Linking Entities in Scientific Metadata

Jian Qin
Syracuse University
jqin@syr.edu

Miao Chen
Syracuse University
mchen14@syr.edu

Xiaozhong Liu
Syracuse University
xliu12@syr.edu

Andrea Wiggins
Syracuse University
awiggins@syr.edu

Abstract

Linked entity data in metadata records builds a foundation for the Semantic Web. Even though metadata records contain rich entity data, there is no linking between associated entities such as persons, datasets, projects, publications, or organizations. We conducted a small experiment using the dataset collection from the Hubbard Brook Ecosystem Study (HBES), in which we converted the entities and their relationships into RDF triples and linked the URIs contained in RDF triples to the corresponding entities in the Ecological Metadata Language (EML) records. Through the transformation program written in XML Stylesheet Language (XSL), we turned a plain EML record display into an interlinked semantic web of ecological datasets. The experiment suggests a methodological feasibility in incorporating linked entity data into metadata records. The paper also argues for the need for change in the scientific as well as the general metadata paradigm.

Keywords: scientific metadata; ecological data; metadata for data sets

1. Research Problem

The term scientific metadata is often used to refer to the data describing the datasets collected or generated from scientific research. A large number of scientific metadata standards and conventions exist in major disciplinary fields, e.g., the Content Standards for Digital Geospatial Metadata (CSDGM, http://www.fgdc.gov/metadata/csdgm/), Ecological Metadata Language (EML, http://knb.ecoinformatics.org/software/eml/), and Darwin Core. Description of scientific datasets proves to be extremely challenging due to their complexities. The metadata schemas include not only entities responsible for data collection, processing, and distribution, but also data for assessing the applicability, quality, and accuracy of a dataset. Information on data files is necessary for user access or physically reading the values in a dataset, which should support the sharing and exchange of data stored in differing physical format and between communities (Gritton et al., 1995). These requirements for scientific metadata lead to standards or schemas that often contain hundreds of metadata elements.

While it is vital for scientific metadata to allow for dataset identification, quality assessment, verifiability, and dissemination, the large, complex metadata standards also create problems for metadata generation. One such problem is duplicated data entry for some entities within the same record or across records. The term *entity* or *entities* in this paper refers to persons, organizations, projects, datasets, subject fields, and publications. In CSDGM, for example, elements for capturing person and institution information appear in at least five of its seven sections at least once. If any of these entities undertook more than one role, e.g., the same entity was both originator and contact, the same entity data would be entered more than once. Similar practice can be found in other scientific metadata standards. Although some metadata editing tools can reduce the repetition in data entry, it does not scale effectively.

Duplicate entity data entry in scientific metadata creates another problem—disconnected entities. Conventional metadata practice follows a workflow that starts with defining a schema, and then develops a data entry interface that will send the data entered to a relational database or

XML record files. In scientific metadata, the entity data is generally embedded in records and requires special programming if any association between an entity and all datasets related to this entity needs to be established. When a user searches for datasets, she or he has to rely on the search options made available by the search interface and her or his familiarity with the subject and datasets to retrieve relevant datasets. Searching for relevant datasets or all related information can be even more difficult if the user knows little about the datasets or the subject domain except that s/he only knows the data needed is in the repository.

The lack of interlinking between entities and datasets not only causes the same entity data to be entered repeatedly, which slows down metadata generation and makes it forever lag behind research data growth, but also affects the use by people who are not the creator nor expert on the dataset topics but need to find and use them. Data-intensive science expects to "move beyond data warehouses and closed systems" and "allow access to data to those outside the main project teams, allow for greater integration of sources, and provide interfaces to those who are expert scientists but not experts in data administration and computation" (Fox & Handler, 2009, p. 147). Accomplishing this goal for eScience requires adding semantics to research datasets. The semantics in this context, according to Fox and Handler (2009), include well-defined and machine-encoded concepts and terms as well as interrelationships among them. Entity data, as an important part of the semantics in eScience, would be a relatively low entry point for building an eScience semantic web.

