Beyond MARC:
Lessons from RDF and Linked Open Data

Myung-Ja Han
Timothy W. Cole
University of Illinois at Urbana-Champaign

- MARC & the History of Library Catalog
- Evolving Role of Cataloging & Metadata
- RDF, Semantic Web & Linked Open Data
- Library Initiatives: MODS/RDF & BIBFRAME
- Other Initiatives: SKOS, FOAF, schema.org
- Collaborative Metadata
- The future: from records to triples?

Some Background & Perspective

Metadata help us (librarians) to Organize, Manage, and Preserve library resources.

Metadata help our users to Discover, Connect, & Use library resources.

Bibliographic Control

Early on this was largely about inventory:

- What do we own?
- Where is it shelved?
- What is its provenance?
- What is it about?

Descriptive cataloging has evolved in concert with our discovery systems:

- Card Catalogs
- Online Public Access Catalogs (OPAC)
- Web [of documents]
- Semantic Web

FRBR Put a Focus on the User

FRBR User Tasks

- Find [Discover]
- Identify
- Select [Evaluate]
- Acquire/Obtain access [Retrieve]

Today we can use metadata to:

- Enumerate & describe our collections, both physical and virtual
- Manage diverse, complex often multi-part digital & physical resources
- Enable discovery, interoperability, linking & interactivity at multiple levels of granularity

Metadata provide multiple views
of the library resources we curate and their
connections to context & other resources.

MARC and History of Library Catalog

Hughes, William, Clerigog Fawr.

Cyfaill y cyfaddain, neu gatorfa o wybodaeth galfyddydawl; yn cynnwys hairnyddiaeth, etc.

Caergybi, H. Jones, 1840. 60 tud.

G1.25

MARC I

00667cam a2200229Ii
4500001000800000005001700008008004100025035002300066040002300089035001300112035001600125090002200141049000900163100003500172245010800207260003800315300001900353500003200372910001200404994001200416910000900428262410820030807121329.0900530
s1841 enk 000 1 eng d a(OCoLC)ocm03535290
aSUCcSUCdOCLdUIU 9ARY-4606 9UC 12037257 aPR4058b.S65 1841
aUIUU1 aIngoldsby, Thomas,d1788-1845.10aSome account of my cousin Nicholas /cby Thomas Igoldsby, esq. ; to which is added, The rubber of life.
aLondon :bRichard Bentley,c1841. a3 v. ;c19 cm. aFirst edition. Sadleir 157. arcp3356 a02bUIU aMARS

MARC II

- Developed in the 1960s
- ISO2709 - Format for Bibliographic Information Interchange on *Magnetic Tape*
- ANSI/NISO Z39.2 – Information Interchange Format
- Used as library's bibliographic standard and holdings standard
- Uses AACR2/RDA as content standards

“The development of MARC was a revolutionary advancement in modern librarianship.”

BUT NOW …

Tennant, R. (2004).
A bibliographic metadata infrastructure for the twenty-first century.
Library Hi Tech, 22(2), 175-181.

MARC Must Die?!

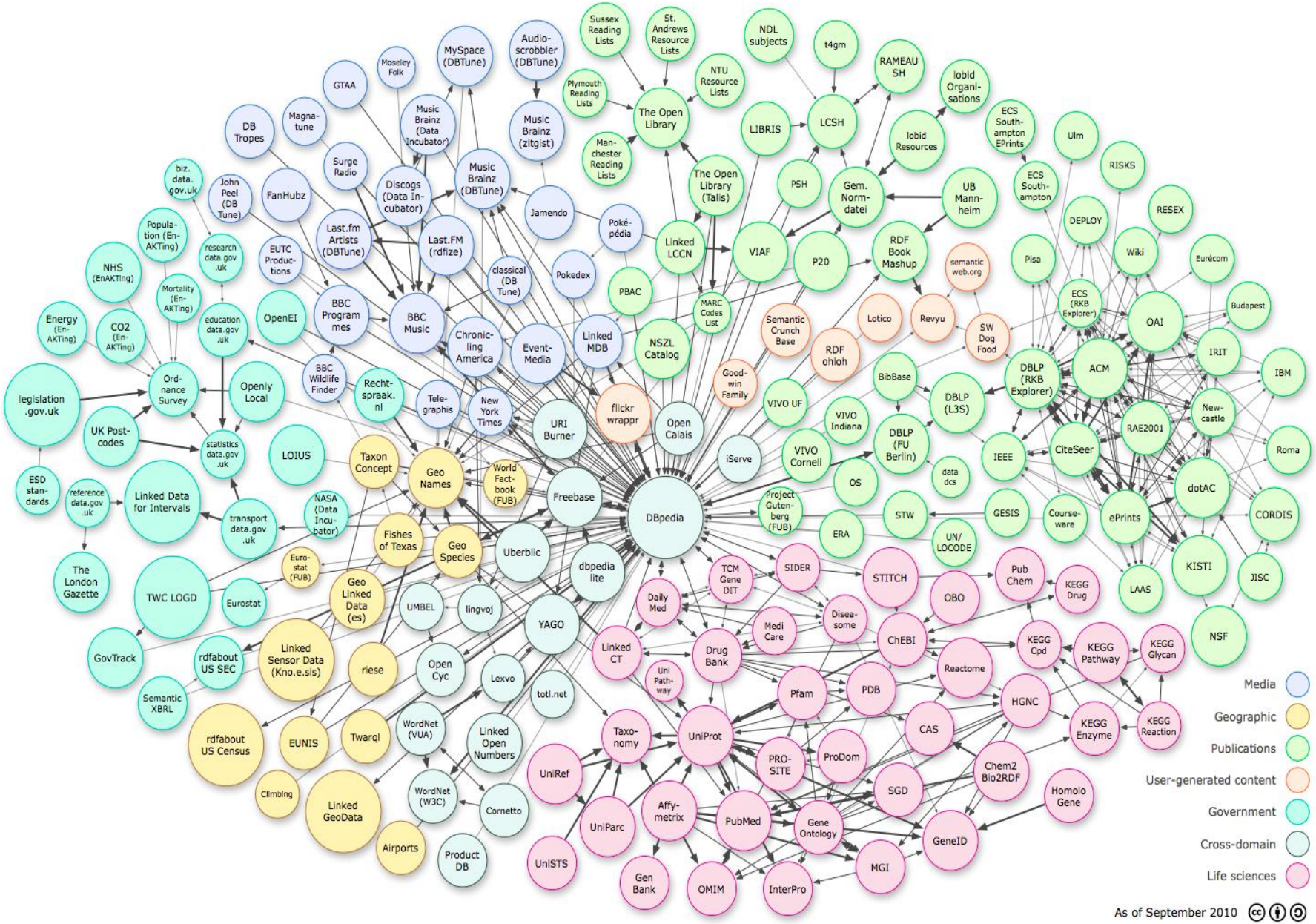
Not Quite YET

- Records in MARC format
  - 311,114,134 MARC records as of 1 Jan 2014*
- Library systems still rely on MARC
- New standards/data model & new descriptive rules are coming
  - this partly reflects a proactive effort to transform metadata records into MARC

*MARC Usage in WorldCat (http://experimental.worldcat.org/marcusage/)

Still, MARC Is No Longer Enough

- Resources in many different formats
  - audio / visual, as well as electronic resources
- Increasing number of digital resources
- Need to have granular levels of metadata
- Something easier to create and share


