Preparing Heritage Institutions' Metadata for the Semantic Web: Case Study of Bibliographic Data

Assoc. Prof. Mirna Willer, PhD

University of Zadar, *Department of Library and Information Science, Zadar, Croatia*

mwiller@unizd.hr

2011 Archival Education and Research Institute (AERI), Simmons College, Boston, Mass., July 10-15, 2011

Content

(1) Context: Semantic Web and Linked Data
   (1) Basic definitions & concepts
   (2) Archival context
(2) IFLA conceptual models and standards in RDF
   (1) ISBD, FRBR/FRAD/FRSAD, UNIMARC/B and UNIMARC/A: representation in RDF
   (2) Namespaces, classes & properties
   (3) Open Metadata Registry: infrastructure
(3) Discussion & comments from the archival community

(1) Context: Semantic Web and Linked Data

Argument (#1)

Heritage institutions possess highly **authoritative and trustworthy descriptive and authority data**; however, they are still closed in **silos** of their local systems or different kinds of union catalogues – whether physical or virtual. These data can be searched and retrieved, and subsequently used in a manner **structured and predefined** by the professionals based on **their assumptions of users needs**.

(1) Context: Semantic Web and Linked Data

Argument (#2)

Semantic Web standards and technologies and **Linked Data** concept provides means of releasing heritage institutions' metadata into the Semantic Web environment; however, ALM professionals have to prepare their **metadata** for such an environment in which **users will be enabled to perform tasks according to their specific needs**, and not necessarily only those traditionally envisioned by them.

(1) Context: Semantic Web and Linked Data

Argument (#3)

World Wide Web (1989/90-) → T. Berners-Lee "universal linked information system"

Information Object /Lifecycle of the [digital]Object = unit of description& its processing

Models & standards = interoperability if not convergence

e.g., IFLA conceptual model for authority data – FRAD: influenced by archival and museum communities → potential interest for these communities & therefore a meeting point / link!

Semantic Web (2001-) → space for further evolution of "our" co-operation: standards, data, serevices to users in the form of LINKING "things"

(1) Context: Semantic Web and Linked Data

"**Talk** of the Semantic Web is reminiscent of the artificial intelligence (AI) rage of two decades ago. There is compelling jargon, hyped promise, over-arching expectations, large sums of money, conferences, grants, and grant research programs. Will we see **a reprise of the collapse of expectations, the so-called *AI-winter***, as well? AI research suffered not so much from lack of ideas, resources, and enthusiasm, as from a plethora of all of these, without a **sufficient intellectual foundation** to support them."

Stuart L. Weibel. Social Bibliography : A Personal Perspective on Libraries and the Semantic Web (2007). Knitting the Semantic Web. Binghampton: Howarth Press. Pp. 227-236.

(1) Context: Semantic Web and Linked Data

(1) Context: Semantic Web and Linked Data

The Semantic Web is a web of data. [...]

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**. (2001)

(1) Context: Semantic Web and Linked Data

Weibel's "**sufficient intellectual foundation**": W3C

In addition to the classic "Web of documents" W3C is helping to build a technology stack to support a "Web of data," the sort of data you find in databases. The ultimate goal of the Web of data is to enable **computers to do more useful work** and to develop systems that can support **trusted interactions** over the network. The term "Semantic Web" refers to W3C's **vision of the Web of linked data**. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as **RDF**, SPARQL, OWL, and SKOS.

