DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2012

Functional and Architectural Requirements for Metadata: Supporting Discovery and Management of Scientific Data

Jian Qin
School of Information Studies
Syracuse University
USA
jqin@syr.edu

Alex Ball
Digital Curation Centre
UKOLN
University of Bath
UK
a.ball@ukoln.ac.uk

Jane Greenberg
School of Information and Library Science
University of North Caroline
Chapel Hill, USA
janeg@email.unc.edu

Abstract

The tremendous growth in digital data has led to an increase in metadata initiatives for different types of scientific data, as evident in Ball's survey (2009). Although individual communities have specific needs, there are shared goals that need to be recognized if systems are to effectively support data sharing within and across all domains. This paper considers this need, and explores systems requirements that are essential for metadata supporting the discovery and management of scientific data. The paper begins with an introduction and a review of selected research specific to metadata modeling in the sciences. Next, the paper's goals are stated, followed by the presentation of valuable systems requirements. The results include a base-model with three chief principles: principle of least effort, infrastructure service, and portability. The principles are intended to support "data user" tasks. Results also include a set of defined user tasks and functions, and applications scenarios.

1. Introduction

We are facing a proliferation of scientific data and increased challenges relating to management and curation. There is a consensus among science data communities that metadata is the foundation for data discovery, use, and preservation. Collaborative efforts specific to digital data started to become more prominent about two decades ago in developing metadata standards for scientific data. Examples include the Content Standard for Digital Geospatial Metadata (CSDGM) published in 1998 that represents a notable achievement in this area, mandated by an executive order (FGDC, 1998). Ball (2009) provides a survey of many efforts that have emerged, and Willis, et al. (*in press*, 2012) account for community efforts from a range of disciplines (e.g., thermodynamics, crystallography, etc.), several of which emerged pre-Web.

Although metadata standards for scientific data have been developed to manage data sets, there is evidence that their application has not fully kept pace with the growth in scientific data. One reason is the complexity and specificity in these metadata standards that makes them "unwieldy to apply" and not "readily available or desirable for searching and browsing" (Riall et al., 2004). A case in point is the geospatial community's support of CSDGM, which has been implemented in a reduced and simplified form ("metadata-lite") in various initiatives (Caplan, 2003, pp. 140–141). The resource requirement to fully implement this scheme is quite extensive.

Scientific data vary greatly from discipline to discipline as well as in formats, types, processing, methods, and requirements. These variations are reflected in how data are sought out and used by different researchers, and this in turn places different and sometimes contradictory requirements on the metadata used to support such activity. For this reason, no single metadata standard can be applied universally to describe all types of scientific data sets and collections. Indeed, almost all data libraries and repositories developed over the last two decades have modified or extended existing metadata standards to suit their local needs. The possibility remains, however, that for specific functions, there may be enough requirements in common

https://doi.org/10.23106/dcmi.952135997

across the majority of scientific data for discrete blocks of metadata to satisfy them. To explore this possibility, two basic questions need to be addressed. First, what functions do metadata standards for scientific data serve? Second, how should metadata standards for scientific data be modeled to support these functions by meeting the associated requirements?

This paper attempts to address the two basic questions through analyzing user tasks and requirements for scientific data. These requirements in turn are translated into the functions and architectural building blocks for metadata. The mapping between user tasks and metadata functions and building blocks presents a methodology or an approach to re-examine the metadata constructs for the management, quality control, discovery, and use of scientific data.

2. Related Research

Modeling metadata is not a new research topic and has been studied extensively over the last decade, with links to library cataloging in related areas of study. General metadata modeling approaches are embedded in the descriptive tradition. Describing resources with metadata has a long history in the library community and well-established objectives and principles. Metadata schemas under such a tradition are expected to be comprehensive, consistent, rational, current, compatible with international standards, adaptable, and easy to use (Danskin, 2009). Stemming from bibliographic control principles, the metadata created with any schema or standard is intended to enable users to find, identify, select, and obtain resources as well as navigate within a catalogue and beyond (IFLA, 2009b). These metadata models have had an impact on modeling the structures of new metadata schemas.