Sussex Reading Lists
St. Andrews Resource Lists
NDL subjects
t4gm
RAMEAU SH
lobid Organi-sations
Plymouth Reading Lists
NTU Resource Lists
The Open Library
LIBRIS
LCSH
lobid Resources
ECS South-ampton EPrints
Ulm
Man-chester Reading Lists
The Open Library (Talis)
PSH
Gem. Norm-datei
UB Mann-heim
ECS South-ampton
RISKS
MySpace (DBTune)
Audio-scrobbler (DBTune)
Moseley Folk
GTAA
Music Brainz (Data Incubator)
Music Brainz (zitgist)
DB Tropes
Magna-tune
Surge Radio
Discogs (Data In-cubator)
Music Brainz (DBTune)
DEPLOY
RESEX
biz. data. gov.uk
John Peel (DB Tune)
FanHubz
Jamendo
Poké-pédia
Linked LCCN
VIAF
P20
RDF Book Mashup
Pisa
Wiki
Eurécom
Popula-tion (En-AKTing)
Last.fm Artists (DBTune)
Last.FM (rdfize)
Pokedex
PBAC
MARC Codes List
semantic web.org
ECS (RKB Explorer)
OAI
Budapest
NHS (EnAKTing)
research data.gov .uk
EUTC Produc-tions
classical (DB Tune)
Semantic Crunch Base
Lotico
Revyu
SW Dog Food
IRIT
Energy (En-AKTing)
Mortality (En-AKTing)
BBC Programmes
BBC Music
Chronic-ling America
Event-Media
Linked MDB
NSZL Catalog
RDF ohloh
BibBase
DBLP (RKB Explorer)
ACM
IBM
CO2 (En-AKTing)
education data.gov .uk
OpenEI
Good-win Family
VIVO UF
DBLP (L3S)
New-castle
legislation .gov.uk
Ord-nance Survey
Openly Local
BBC Wildlife Finder
Recht-spraak. nl
Tele-graphis
New York Times
URI Burner
flickr wrappr
Open Calais
VIVO Indiana
DBLP (FU Berlin)
RAE2001
Roma
UK Post-codes
statistics data.gov .uk
LOIUS
iServe
VIVO Cornell
IEEE
CiteSeer
dotAC
ESD stan-dards
reference data.gov .uk
Linked Data for Intervals
Taxon Concept
Geo Names
World Fact-book (FUB)
Freebase
DBpedia
Project Guten-berg (FUB)
OS
data dcs
GESIS
Course-ware
ePrints
CORDIS
transport data.gov .uk
NASA (Data Incu-bator)
Fishes of Texas
Geo Species
Uberblic
dbpedia lite
ERA
STW
UN/ LOCODE
The London Gazette
Euro-stat (FUB)
TWC LOGD
Eurostat
Geo Linked Data (es)
UMBEL
lingvoj
YAGO
TCM Gene DIT
SIDER
STITCH
Pub Chem
KEGG Drug
OBO
LAAS
KISTI
JISC
Daily Med
Medi Care
Disea-some
ChEBI
KEGG Cpd
KEGG Pathway
KEGG Glycan
NSF
GovTrack
rdfabout US SEC
Linked Sensor Data (Kno.e.sis)
riese
Open Cyc
Lexvo
totl.net
Linked CT
Uni Path-way
Drug Bank
Pfam
PDB
Reactome
HGNC
Semantic XBRL
rdfabout US Census
EUNIS
Twarql
WordNet (VUA)
Linked Open Numbers
Taxo-nomy
UniProt
CAS
KEGG Enzyme
KEGG Reaction
Climbing
Linked GeoData
WordNet (W3C)
Cornetto
PRO-SITE
ProDom
Chem2 Bio2RDF
UniRef
SGD
Homolo Gene
Airports
Product DB
UniParc
Affy-metrix
PubMed
Gene Ontology
GeneID
UniSTS
Gen Bank
OMIM
InterPro
MGI
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2010

60+ years	**Cards Catalog to MARC**
40+ years	**AACR to RDA**
30+ years	**MARC to MARCXML**
10+ years	***Semantic web and Linked Data***

Evolving Role of Cataloging & Metadata

Changing Environment I

- Libraries have lost their place as primary information providers

  (Karen Coyle & Diane Hillman, 2007)
- The printed book is no longer the only major vehicle for scholarly communication

  (Mark Sandler, 2005)
- Increase in Web-based information resources

Changing Environment II

- Records contain intellectual content
- Records are harvested and converted
- Records are created on many different levels
- Too many metadata schemes to consider
- Records become ***'resources'***

Jenn Riley (2010).
Seeing Standards: A Visualization of the Metadata Universe.

Who Creates Metadata?

Who Works with Metadata?

- Data Analyst
- Digital Humanist (Scholars and Researchers)
- Data Architect
- Portal developer
- Discovery service programmer

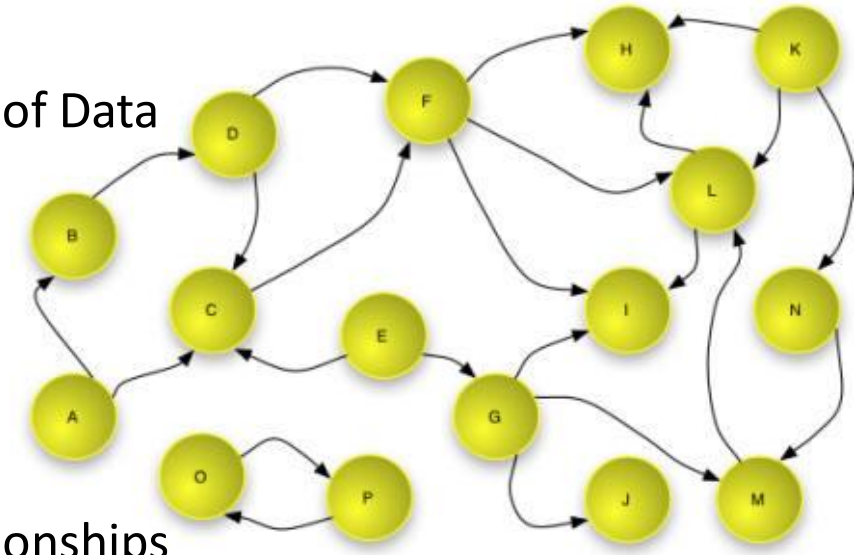
DATA

RESOURCE

RDF, Semantic Web & Linked Open Data

The Web is a Graph

- Classically, the nodes of the Web Graph are 'document-like-objects', i.e., unstructured and semi-structured resources
- The edges of the Web Graph are directed links from one Web 'document' to another
- The Semantic Web: towards a Web of Data
  - A warehouse of actionable distributed data
  - Supports recombination & integration of distributed data
  - Supports reasoning about relationships between resources & the data they contain
  - Relies on metadata, link typing, identifiers & authorities



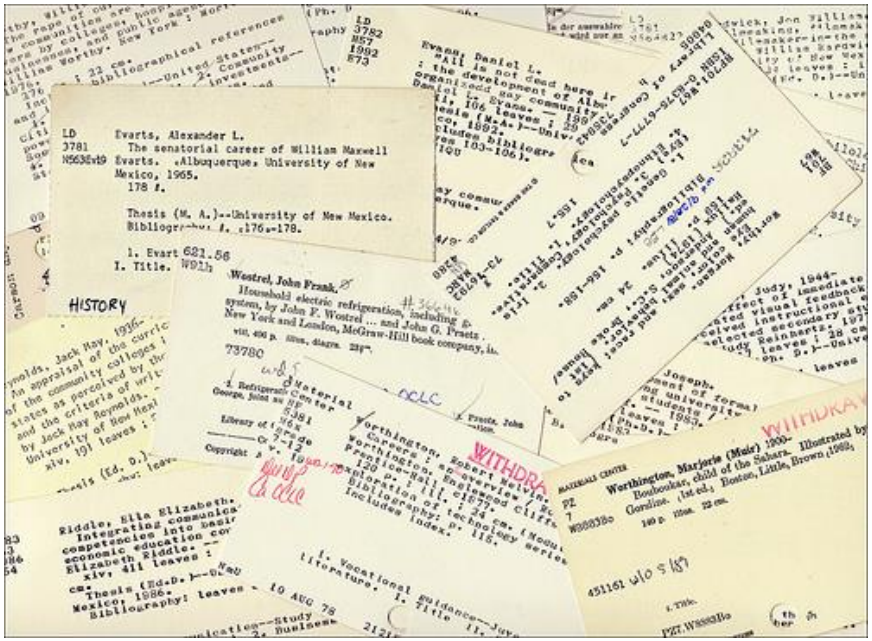
Lorcan Dempsey, The Recombinant Library: Portals and People