(1) Context: Semantic Web and Linked Data

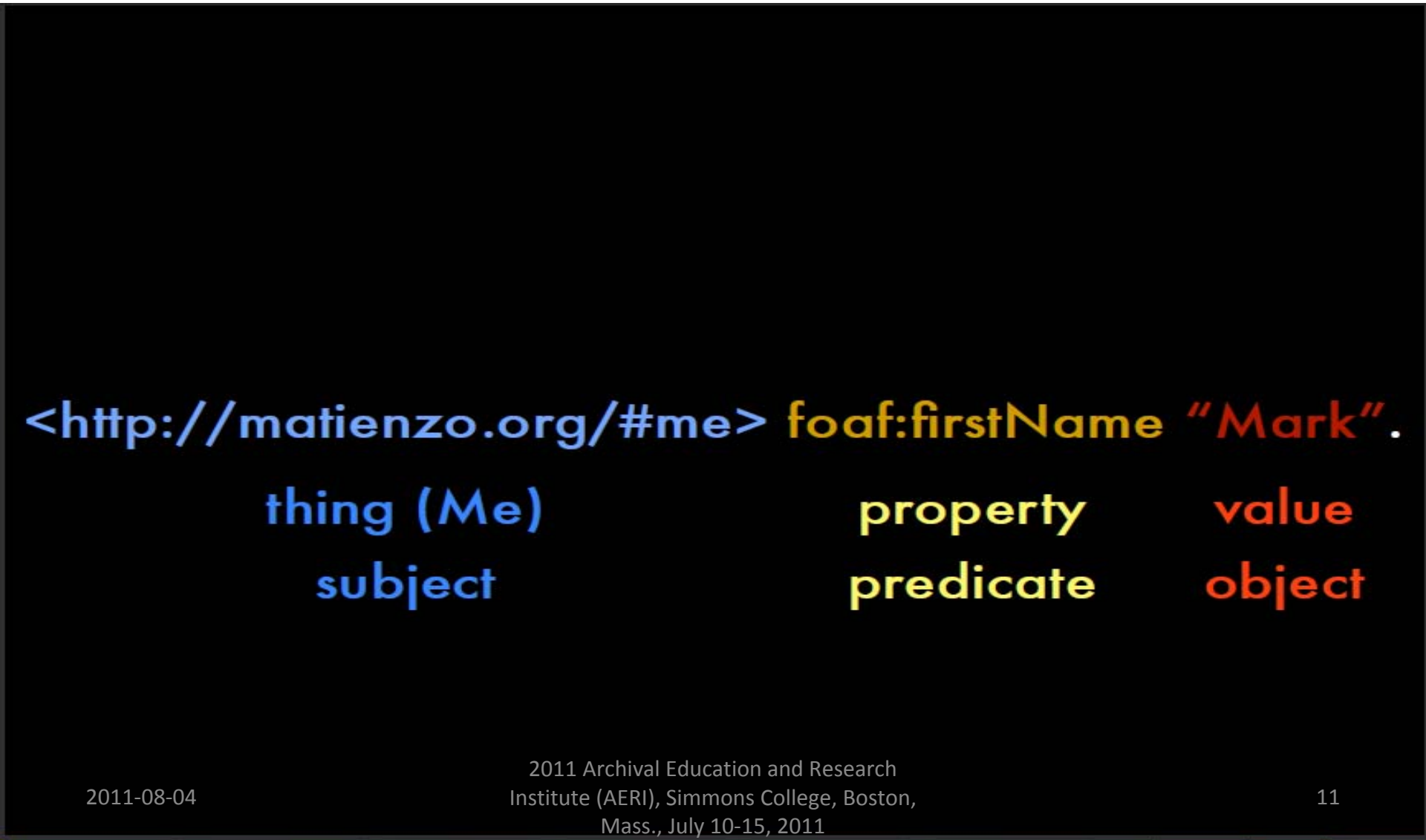
Inference [machine-understandable data]

inference on the Semantic Web can be characterized by **discovering new relationships**. On the Semantic Web, data is modeled as a **set of (named) relationships between resources**. "Inference" means that automatic procedures can generate new relationships based on the data and based on some additional information in the form of a vocabulary, e.g., a set of rules. Whether the new relationships are **explicitly added** to the set of data, or are returned at query time, is an implementation issue.