A common approach to model metadata schemas is entity-based modeling. This method focuses on identifying entities and relationships in a domain. Typical entities include *agent* or *person/corporate body*, *event*, *place*, and *object*, while "is-a," "is-part-of," and "contains" are examples of general relations (Lagoze & Hunter, 2001; IFLA, 2009a; Rust & Bide, 2000). In the science metadata domain, the important entities are those related to *investigation* (study or project), *investigator*, *topic*, *publication*, *sample*, *dataset*, *data file*, and *parameter* (Matthews et al., 2009). These models represent abstract views of the entities and their attributes and relations surrounding the creation, publication, and management of resources. Entity-based metadata models provide semantics and structures that assist in developing metadata schemas.

Haslhofer and Klas (2010) state that metadata applications need three building blocks: schema definition language, metadata schema, and metadata instance. Depending on the function of metadata elements, they can be grouped into different types: administrative, descriptive, preservation, structural, technical, and use (Gill et al., 2008; NISO, 2004) and other domain-dependent types such as educational and geotemporal. The grouping of metadata elements does not involve modeling the structure of a metadata schema, rather, it is merely based on an attribute or element's role in representing the resource.

Another modeling method is to conceptually define the levels of representation of resources and then translate the levels of representation into appropriate metadata construct levels. In developing an application profile for the DRIADE project, Carrier et al. (2007) adopted a three-level approach for moving forward their application profile model. The level one application profile is intended for initial repository implementation, as with most existing application profiles. The second level extends level-one functionalities by capturing the complex relationships that exist among data objects and supporting expanded usage, interoperability, preservation, and administration. The third level supports "next generation" a.k.a. *NextGen*/Web 2.0 functionalities in the repository, such as personalization, social tagging, syntactic interoperability for data, data and collection visualization, and user feedback. Taking a slightly different angle, Takeda (2009) in a report about the Institutional Data Management Blueprint (IDMB) project models the metadata into three levels of findability: core metadata that helps users find authors, publishers, disciplines, and date; discipline metadata that assists users in finding the right sub-domains, projects, funders, and techniques; and project metadata that

https://doi.org/10.23106/dcmi.952135997

contains detailed dataset and its context. In both DRIADE and IDMB cases, the first level of metadata is considered as the place for general description of data while the second and third levels of representation involve more in-depth, specialized representations, which echoes what Keith Jeffery mentioned in his "Data surgery" presentation that organizing content is based on a three-level model—a top level based on a general purpose metadata element set such as Dublin Core, a second level focusing on contextual metadata as reflected in entities, and a third level covering more granular, detailed information (Boyd, 2012). As for what the third-level metadata exactly is, there has been a lack of discussion in literature.

Metadata standards containing large numbers of elements and with complicated structures have consistently run into the barriers of high costs in implementation and steep learning curves for metadata contributors. A complex, deep-layered structure also makes automatic metadata generation extremely difficult if not impossible, a hurdle for metadata generation to forever lag behind the pace of data growth. While the description tradition developed over the last hundred years still reigns, different approaches have been proposed to model metadata schemas. Scientific data is a new species in the metadata description land and hence new lenses need to be used to examine it thoroughly, starting from its foundation and extending into the new research conditions and requirements in cyberinfrastructure.

3. Goals

This paper addresses two questions fundamental to metadata for scientific data: (1) what functional requirements should metadata standards for scientific data support? (2) How should metadata standards for scientific data be modeled to support the functional requirements?

Although research in scientific data management recognizes the role and importance of metadata, there are gaps between the properties of scientific data required in the e-science environment and functional requirements for metadata, as shown in the "Related Research" section. In exploring answers to the first question, we start from the requirements for scientific data: what properties are expected of scientific data in the cyberinfrastructure-enabled research environment and how such expectations affect the metadata modeling. The analysis and results are based on authors' research projects and experience related to scientific data management.

