

Instantiation: metadata models, rationales and realities for knowledge organization

NASKO/Toronto. June 17, 2011

classification concept concerned considered culture databases
different disciplines documents domain example field fjordback
form hjerland http important indexing information journal
knowledge ko library lis literature order
organization principles recall records related research
science scientific sense social sondergaard subject system term
traditions view

Tag-cloud from Hjørland's
Lifeboat, 2007

Jane Greenberg (janeg@email.unc.edu)
Metadata Research Center/School of Info. + Lib. Sci.
University of North Carolina at Chapel Hill

Overview

1. Present some questions
2. Models, metadata models
 - Introduce the problem
 - Context ~ DCAM and RDF very briefly
3. Life-cycle modeling as a framework
4. Eat some of our own dog-food!
5. Have dialog on advancing the study instantiation, metadata models..



Questions

Assumption: We should study instantiation

1. Are we doing a good job as a community studying instantiation? Why/why not? What are the obstacles, if any?
2. How might we advance our approach the study of instantiation?
3. What methods are useful for studying instantiation?
4. If you had all the funding in the world, how might you proceed with studying instantiation?



Models

- Help us to conceptualize
 - Provide a map
 - Provide a means of doing something, output
 - Can serve as guidelines
-
- What is a model in your life?
 - Name a model that impacts your every day
-
- Models can be ambiguous, personal, etc.

How do researchers evaluate said models?

...



How do researchers evaluate said models?

- Sit around a table and discuss,... and...discuss more?
- Compare, but how?
- User studies (Does it make work?)
- Experiment (with and w/out the system), measure outcomes (**productivity, ROI**)
- Gather data...



Metadata models..

... carrier, container

Metadata models ~ interoperability

Semantic Interoperability

- **Feature:** Definition/meaning, vocabulary
- **Facilitated by:** Metadata registries, application profiles, and mrosswalks

Structural Interoperability

- **Feature:** Modeling/frameworks, packaging
- **Facilitated by:** RDF, FRBR, CIDOC, METZ, OAIS, DCAM

Syntactical Interoperability

- **Feature:** Encoding
- **Facilitated by:** X/HTML, XML, MARC, punctuation (ISBD)

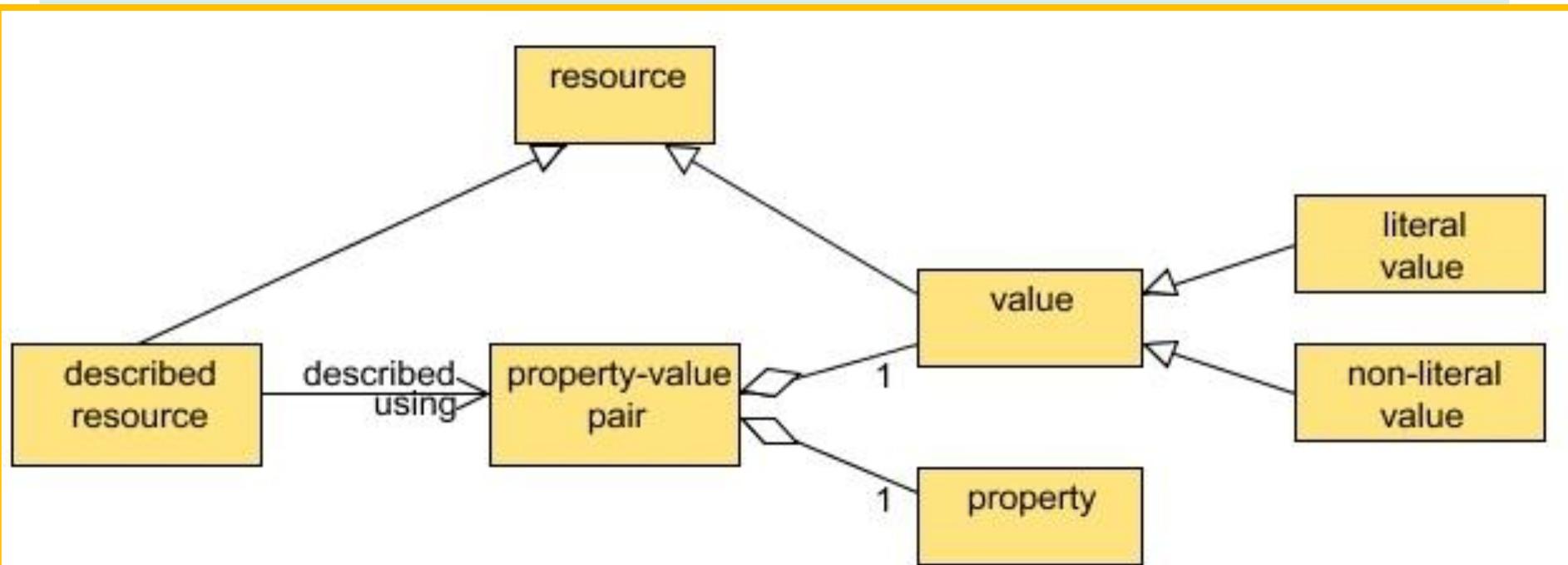
Metadata models predicated on ...

