standard' that describes resources
- Based on these conceptual models
 - **Functional Requirements for Bibliographic Records (FRBR)**
 - **Functional Requirements for Authority Data (FRAD)**

3

One of the fundamental differences between AACR2 and RDA, is that RDA is a “content standard,” which means the guidelines focus on describing the intellectual content of the resource and minimize the need to provide descriptions based on particular formats. AACR2 descriptions emphasized multiple formats which had begun to complicate the cataloguing process.

RDA has also incorporated the conceptual models FRBR (the Functional Requirements for Bibliographic Records) and FRAD (the Functional Requirements for Authority Data). For the purposes of this morning's introduction I will only talk a bit about FRBR.

How many of you are familiar with FRBR?

Introduction to FRBR

- In 1998 **International Federation of Library Associations** (IFLA) released Final report
- In 2007 **RDA** adopted this conceptual model and reorganized draft text

4

FRBR was developed by a study group of the International Federation of Library Associations and published in 1998. So it's been influencing the library community for over a decade and a half now. It was developed by analyzing existing bibliographic records and authority data, defining the entities, listing their attributes, and identifying relationships between the entities to develop what is called an "entity-relationship model."

In 2007, two years after the first RDA draft had been presented, the new JSC, the Joint Steering Committee for the Development of RDA, decided to use the FRBR model to address the structural problems that had been identified in AACR2 and reorganized RDA using the FRBR principles.

Introduction to FRBR

The aim of FRBR was to,

*"... produce a framework that would provide a clear, precisely stated, and commonly shared understanding of what it is that the bibliographic record aims to provide information about, and what it is that we expect the record to achieve in terms of **answering user needs.**"*

Functional Requirements for Bibliographic Records: Final Report

5

FRBR is important because it has provided the cataloguing community with a consistent language that allows us to frame discussions around what bibliographic data is and how it is used.

The aim of FRBR was to,

"... produce a framework that would provide a clear, precisely stated, and commonly shared understanding of what it is that the bibliographic record aims to provide information about, and what it is that we expect the record to achieve in terms of answering user needs."

One of the key aspects of FRBR, that I've emphasized at the end of this quote from the Final Report, and which echoes back to the Statement of International Cataloguing Principles, is the importance for bibliographic information to answer "user needs."

To Support Resource Discovery

- RDA is a set of guidelines and instructions on **formulating data to support resource discovery**. -- RDA 0.0
- *"The purpose of recording data is to support the user tasks."*-- Chris Oliver, *Introducing RDA*

6

And user needs are an important part of the new cataloguing rules as seen in the opening remarks of RDA:

RDA aims to provide, "a set of guidelines and instructions on formulating data to support resource discovery."

And Chris Oliver also reminded us in her recent book, 'Introducing RDA,' that the "purpose of recording data is to support the user tasks."

So, what are the user tasks? What do users need to do to find resources?

The FRBR User Tasks

- **Find** – to find entities corresponding to the users search criteria
- **Identify** – to confirm that the entity sought is the same as the entity described
- **Select** – to select an entity appropriate to the needs of the user
- **Obtain** – to acquire or gain access to the entity

7

There are the four user tasks identified in FRBR:

- **Find** – to find entities corresponding to the users search criteria
- **Identify** – to confirm that the entity sought is the same as the entity described
- **Select** – to select an entity appropriate to the needs of the user
- **Obtain** – to acquire or gain access to the entity

The Entity-Relationship Model

- The three components of an entity-relationship data model
 - **Entities** – the things that users are interested in: resources; people; subjects
 - **Attributes** – the characteristics of these things
 - **Relationships** – relationships between these things

8

A few slides back I mentioned that FRBR is an “entity-relationship model.” There are three parts to an entity-relationship model:

-the **entities** themselves, the things that people are looking for, the products of intellectual or artistic creation, the short story, the research paper, the film. Entities can also be the people, or groups of people, responsible for the creation of these things; and entities can be the subjects or concepts about those things; so, resources, people and subjects.