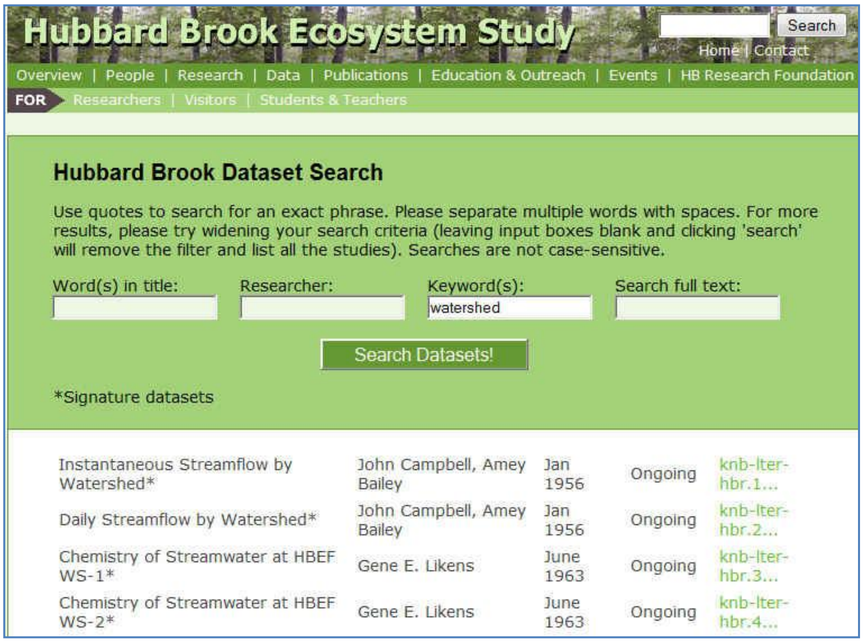
New advances in semantic web technologies have provided a fertile ground for interlinking scientific data and metadata. Linked data, a concept proposed by Tim Berners-Lee (Berners-Lee, 2009), has gained a wide acceptance in the last couple of years. It is defined as the practice of connecting data that was not previously linked. Using Uniformed Resource Identifiers (URIs) and Resource Description Framework (RDF), data may be linked through exposing, sharing, and connecting pieces of data, information, and knowledge (Wikipedia, 2010). Linked data as an emerging semantic web technology has promising applications for scientific metadata in both enhancing metadata creation effectiveness and promoting smart resource discovery.

Numerous linked data sets have been published so far (W3C, 2010) but not many applications have been reported. The purpose of this study is not simply to add another linked data set to the vast existing and still growing collections, but rather, to apply the linked data to metadata generation and resource discovery through an experiment with a small ecological dataset collection. The experiment focuses on the question of how we can build an interlinking network of researchers, institutions, projects, datasets, and publications in a domain, and more importantly, how we can associate the linked data with metadata. In the following sections, we will describe 1) the ecological dataset collection and the entity database we built based on the datasets, 2) the Ecological Metadata Language (EML) and requirements for ecological metadata, and 3) an experiment constructing the RDF data set with the relational database and linking it to the metadata records. We will also discuss the implications of this project for ecological metadata generation as well as for scientific metadata in general.

2. The Ecological Dataset Collection

The Hubbard Brook Ecosystem Study (HBES, http://hubbardbrook.org/) is one of the long-term ecological research sites around the country. Within this 3,160 hectare reserve, the Hubbard Brook Experimental Forest offers a great potential for multifaceted ecosystem research. Launched in 1960, HBES has six principal organizational partners (USDA Forest Service, Cornell, Dartmouth, Syracuse, Yale, the Institute of Ecosystem Studies (IES), and the U.S. Geological Survey) and 10 other organizational participants. Since 1963, approximately 2000 publications have been produced through HBES and over 300 datasets are made available on its website (Hornbeck, 2001). Many datasets that started in the 1950's and 1960's are still ongoing. The HBES dataset collection has a simple search engine (Figure 1) with options to search by title, researcher, keyword, or full text. The result display shows title, investigator(s), date, status, and a link to view detailed metadata.

In 2009 we collected data about projects, persons, publications, subject interests, and datasets from HBES website. Person and project information was verified against the Long-Term Ecological Research (LTER) Directory (http://search.lternet.edu/dir.php) when necessary. Our original goal was to study the interactions among the scientists, students, publications, research fields, and datasets in the HBES community (a work in progress), but found it to be also a perfect case for studying the methodology for turning relational databases into a semantic web for scientific metadata because of its relatively small size and ingredients available for a methodological exploration.


Hubbard Brook Ecosystem Study
Search
Home | Contact
Overview | People | Research | Data | Publications | Education & Outreach | Events | HB Research Foundation
FOR Researchers | Visitors | Students & Teachers
Search Datasets!

Hubbard Brook Dataset Search

Use quotes to search for an exact phrase. Please separate multiple words with spaces. For more results, please try widening your search criteria (leaving input boxes blank and clicking 'search' will remove the filter and list all the studies). Searches are not case-sensitive.

Word(s) in title:
Researcher:
Keyword(s): watershed
Search full text:

*Signature datasets

Instantaneous Streamflow by Watershed*	John Campbell, Amey Bailey	Jan 1956	Ongoing	knb-lter-hbr.1...
Daily Streamflow by Watershed*	John Campbell, Amey Bailey	Jan 1956	Ongoing	knb-lter-hbr.2...
Chemistry of Streamwater at HBEF WS-1*	Gene E. Likens	June 1963	Ongoing	knb-lter-hbr.3...
Chemistry of Streamwater at HBEF WS-2*	Gene E. Likens	June 1963	Ongoing	knb-lter-hbr.4...