**Resource
= Object**

- W3C/IETF definition of resource is
 - “...anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other *resources*. Not all *resources* are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered *resources*.”
- i.e. a *resource* is “anything”
 - physical things (books, cars, people)
 - digital things (Web pages, digital images)
 - conceptual things (colours, points in time, subjects)

DCMI abstract model

- A *description* = one or more *statements* (about one, only one, *resource*)
 - *Statement* = a *property-value pair*
- Each *property-value pair* is made up of one *property* and one *value*.
- Each *value* is either a *literal value* or a *non-literal value*:
 - A *literal value* is a *value* which is a *literal*
 - A *non-literal value* is a *value* which is a physical, digital or conceptual entity



Review: RDF Triples



`http://www.example.org/index.html` has a **creator** whose value is **Sally Smith**

The RDF terms for the various parts of the statement are:

- the *subject* is the URL `http://www.example.org/index.html`
- the *predicate* is the word "creator"
- the *object* is the phrase "Sally Smith"

How do researchers (“we”) evaluate said models?

1. Sit around a table and discuss,...
and...discuss more? Theorize? Sampling...
lot of shoe-horning (fitting in)
2. Compare...
some of...*my model is better than your model*
3. User studies, does this make sense?
4. Experiment (with and w/out the system),
measure outcomes (productivity, ROI)



Life-cycle modeling for studying instantiation



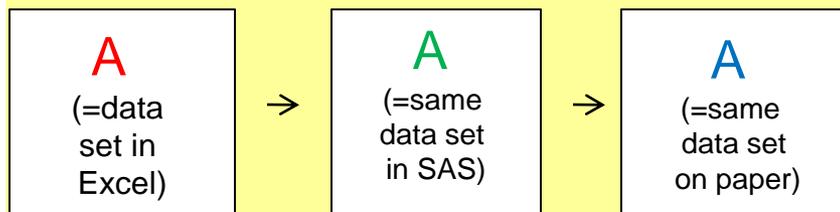
Life-cycle as a framework for study

- Dryad – applied and basic bio-sciences
 - **Data-reuse/re-purpose** (What is data, or the data object?)
- Lifecycle-modeling a *theoretical framework* (Greenberg, 2009)
 - Faceted term mapping experiment, 12 vocabularies 600 terms; survey 400 evolutionary biologists
 - Demonstrated I-s-m **automatic propagation, metadata inheritance**, and **value system adoption** (living organisms adopt value)
- Data-life cycle study, 22 biologist
 - Multi-method, scenario assessment; survey

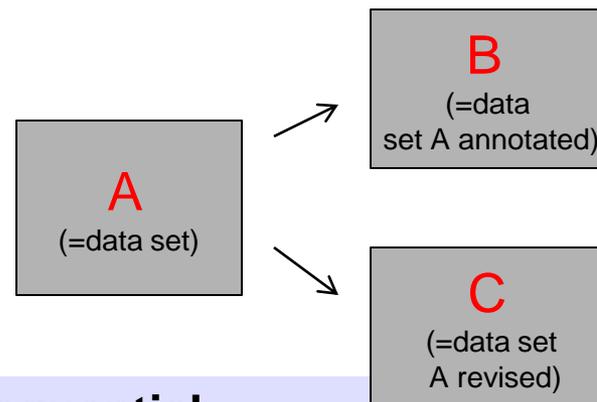
Bibliographic relationship (Smiraglia, 1999,.....n; Coleman, 2002; Tillett, 1992, 1992; DCMI relationship, DataCite)