-the **attributes** are the characteristics we use to describe those resources, people or subjects

-and finally there are the **relationships** that identify the connections that relate these things to one another, to other things, people and subjects.

The FRBR Entities

- **Group 1** – entities that are the results of intellectual or artistic effort:
 - **Work; Expression; Manifestation; Item**
- **Group 2** – entities responsible for intellectual or artistic work:
 - **Person; Corporate Body; Family**
- **Group 3** – entities that are subjects of intellectual or artistic work:
 - **Concept; Object; Event; Place**

9

There are three groups of entities defined in the FRBR model:

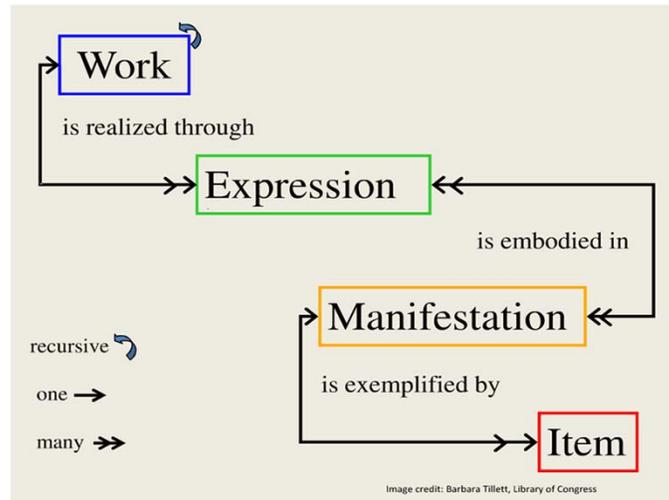
The Group 1 entities are the resources themselves, the results of intellectual or artistic effort. These are described in FRBR as the **Work**, the **Expression**, the **Manifestation** and the **Item**. These have been collectively referred to as **WEMI**.

The Group 2 entities represent those responsible for the intellectual or artistic effort. These entities can be a **Person**, a **Corporate Body**, or a **Family**.

And finally Group 3 entities are the subjects of intellectual or artistic effort: a **Concept**; an **Object**; an **Event**; or a **Place**. And any Group 1 and Group 2 entity can be subjects handled by the Group 3 entities.

This last group, the Group 3 entities have so far not been fully developed in RDA, but placeholder chapters have been included for future development of this area.

FRBR Group 1 Entities



This is the diagram that Barbara Tillett from the Library of Congress developed to illustrate the FRBR Group 1 entities. The **Work** is the most abstract concept and is realized through an **Expression**. The **Expression** is then embodied in a **Manifestation**. It's the **Manifestation** that cataloguers typically deal with, e.g. a book, electronic book, journal, etc. The **Manifestation** is then exemplified by the **Item**. The **Item** is where local information about the particular copy held by the library is recorded, e.g. a damaged copy, where the copy is located, whether it can be used outside of the library, etc.

These entities are collectively this is known as **WEMI**, i.e. Work, Expression, Manifestation and Item. Keep this in mind when we look at BIBFRAME, which we will do shortly.

RDA and FRBR

"The RDA and FRBR efforts have been one of the key contributions in re-focusing cataloging efforts from 'strings to things' and in providing a set of base line functional requirements for supporting the future of cataloging. The holistic approach to retrieval and access as defined by the FRBR work has been a guiding principal to the model proposed in this document."

Library of Congress. 2012. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services , p. 37
<<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>

OK, so we've got a new set of cataloguing rules or guidelines, which is good. But really, the most important thing about the transition to RDA has been what Sarah talked about earlier: RDA has started us down a road that will lead us away from a record-centric view of the bibliographic universe to a view that focuses on our data.