FIG. 1. Dataset collection search interface at HBES (http://hubbardbrook.org/data/dataset_search.php)

3. Metadata for the Datasets

The datasets at HBES are described using the Ecological Metadata Language (EML), a metadata specification for the ecology discipline. EML is structured in modules (Figure 2) and each module is defined by an XML schema. The EML root is a wrapper that encapsulates all metadata content in a single EML document. An EML record contains one of the four modules immediately below the root and can import one or more of the sibling modules, supporting modules, data organization modules, or entity types to add more details about the dataset. EML allows for reuse of elements and data through references. For example, a project module from the supporting module group can be referenced in the dataset module to provide the larger context in which the dataset was created. Although referencing between modules and elements allows for reuse of element definitions and can save time in data entry, it does not automatically establish interlinking between entities without special programming.

The XML formatted metadata records offer many advantages for data administration and presentation tasks. Its structures and encoding lay down the foundation for scientific metadata to go beyond “data warehouses and closed systems.” Current EML records can be transformed from XML format to HTML or XHTML format by using the schema-based programs written in XML Stylesheet Language (XSL). For instance, entity data such as creator, organization, and dataset (Box 1) is encoded with EML tags, but in the browser the names are not linked to anywhere. This lack of interlinking makes it particularly cumbersome and difficult to locate other datasets from the same or related projects, or persons associated with these datasets or projects without leaving the record on display to initiate a new search.

Long-term ecology data contains evidence of environmental influences on ecosystem changes and has great value for interdisciplinary research and policy making. Providing interlinked metadata for ecological datasets and scientific research datasets in general would provide easier access to datasets for non-subject expert users to obtain information on who were involved in which research projects, what publications were the results from which projects and datasets, and what projects were associated with which subjects. In order to accomplish this, scientific metadata must include:

- URI-identified entities;
- Relationships between these entities; and
- Relationships between the entities and metadata records.

The relational database we have built contains the mappings between the entities related to the 300+ datasets available on the HBES website. While the database content is not RDF, it is possible to convert the entities and relationships into RDF triples by using computer programs. Among the tools we examined, we found D2R server (http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/) suitable for our need. D2R server facilitates the transformation of relational database data into RDF triples and publishes the results on the semantic web. "D2R Server uses a customizable D2RQ mapping to map database content into this format, and allows the RDF data to be browsed and searched – the two main access paradigms to the Semantic Web" (Bizer & Cyganiak, 2009). This feature of D2R server makes it an ideal tool for our purpose—converting the content of a relational database into RDF triples. If we incorporate the RDF triples with EML metadata records, not only can the entities and relationships be browsed and searched, but a new venue for building a semantic web for domain specific scientific metadata is enabled.

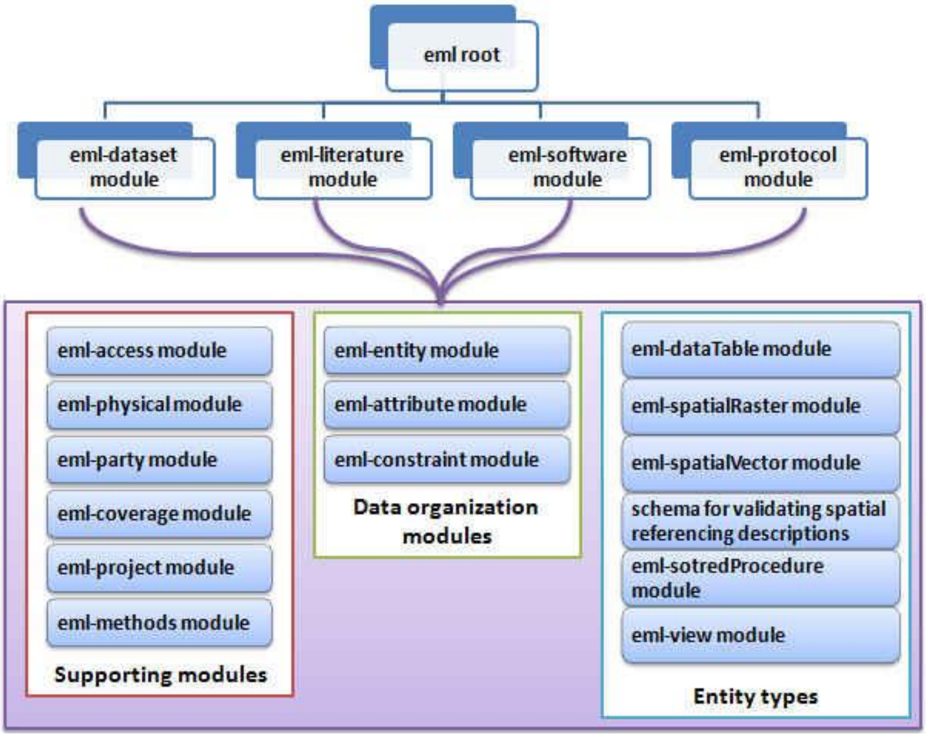
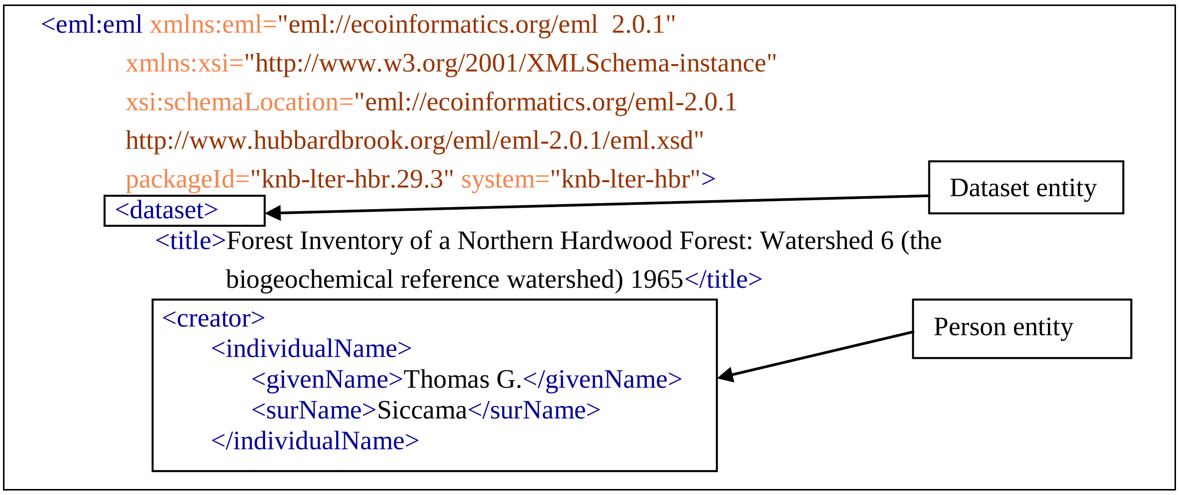