The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

RDF – a language for representing information about resources on the Semantic Web

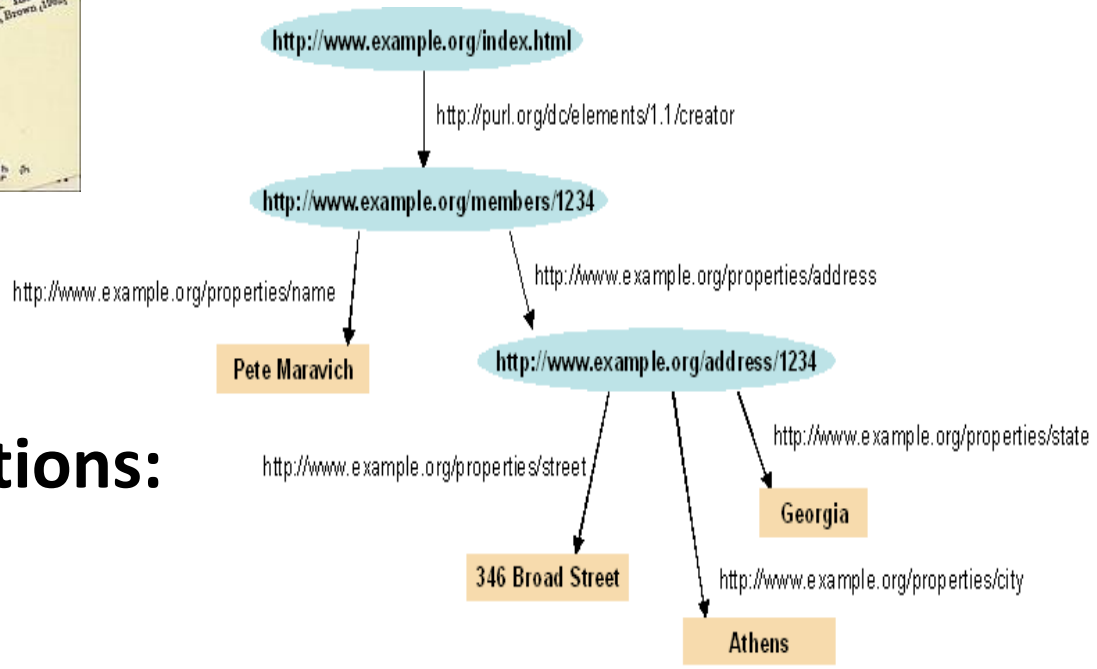
RDF is the foundation of the Semantic Web

- Allows you to express statements about resources, content, and data on the Web
- Statements consist of a subject, a predicate (property) and an object (value)
  - **My book**'s **title** is ***Metadata is Easy***.
- RDF models these statements as a graph of triples



RDF:

- subject – predicate – object
- graph / triple based
- Open world



Library Descriptive Traditions:

- Record oriented
- Closed world

Strengths and Weaknesses of RDF

- Triples / graphs are ‘easy’ to reason over
- Triples / graphs are ‘easy’ to create
- Triples / graphs are ‘easy’ to extend and combine
- RDF & Linked Open Data support distributed metadata, including metadata that are constantly being enriched, augmented

- Triples-based UI, query logic can be harder
- Hard to validate, e.g., for completeness, logical consistency
- The Open World axiom limits assumptions / inferences
- Triples in isolation can be ambiguous, misinterpreted

Interface Design Issues

- Less uniformity of metadata for each item
  - UI must not break when attributes are missing
  - UI must be flexible enough to show attributes when present
  - UI functionality must deal with heterogeneity

- UI must deal with sub-graphs
  - May need to retrieve information from external sources
  - UI must work asynchronously & degrade gracefully
  - Need to be tolerant of conflicting metadata values
  - Need to be tolerant of links that lead to infinite loops

MODS/RDF & BIBFRAME

MODS RDF I

- MODS is (loosely) based on MARC
- Widely used to describe objects in Libraries, Archives, and Museums
- MODS community requested LC for explorations into Linked Open Data
  - MAD RDF was already developed and used since March 2011, updated in May 2012. http://www.loc.gov/standards/mads/rdf/

MODS RDF II

- Can be used to
  - create born-RDF MODS, or
  - create an RDF description corresponding to an existing MODS XML record
- XSLT is available to create MODSRDF from MODSXML

  (http://www.loc.gov/standards/mods/modsrdf/xsl-files/modsrdf.xsl)

MODS RDF III

- MODS RDF developments
  - Ontology Primer
  - MODS RDF namespace document
  - Examples of MODS RDF records (http://www.loc.gov/standards/mods/modsrdf/examples/)
- However,
  - MODS in RDF work is still a draft and work in progress (last updated in June 2013)
  - XSLT is not complete

Bibliographic Framework Initiative

- Library of Congress issues its initial plans for its Bibliographic Framework Transition Initiative (May, 2011)
- To determine a transition path for the MARC 21 exchange format in order to reap the benefits of newer technology while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades.

http://www.loc.gov/bibframe/

Requirements for a New Bibliographic Framework Environment

- Broad accommodation of content rules and data models
- Provision for types of data that logically accompany or support bibliographic description
- Accommodation of textual data, linked data with URIs instead of text, and both
- Consideration of the relationships between and recommendations for communications format tagging, record input conventions, and system storage/manipulation