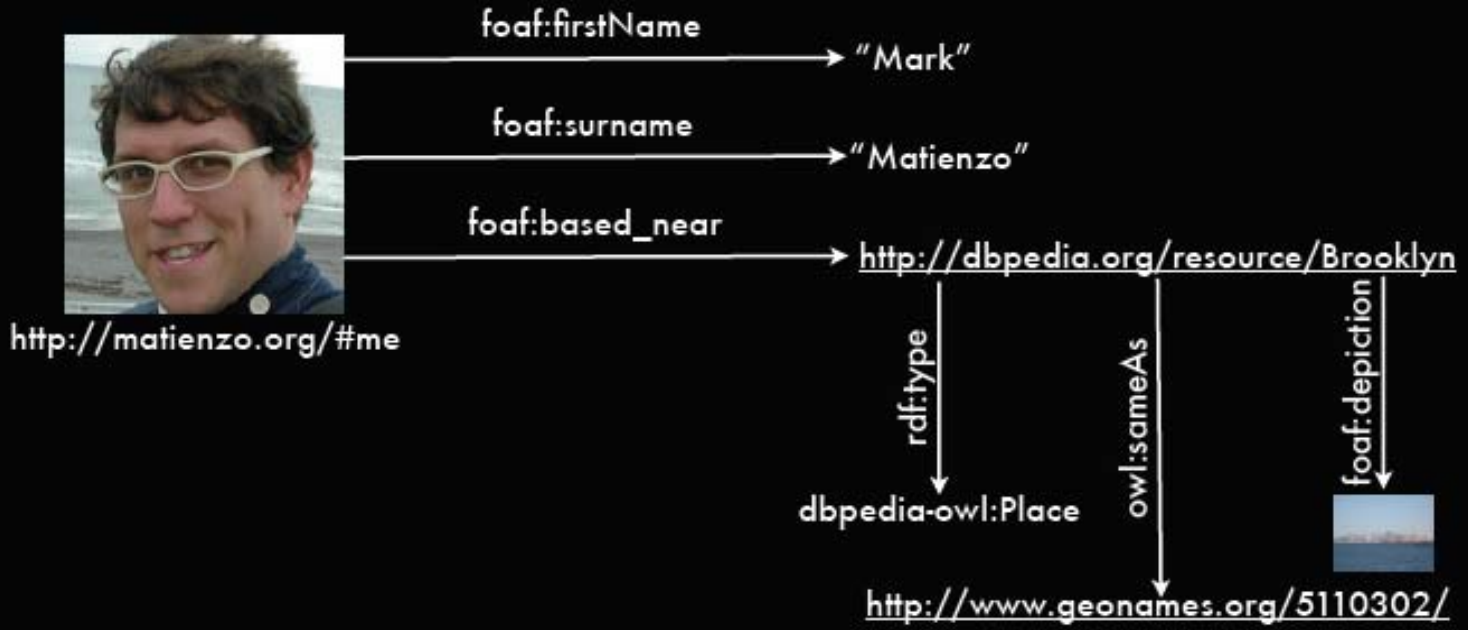
RDF: Resource Description Framework [format/standard]

is a language for representing information about resources in the World Wide Web → **triple**: subject – predicate – object