In their report on bibliographic data modeling the BIBFRAME folks characterized it this way:

"The RDA and FRBR efforts have been one of the key contributions in re-focusing cataloging efforts from 'strings to things' and in providing a set of base line functional requirements for supporting the future of cataloging. The holistic approach to retrieval and access as defined by the FRBR work has been a guiding principal to the model proposed in [BIBFRAME model]."

We'll talk later about this idea of moving from "strings to things" later on this morning.

However, although we've set off down the data-centric road we need to realize that because we are still so heavily invested in the MARC format we are somewhat stuck in a record-centric view of our bibliographic world.

Our Friend MARC

LC Control No.: 78005880
LCN Permalink: <http://lcn.loc.gov/78005880>
000 00914pam a2200277 i 450
001 3681622
005 19781228000000.0
008 780322s1978 nyu b 001 0 eng
035 __ |9 (DLC) 78005880
906 __ |a 7 |b cbc |c orignew |d 1 |e ocip |f 19 |g y-gencatlg
010 __ |a 78005880
020 __ |a [0698307011](#)
040 __ |a DLC |c DLC |d DLC
043 __ |a n-us--
050 00 |a KF9725 |b .S74
082 00 |a 345/ 73/077
100 1_ |a Stevens, Leonard A.
245 10 |a Death penalty : |b the case of life vs. death in the United States / |c by Leonard A. Stevens ; foreword by Michael Meltser.
260 __ |a New York : |b Coward, McCann & Geoghegan, |c c1978.
300 __ |a 159 p. : |c 24 cm.
490 0_ |a Great constitutional issues
504 __ |a Bibliography: p. 154-155.
500 __ |a Includes index.
650 0_ |a Capital punishment |z United States.
600 10 |a Furman, William Henry.
991 __ |b c-LL |h KF9725 |i .S74 |t Copy 1 |w BOOKS

CALL NUMBER: [KF9725 .S74](#)
Copy 1

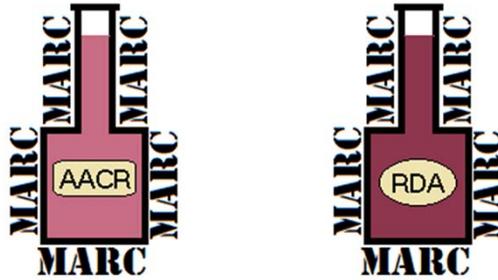
The MARC format has certainly been our friend. And, it is a friend that we know well and a friend that our library systems still rely on. On the screen here is one view of a MARC record, something you might encounter when you click the MARC or Staff view in the catalogue. When cataloguers work in the library system they are presented with a similar view which they use to create new records and to modify existing ones.

Our Friend MARC

```
00914pam a2200277 i
450000100080000000500170000800800410002503500210006690600450008
701000170013202000150014904000180016404300120018205000170019408
200160021110000240022724501260025126000520037730000210042949000
320045050400300048250000200051265000390053260000270057199100380
0598368162219781228000000.0780322s1978 nyu b 001 0
eng 9(DLC) 78005880 a7bcbccorignewdleocipf19gy-gencatlg
a 78005880 a0698307011 aDLCcDLCdDLC an-
us--00aKF9725b.S7400a345/.73/0771 aStevens, Leonard A.10aDeath
penalty :bthe case of life vs. death in the United States /cby
Leonard A. Stevens ; foreword by Michael Meltsner. aNew York
:bCoward, McCann & Geoghegan,cc1978. a159 p. ;c24 cm.0 aGreat
constitutional issues aBibliography: p. 154-155. aIncludes
index. 0aCapital punishmentzUnited States.10aFurman, William
Henry. bc-LLhKF9725i.S74tCopy lwBOOKS
```

However, this is really what a MARC record looks like. In fact this is displayed using word wrap, it's really one long string of letters and numbers. This is what exists behind the scenes and is something much more complex than what we usually encounter in our day to day library activities. This is what developers work with and this is what makes it difficult, especially for non-library developers, to create web-based applications that interact with our bibliographic data.