The second question seeks the methodology and conceptualization in developing metadata models for scientific data. As with many other domains, metadata models for scientific data need to be not only "scientifically" built but also easy to use and useful, and should allow the data to be cited (Smith, 2009). The scientific-ness and usefulness are not always in harmony nor easy to balance. Our method for addressing this question is to analyze the data user tasks and map the tasks with both functional and architectural metadata, which can then be used to derive minimal metadata models for specific data user tasks.

4. Properties of Scientific Data and Requirements for Metadata

Metadata for scientific data can be considered as mission-critical in scientific data discovery, use, and citation. Research conducted in the cyberinfrastructure environment needs scientific data to have the following "e-science properties":

1. *Verifiable*: datasets should bear provenance metadata that allow researchers to trace them back to the raw data for quality control and data reuse purposes. The verifiability ensures the validity of research and allows researchers other than the data owner to repeat the study using the same data.
2. *Interworkable* (NSF, 2011): datasets should contain sufficient metadata to facilitate data discovery, selection, aggregation or filtering, and reuse. The interworkability of data types and related metadata should be built to accommodate researchers generating data from the very beginning and throughout the research lifecycle.
3. *Analyzable*: datasets should be in a state that requires minimal data manipulation in order to proceed with science research. This property implies that the data management

https://doi.org/10.23106/dcmi.952135997

system prepares the data to be ready for analysis based on science requirements for one or more research communities. Such analysis-ready datasets would be of appropriate type and include necessary documentation and/or metadata delivered as part of infrastructure services.

4. *Interoperable*: datasets should conform to standards so that they can be communicated and processed by different systems and software tools. Such interoperability ensures that the verifiability, interworkability, and analyzability of data will not only transcend space and time, but also reach across the practices of the research communities that need to use or reuse them (Qin et al., 2011).

These properties of scientific data can be translated into functional requirements for metadata used to describe and represent scientific datasets, which Greenberg (2009) summarizes as:

- Resource discovery and use,
- Data interoperability,
- Automatic and semi-automatic metadata generation,
- Linking of publications and underlying datasets,
- Data/metadata quality control, and
- Data security.

Incorporating the "e-science properties" of scientific data and functional requirements for scientific metadata, we developed four areas of requirements for scientific data description and representation (FIG. 1). The requirement model in FIG. 1 is a combination of metadata operation (data management), user tasks, and research requirements for data. From a functional view, data management functions build the foundation for other areas of functions through a wide range of activities from data storage, transformation, and organization to metadata generation to preservation of data for long-term access. These activities produce metadata artifacts – metadata schemas, metadata description sets, terminology, and best practice guidelines – necessary for data quality control, data discovery, and data use. Each of these three areas has its own specific requirements and can be mapped into the types of metadata mentioned earlier in the related literature. This functional view of metadata is the most familiar to the library/metadata community.

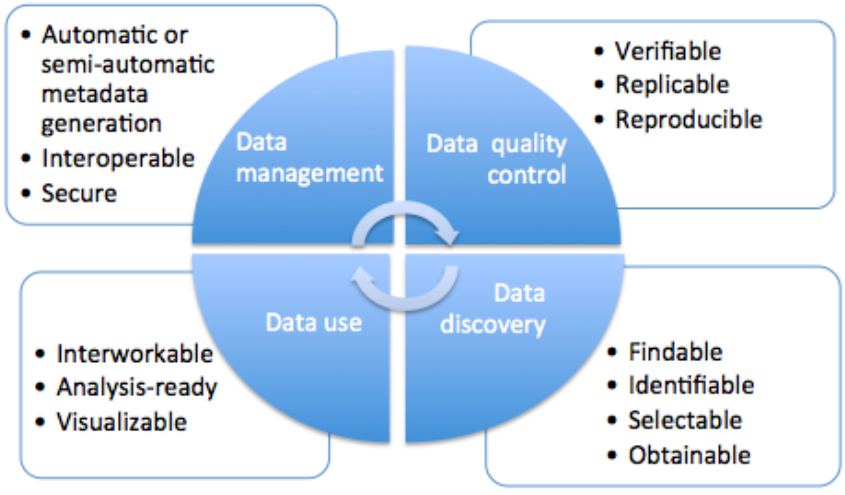

FIG. 1. Metadata requirements for scientific data in support of data management, data quality control, data discovery, and data use

The architectural view, however, is much less frequently mentioned. The so-called "architectural view" of metadata sees metadata attributes as building blocks that form a

https://doi.org/10.23106/dcmi.952135997

comprehensive representation of data or information objects. The architectural view of scientific data is illustrated in FIG. 2.