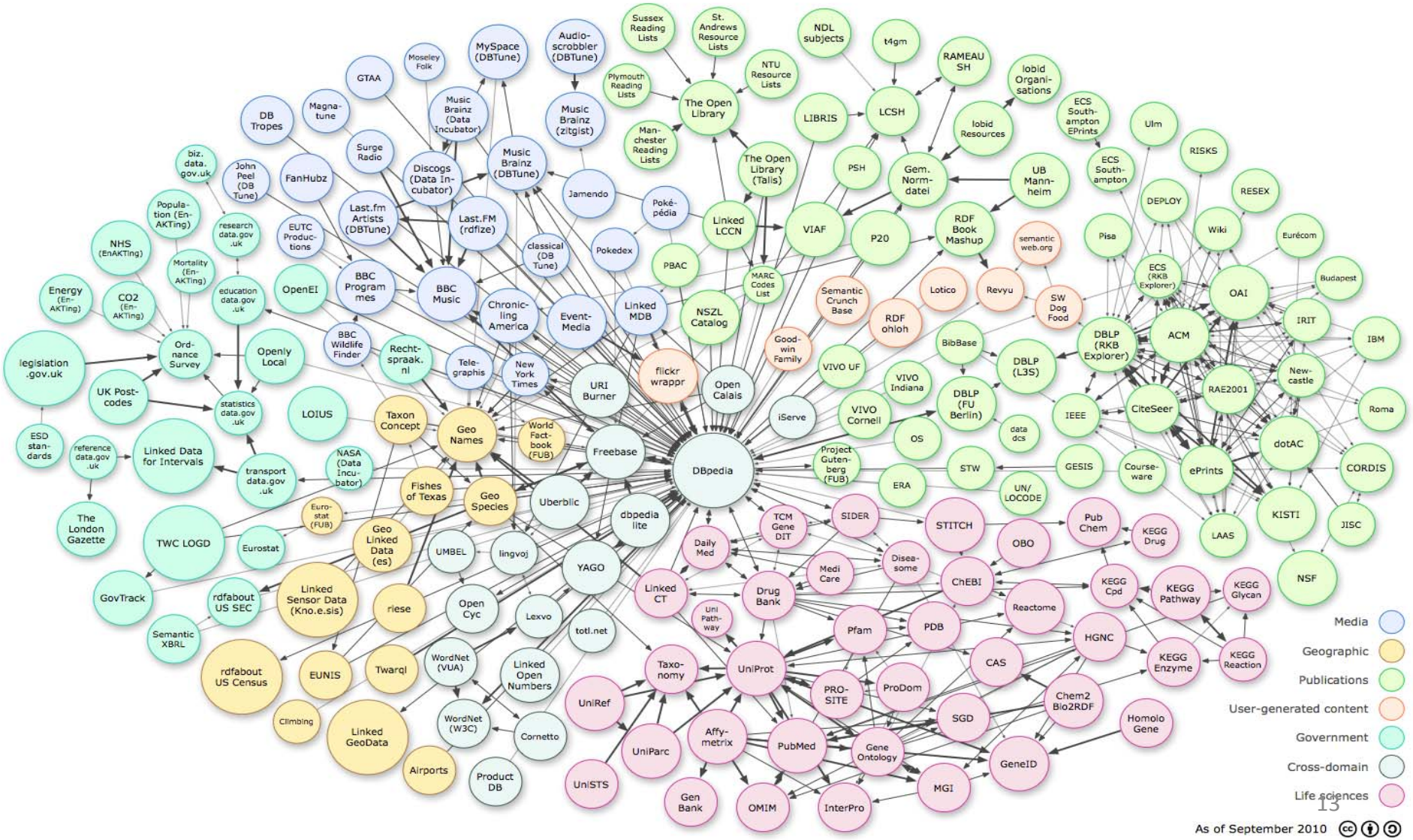
Mark A. Matienzo. Linked Data and Archival Description:
Confluence, Contingencies, and Conflicts
Cornell University Metadata Group, February 19, 2010



An RDF Graph


foaf:firstName
"Mark"
foaf:surname
"Matienzo"
foaf:based_near
http://dbpedia.org/resource/Brooklyn
http://matienzo.org/#me
rdf:type
dbpedia-owl:Place
owl:sameAs
foaf:depiction
http://www.geonames.org/5110302/

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/



Mark A. Matienzo. Linked Data and Archival Description:
Confluence, Contingencies, and Conflicts
Cornell University Metadata Group, February 19, 2010

M. A. Matienzo's conclusions:

- EAD as/and Linked/Linking Data → EAD isn't up for the task
- EAD is document-centeric standard, not a data-centeric standard.
- EAD does not allow elements from other namespaces.
- Attributes are totally insuficient.
- Linking elements do not describe why we make links in a machine-friendly way.

What about archival description as RDF, reusing library-focused ontologies, etc.?

Mark A. Matienzo. Linked Data and Archival Description:
Confluence, Contingencies, and Conflicts
Cornell University Metadata Group, February 19, 2010

Future Work and Improvements:

- Identify subsets/types of relationships that are easy to model in RDF.
- Expose data in public descriptive systems as best we can (e.g. RDFa).
- Add meaningful links to online descriptive apparatuses.
- Revise our descriptive standards to allow better recombination of metadata.
- AND:

Follow the lead of our library and museum colleagues and develop conceptual models of archival description.