New Wine in an Old Bottle



We are in a transition period at the moment. Although RDA has set a course away from a record-centric data model the potential of RDA cannot be realized in the current record-centric approach that MARC provides us. What we've created is a new RDA wine, but it's poured into the same old MARC bottle.

This is where the Bibliographic Framework Initiative comes in.

The BIBFRAME Initiative

“[BIBFRAME] is the foundation for the future of bibliographic description that happens on, in, and as part of the web and the networked world we live in. It is designed to integrate with and engage in the wider information community while also serving the very specific needs of ... libraries and similar memory organizations.”

Library of Congress. 2012. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services <<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>

The Library of Congress, again taking the lead, launched the Bibliographic Framework Initiative about three years ago, in May 2011. This was just as institutions began to implement RDA in their libraries around the world.

The new BIBFRAME model intends to replace the MARC format and help us to transition away from the current record-centric view. But more than that, it will also provide the *“foundation for the future of bibliographic description that happens on, in, and as part of the web and the networked world we live in. It is designed to integrate with and engage in the wider information community while also serving the very specific needs of ... libraries and similar memory organizations.”*

This makes me think of Tim Berners-Lee’s evolution that I introduced earlier.

"It isn't the records, it is the networks that connect them which are interesting."

It's a similar shift in thinking.

It's not the records that are important, it's the network that's interesting, it's the connections.

The BIBFRAME Initiative

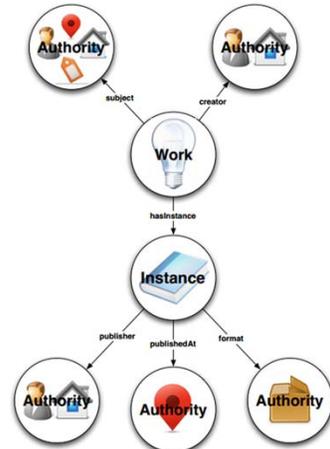
“The BIBFRAME model is intentionally designed to coordinate the cataloging and metadata that libraries create with these efforts, and connect with them. In short, the BIBFRAME model is the library community’s formal entry point for becoming part of a much larger web of data.”

Library of Congress. 2012. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services <<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>

As we all know, people begin their search for information with a search engine, or they discover something through one of the social networks they belong to. Libraries were once a large part of this discovery process assisting library users find information and the answers to their questions. In the current environment much of the information that we can provide is locked up in one of our record-centric information systems.

“The BIBFRAME model is intentionally designed to [take advantage of the current network and] coordinate the cataloging and metadata that libraries create with the [other services available on the Web], and connect with them. In short, the BIBFRAME model is the library community’s formal entry point for becoming part of a much larger web of data.”

BIBFRAME Approach



- Work
- Instance
- Authority
- Annotation

Figure 1: A graphical representation of the BIBFRAME Linked Data model defining the relation between Work and Instance resources and their contextualization to Web addressable Authority resources.

Library of Congress. 2012. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services <<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>

In some respects the proposed BIBFRAME model is similar to the FRBR Group 1 entities, which, as you recall are Work, Expression, Manifestation and Item.

The BIBFRAME model reimagines this framework as: Work; Instance; Authority; and Annotation.

Creative Work



The **Work**, or what BIBFRAME calls, the **Creative Work** is the “conceptual essence” of the resource to be described, catalogued and ultimately discovered. This is very close to how FRBR view of the **Work** as an “abstract notion” that is not realized until the first **Expression** is created.

It is the **Work** that allows us to conceptually connect what might be a textual print version, with a visual film version, with a performed theatre version. It enables us to identify that this is Shakespeare’s “Hamlet” or this is Stravinsky’s “Rite of Spring.”