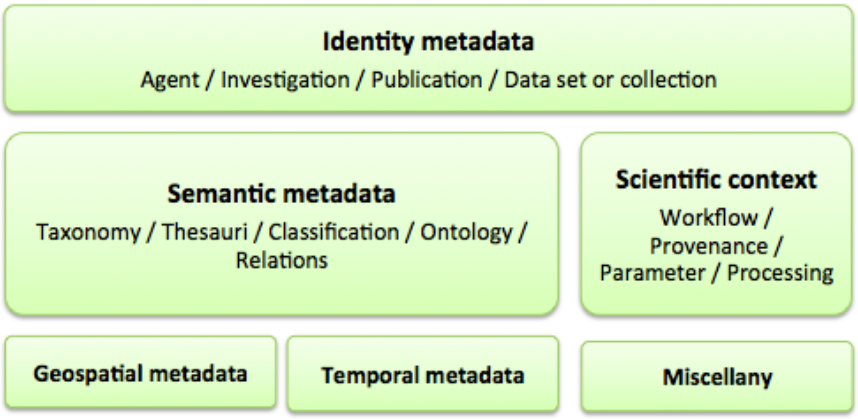

FIG. 2. An architectural view of metadata requirements

Identity metadata includes the entities that have been discussed in metadata models (see section 2 for reference). Each of these entities has its own set of metadata elements for description purpose, for example, a person entity has name, role, affiliation, contact information, and may be identified by a standard identifier system such as ORCID[1] and ResearcherID[2]. An event entity would have a name, time and place of occurrence, description, type, keywords, and other attributes. It would also have an identifier conforming to some standard system. When using Dublin Core to describe publications and web resources, one or more identifiers can be used to uniquely identify the resource, in addition to other descriptive elements. The unique identifiers may come from standard identifier systems such as Digital Object Identifier (DOI)[3], Uniform Resource Identifier (URI),[4] Handle System,[5] and/or Universal Numeric Fingerprint (UNF)[6]. Another example is data citation in which only identity metadata is needed in the actual citation with the identifier or identifiers pointing to where the dataset or data collection is located. The DataCite Metadata Schema,[7] a Dublin Core compliant metadata schema, is designed for just this purpose. Identity metadata builds the basis for making data identifiable and readily findable when such identities are known. No matter which type of entities we deal with, a common theme is the use of standard identifiers that can uniquely identify the entity globally. This is a necessary condition as well as the foundation for the Semantic Web envisioned by Berners-Lee (Berners-Lee, Hendler, & Lassila, 2001).

Semantic metadata for scientific data plays two roles: one is as the subject identifier for data and the other as the subject grouping criteria and linking mechanism for data with similar subject content. Large semantic tools such as the Unified Medical Language System (UMLS) and Library of Congress Subject Headings (LCSH) have been converted into several encoding formats, including the format of Resource Description Framework (RDF), which makes it possible for metadata tools to utilize the semantic tools for much more flexible and extensive representation and linking. For instance, we may use a subject term's URI to represent the subject

[1] http://www.orcid.org
[2] http://www.researcherid.com
[3] http://www.doi.org/index.html
[4] http://www.w3.org/Addressing/
[5] http://www.handle.net/
[6] http://thedata.org/book/unf
[7] http://schema.datacite.org/

https://doi.org/10.23106/dcmi.952135997

content of a dataset, rather than the term in natural language form. If both scientific datasets and semantic sources support standard identification schema such as URI, the implicit relations between datasets and the link target (subject term or other entities) can be made explicit as RDF links. A large number of scientific semantic data in the form of linked data are being developed (Bizer, Heath, & Berners-Lee, 2009). Its potential for promoting interdisciplinary scientific discovery and data is still to be explored and deployed (Bechhofer et al., 2011).