(2) IFLA conceptual models and standards in RDF

(1) Representation in RDF

(1) 2006: ISBD Review Group. Material Designations Study Group → ISBD/XML

(2) 2007: FRBR Review Group → FRBR in RDF

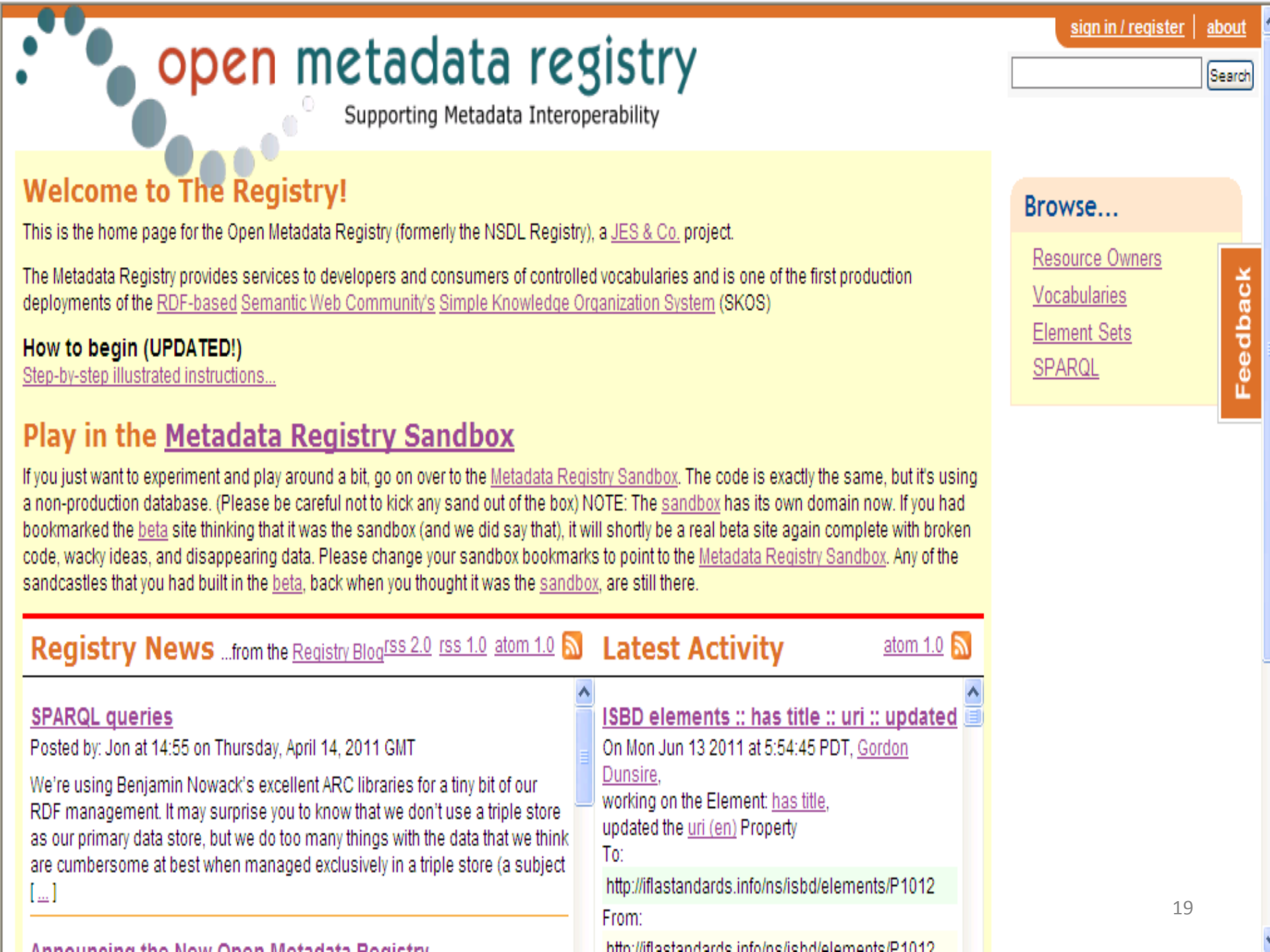
(3) 2008-2011: ISBD Review Group. ISBD/XML Study Group → 3 year project: ISBD in RDF

(4) 2009-: FRBR Review Group → FRAD/FRSAD in RDF

(5) IFLA . Namespaces Task Group. IFLA namespaces - requirements and options

(6) 2011: proposal for UNIMARC/B and UNIMARC/A in RDF

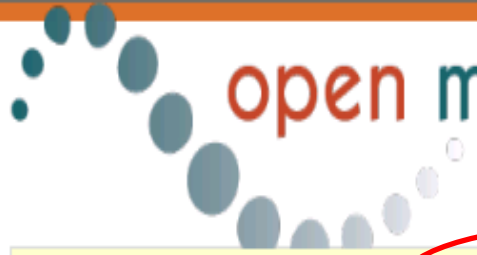
Technical infrastructure: Open Metadata Registry