FIG. 2. EML structure and modules

4. Experiment

The purpose of this experiment was to link the entity data in a relational database with EML metadata records by converting the entity data into RDF triples, or linked data, for the HBES datasets. The methodological procedures used in this experiment would be useful for larger scale application.

Box 1: Entity data examples in an XML coded metadata record

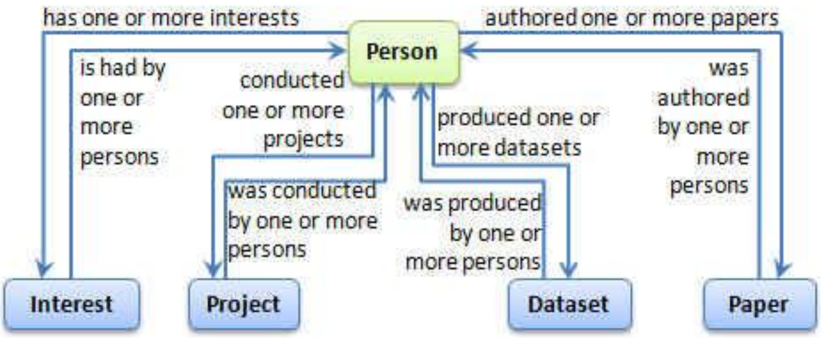


```
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml  2.0.1"
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:schemaLocation="eml://ecoinformatics.org/eml-2.0.1
         http://www.hubbardbrook.org/eml/eml-2.0.1/eml.xsd"
         packageId="knb-lter-hbr.29.3" system="knb-lter-hbr">
   <dataset>
      <title>Forest Inventory of a Northern Hardwood Forest: Watershed 6 (the
            biogeochemical reference watershed) 1965</title>
      <creator>
         <individualName>
            <givenName>Thomas G.</givenName>
            <surName>Siccama</surName>
         </individualName>
```

4.1. Preparing Data

Two sets of data were critical for the experiment: 1) entities and their relationships and 2) EML records in XML format. The relational database for HBES contains five entities—person, subject interest, project, dataset, and paper and the four entities at the bottom of Figure 3 have a many-to-many relationship with the person entity. In fact, more relationships could have been established between project and dataset, project and paper, etc. Since the dataset collection was small enough for this experiment, we concentrated only on the relationships indicated in Figure 3. Other relationships (e.g., project→dataset or vice versa) could be derived indirectly for such a small collection of datasets, e.g., project-dataset relationship could be derived through person entity.

URIs obtained from the database to RDF conversion process were embedded in the entity elements with matching values. We downloaded all 126 EML records available on the HBES site, which include both metadata and dataset(s).



has one or more interests
authored one or more papers
Person
is had by one or more persons
conducted one or more projects
produced one or more datasets
was authored by one or more persons
was conducted by one or more persons
was produced by one or more persons
Interest
Project
Dataset
Paper

FIG. 3. Entities and relationships in the HBES database

4.2. Converting Relational Entity Data into RDF Triples

We converted the entities and relationships in the relational database into RDF triples by using the D2R package. While the relationships between entities were preserved as they existed in the database, we turned each table into a class, each column in the table into a class property, and each value of a column into an instance. A URI was assigned to each class, property, and instance. The example in Box 2 shows the URIs assigned to a project and a person based on the ID information from the database.