Data object relationships

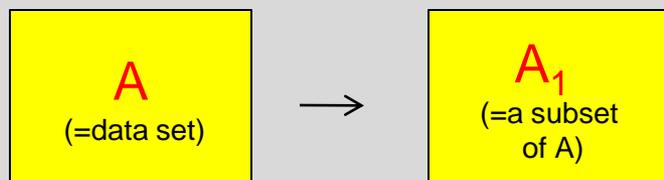
Equivalence



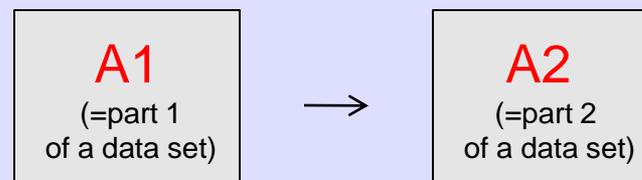
Derivative



Whole-part



Sequential



Instantiation

Scenario: Sherry collects data on the survival and growth of the plant *Borrichia frutescens* (the bushy seaside tansy)... back at the lab she enters the exact same data into an excel spreadsheet and saves it on her hard drive.

Question: What is the relationship between Sherry's paper data sheet and her excel spreadsheet?

Answer: Equivalent | Derivative | Whole-part | Sequential
(circle one)

Findings (22 participants)

- In general, more seasoned scientists better grasp
- Sequential data presented the most difficulty (less seasoned sci.)
- Unanimous support: “very → extremely important”

Instantiation example 2

Each year Tom, a staff member at the Georgia Coastal Marine Lab, collects data on the abundance of marine invertebrates at several study sites in salt marsh habitats on Sapelo Island, GA. These data are released annually as part of the Georgia Coastal Ecology Long Term Ecological Research Program.

Question: What is the relationship among each data set released annually?

Answer: Equivalent | Derivative | Whole-part | Sequential
(circle one)

DataCite, ver 2.1

http://www.datacite.org/schema/DataCite-MetadadataKernel_v2.1.pdf

- dcterms:relation
- dcterms:conformsTo:
- dcterms:isReferencedBy
- dcterms:references
- dcterms:isVersionOf
- dcterms:hasVersion
- dcterms:isFormatOf
- dcterms:hasFormat
- dcterms:isPartOf
- dcterms:hasPart
- dcterms:isReplacedBy dcterms:replaces
- dcterms:source



Overview

1. Present some questions
2. Models, metadata models
 - Introduce the problem
 - Context ~ DCAM and RDF very briefly
3. Life-cycle modeling as a framework
4. Eat some of our own dog-food!
5. Have dialog on advancing the study instantiation, metadata models..



Explore the relationship between original data set (A) and data set generated (B)?

1. Jim has a fairly extensive data set on *Drosophila bifurca* (fruit fly), gathered over a three year period, and organized by organ system (e.g., adipose, circulatory, excretory, muscle, digestive, etc.). He wants to make this entire dataset (A) accessible to students who are learning about life-cycle stages. Before doing this, he inserts image links to FlyBase, and also makes several revisions, updates, and annotation to the data, and then make the data available (B).
2. Margaret studies the relationship between social tagging and controlled vocabulary. Louise sends Margaret the excel spreadsheets from her recent study on tags and retrieval effectiveness in OPACs (A). Margaret uses the set of tags Louise collected conduct a new study--mapping the tags to LCSH, and produced a new data set (B).
3. **For you/your group:** Consider a scenario and a relationship, real or imagined, and identify the relationship illustrating data instantiation?

Questions

Assumption: We should study instantiation

1. Are we doing a good job as a community studying instantiation? Why/why not? What are the obstacles, if any?
2. How might we advance our approach the study of instantiation?
3. What methods are useful for studying instantiation?
4. If you had all the funding in the world, how might you proceed with studying instantiation?
5. Other... KO...aspects?

