Metadata's Ontological Turn

presented by John Huck at Code4Lib Alberta, November 25th, 2016

Dublin Core & RDF: Metadata chums

Metadata: "Data about data"

- Structured descriptive information about a (digital) resource
- Standardized through schemas, profiles
- Often harvested from one system for re-use in another
- Different kinds of metadata may support different tasks:
  - Search and discovery
  - Evaluating resources for use (characteristics, access conditions)
  - Long term preservation

Resource Description Framework (RDF)

- *"Key industry players collaborate to develop **interoperable metadata for the web**"* (World Wide Web Consortium 1997)
- A schema-neutral data structure, where the **triple** is the basic unit

subject	**predicate**	**object** .
<someResource>	<hasRelationTo>	<anotherResource> .
<someResource>	<hasProperty>	"a literal value string" .

1995 - 2000

	DCMI	W3C	RDF	RDF Schema
1995	1st Workshop			
1997	RDF draft presented in Helsinki	"Metadata Activity" begins	1st Working Draft	
1998				1st Working Draft
1999	Element Set		RDF	
2000	Terms Set			

Grand visions of the semantic web

The semantic web (Berners-Lee, et al. 2001)

- XML, RDF, and ontologies as the "basic components" of the semantic web.

"For web researchers" an ontology is comprised of *"a taxonomy and a set of inferencing rules. The taxonomy defines classes of objects and relations among them."*

"Adding logic to the web—the means to use rules to make inferences, choose courses of action and answer questions— is the task before the Semantic Web community at the moment."

2001 - 2005

	DCMI	W3C	RDF	RDF-S	OWL
2001	First proposal for DC in RDF	"Semantic Web Activity" begins			Working Group begins
2002					1st working draft
2003					
2004			RDF (Revised)	RDFS	OWL
2005	DCMI Abstract Model				

How do ontologies work in RDF Schema?

A **predicate** may be given a **domain**, a **range**, or both:

```
wikidata:Q2806736   ex:playsFor     wikidata:Q205973  .
ex:playsFor         rdfs:domain     ex:Athlete  .
ex:playsFor         rdfs:range      ex:Team  .
```

And new **classes** can be defined as **sub-classes** of existing classes:

```
ex:Athlete          rdfs:subClassOf   foaf:Person  .
```

What about Web Ontology Language (OWL)?

- More complex structures for defining things like value restrictions and cardinality
  - e.g., owl:Restriction class
- More sophisticated ways of defining properties
  - e.g., owl:InverseFunctionalProperty, owl:SymmetricProperty

Linked data: Vision correction

Linked data (Berners-Lee, 2006)

"A surprising amount of data isn't linked in 2006, because of problems with one or more of the steps."

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 (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

2006 - 2014

	DCMI	W3C	RDF	RDF-S	OWL	SKOS
2006						WG begins
2008	DC in RDF + Domains & Ranges					1st Working Draft
2009					OWL 2	SKOS
2013		"Data Activity" begins				
2014			RDF 1.1	RDFS 1.1		

Meanwhile, back in metadata-land

Slow boil?

"What do RDF, Linked Data, and Semantic Web have to do with the majority of current implementers of digital collections? ***Not much, if anything, at the present time,*** *other than having an awareness of the basic concepts and vision."* (Miller 2011, p. 321)

Movin' on up

"LOD [Linked open data] allows for the structured metadata created and maintained by LAM [libraries, archives, museums] institutions to be shared in such a way that the general community can interact and enrich the data." (Yoose & Perkins 2013)

Developing vocabularies for metadata

	BIBFRAME	DDI-RDF Discovery Vocabulary (Disco)
2011		Work Began
2012	BIBFRAME 1.0 Released	
2013		First Draft Released
2014		
2015		Latest Draft Released
2016	BIBFRAME 2.0 Released	

Lessons from BIBFRAME 1.0

- Reuse existing vocabularies, especially from the core RDF standards (Sanderson 2015)
- Properties are meant to be reused throughout an ontology: can't be limited to a specific class (Baker, et al. 2014)
- OWL axioms are not data validation constraints (Baker, et al. 2014)
- Mellon funded LD4L (Linked Data for Libraries) is currently developing a bibliographic ontology

DDI-RDF Discovery Vocabulary (Disco)

- Contributions from 26 people from 12 countries (Vompras, et al. 2013)
- Input from domain experts, semantic web experts, DDI experts
- Implements only a small subset of the DDI metadata standard
- Designed to support specific search and discovery use cases
- Optimized for SPARQL queries
- Re-uses standard semantic web vocabularies

Minimal ontological commitment

"*An ontology should make as few claims as possible about the world being modeled, allowing the parties committed to the ontology freedom to specialize and instantiate the ontology as needed.*" (Gruber in Baker, et al. 2013)

- Design of SKOS followed this principle
- "Only define what matters" (Sanderson 2015)
- "an RDF vocabulary is more reusable the fewer constraints it defines" (Baker, et al. 2014)

Anthropology had an ontological turn too!

Thank you!

john.huck@ualberta.ca
Twitter: @jhuck_AB

References & Resources Consulted

Allemang, D., & Hendler, J. (2011). *Semantic web for the working ontologist: Effective modeling in RDFS and OWL* (2nd ed.). Amsterdam: Morgan Kaufmann.

Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., & Summers, E. (2013). Key choices in the design of Simple Knowledge Organization System (SKOS). *Web Semantics: Science, Services and Agents on the World Wide Web*, *20*, 35-49. http://dx.doi.org/10.1016/j.websem.2013.05.001

Baker, T., Coyle, K., & Petiya, S. (2014). Multi-entity models of resource description in the semantic web: A comparison of FRBR, RDA, and BIBFRAME. *Library Hi Tech, 32*(4), 562-582. DOI:10.1108/LHT-08-2014-0081. Open access preprint retrieved from http://kcoyle.net/LHTv32n4preprint.pdf

Berners-Lee, T. (2006). *Linked data*. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*(5), 34-43.

Bosch, T., Cyganiak, R., Gregory, A., Wackerow, J. (2013). DDI-RDF discovery vocabulary: A metadata vocabulary for documenting research and survey data. 2013. *Proceedings of the WWW2013 Workshop on Linked Data on the Web*. Retrived from http://ceur-ws.org/Vol-996/papers/ldow2013-paper-12.pdf

Coyle, K. (2016). *FRBR before and after: A look at our bibliographic models.* Chicago: ALA Editions. Retrieved from http://kcoyle.net/

Golub, A. (2013, November 27). Ontology as the major theme of AAA 2013. Retrieved November 23, 2016, from http://savageminds.org/2013/11/27/ontology-as-the-major-theme-of-aaa-2013/

Henare, A., Holbraad, M., & Wastell, S. (2007). Introduction: Thinking through things. In A. Henare, M. Holbraad & S. Wastell (Eds.), *Thinking through things: Theorising artefacts ethnographically* (pp.1-31). London: Routledge.

Lowe, E.J. (1995). Ontology. In T. Honderich (Ed.), *The Oxford companion to philosophy* (p. 634). Oxford: Oxford University Press.

Miller, S. (2011). Metadata for digital collections. New York: Neal-Schuman.

Paleček, M., & Risjord, M. (2012). Relativism and the ontological turn within anthropology. *Philosophy of the Social Sciences, 43*, 3-23. doi:10.1177/0048393112463335. Pre-print retrieved from http://lms.ff.uhk.cz/pool/download_16.pdf

Sanderson, R. (2015). Analysis of the BIBFRAME Ontology for Linked Data Best Practices. Prepared for: Office of Network Standards, The Library of Congress. Retrieved from https://docs.google.com/document/d/1dIy-FgQsH67Ay0T0O0ulhyRiKjpf_I0AVQ9v8FLmPNo/edit?usp=sharing

van Hooland, S., & Verborgh, R. (2014). *Linked data for libraries, archives and museums*. Chicago: Neal-Schuman.

Vompras, J., Gregory, A., Bosch, T., Capadisli, and Wackerow, J. (2013). Scenarios for the DDI-RDF discovery vocabulary. DDI Working Paper Series – Semantic Web 2. https://doi.org/10.3886/DDISemanticWeb02

World Wide Web Consortium publishes public draft of Resource Description Framework (RDF). (1997, October 3). World Wide Web Consortium. Retrieved from https://www.w3.org/Press/RDF

Yoose, B., & Perkins, J. (2013). The linked open data landscape in libraries and beyond. *Journal of Library Metadata*, *13*(2-3), 197-211. http://dx.doi.org/10.1080/19386389.2013.826075

Zeng, M., & Qin, J. (2008). *Metadata*. New York: Neal-Schuman.

DDI-RDF Discovery Vocabulary: A vocabulary for publishing metadata about data sets (research and survey data) into the Web of Linked Data. Unofficial Draft 26 March 2015 http://rdf-vocabulary.ddialliance.org/discovery.html

Resources about the standards

https://www.w3.org/ (various pages)

http://dublincore.org/

https://en.wikipedia.org/wiki/Resource_Description_Framework

https://en.wikipedia.org/wiki/RDF_Schema

https://en.wikipedia.org/wiki/Web_Ontology_Language

https://en.wikipedia.org/wiki/Simple_Knowledge_Organization_System

Presentation Abstract

Metadata's ontological turn

Ever since continental philosophy took its "linguistic turn" at the beginning of the 20th Century, "turns" have been declared whenever a set of preoccupations comes to dominate the discourse and practice of a given field and alters its prevailing course. The "spatial turn" in the humanities is one example.

The library world is in the midst of a long march towards a new linked data and semantic web environment, as evidenced by such major undertaking as the development of an RDF encoding for library data, BIBFRAME. Authority files for names, topics, places, works and other entities in the bibliographic world have been converted to linked data vocabularies. "Things not strings" is the current mantra.

Whether one speaks of linked data or the semantic web, the organizing principle that lies beneath all of this is that of ontology. In philosophy, ontology may be characterized as the study of what kinds of things there are in the world and their relations. While it is common to refer to a set of linked data predicates as an ontology in the same way that you might talk about a metadata schema,

in fact, when the predicates are modelled in RDFS, SKOS or OWL, and are then employed in triples, they impose an ontological structure on the entities that they organize, and the resulting graph constitutes an instantiated ontology of some neighbourhood of human knowledge upon which formal logical operations may be performed.

What I am calling this ontological turn in metadata raises many questions:

- What is the theoretical relationship between information and ontology?
- Is the ontological modelling of information inevitable?
- Is ontology a more advanced form of metadata?
- What implications does it have for metadata practices?
- What practical constraints are entailed by an ontological framework?
- What types of information are well suited to an ontological framework and what types are not?

This presentation explores some of the questions raised, making observations on vocabularies like Dublin Core, drawing on commentary from the LIS community, and makes a comparison with another field that has recently undergone an "ontological turn," namely Anthropology.