Box 2: Portion of RDF triples generated from the relational database

```
<rdf:Description rdf:about="http://hubbardbrook.org/data/people-projects/103?output=rdfxml">
   <rdfs:label>RDF Description of people-projects #103</rdfs:label>
   <foaf:primaryTopic>
     <vocab:people-projects rdf:about="http://hubbardbrook.org/resource/people-projects/103">
       <vocab:people-projects_projectID rdf:resource="http://hubbardbrook.org/resource/projects/p34"/>
       <vocab:people-projects_personID rdf:resource="http://hubbardbrook.org/resource/people/jsteinweg"/>
       <rdfs:label>people-projects #103</rdfs:label>
       <vocab:people-projects_ID rdf:datatype="http://www.w3.org/2001/XMLSchema#int"> 103 </vocab:people-
            projects_ID>
       </vocab:people-projects>
   </foaf:primaryTopic>
 </rdf:Description>
</rdf:RDF>
```

Figure 4 on the following page shows a portion of the projects and datasets associated with researcher Charles Driscoll. Starting from Driscoll's URI, one can navigate away from any of the person-project, person-dataset, etc. relationships to find related projects, datasets, persons, subject interests, and papers. Although the entity relationships are transformed from a database to RDF triples, this is only the first step toward a semantic web for scientific datasets. The next challenge is how to "plant" the URIs into metadata records so that metadata records will be displayed with linked entities.

4.3. Transforming EML Records

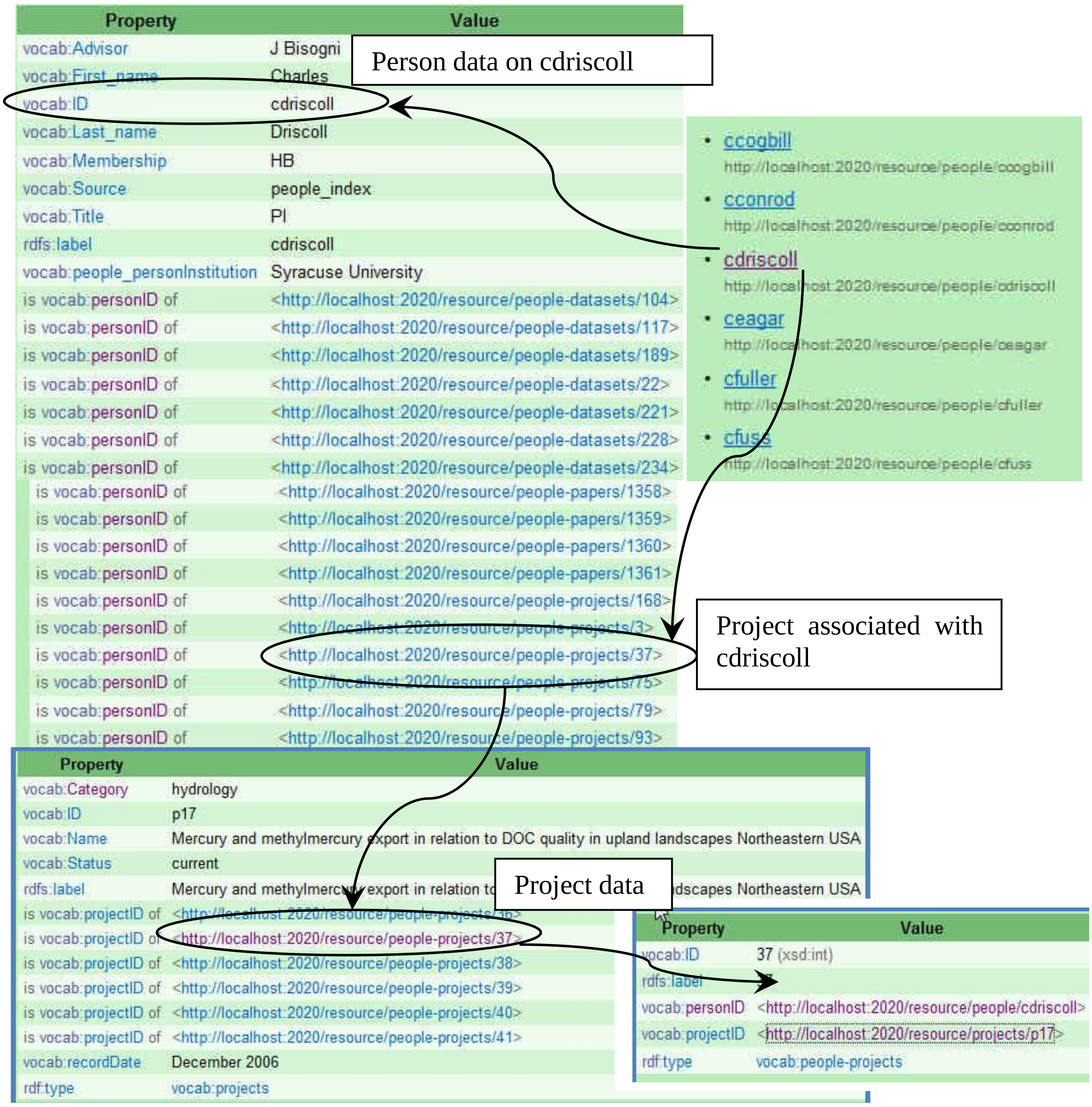
Planting URIs in EML records is essentially an issue of XML transformation. Technically, the transformation needs to go through a two-stage process to generate a user access page showing linked entities. During the first stage, we wrote a transformation program using the XML Stylesheet Language (XSL) to add the URIs generated from the D2R software to their corresponding entities, as shown in Box 3. The URI patterns are predefined based on domain name (root) and category layers after the root. Our test server has a temporary root (http://localhost:2020/). The URI example in Box 3 contains a relative path, i.e., the root was omitted.