sign in / register | about
Search
open metadata registry
Supporting Metadata Interoperability
Welcome to The Registry!
This is the home page for the Open Metadata Registry (formerly the NSDL Registry), a JES & Co. project.
The Metadata Registry provides services to developers and consumers of controlled vocabularies and is one of the first production deployments of the RDF-based Semantic Web Community's Simple Knowledge Organization System (SKOS)
How to begin (UPDATED!)
Step-by-step illustrated instructions...
Play in the Metadata Registry Sandbox
If you just want to experiment and play around a bit, go on over to the Metadata Registry Sandbox. The code is exactly the same, but it's using a non-production database. (Please be careful not to kick any sand out of the box) NOTE: The sandbox has its own domain now. If you had bookmarked the beta site thinking that it was the sandbox (and we did say that), it will shortly be a real beta site again complete with broken code, wacky ideas, and disappearing data. Please change your sandbox bookmarks to point to the Metadata Registry Sandbox. Any of the sandcastles that you had built in the beta, back when you thought it was the sandbox, are still there.
Registry News ...from the Registry Blog rss 2.0 rss 1.0 atom 1.0
SPARQL queries
Posted by: Jon at 14:55 on Thursday, April 14, 2011 GMT
We're using Benjamin Nowack's excellent ARC libraries for a tiny bit of our RDF management. It may surprise you to know that we don't use a triple store as our primary data store, but we do too many things with the data that we think are cumbersome at best when managed exclusively in a triple store (a subject [...]
Latest Activity atom 1.0
ISBD elements :: has title :: uri :: updated
On Mon Jun 13 2011 at 5:54:45 PDT, Gordon Dunsire,
working on the Element: has title,
updated the uri (en) Property
To:
http://iflastandards.info/ns/isbd/elements/P1012
From:
Browse...
Resource Owners
Vocabularies
Element Sets
SPARQL
Feedback

Element Sets

Label	Owner	Last Updated	Updated by	Actions
academics	Dapme Passah	2010-11-18	dapme	
Application Profile for Ceramic Resources	Chandana Patra	2010-02-02	chandana	
Application Profile: Ceramic Repository	Chandana Patra	2009-08-20	chandana	
beefschema	beefsteak	2008-09-27	beefsteak	
Dublin Core Element Set	Dublin Core Metadata Initiative	2008-05-29	Jon Phipps	
FOAF (Friend Of A Friend)	Metadata Management Associates	2008-08-05	Jon Phipps	
FRAD model	Gordon Dunsire	2010-02-15	Gordon Dunsire	
FRBR Entities for RDA	Metadata Management Associates	2010-09-17	DianeH	
FRBRer model	Gordon Dunsire	2011-03-18	Gordon Dunsire	
FRBRer open model	Gordon Dunsire	2010-10-04	Gordon Dunsire	
FRSAD model	Gordon Dunsire	2010-07-24	Gordon Dunsire	
ISBD elements	Gordon Dunsire	2010-06-10	Gordon Dunsire	
KACES	Sam Oh	2008-12-17	sam.oh	