Scientific context, geospatial, and temporal metadata fulfill the requirements for data verifiability, replicability, and reproducibility. It is important to point out that these types of metadata are not necessarily exclusive from semantic metadata, and, in fact, can convey subject-related aspects of data. These areas of metadata describe the science aspects of the data and can be separate description units while maintaining proper associations with identity metadata. Examples of these types include the method and protocol elements in the Ecological Metadata Language (EML)[8] and the lineage element in CSDGM. Provenance metadata is another term used to describe the metadata about scientific context. In research fields that are highly computational from data capture to analysis—for example, gravitational wave research that uses computer scripts for data processing, calibration, and analysis—provenance metadata needs to be captured by workflow management systems and becomes part of the documentation for the analysis job run. The miscellany metadata includes elements that do not fit into any of the other blocks: file size, storage medium and dissemination medium (for offline data) are typical examples.

There is no doubt that metadata is invaluable in supporting data discovery, data quality control, and data use. The metadata application examples in the Dryad project[9] and DataCite[10] demonstrate that not all user tasks require full-fledged metadata representation and that we should reconsider the approach of one huge metadata schema with hundreds of elements to represent the full diversity of scientific data attributes. The Dublin Core Metadata Initiative (DCMI) is built upon the same fundamental idea, combined with the notion of allowing for interdisciplinary discovery. Making metadata standards portable and mutually integrable can be a solution to the barriers for metadata standard adoption. This line of thought leads to a key question that needs to be addressed: How can a model for metadata specific to scientific data meet the requirements for data management, discovery, quality control, and use while remaining easy to use and economic to maintain? To move forward with this idea, we present a series of principles. They have emerged from discussion in the DC-Science and Metadata community and the discussions that have taken place at a host of data curation and digital data conferences and workshops (e.g., RDAP, IDCC, ASIST).

5. Modeling Metadata for Data User Tasks

Based on the requirement model in FIG. 1 & 2, we propose three principles in modeling metadata for scientific data:

1. *The least effort principle*: the metadata is carefully designed to minimize redundancies in data entry, e.g., entities such as researchers, institutions, or funding agencies should utilize existing databases to populate if possible. Efforts in building entities as identifiable objects or linked data are underway. ORCID and DOI are two examples mentioned earlier in this paper. In principle, entities should be created once and reused whenever and wherever they are needed. With rich legacy entity data collections in indexing and catalog database systems, using semantic technologies to convert them into linked datasets will greatly benefit metadata creation for scientific data and reduce the time and duplicate efforts in having to reenter the information for each metadata record.

[8] http://knb.ecoinformatics.org/software/eml/
[9] http://datadryad.org/
[10] http://datacite.org

https://doi.org/10.23106/dcmi.952135997

2. *The infrastructure service principle*: architectural building blocks may well be considered as metadata infrastructure. The identity, semantic, scientific context, and geospatial-temporal metadata should be established as infrastructure services for scientific data. Creating metadata for scientific data often needs to associate complex entities with datasets and collections. Many of the metadata values have existed in databases and other sources, such as geographic data for countries, states, counties, and landscape features; time data in various scales and formats; investigations and studies as well as the researchers. An important task in building such infrastructure services for metadata generation/creation for scientific data lies in converting what we already built in the last 40 years into the formats and structures suitable for the new e-science environment.
3. *The portable principle:* Architectural blocks of metadata attributes should remain independent while being flexible enough to allow multiple portable schemas to be merged together to meet specialized representation needs. This principle means that metadata properties are modeled by ontological methods, which will then be encoded in the formats that support linking and reuse. As metadata infrastructure services become established, the main task of developing metadata schemas will shift to modeling the domain of interest and assembling description sets by drawing on existing metadata and creating new if none exits.