Box 3: Individual name without and with URI added

Original EML record without URI	URI added to individual name element

<individualName>
    <givenName>Thomas G.</givenName>
    <surName>Siccama</surName>
</individualName>
	<individualName>
	    <givenName>Thomas G.</givenName>
	    <surName>Siccama</surName>
	    <personURI>page/people/tsiccama</personURI>
	</individualName>

At the second stage, we wrote another XSL program to transform the EML records with inserted URIs into the HTML format so that a web browser can display the XML records with



Person data on cdriscoll

Property	Value
vocab:Advisor	J Bisogni
vocab:First_name	Charles
vocab:ID	cdriscoll
vocab:Last_name	Driscoll
vocab:Membership	HB
vocab:Source	people_index
vocab:Title	PI
rdfs:label	cdriscoll
vocab:people_personInstitution	Syracuse University
is vocab:personID of	<http://localhost:2020/resource/people-datasets/104>
is vocab:personID of	<http://localhost:2020/resource/people-datasets/117>
is vocab:personID of	<http://localhost:2020/resource/people-datasets/189>
is vocab:personID of	<http://localhost:2020/resource/people-datasets/22>
is vocab:personID of	<http://localhost:2020/resource/people-datasets/221>
is vocab:personID of	<http://localhost:2020/resource/people-datasets/228>
is vocab:personID of	<http://localhost:2020/resource/people-datasets/234>
is vocab:personID of	<http://localhost:2020/resource/people-papers/1358>
is vocab:personID of	<http://localhost:2020/resource/people-papers/1359>
is vocab:personID of	<http://localhost:2020/resource/people-papers/1360>
is vocab:personID of	<http://localhost:2020/resource/people-papers/1361>
is vocab:personID of	<http://localhost:2020/resource/people-projects/168>
is vocab:personID of	<http://localhost:2020/resource/people-projects/3>
is vocab:personID of	<http://localhost:2020/resource/people-projects/37>
is vocab:personID of	<http://localhost:2020/resource/people-projects/75>
is vocab:personID of	<http://localhost:2020/resource/people-projects/79>
is vocab:personID of	<http://localhost:2020/resource/people-projects/93>

- ccogbill
  http://localhost:2020/resource/people/ccogbill
- cconrod
  http://localhost:2020/resource/people/cconrod
- cdriscoll
  http://localhost:2020/resource/people/cdriscoll
- ceagar
  http://localhost:2020/resource/people/ceagar
- cfuller
  http://localhost:2020/resource/people/cfuller
- cfuss
  http://localhost:2020/resource/people/cfuss

Project associated with cdriscoll

Property	Value
vocab:Category	hydrology
vocab:ID	p17
vocab:Name	Mercury and methylmercury export in relation to DOC quality in upland landscapes Northeastern USA
vocab:Status	current
rdfs:label	Mercury and methylmercury export in relation to [illegible] dscapes Northeastern USA
is vocab:projectID of	<http://localhost:2020/resource/people-projects/36>
is vocab:projectID of	<http://localhost:2020/resource/people-projects/37>
is vocab:projectID of	<http://localhost:2020/resource/people-projects/38>
is vocab:projectID of	<http://localhost:2020/resource/people-projects/39>
is vocab:projectID of	<http://localhost:2020/resource/people-projects/40>
is vocab:projectID of	<http://localhost:2020/resource/people-projects/41>
vocab:recordDate	December 2006
rdf:type	vocab:projects

Project data

Property	Value
vocab:ID	37 (xsd:int)
rdfs:label	37
vocab:personID	<http://localhost:2020/resource/people/cdriscoll>
vocab:projectID	<http://localhost:2020/resource/projects/p17>
rdf:type	vocab:people-projects