Browse...
Resource Owners
Vocabularies
Element Sets
SPARQL



open metadata registry

Supporting Metadata Interoperability

Element Sets: Show detail for FRBRer model

Detail | Elements | History | Maintainers

Label	Type	URI	Status	Updated	Last Updated by	Actions
Work	class	.../ns/fr/frbr/frbrer/C1001	Published	2010-09-30 15:17	Gordon Dunsire	
Expression	class	.../ns/fr/frbr/frbrer/C1002	Published	2011-05-02 6:57	Gordon Dunsire	
Manifestation	class	.../ns/fr/frbr/frbrer/C1003	Published	2011-05-02 6:52	Gordon Dunsire	
Item	class	.../ns/fr/frbr/frbrer/C1004	Published	2010-09-30 15:18	Gordon Dunsire	
Person	class	.../ns/fr/frbr/frbrer/C1005	Published	2010-09-30 15:19	Gordon Dunsire	
Corporate Body	class	.../ns/fr/frbr/frbrer/C1006	Published	2010-09-30 15:19	Gordon Dunsire	
Concept	class	.../ns/fr/frbr/frbrer/C1007	Published	2010-09-30 15:20	Gordon Dunsire	
Object	class	.../ns/fr/frbr/frbrer/C1008	Published	2010-09-30 15:20	Gordon Dunsire	
Event	class	.../ns/fr/frbr/frbrer/C1009	Published	2010-09-30 15:10	Gordon Dunsire	
Place	class	.../ns/fr/frbr/frbrer/C1010	Published	2010-09-30 15:19	Gordon Dunsire	
is realized through	property	.../ns/fr/frbr/frbrer/P2001	Published	2010-09-30 15:07	Gordon Dunsire	
is realization of	property	.../ns/fr/frbr/frbrer/P2002	Published	2010-09-30 15:10	Gordon Dunsire	
is embodied in	property	.../ns/fr/frbr/frbrer/P2003	Published	2010-09-30 15:11	Gordon Dunsire	
is embodiment of	property	.../ns/fr/frbr/frbrer/P2004	Published	2010-09-30 15:11	Gordon Dunsire	
is exemplified by	property	.../ns/fr/frbr/frbrer/P2005	Published	2010-09-30 15:13	Gordon Dunsire	
is exemplar of	property	.../ns/fr/frbr/frbrer/P2006	Published	2010-09-30 15:12	Gordon Dunsire	
is created by (corporate body)	property	.../ns/fr/frbr/frbrer/P2007	Published	2010-09-30 15:12	Gordon Dunsire	

Browse...

Resource Owners
Vocabularies
Element Sets
SPARQL

Feedback

Element Sets: FRBRer model
Elements: Person

Detail | Statements | History

Metadata +

Detail

Label:	Person
Name:	Person
URI:	http://iflastandards.info/ns/fr/frbr/frbrer/C1005 (RDF)
Description:	An individual.
Comment:	
Type:	class
Status:	Published
Language:	English
Note:	

List | Get RDF

sign in / register | about

Search

Element Sets: FRAD model
Elements: Family

Detail | Statements | History

Metadata +

Detail

Label:	Family
Name:	Family
URI:	http://iflastandards.info/ns/fr/frad/C1002 (RDF)
Description:	Two or more persons related by birth, marriage, adoption, civil union, or similar legal status, or who otherwise present themselves as a family.
Comment:	
Type:	subclass
Parent:	Bibliographic Entity
Status:	New-Proposed
Language:	English
Note:	

Browse...

- Resource Owners
- Vocabularies
- Element Sets
- SPARQL

Feedback

List | Get RDF

(2) IFLA models and standards in RDF

Questions for interoperability or linking classes and properties of ALM models/standards represented in RDF:

- Namespaces: re-use of exisitng namespaces or represent "own" elements and link to existing classes & properties
- Classes [Subject of the RDF triple] & Properties [relationship or predicate in RDF]
- Controlled vocabularies: VIAF, LCSH, DDC, etc.

→ Argument (#4): representing models/standards in RDF & liaising with similar research in the field by other interested parties **informed back** the the reconsideration and possible development of IFLA standards themselves & their representations in RDF.

(Some) Background Papers

Bizer, Christian; Tom Heath, Tim Berners-Lee. Linked data – the story so far. // International Journal on Semantic Web and Information Systems (IJSWIS), vol. 5, issue 3. (2009). Pre-print available at: http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf

Dunsire, Gordon. UNIMARC, RDA and the Semantic Web. // International Cataloguing and Bibliographic Control (ICBC), vol. 39, no. 2 (April/June 2010). Based on a paper presented to the World Library and Information Congress: 75th IFLA General Conference and Assembly, 23-27 August 2009, Milan, Italy; available at: www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf

Dunsire, Gordon; Mirna Willer. Standard library metadata models and structures for the Semantic Web. // Library hi tech news, vol. 28, no. 3. (2011), pp. 1-12. Available at: http://dx.doi.org/10.1108/07419051111145118

Dunsire, Gordon; Mirna Willer. UNIMARC and linked data. // World Library and Information Congress: 77th IFLA General Conference and Assembly, 13-18 August 2011, San Juan, Puerto Rico. Available at: http://conference.ifla.org/sites/default/files/files/papers/ifla77/187-dunsire-en.pdf

Willer, Mirna; Gordon Dunsire, and Boris Bosančić. ISBD and the Semantic Web. // JLIS.it: Journal of Library and Information Science. Italy, vol. 1, no. 2, (2010) pp. 213-236. Available at: http://dx.doi.org/10.4403/jlis.it-4536

Discussion & comments from the archival community