Currently, at least three factors motivate the thinking about modeling metadata for scientific data out of the descriptive tradition. First, scientific data is a diverse and dynamic domain. Nevertheless, an approximate consensus has arisen on the entities associated with scientific data and their respective attributes among scientists and data/information scientists. This consensus creates a common ground for building a metadata infrastructure to support portable, dynamic metadata schemes. Second, the large, comprehensive metadata standards for scientific data have proven to be difficult to use and, with a whole suite of artifacts to consider (e.g., metadata specifications, schemas, instances), expensive to implement and maintain. It will make more sense to develop specific goal-oriented metadata schemes for specific user tasks. Smaller, more specific metadata schemes will likely increase the full adoption of a scheme and hence reduce duplication in creating the same metadata elements and values in different local contexts. They can also increase the portability of metadata within the infrastructure, whether used alone or merged with others. Finally, the metadata infrastructure does not have to be built from scratch; many such infrastructure services already exist, for example, the name authority database and geographic controlled vocabulary at the Library of Congress and the naming systems for geographic, planetary nomenclature, and geologic terms at U.S. Geological Survey (http://www.usgs.gov/pubprod/reference.html).

Based on a typical research life cycle, we define 10 data user tasks (TABLE 1). The first four user tasks are the same as the ones defined by the library cataloging standards. The remaining six tasks are unique to scientific research. We view the metadata required to perform the user tasks from two perspectives: the function embedded in a type of metadata and the architectural building block of metadata attributes needed to support the metadata function. What needs to be pointed out is that a metadata scheme may target only one task but at the same time can perform other tasks either as a requisite for the primary task or as a side job. For example, the DataCite metadata schema is designed for one task—citing datasets—but incorporates additional, optional elements to allow it to perform the tasks of discovering, identifying, and locating datasets as well. In this sense, we divide data user tasks into:

- Generic tasks: discovery, identify, select, obtain
- Scientific tasks: verify, analyze
- Data tasks: manage, archive
- Dissemination tasks: publish, cite

https://doi.org/10.23106/dcmi.952135997

It is clear that identity metadata is ubiquitous in all user tasks, which makes it a requisite for other tasks in order to be performed. This can be considered as evidence that a metadata infrastructure for scientific data, if designed properly, can benefit the metadata creation and other scientific data user tasks tremendously.

TABLE 1. Mapping data user tasks with metadata functions and architectural building blocks

Data user tasks	Metadata function	Architectural building block
Discover	Descriptive metadata	Identity and semantic metadata
Identify	Descriptive metadata	Identity metadata
Select	Descriptive, technical metadata	Identity, semantic, scientific context, geospatial, temporal, miscellany metadata
Obtain	Descriptive metadata	Identify metadata
Verify	Descriptive metadata	Scientific context metadata
Analyze		Scientific context, geospatial, and temporal metadata
Manage	Descriptive, administrative, structural, and technical metadata	Identify, semantic, scientific context, geospatial, temporal, miscellany metadata
Archive	Descriptive, administrative, structural, and technical metadata	Identify, semantic, scientific context, geospatial, temporal, miscellany metadata
Publish	Descriptive metadata	Identity, semantic, scientific context, geospatial, and temporal metadata
Cite	Descriptive metadata	Identify metadata

6. Application Scenarios

As part of our analysis we present two scenarios to demonstrate the possible application of any of the blocks of metadata supporting scientific data.

Scenario 1: Emphasis: Cross-domain discovery and verifiability

A researcher is interested in a particular type of measurement made within a defined geographical area. The researcher chooses a data repository aggregator or cross-search service and searches for relevant data; search entry points include geographical area (e.g. latitude/longitude 'square'), time period, the field of research and the variable or keyword measured. The search returns a list of possible datasets, each accompanied by a brief abstract or summary, alongside suggestions for filtering the result list (e.g., by date, publisher or creator). The researcher may narrow down the result set by choosing one or more filtering criteria.
Each item in the result set links through to a fuller catalog record for the dataset: information on entities related to the dataset, spatiotemporal resolution, data quality, provenance and data collection methodology that the researcher can use to assess if the data are suitable. Where available, the researcher makes use of preview images or data to make comparisons and gain a preliminary understanding of the data. The researcher uses details of how to access the data, also part of the detailed records, to obtain a copies of the most relevant and useful datasets in a suitable format.