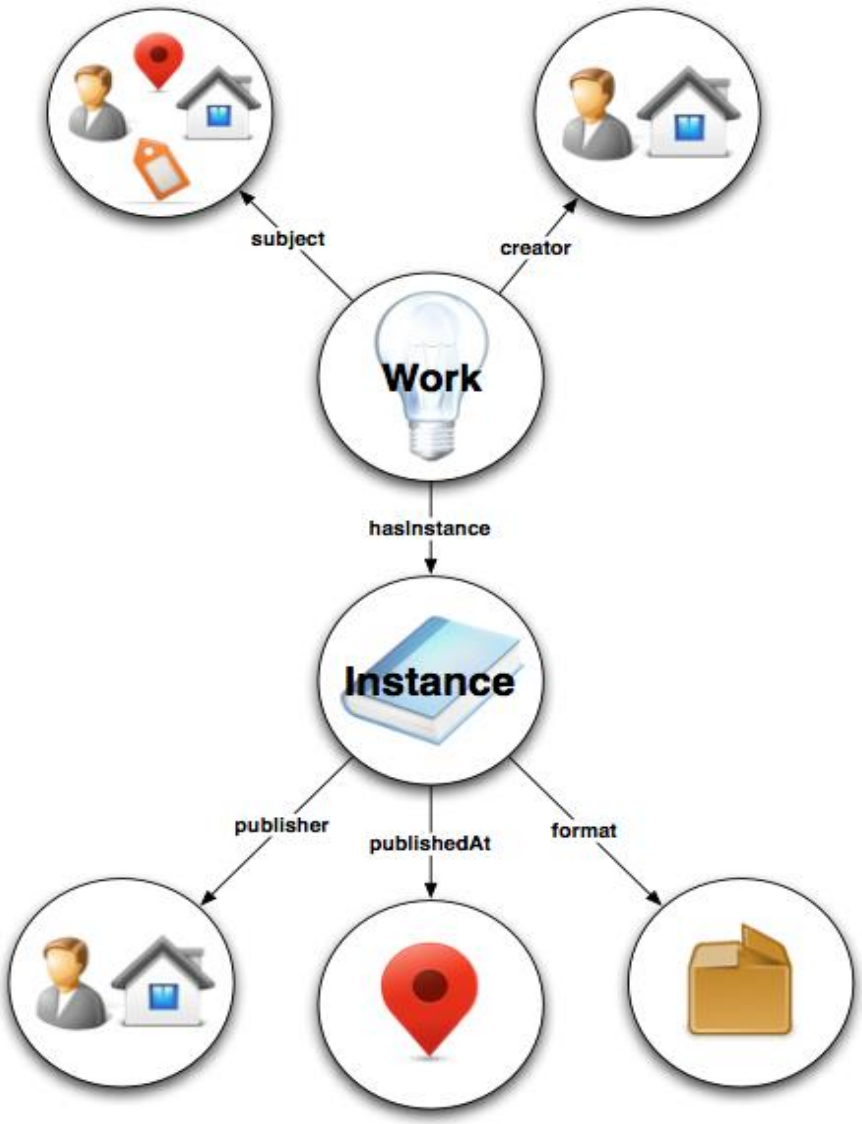
- Consideration of the needs of all sizes and types of libraries, from small public to large research
- Continuation of maintenance of MARC until no longer necessary
- Compatibility with MARC-based records
- Provision of transformation from MARC 21 to a new bibliographic environment

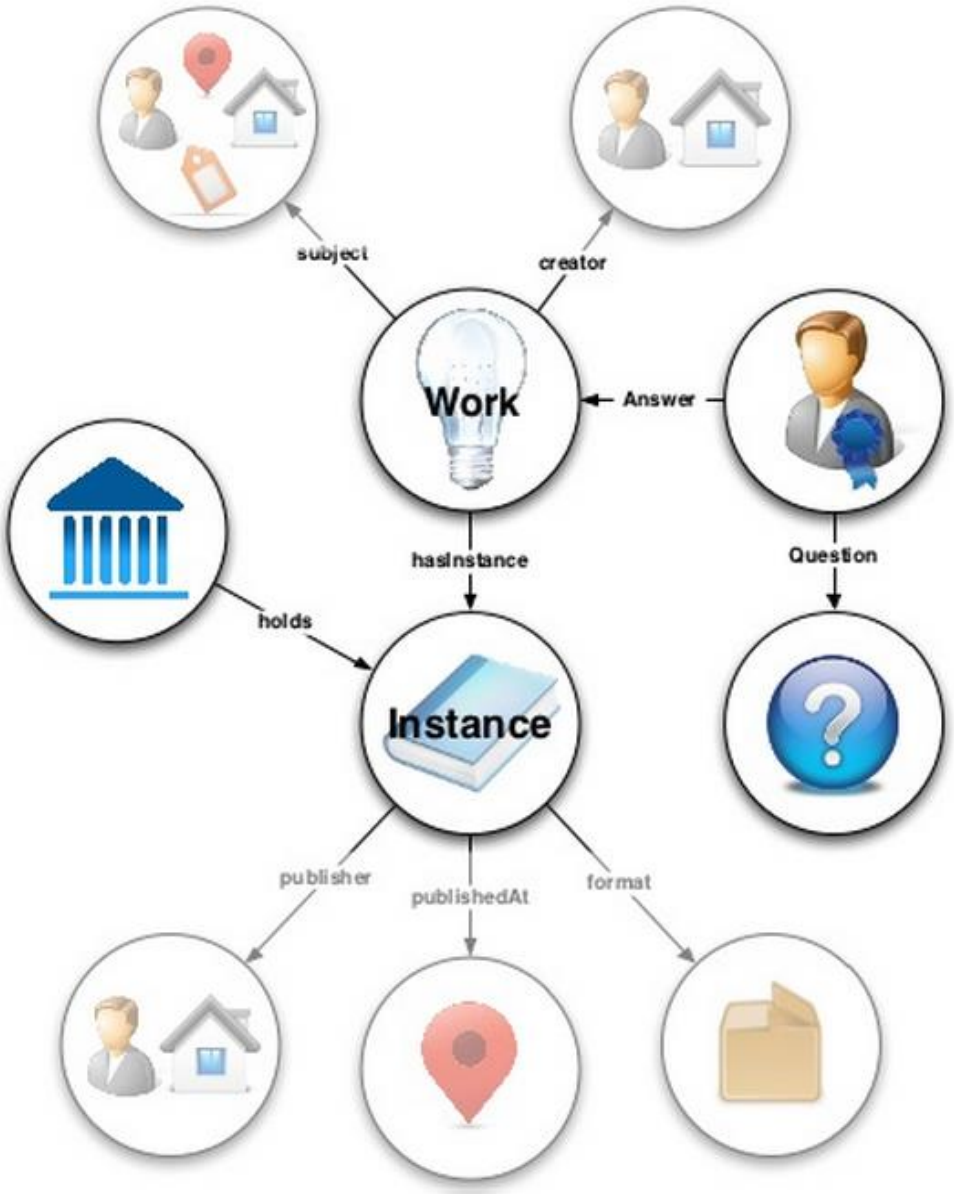
Work Areas

- Develop possible interaction scenarios within the information community
- Develop domain ontologies for the description of resources and related data in scope
- Organiz prototyping and reference implementations
- Analyze related initiatives (RDA, FRBR, ...)
- Analyze MARC and its data model

BibFrame Model Core Classes

- **Creative Work** - a resource reflecting a conceptual essence of the cataloging item.
- **Instance** - a resource reflecting an individual, material embodiment of the Work.
- **Authority** - a resource reflecting key authority concepts that have defined relationships reflected in the Work and Instance. Examples of Authority Resources include People, Places, Topics, Organizations, etc.
- **Annotation** - a resource that decorates other BIBFRAME resources with additional information. Examples of such annotations include Library Holdings information, cover art and reviews.


subject
creator
Work
hasInstance
Instance
publisher
publishedAt
format


subject
creator
Work
Answer
hasInstance
Question
holds
Instance
publisher
publishedAt
format

BibFrame, So Far…

- BIBFRAME Annotation Model (Updated - 26 August 2013)
- BIBFRAME Use Cases and Requirements (NEW! - 14 August 2013)
- On BIBFRAME Authority - Discussion Paper (Updated - 15 August 2013)
- BIBFRAME Resource Types Discussion Paper (17 June 2013)
- Vocabulary updates (Ongoing)
- MARC21 to BIBFRAME Transformation updates (Ongoing)

MARC in BibFrame Model

- Comparison Service
- Transformation Service

- *Try these services to see how MARC records look different in BibFrame model.

