Ontological Ownership: Empowerment and Sustainability

Jane Greenberg, Angela Murillo
Metadata Research
Center, School of Information and Library Science,
University of North Carolina at Chapel Hill
janeg@email.unc.edu; amurillo@email.unc.edu

John A. Kunze
UC Curation Center, California Digital Library
jak@ucop.edu

ABSTRACT
Positive impacts associated with urban housing/home ownership programs motivate us to study this topic in relation to ontologies. This paper reviews ontological dependence and presents early work underway in the DataONE Preservation and Metadata Working Group (PAMWG) to collectively leverage existing metadata schemes and ontologies. The paper introduces a high-level set of functional requirements and the stackoverflow model that may be used detect highly rated metadata or ontological properties to from a loose cannon for describing scientific data. The long term goal is to establish community identity and rhythm supporting a sustainable ontology/metadata driven workflow.

Keywords
Ontology, metadata, ownership, empowerment, sustainability, functional requirements.

INTRODUCTION
Ownership and related outcomes may motivate new ontological approaches. Urban housing/home ownership programs provide a motivating context. Government programs supporting dwelling ownership have promoted sustainable positive living experiences (Reingold, 1995; Rossi and Weber, 1996). Residents that retain ownership in public housing have been known to maintain community gardens, hold barbecues and fund raisers, and participate in neighborhood activities. The success is attributed to a simple equation of ownership, leading to empowerment, collective maintenance, and a sustainable environment.

Ownership and empowerment are fundamental tenants shaping the COPD ontology framework (Greenberg, et al, 2010). The Data Observation Network for Earth (DataONE) National Science Foundation (NSF) DataNet requires a sustainable ontological infrastructure. Collective, community driven ownership may support this goal and help limit the proliferation of schemes that overlap in purpose and scope. This paper reviews ontological dependence and presents early work underway in the DataONE Preservation and Metadata Working Group (PAMWG) to collectively leverage existing metadata schemes and ontologies.

Advances in Classification Research, 2012, October 26, 2012, Baltimore, MD, USA.

ONTOLOGIES: A BASIC NEED
Ontologies are not physical structures like a dwelling; and it is unrealistic to draw a true parallel between ontologies and the necessity of housing (shelter) as positioned in Maslow's hierarchy of needs (1943). Even so, we can agree that ontologies are essential for information system operations.

Ontological Dependence
Reliance on ontologies varies among information systems. Ontologies provide basic semantic structure and aid indexing, search, and retrieval. A common, low-level ontology is the Dublin Core metadata standard used in many institutional repositories. Slightly more sophisticated are library catalogs, abstracting/indexing databases, and digital libraries that use a basic semantic framework and terminological tools (e.g., thesauri, subject heading systems, etc.) that are viewed as simple ontologies (McGuinness, 2003).

Complex, multi-functional information systems use ontologies to support search and retrieval, as well as actionable processes. One example is Nitzsche, et al's (2007) generic ontology framework supporting executable business processes such as booking travel. Another example is the Biological Process Ontology Guidelines (the Gene Ontology, 2012) mapping biological processes and detecting cellular components and molecular functions.

Ownership, Maintenance, and Cost
Ontologies are generally owned by national and federal agencies, or select, privatized groups. These systems are maintained by a smaller, sometimes seemingly exclusive cluster of people. A chief reason for the exclusivity is that it is costly financially and time-wise for individuals to engage in ontology work following current development and maintenance practices.

Ownership has motivated maintenance and cost-savings with housing programs. Ontology design may benefit from

such a model. There are vocabulary programs supporting user contributions, such as the Government of Canada Core Subject Thesaurus.[1] A current limitation many efforts is candidate term review is not a community effort. Transparency and community driven review may expedite ontological growth, use, and sustainability, and reduce current associated costs.

CHANGE: TECHNOLOGY, COMMUNICATION, AND ONTOLOGICAL APPROACHES

Networked technology has transformed communication processes. The change presents a new and potentially cost-effective infrastructure for engaging experts and committed users in ontology design. Additionally, linked data developments invite questions about the best means for exploiting ontologies. Perhaps, most profound, are the variety of metadata/vocabulary registries (Murillo, 2012) and the lack of cohesive knowledge about use (heavy use to no use). All of these factors motivate study of ontology approaches from different angels, including the Data Observation Network for Earth (DataONE)[2] community.

DataONE

DataONE is a community and a distributed framework providing steps toward a sustainable cyberinfrastructure. DataONE seeks to meet the needs of science and society by developing an innovative, persistent, and robust environment for observational data about the Earth. DataONE represents a range of disciplines specific to the Earth (e.g., ecology, biology, geology, astronomy, etc., and the many sub-disciplines). A series of working groups are investigating new approaches and models to further DataONE's mission. The DataONE Preservation and Metadata Working Group (PAMWG) is one group addressing metadata specific issues.

THE DataONE PRESERVATION AND METADATA WORKING GROUP(PAMWG)

The DataONE PAMWG was established to address preservation and metadata challenges within the DataONE environment. DataONE does not store the 'scientific data,' rather the emphasis is on metadata supporting the discovery, access, and use of scientific data, regardless of where the data resides.