FIG. 4. Example result for person and related data with URIs

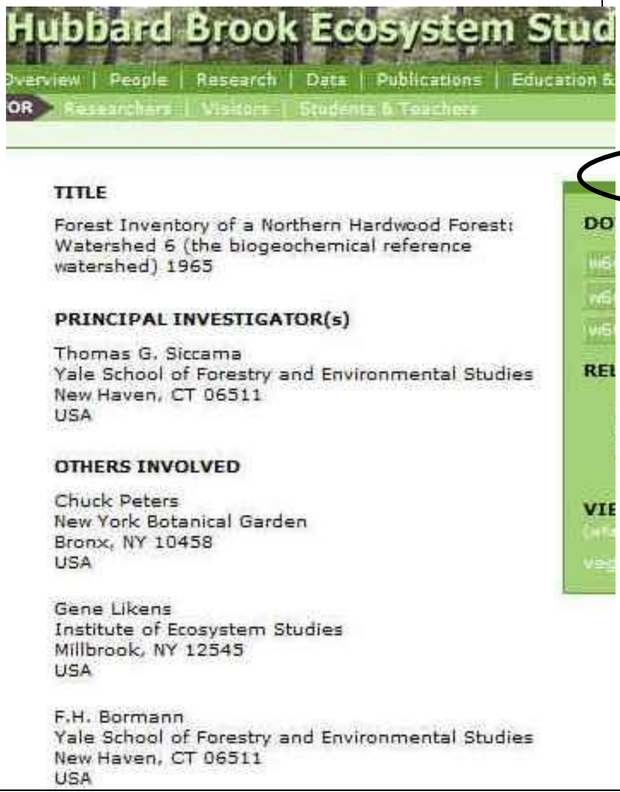
linked entities. The XSL program was designed so that when a user selects a specific person or project, the target URI will be sent to the server. The server will then search the RDF triple collection and return all the related information. Box 4 presents the original display of the EML record on the HBES website (left column) in which the entity data is plain text without hyperlinks, while the RDF-enabled display converted person names and dataset title (backed by the dataset's URI) as clickable links. When a user selects the principle investigator "Thomas G. Siccama" (the right column), the server will find his URI (http://localhost:2020/people/tsiccama) and bring up all related projects, datasets, papers, persons, and subject interests, which is similar to the result shown in Figure 4.

5. Discussion and Conclusions

This experiment, though on a small scale, presents some interesting methodological and theoretical inspirations for building a semantic web for scientific datasets. Metadata in the digital era has inherited a descriptive tradition from library and information science. Current metadata practice follows a cycle of developing a metadata schema or standard, then the tool for metadata creation, and the rest is in catalogers' hands. This tradition, as it is reflected in metadata system design, weights description of resources and metadata management more than link description (metadata) and data between associated entities. Two major metadata tools, Metacat (Jones et al., 2001) and Morpho (http://knb.ecoinformatics.org/morphoportal.jsp) were developed for creating EML records early in the 2000s. The data entry and entity linking in these tools were not given as much attention as making a functional system. The Knowledge Network for Biocomplexity (http://knb.ecoinformatics.org/index.jsp) and HBES are both examples of the resource description and management paradigm.

Box 4: Original and RDF-enabled displays of EML record

Original display of EML record	RDF-enabled display of EML record



Hubbard Brook Ecosystem Stud

Overview | People | Research | Data | Publications | Education &
OR Researchers | Visitors | Students & Teachers

TITLE

Forest Inventory of a Northern Hardwood Forest: Watershed 6 (the biogeochemical reference watershed) 1965

PRINCIPAL INVESTIGATOR(s)

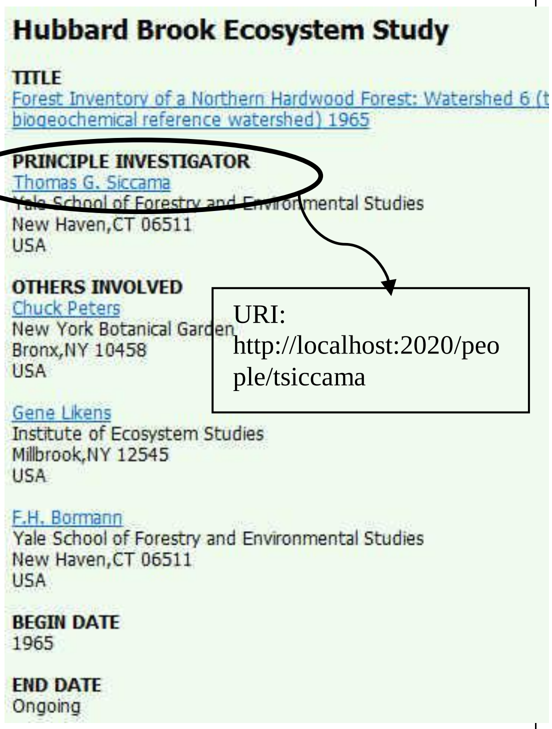
Thomas G. Siccama
Yale School of Forestry and Environmental Studies
New Haven, CT 06511
USA