From the Web: FOAF, SKOS, schema.org

Friend of a Friend (FOAF)

- RDF-compatible semantics (classes & properties)
  - designed to enable "…a Web of machine-readable pages describing people, the links between them and the things they create and do…."
  - FOAF Project (http://www.foaf-project.org/) started in 2000
  - FOAF Vocabulary Spec 0.99 published 14 January 2014 (*Paddington edition*)

FOAF Classes (selected)

Classes (*unstable*, *testing*, *stable* & *archaic*)

- Agent
- Group (sub-class of Agent)
- Image
- OnlineAccount
- OnlineChatAccount (sub-class of OnlineAccount)
- Organization (sub-class of Agent)
- Person (sub-class of Agent)
- Project

FOAF core properties (selected)

Properties (core)

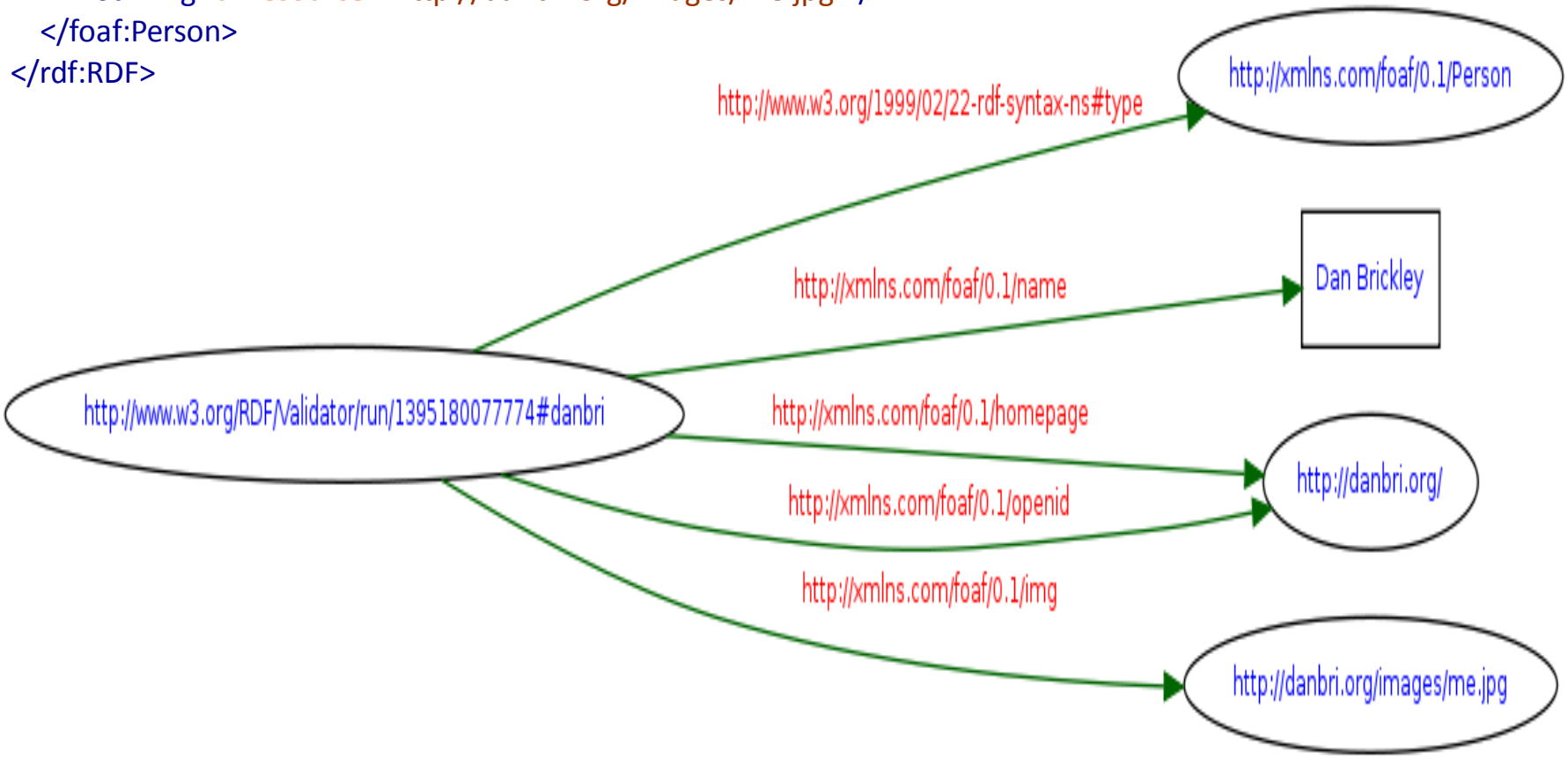
- name
- familyName, givenName (domain:Person)
- knows (domain:Person, range:Person)
- img (domain: Person, range: Image)
- depiction (range: Image), super-property of img
- age (domain: Agent), see also birthday (not core)
- member (domain: Group, range: Agent)

FOAF social web properties (selected)

Properties (Social Web)

- nick
- mbox (domain: Agent)
- homepage (range: Document)
- workplaceHomepage (domain:Person, range:Document)
- topic (domain: Document), topic_interest (domain:Agent)
- interest (domain: Agent, range: Document)
- account (domain: Agent, range: OnlineAccount)
- publications (domain: Person, range: Document)
- openid, jabberID (domain: Agent)