Scenario 2: Emphasis: Creating metadata description sets

A researcher has just received a grant from a funding agency for her research project. Staff at her institution's data repository are notified by the project module, which has been created as one of the infrastructure services that keeps track of funded projects within the institution. The staff retrieve the data management plan prepared for the proposal and, by consulting the researcher, they identify the metadata model needed for organizing and managing the anticipated datasets and products and then configure the metadata submission interface. The data repository staff help the researcher locate necessary entity data (team members, previous related projects and

https://doi.org/10.23106/dcmi.952135997

publications, etc.) and embed frequently-used entities and their URIs or DOIs in the metadata entry interface so that the researcher's team can save time and minimize errors in data entry. New entities such as team members who do not have an identity metadata record will be created if needed.

The scenario activity is useful for conceptualizing where metadata supports particular functions. It can be a time-consuming task to consider the full range of scenarios, , but doing so helps to characterize how the functions should be supported. More work in this area will help to confirm the work in this paper, and identify areas requiring more attention as well.

7. Conclusion

This paper reports on exploratory work examining metadata functions and modeling for scientific data. This work is presented in the context of systems requirements that are essential for metadata supporting the discovery and management of scientific data. The results include a base-model with three chief principles: principle of least effort, infrastructure service, and portability. Results also include a set of defined user tasks and functions. Finally, two application scenarios are presented.

This paper is limited in that the work presented is based on factors gleaned from scholarly and scientific research, and discussions at research conferences and workshops. The authors recognize this limitation, but the sources informing the principles and models presented here are valid, and important venues for the exploring a range of topics relating to data curation, including the role and functionality of metadata for supporting the discovery and management of scientific data. By laying down a metadata modeling base here, supported by principles and examples of functions and metadata types, this paper has made a contribution and provides base-level modeling that can serve as a source in future research. The work presented may also help guide further qualitative research in this area, and ultimately form the design of instruments that could assist in gathering more empirical data to support or modify the proposed models. Next steps by these authors will consider these ideas, and seeks to gather more data to aid in the development of a sustainable and informative model supporting the discovery and management of scientific data.

References

Ball, Alex. (2009). Scientific Data Application Profile Scoping Study Report. University of Bath, UKOLN. Retrieved, June 28, 2012, from http://www.ukoln.ac.uk/projects/sdapss/papers/ball2009sda-v11.pdf

Bechhofer, Sean, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, . . . Carole Goble. (2011). Why linked data is not enough for scientists. Future Generation Computer Systems. Advance online publication. Retrieved, June 28, 2012, from http://dx.doi.org/10.1016/j.future.2011.08.004

Berners-Lee, T., James Hendler, and Ora Lassila. (2001). The Semantic Web. Scientific American, May 2001, p. 29-37.

Bizer, Christian, Tom Heath, and Tim Berners-Lee. (2009). Linked data: The story so far. International Journal on Semantic Web and Information Systems 5(3), 1–22.

Boyd, David. (2012). CERIF tutorial and UK data surgery. Blog posted on February 16, 2012. Retrieved, June 28, 2012, from http://data.blogs.ilrt.org/2012/02/16/cerif-tutorial-and-uk-data-surgery/

Caplan, Priscilla. (2003). Metadata fundamentals for all librarians. Chicago: American Library Association.