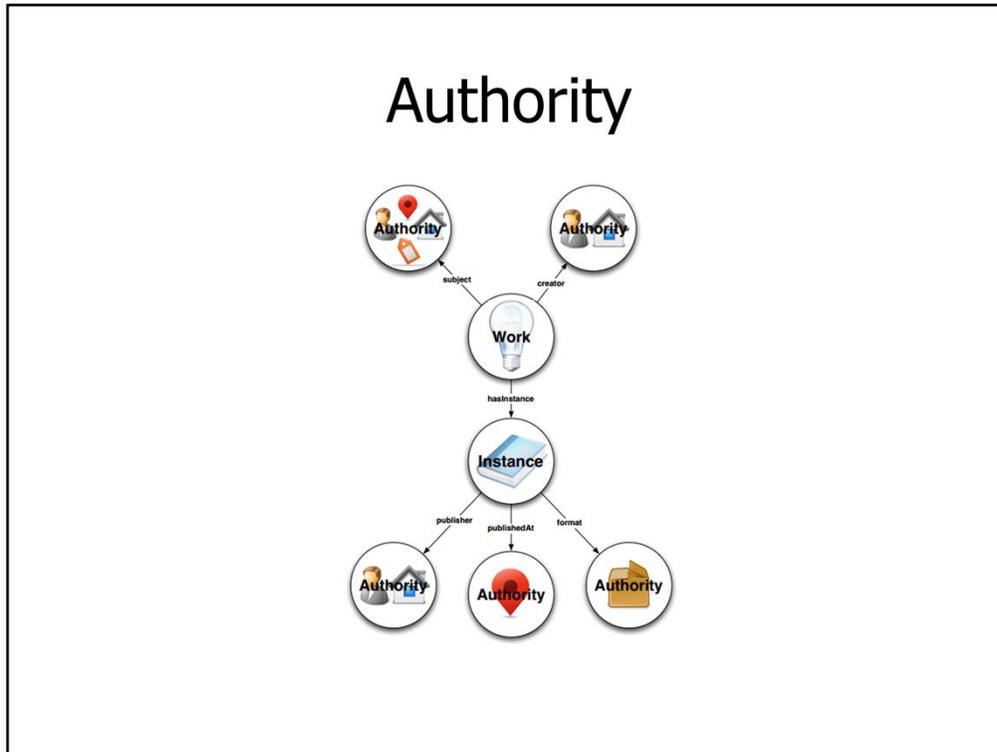
In BIBFRAME speak the **Work** is “a *Web based control point that reflects both commonality of content between and among the various Instances associated with the Work as well as a reference point for other Works.*”

Instance



In the BIBFRAME model it is the **Instance** that embodies the **Creative Work**. Compared to the FRBR Group 1 entities this essentially collapses the **Expression** and **Manifestation** entities into this single **Instance** entity. An **Instance** is the resource that a cataloguer would be describing, so for example, the electronic book copy of “Hamlet” or the vinyl recording of the “Rite of Spring.”