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
                xmlns:dc="http://purl.org/dc/elements/1.1/">
  <foaf:Person rdf:about="#danbri" xmlns:foaf="http://xmlns.com/foaf/0.1/">
    <foaf:name>Dan Brickley</foaf:name>
    <foaf:homepage rdf:resource="http://danbri.org/" />
    <foaf:openid rdf:resource="http://danbri.org/" />
    <foaf:img rdf:resource="http://danbri.org/images/me.jpg" />
  </foaf:Person>
</rdf:RDF>
```



SKOS Introduction

- Grew out of series of EU initiatives (DESIRE II, LIMBER, SWAD, …)
  - Initial release of SKOS Core was in 2003
  - Picked up by W3C Semantic Web Activity in 2004
  - W3C Recommendation in 2009
- Alignment between ISO 25964 (the International Std. for Thesauri…) and SKOS & MADS/RDF
- SKOS is “ a common data model for sharing and linking knowledge organization systems via the Web.”

SKOS Classes

Classes

- ConceptScheme (aggregation of Concepts)
- Concept

- Collection (group of Concepts, for convenience)
- OrderedCollection (sub-class of Collection)

SKOS Properties (selected)

Properties

- inScheme
- notation
- prefLabel, altLabel
- broader, narrower (direct hierarchical)
- broaderTransitive, narrowerTransitive (indirect hierarchical)
- Note (changeNote, definition, editorialNote, example, historyNote, scopeNote)
- broadMatch, closeMatch, exactMatch, narrowMatch,….

Schema.org

The next generation's Dublin Core?

Semantics (Classes [aka types] & Properties) designed to be used with HTML microdata or RDFa formats

- Data model is compatible (mostly) with RDF Schema
- Microdata formats compatible (mostly) with RDFa 1.1 Lite

Microformats, RDFa, microdata, …

- Microformats – an attempt to make human-readable HTML more machine processible
  - Add attributes to HTML to convey metadata about the information contained in the HTML
- Microdata – more sophisticated microformat
  - Coincident with introduction of schema.org
  - Developed in concert with HTML 5
  - Anticipated much of what was ratified as RDFa 1.1 Lite
- Tools are available now for validating microdata and/or RDFa

Schema.org types (selected)

- Thing
  - Event
  - Organization
  - Person
  - Place
  - CreativeWork
    - Article
    - Book
    - Map
    - WebPage
      - AboutPage
      - ContactPage
      - SearchResultsPage

Schema.org Book Properties (selected)

From Thing:

- name
- url

From CreativeWork:

- about
- author / creator
- dateCreated, dateModified, datePublished
- inLanguage
- genre

From Book:

- isbn
- numberOfPages

Library (OCLC) extensions

Classes:

- Carrier
- Image
- ComputerFile
- Newspaper
- Periodical

Properties:

- hasCarrier
- placeOfPublication
- oclcnum


OCLC WorldCat
Search
Advanced Search
Find a Library
<< Return to Search Results
Cite/Export
Print
E-mail
Share
Permalink
Add to list
Add tags
Write a review
Rate this item:
XML for Catalogers and Metadata Librarians
XML for catalogers and metadata librarians
Author: Timothy W Cole; Myung-Ja K Han
Publisher: Santa Barbara, California : Libraries Unlimited, an imprint of ABC-CLIO, LLC, [2013]
Series: Third millennium cataloging
Edition/Format: Book : English View all editions and formats
Database: WorldCat
Summary: "This book provides a foundation of knowledge for catalogers, metadata librarians, and library school students on the Extensible Markup Language (XML)--one of the most commonly listed qualifications in today's cataloger and metadata librarian job postings. How are today's librarians to manage and describe the ever-expanding volumes of resources, in both digital and print formats? The use of XML in cataloging and Read more...
Rating: (not yet rated) 0 with reviews - Be the first.
Subjects
XML (Document markup language)
Cataloging -- Data processing.
Metadata.
View all subjects
More like this
User lists
Similar Items
Nearby libraries
to 61801
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801, United States
< 1 m / km
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801, United States
< 1 m / km
University of Illinois, Agriculture ARL Recon
Librarian? Claim your library
Borrow / obtain a copy
Buy it

owl:sameAs	<info:oclcnum/816512986>
rdf:type	schema:Book
rdfs:seeAlso	<http://www.worldcat.org/isbn/1598845195> owl:sameAs <urn:isbn:1598845195> rdf:type schema:ProductModel rdf:type schema:CreativeWork schema:isbn "1598845195" schema:model <http://www.worldcat.org/oclc/816512986>
rdfs:seeAlso	<http://www.worldcat.org/isbn/9781598845198> owl:sameAs <urn:isbn:9781598845198> rdf:type schema:ProductModel rdf:type schema:CreativeWork schema:isbn "9781598845198" schema:model <http://www.worldcat.org/oclc/816512986>
schema:author	<http://viaf.org/viaf/27338808> madsrdf:isIdentifiedByAuthority <http://id.loc.gov/authorities/names/n83184882> rdf:type schema:Person rdfs:label "Cole, Timothy"
schema:datePublished	"2013"
schema:inLanguage	"en"
schema:name	"Xml for catalogers and metadata librarians."

Collaborative Metadata

The new cooperative cataloging

- Folksonomies in addition to or instead of taxonomies
  - Built from the bottom up
  - Users coin terms (tag) independently as they use a resource
  - Changes in language are immediately reflected
- Issues
  - Polysemy (ambiguity) and Synonymy (redundancy)
  - Lack of hierarchy