Carrier, Sarah, Jed Dube, Jane Greenberg. (2007). The DRIADE project: Phased application profile development in support of open science. Proceedings of the International Conference on Dublin Core and Metadata Applications 2007, 35-42. Retrieved, June 28, 2012, from http://www.dcmipubs.org/ojs/index.php/pubs/article/viewFile/39/19

Danskin, Alan. (2009). Statement of objectives and principles for RDA. Retrieved, June 28, 2012, from http://www.rda-jsc.org/docs/5rda-objectivesrev3.pdf

FGDC. (1998). Content Standard for Digital Geospatial Metadata (Document no. FGDC-STD-001-1998). Retrieved, June 28, 2012, from http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/

Gill, Tony, Anne J. Gilliland, Maureen Whalen, and Mary S. Woodley. (2000-2008). Introduction to Metadata: Pathways to Digital Information. Online Version 3.0. Edited by Murtha Baca. Los Angeles: J. Paul Getty Trust,

https://doi.org/10.23106/dcmi.952135997

Getty Research Institute. Retrieved, June 28, 2012, from http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html

Greenberg, Jane. (2009). Metadata research supporting the Dryad Data Repository. Presentation at the Metadata Working Group, Cornell University Library. Retrieved, June 28, 2012, from http://dspace.library.cornell.edu/bitstream/1813/12247/1/DryadCornell.pdf

Haslhofer, Bernhard and Wolfgang Klas. (2010). A survey of techniques for achieving metadata interoperability. ACM Computing Surveys, 42(2): article 7. Retrieved, June 28, 2012, from http://eprints.cs.univie.ac.at/79/1/haslhofer08_acmSur_final.pdf

IFLA. (2009a). Functional Requirements for Bibliographic Records: Final Report. Retrieved, June 28, 2012, from http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

IFLA. (2009b). Statement of international cataloguing principles. Retrieved, June 28, 2012, from http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf

Lagoze, Carl and Jane Hunter. (2001). The ABC ontology and model. Journal of Digital Information 2(2). Retrieved, June 28, 2012, from http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/

Matthews, Brian, Shoaib Sufi, Damian Flannery, Lauren Lerusse, Tom Griffin, Michael Gleaves, and Kerstin Kleese. (2009). Using a core scientific metadata model in large-scale facilities. The 5th International Digital Curation Conference, December 2009. Retrieved, June 28, 2012, from http://epubs.cclrc.ac.uk/bitstream/4837/DCC09-CSMD-final.pdf

NISO. 2004. Understanding Metadata. Bethesda, MD: NISO Press. Retrieved, June 28, 2012, from http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

NSF. (2011). The "Earth Cube": Towards a national data infrastructure for Earth system science. Retrieved, June 28, 2012, from http://www.nsf.gov/geo/earthcube/geo-oci-earthcube-webinar-july11-2011.pptx

Qin, Jian, John D'Ignazio, and Suzanne Baldwin. (2011). A workflow-based knowledge management architecture for geodynamics data. A White paper submitted to NSF GEO/OCI EarchCube Charrette meeting: Retrieved, June 28, 2012, from http://earthcube.ning.com/page/whitepapers

Riall, Rebecca L., Fausto Marincioni, and Frances L. Lightsom. (2004). Content metadata standards for marine science: A case study: Chapter: Evolution. USGS Open-File Report 2004-1002. Retrieved, June 28, 2012, from http://pubs.usgs.gov/of/2004/1002/html/evol.html

Rust, Godfrey and Mark Bide. (2000). The <indecs> Metadata Framework: Principles, model and data dictionary. <indecs> Framework. Retrieved, June 28, 2012, from http://www.indecs.org/pdf/framework.pdf

Smith, Vincent S. (2009). Data publication: Towards a database of everything. BMC Research Notes, 2:113. Retrieved, June 28, 2012, from http://www.biomedcentral.com/1756-0500/2/113

Takeda, Kenji. (2010). Institutional data management blueprint project: Initial findings report. University of Southampton. Retrieved, June 28, 2012, from http://eprints.soton.ac.uk/195155/.

Willis, Craig, Jane Greenberg, and Hollie White. (in press, 2012). Analysis and Synthesis of Metadata Goals for Scientific Data. Journal of the American Society for Information Science and Technology. Retrieved, June 28, 2012, from http://ils.unc.edu/mrc/wp-content/uploads/2012/04/metadata-for-scientific-data-preprint.pdf

https://doi.org/10.23106/dcmi.952135997