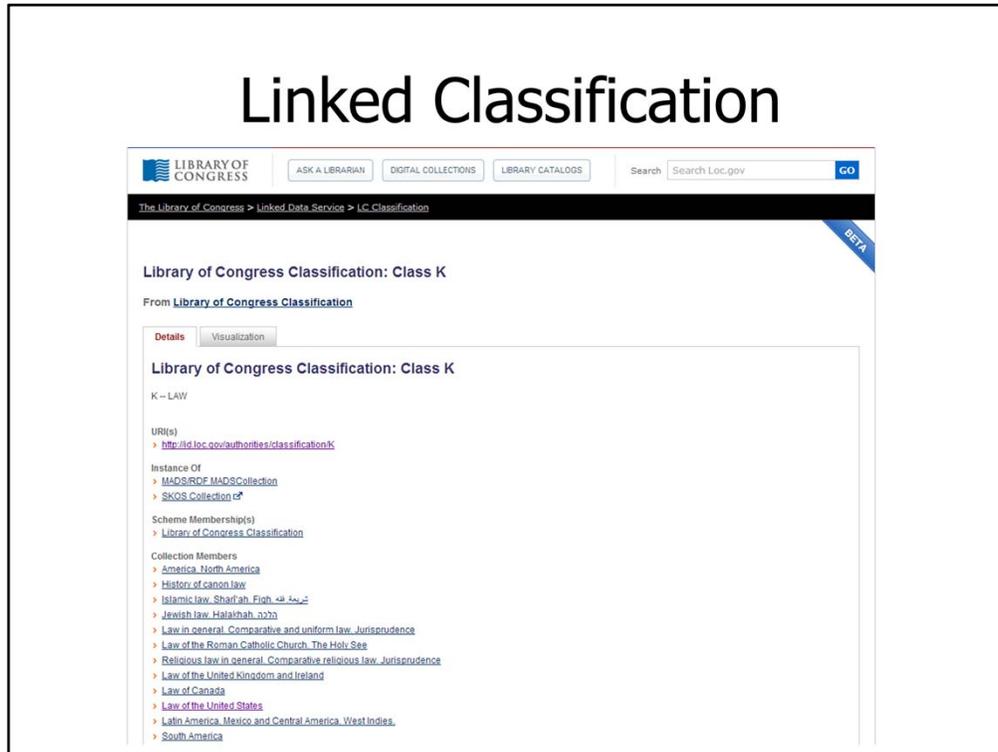
BIBFRAME defines it like this: “[An Instance is] a *Web based control point that includes properties specific to the materialization as well as contextual relationships to appropriate BIBFRAME Authorities related to the publication, production, distribution of the material resource.*”



The BIBFRAME **Authority** is similar to what cataloguer’s usually think of as authority control and is what is defined as an access point for the resource. Authorities in cataloguing also help disambiguate similar terms and entities so that you know which Jane Smith you are looking at compared to the Jane Smith you might be looking for. Authorities also provide additional information like scope notes for subject headings and cross-references between narrower, broader and related terms providing users with the additional information they need to contextualize what they have found and assist with their resource choices.

The BIBFRAME **Authority** provides “*key concepts that are the target of defined relationships reflected in the Work and Instance.*” Examples of what they call **Authority Resources** include things like People, Places, Topics, Organizations, etc.