  - Longevity (or lack there of)
  - Authority / authorized forms
  - More used resources get used more, get more tags, …

http://dx.doi.org/10.1108/07378830910942928 Tom Steele, U of Oklahoma

Positive experiences – sort of

- PennTags (& MTagger)

  http://tags.library.upenn.edu/help/

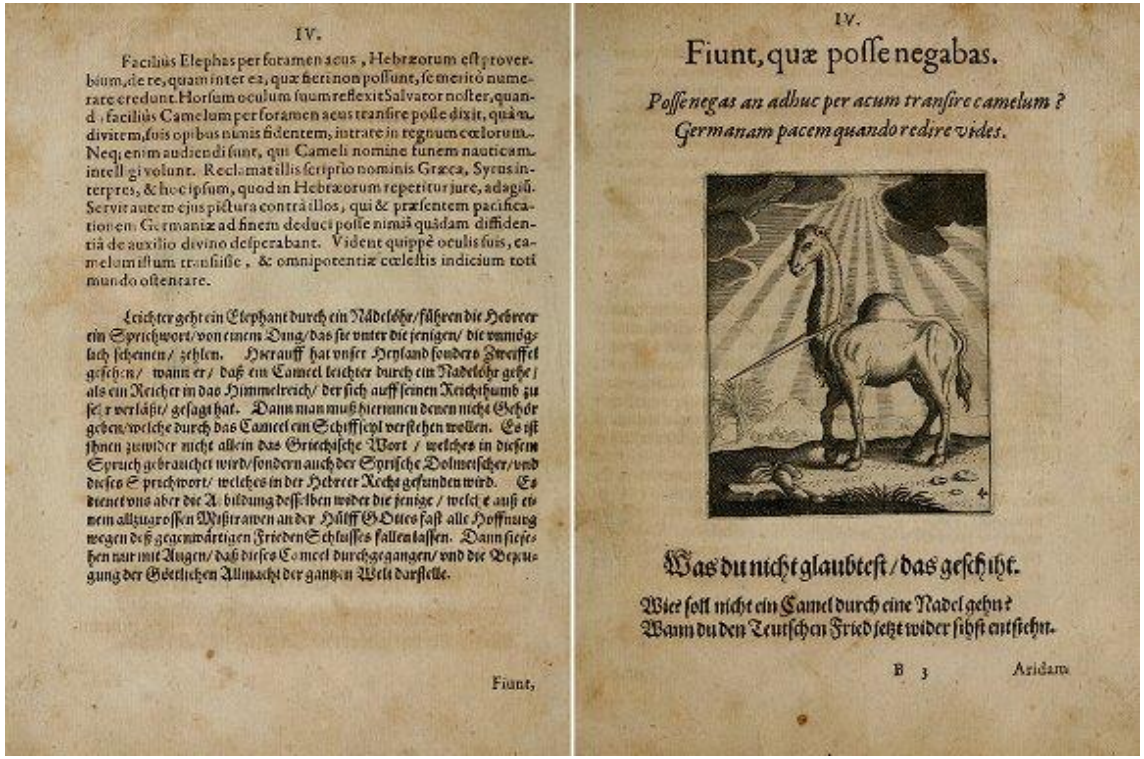
  Margaret Kipp & Grant Campbell. 2010. Searching with Tags: Do Tags Help Users Find Things? *Knowledge Organization* 37 (4): 239-55. Preprint: http://eprints.rclis.org/15033/1/kippandcampbell-2010-searchingwithtags-ko374.pdf

- LibraryThing

  Carrie Pirmann. *Using Tags to Improve Findability in Library OPACs: A Usability Study of LibraryThing for Libraries,* http://hdl.handle.net/2142/18918

- Library of Congress in Flickr Commons

  For the Common Good: The Library of Congress Flickr Pilot Project. 2008. http://www.loc.gov/rr/print/flickr_report_final.pdf



IV.

Facilius Elephas per foramen acus, Hebræorum est proverbium, de re, quam inter ea, quæ fieri non possunt, se meritò numerare credunt. Horsum oculum suum reflexit Salvator noster, quando facilius Camelum per foramen acus transire posse dixit, quàm divitem, suis opibus nimis fidentem, intrare in regnum cœlorum. Neq; enim audiendi sunt, qui Cameli nomine funem nauticam intelligi volunt. Reclamat illis scriptio nominis Græca, Syrus interpres, & hoc ipsum, quod in Hebræorum repetitur jure, adagiũ. Servit autem ejus pictura contrà illos, qui & præsentem pacificationem Germaniæ ad finem deduci posse nimiũ quãdam diffidentiã de auxilio divino desperabant. Vident quippè oculis suis, camelum istum transisse, & omnipotentiæ cœlestis indicium toti mundo ostentare.

Leichter geht ein Elephant durch ein Nädelöhr/ führen die Hebreer ein Sprichwort/ von einem Ding/ das sie unter die jenigen/ die unmöglich scheinen/ zehlen. Hierauff hat unser Heyland sonders Zweiffel gesehen/ wann er/ daß ein Camel leichter durch ein Nadelöhr gehe/ als ein Reicher in das Himmelreich/ der sich auff seinen Reichthumb zu sehr verläßt/ gesagt hat. Dann man muß hierinnen denen nicht Gehör geben/ welche durch das Camel ein Schiff seyl verstehen wollen. Es ist ihnen zuwider nicht allein das Griechische Wort / welches in diesem Spruch gebraucht wird/ sondern auch der Syrische Dolmetscher/ und dieses Sprichwort/ welches in der Hebreer Recht gefunden wird. Es dienet uns aber die Abbildung desselben wider die jenige / welche auß einem allzugrossen Mißtrauen an der Hülff GOttes fast alle Hoffnung wegen deß gegenwärtigen Friedens Schlusses fallen lassen. Dann sie sehen nur mit Augen/ daß dieses Camel durchgegangen/ und die Bezeugung der Göttlichen Allmacht der gantzen Welt darstelle.

Fiunt,

IV.

Fiunt, quæ posse negabas.

Posse negas an adhuc per acum transire camelum?
Germanam pacem quando redire vides.

Was du nicht glaubtest/ das geschiht.

Wie soll nicht ein Camel durch eine Nadel gehn?
Wann du den Teutschen Fried setzt wider sihst entstehn.

B 3 Aridam

How expert is the crowd?

Motto-Transcribed (la):
Fiunt, quae posse negabas.

Motto-Transcribed (de):
Was du nicht glaubtest/ das geschiht.

Motto_Normalized (de):
Was du nicht glaubtest, das geschieht.

From Book Entitled:
Meditationes emblematicae de restaurata pace Germaniae, 1649, Francofurti

Counterpoint – Cory Doctorow's *Metacrap*

- People lie
- People are lazy
- People are stupid
- Mission: Impossible -- know thyself
- There's more than one way to describe something
- Schemas aren't neutral
- Metrics influence results

The Future of Library Metadata

Linked Open Data, BIBFRAME, schema.org, ???

LD4L – Linked Data for Libraries

- Exploration & Research
  - How should libraries contribute Linked Data
  - How might libraries want to use Linked Data
- Cornell, Harvard & Stanford supported by 2-year, $1 million grant from Mellon Foundation
- Workshop planned for midpoint of grant
- Focused (for now) on MARC → BIBFRAME....

Linked Data for Libraries

Pages

PAGE TREE

- Project Proposal
- Project Timeline
- Planned LD4L Workshop
- Project Team
- Communications and Outreach
- Related Projects
- Working Documents

Space tools

Linked Data for Libraries (LD4L)

Created by Tim Donohue, last modified by Dean B. Krafft on Feb 06, 2014

Welcome to the project wiki-space for the Linked Data for Libraries (LD4L) project. The project is a collaboration of the Cornell University Library, the Harvard Library Innovation Lab, and the Stanford University Libraries, and is funded by a nearly $1 million two-year grant from the Andrew W. Mellon Foundation.

The goal of the project is to create a Scholarly Resource Semantic Information Store (SRSIS) model that works both within individual institutions and through a coordinated, extensible network of Linked Open Data to capture the intellectual value that librarians and other domain experts and scholars add to information resources when they describe, annotate, organize, select, and use those resources, together with the social value evident from patterns of usage.

Our intent is to do so using existing ontologies and Open Source technology.

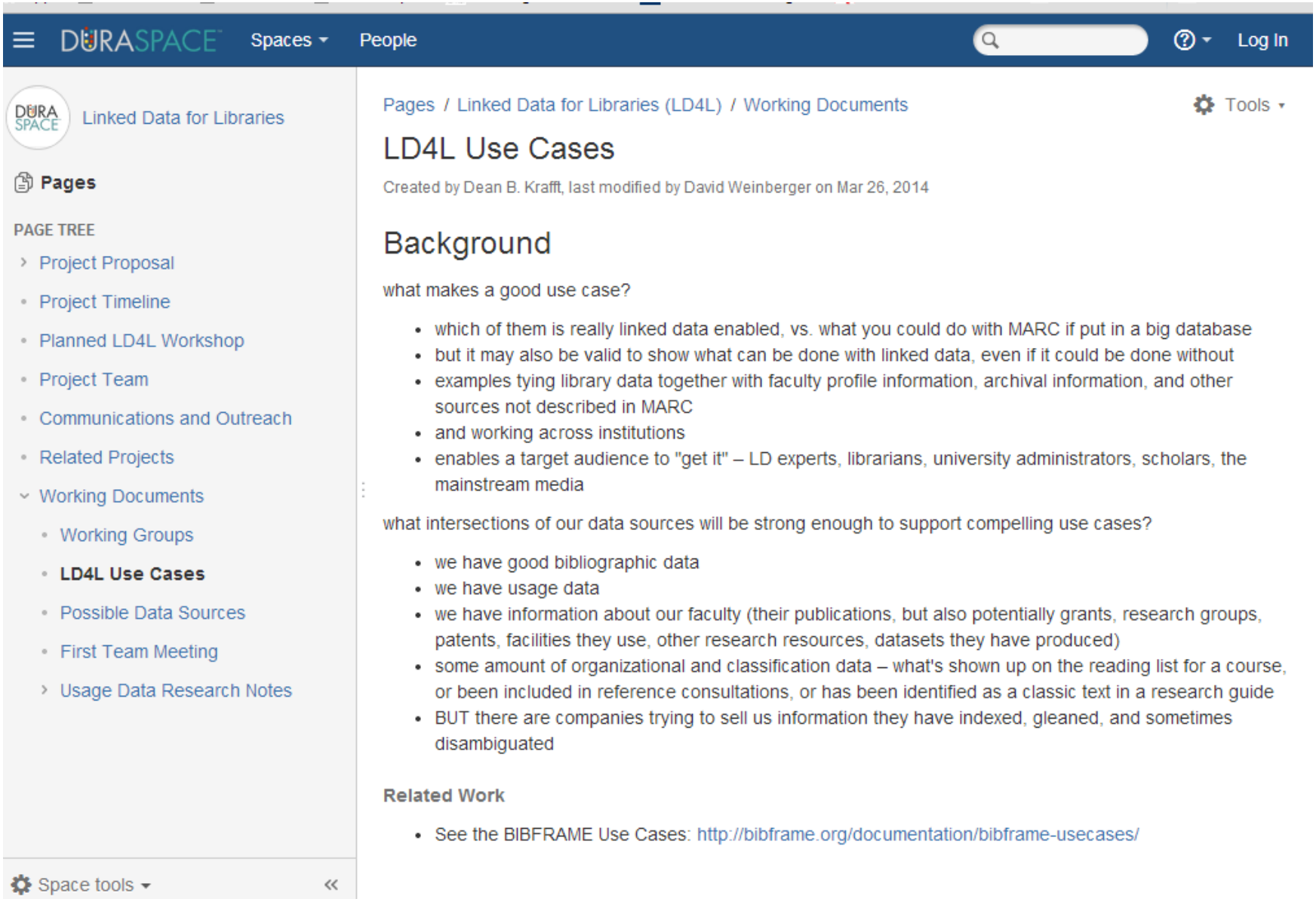
Project Pages

- Project Proposal:
  - Why Linked Data?
  - Previous Library-related Linked Data work at Cornell, Harvard, and Stanford