OTHERS INVOLVED

Chuck Peters
New York Botanical Garden
Bronx, NY 10458
USA

Gene Likens
Institute of Ecosystem Studies
Millbrook, NY 12545
USA

F.H. Bormann
Yale School of Forestry and Environmental Studies
New Haven, CT 06511
USA



Hubbard Brook Ecosystem Study

TITLE
Forest Inventory of a Northern Hardwood Forest: Watershed 6 (t
biogeochemical reference watershed) 1965

PRINCIPLE INVESTIGATOR
Thomas G. Siccama
Yale School of Forestry and Environmental Studies
New Haven,CT 06511
USA

OTHERS INVOLVED
Chuck Peters
New York Botanical Garden
Bronx,NY 10458
USA

Gene Likens
Institute of Ecosystem Studies
Millbrook,NY 12545
USA

F.H. Bormann
Yale School of Forestry and Environmental Studies
New Haven,CT 06511
USA

BEGIN DATE
1965

END DATE
Ongoing

URI: http://localhost:2020/people/tsiccama

The EML standard was developed to support data discovery, interpretation and appropriate use, and automated use of data (Michener, 2006). In a typical end-to-end flow of in situ environmental sensor data, which includes raw data ingestion, quality assurance/quality control (QA/QC), data integration, analysis & forecasting, and published data products, metadata could be added at each of these stages (Michener, 2006). From a data entry standpoint, the many possible metadata input points in the data flow increases not only the number of entities, but also possibilities of duplicate data entry for person names, their affiliations, responsible agencies, subject topics, and associated publications and projects. Linked entity data for a science domain could lay the foundation for eliminating or reducing duplicate data entry while enhancing the linkage between associated entities.

Thus far the linked entity data remains a problem of metadata tradition. The description paradigm favors a separation of metadata specification from implementation. In reality, however, there is no such a thing of absolute separation between metadata specification and implementation. It is not uncommon that metadata standards or specifications offer recommended

XML encoding schemas, which are in fact a kind of implementation. As technology advances, things that were deemed impossible or inappropriate in the old technology environment may need to be re-studied to see whether the changed conditions have also reversed their paths. Current technological capabilities have made it possible to break the description paradigm to incorporate some new approaches such as linked entity data. Such new approaches would imply a break of traditional metadata cycle – defining specifications and possibly encoding schemas, building data entry interface, and input metadata – to bring in multiple modes for building metadata architecture. In addition to metadata specifications, the multimode paradigm may involve designing the specifications, encoding, and implementation with multiple technologies and semantic systems. This approach would greatly enhance the effectiveness and usefulness of domain specific metadata.

As a linked data experiment, this project raises more research questions than it answers. For example, how can the person entity data use FOAF (Friends-Of-A-Friend, http://www.foaf-project.org/) syntax? A similar question can also be asked for other entities. Utilizing and linking to data already made available from the ecological research datasets and their metadata records would be our next stage of research. We also hope to apply the approaches and methods of our experiment to a larger scale ecological data repository. Such future research would be important not only for validating our methodology but also for advancing metadata creation and services.

References

Berners-Lee, Tim. (2009, 6 18). *Linked data.* Retrieved April 1, 2010, from Design Issues: Architectural and Philosophical Points. Retrieved July 11, 2010, from http://www.w3.org/DesignIssues/LinkedData.html.

Bizer, Chris & Richard Cyganiak. (2009, 8 10). *D2R server: Publishing relational databases on the semantic web.* Retrieved July 11, 2010, from http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/.

Fox, Peter & James Handler. (2009). Semantic eScience: Encoding meaning in next-generation digitally enhanced science. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The Fouth Paradigm: Data-Intensive Scientific Discovery* (pp. 147-152). Edmond, WA: Microsoft Research.

Gritton, Bruce, Richard Dugdale, Thomas Duncan, Robert Evans, Terrence Joyce, & Victor Zlotnicki. (1995). Report of the ocean sciences data panel. In *Study on the Long-Term Retention of Selected Scientific and Technical Records of the Federal Government: Working Papers.* (pp. 86-104). Washington, D.C.: National Academy Press.

Hornbeck, Jim. (2001, May). *Events leading to establishment of the Hubbard Brook Experiment Forest.* Retrieved July 11, 2010, from Hubbard Brook Ecosystem Study: http://hubbardbrook.org/overview/HBEF_establishment.htm.

Jones, Matthew B., Chad Berkley, Jivka Bojilova, & Mark Schildhauer. (2001). Managing scientific metadata. *IEEE Internet Computing , 5* (5), 59-68.

Michener, William. K. (2006). Meta-information concepts for ecological data management. *Ecological Informatics , 1*, 3-7.

W3C. (2010, 3 23). *SWEO Community Project: Linking Open Data on the Semantic Web.* Retrieved July 11, 2010, from TaskForces / CommunityProjects / LinkingOpenData/DataSets: http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets

Wikipedia. (2010, 3 2). *Linked data.* Retrieved July 11, 2010, from Wikipedia: The Free Encyclopedia: http://en.wikipedia.org/wiki/Linked_Data