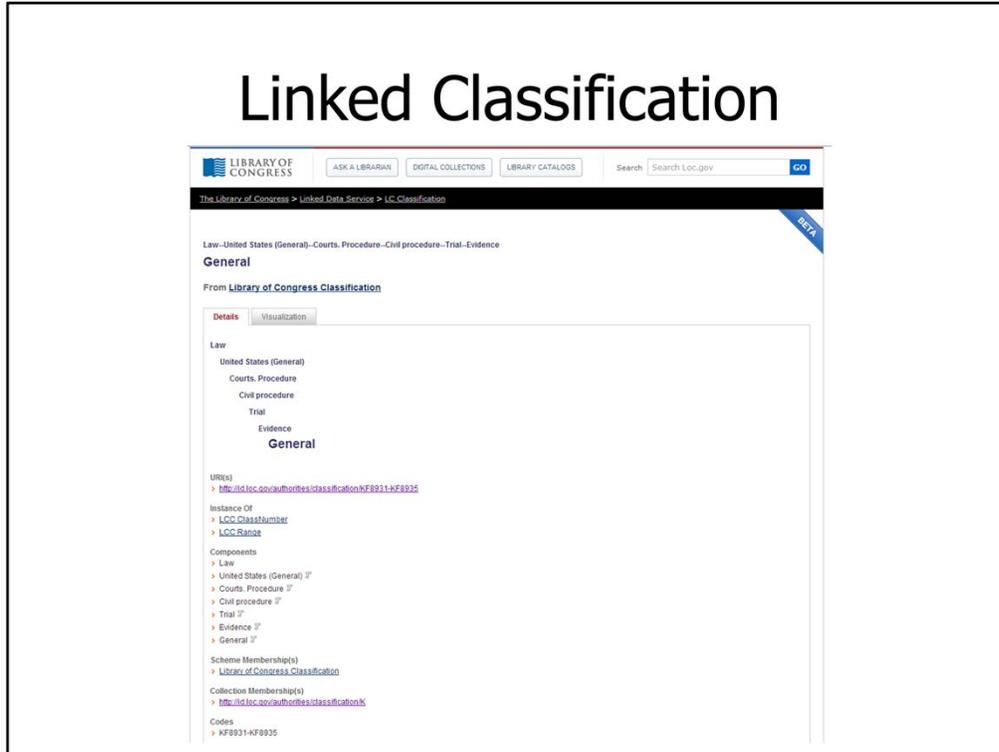
Linked Classification



The last thing I want to touch on briefly is the Library of Congress Classification scheme presented as linked data. About a year ago the Library of Congress released Class K of the Library of Congress Classification as linked data. The general scheme began to appear in beta form in June of 2012. Class K and its subclasses is the largest of the classes in LCC.

I immediately started to think of the implications of this and our own KF Modified classification scheme that is based on the Law of the United States.

Linked Classification



I'm happy to report that Sarah and I have received a research grant from CALL to explore transforming KF Modified to a linked data format.

We hope to create something similar to this HTML version of LCC KF8932-KF8935.

Linked Classification

```
<rdfa:label xmlns:rdfa="http://www.w3.org/2000/01/rdfa-schema#">Civil procedure</rdfa:label>
</madsrdf:Topic>
<madsrdf:Topic>
<rdfa:label xmlns:rdfa="http://www.w3.org/2000/01/rdfa-schema#">Trial</rdfa:label>
</madsrdf:Topic>
<madsrdf:Topic>
<rdfa:label xmlns:rdfa="http://www.w3.org/2000/01/rdfa-schema#">Evidence</rdfa:label>
</madsrdf:Topic>
<madsrdf:Topic>
<madsrdf:authoritativeLabel>General</madsrdf:authoritativeLabel>
</madsrdf:Topic>
</madsrdf:componentList>
<loc:useGuideTable>
<loc:GuideTable rdf:about="http://id.loc.gov/authorities/classification/ct94006200">
<rdf:type rdf:resource="http://id.loc.gov/ontologies/loc#Range"/>
<madsrdf:authoritativeLabel xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">Table of form divisions (5 nos.)</madsrdf:authoritativeLabel>
<rdf:label xmlns:rdfa="http://www.w3.org/2000/01/rdfa-schema#">Law--Table of form divisions (5 nos.)</rdfa:label>
<madsrdf:code xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">1-5</madsrdf:code>
</loc:GuideTable>
</loc:useGuideTable>
<madsrdf:hasBroaderAuthority xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">
<loc:ClassNumber rdf:about="http://id.loc.gov/authorities/classification/KF8931-KF8969">
<rdf:type rdf:resource="http://id.loc.gov/ontologies/loc#Range"/>
<madsrdf:authoritativeLabel>Evidence</madsrdf:authoritativeLabel>
<rdfa:label xmlns:rdfa="http://www.w3.org/2000/01/rdfa-schema#">Law--United States (General)--Courts. Procedure--Civil procedure--Tri
<madsrdf:code>KF8931-KF8969</madsrdf:code>
</loc:ClassNumber>
</madsrdf:hasBroaderAuthority>
<madsrdf:hasNarrowerAuthority xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">
<loc:ClassNumber rdf:about="http://id.loc.gov/authorities/classification/KF8932.4-KF8932.5">
<rdf:type rdf:resource="http://id.loc.gov/ontologies/loc#Range"/>
<madsrdf:authoritativeLabel>Court decisions</madsrdf:authoritativeLabel>
```

By transforming it to the RDF model similar to this excerpt of the underlying RDF coding.

It will be interesting to see how we can use this classification as linked data to organize legal resources on the Web.

Thank you for your attention!

tknight@osgoode.yorku.ca

Twitter: @freemoth