  - Rationale for the project
  - Project Description
  - Expected Outcomes
  - Intellectual Property and Sustainability
- Project Timeline
- Planned LD4L Workshop (tentatively set for January 2015)
- Project Team
- Communications and Outreach (Press releases, Presentations, Social media, Blog Posts, etc.)
- Related Projects

Working Documents

- Working Groups
- Workshop Planning
- LD4L Use Cases
- Possible Data Sources
- First Team Meeting

https://wiki.duraspace.org/pages/viewpage.action?pageId=41354028



https://wiki.duraspace.org/display/ld4l/LD4L+Use+Cases

The Relationship between BIBFRAME and OCLC's Schema.org 'Bib Extensions' Model: A Working Paper -- Carol Jean Godby, for OCLC Research
http://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf

1. Schema.org and BIBFRAME

In the past year, OCLC researchers have tried to make the case that Schema.org is a suitable foundation for the description of library resources. The most important argument is that the library community cannot afford to ignore Schema.org because it has been defined by Google, Yahoo!, Bing and Yandex to be the standard of choice for the publication of structured data and intelligent consumption of web resources that the major search engines commit to recognizing. Though Schema.org was not designed as a replacement for library standards, OCLC's linked data experts and many other library technology experts have concluded that the ontology defines a reasonably coherent commonsense model with classes and properties that are important for simple descriptions of bibliographic resources managed by libraries—including *creative work*, *person*, *author*, *director*, *place*, *organization*, *publisher*, *copyright date*, *book*, *ISBN*, and so on. These concepts can be serialized in a variety of forms and are compatible with the modeling philosophy promoted by the Semantic Web community.

But anyone who has examined Schema.org could easily enumerate where it falls short in describing the domain of library resources and services. There is no representation of the FRBR Group I concepts Work, Expression, Manifestation and Item. There is no clear distinction between *content* and *carrier*. Very few relationships among creative works have been defined. There is no concept of *collection* or *series*. And there are no models of transactions involving library resources and the organizations that provide or receive them, such as libraries, universities, publishers, e-content aggregators, and data service providers. Schema.org might suffice for a description that is equivalent in detail to a Dublin Core record, and is perhaps incrementally better, but the designers never intended it to be a global ontology. Rather, Schema.org is proposed as a starting point, or a scaffolding to which a more detailed ontology designed by interested communities of practice would be attached.

....

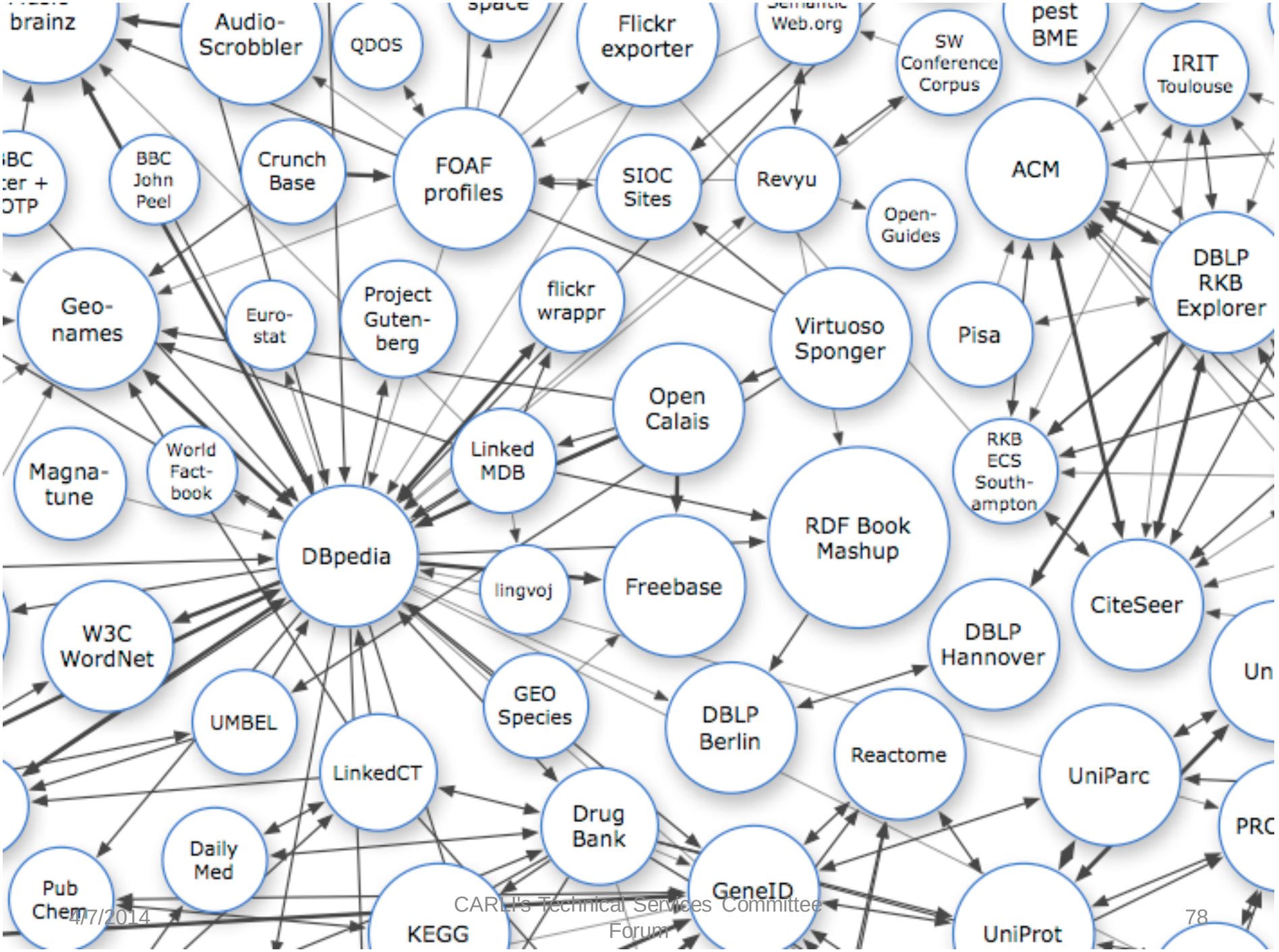
The cumulative effect of these three activities--defining a small number of properties whose role in model-building is essentially 'meta-,' addressing the need to represent FRBR Group I concepts, and eliminating redundant vocabulary—is that the OCLC and BIBFRAME models are now syntactically more compatible but semantically more complementary. This is a step forward. As we see it, the two models have different, but overlapping target audiences. BIBFRAME is designed for practitioners in the library community who create and manage descriptions that facilitate access to library resources. And Schema.org, with enhancements being discussed by the Schema Bib Extend community, provides a simplified description of library resources that can be integrated with related objects on the wider web and discovered by general-purpose search engines. As Richard Wallis said in a recent interview with Semanticweb.com:

The idyllic future of metadata?

- Many users know more about certain attributes of the resources we curate than we do.
- RDF enables the integration of metadata statements from multiple sources
- RDF-based tagging and annotation tools are emerging that allow us to capture user metadata augmentation.
- Therefore (perhaps) metadata designs of the future will be built around metadata enrichment by end-users, e.g. a Folksonomies model

HathiTrust Research Center

- Currently HT metadata is MARCXML
- Scholars tell us HT & HTRC metadata is inadequate
- For their own research they spend time & effort augmenting / enriching HT metadata
- They want to contribute metadata back to HT
- HTRC is looking at conversion to RDF-based metadata model as a way to support this use case and at the same time make use of LOD sources


brainz
Audio-Scrobbler
QDOS
Flickr exporter
Web.org
SW Conference Corpus
pest BME
IRIT Toulouse
BBC John Peel
Crunch Base
FOAF profiles
SIOC Sites
Revyu
ACM
Open-Guides
Geo-names
Euro-stat
Project Guten-berg
flickr wrappr
Virtuoso Sponger
Pisa
DBLP RKB Explorer
Open Calais
Magna-tune
World Fact-book
Linked MDB
RKB ECS South-ampton
DBpedia
RDF Book Mashup
lingvoj
Freebase
CiteSeer
W3C WordNet
DBLP Hannover
UMBEL
GEO Species
DBLP Berlin
Reactome
UniParc
LinkedCT
Drug Bank
Daily Med
Pub Chem
GeneID
KEGG
UniProt