Domain specific and sub-domain ontologies are significant for DataONE. There is, however, a growing realization of duplicate efforts stemming from traditional, approaches (Willis, et al, 2012). Cohesive, collective ontology development could aid DataONE's interoperability goals and potentially reduce and collapse duplication. PAMWG members are cognizant of this challenge and exploring a way data creators, users, system developers, curators, and repository mangers may collectively engage in metadata/ontology development. An overriding question is how to leverage existing domain-specific ontology richness, and, at the same time, move toward a transdisciplinary framework that is community driven and transparent. The first step has been to establish functional requirements supporting an empowered approach (Table 1).

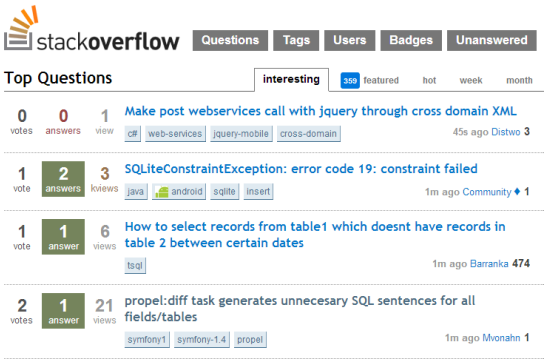
Table 1: Functional Requirements for the PAMWG

Low barrier for contributions.
Transparency in the review process.
Collective team review, with rotating responsibilities among community members (scientists, developers, organizations, curators, etc.)
Consideration of elders (experts) to guide the review process and maintain thoughtful, balanced discussion.
Voting capacity of all users on the candidacy of terms submitted and their use.
Collective ownership of any user or organization.
Stakeholder engagement in the design and review process.

Table 1 presents requirements that are being fleshed out following the September 2012 DataONE All-Hands meeting. PAMWG members advocate an approach combining "crowd sourcing" and "expert feedback." The end goal is a community endorsed lightly supervised cannon.

A second step has been preliminary exploration of the stackoverflow model as a means for engaging DataONE community members to vote on ontology/metadata properties (Figure 2).

Figure 2: The Stackoverflow Model



The stackoverflow model can be used to detect highly rated metadata or ontological properties, and then form a loose cannon of properties for describing scientific data. Launching this approach in an open environment may empower users (scientists, system developers, curators, etc.), and engage them in a more robust manner. Collective activity, suggesting terms and sharing insights, may help eliminate duplicate efforts across domains and support transdiciplinary work. The tenants articulated at this stage are:

[1] Government of Canada Core Subject Thesaurus. Suggest a new thesaurus term: http://www.thesaurus.gc.ca/default.asp?lang=En&n=5B88165E-1.

[2] DataONE: http://www.dataone.org/.

- Anyone can look up terms.
- Anyone can propose and refine their terms.
- Strong terms rise, weak terms decline, due to voting.

CONCLUSION

This paper introduces an ontological design approach relying ownership. The work is motivated by the potential benefits of empowerment and sustainability. A presentation of this work at the SIG/CR can motivate discussion about the range functional requirements for an open, collective approach to ontology design. Dialog on this topic can further provide grounding for a pre-proof-of-concept/Beta system and inform eventual assessment. The long term goal is to establish community identity and rhythm supporting a sustainable ontology/metadata driven workflow in DataONE and applicable to the larger ontology/metadata driven environment.

ACKNOWLEDGMENTS

The work in this paper is supported, in part, by the U.S. National Science Foundation (Grant #OCI-0830944).

REFERENCES

DataONE Metadata and Preservation Working Group (PAMWG): http://www.dataone.org/working_groups/data-preservation-metadata-and-interoperability-working-group.

The Gene Ontology. (1999-2012). *Biological Process Ontology Guidelines*: http://www.geneontology.org/GO.process.guidelines.shtml.

Greenberg, J. J., Mostafa J., Deshmukh, R., Huang, L., La Vange, L., Carretta, E., and O'Neal, W. (2010). The COPD Ontology and Toward Empowering Clinical Scientists as Ontology Engineers. *Journal of Library Metadata, 10* (2-3): 3-187.

Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review, 50:* 370-396.

McGuinness, D. L. (2003). Ontologies Come of Age. In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, 2003.

Murillo, A., P. Greenberg, J., Kunze, J., and Boone, J. (2012). Components of Successful Metadata Registry Frameworks. DataONE All-hands 2012 Meeting, Albuquerque, New Mexico, September 17-20, 2012. https://notebooks.dataone.org/metadata-registry/author/unset-unset-12/.

Nitzsche, J., Wutke, D., and van Lessen, T. (2007). An Ontology for Executable Business Processes. In M. Hepp, K. Hinkelmann, D. Karagiannis, R. Klein, N. Stojanovic (eds.): *Semantic Business Process and Product Lifecycle Management. Proceedings of the Workshop SBPM 2007, Innsbruck, April 7, 2007, CEUR Workshop Proceedings*, online CEUR-WS.org/Vol-251/ [Also available online at: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-251/paper8.pdf.]

Rossi, P. H. and Weber, E. (1996). The Social Benefits of Home Ownership: Empirical Evidence from National Surveys. *Housing Policy Debate, 7* (1):1-35.

Reingold, D. A. (1995). Public housing, Home Ownership, and Community Participation in Chicago's Inner City. *Housing Studies, 10* (4): 445-469.

Willis, C. Greenberg, J., and White, H. (2012). Analysis and Synthesis of Metadata Goals for Scientific Data. *Journal of the American Society for Information Science and Technology, 63* (8): 1505–1520.