Proceedings of the 7th International Conference on HydroScience and Engineering

Philadelphia, USA September 10-13, 2006 (ICHE 2006)

ISBN: 0977447405

Drexel University

College of Engineering

WHEN HYDROSPHERES COLLIDE: LESSONS IN PRACTICAL ENVIRONMENTAL ONTOLOGIES

John Graybeal[1] and Luis Bermudez[2]

ABSTRACT

The Marine Metadata Interoperability Project was created in 2004 with NSF funding. Its mission was to create a community of metadata-aware scientists and data managers, and provide leadership toward interoperable metadata solutions. Recently, MMI brought together a number of international participants, with the eventual objective of developing a fully rationalized ontology of data source types ("sensors").

First, however, the team focused on an ontology for environmental science platforms, with a particular focus on marine platforms. The team believed a platforms ontology was simpler, and knew it was a needed reference in the sensor ontology effort. In addition, the platform ontology could be put to use fairly quickly by a number of interested data system developers. The objectives, and the activities and associated documents are at the site http://marinemetadata.org/sourcesont.

This paper describes the reality of working with words in a computational context, from the point of view of a computer scientist and data manager who is not an ontologist. It also provides an alternative view on approaches to developing an upper ontology for a given topic.

1. INTRODUCTION

At its MMI workshop last year, called Advancing Domain Vocabularies, MMI brought together domain experts in 6 different disciplines, asking them to help create mappings between different science domain vocabularies. The results of this effort were varied, and were published in a workshop report (Graybeal, Watson et al. 2006).

More recently, MMI brought together a number of international participants to develop a fully rationalized ontology of data source types. As with all MMI projects, involvement is open to all interested volunteers, and progress is documented on the web site for community consideration and education. This approach exposes the technical experiences of the projects, which provides a learning opportunity for other technical readers, as well as letting them catch up easily if they join the project later.

As a first objective, the team set out to create an ontology for platforms, as it would be simpler and could be used as a reference in the sensor ontology effort. In addition, the platform ontology

[1] Principal Investigator, Marine Metadata Interoperability Project; Software Engineer, Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA (graybeal@mbari.org)
[2] Technical Lead, Marine Metadata Interoperability Project; Software Engineer, Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA (bermudez@mbari.org)

could be put to use fairly quickly by a number of interested data system developers. The objectives, and the activities and associated documents are at the site http://marinemetadata.org/sourcesont.

2. TRAINING AND TOOLS

The participants without ontological experience (most of the group) were trained in ontological matters using the "Pizza Ontology" example of Protege. This step required two 2-hour sessions, and helped everyone appreciate the tools we used and tradeoffs we considered.

As all the collaborative development was conducted on telecons, a sharable environment was necessary. The team used the tool Protege, running on the collaboration environment WebEx. This combination proved serviceable, but with some minor hiccups due to application issues (usually slowness in WebEx). The shared view of the technical lead's ontology construction proved valuable for the other participants, as it was much easier to divine intent and review the actual implementation in real time

Unlike the previous ontology mapping exercise mentioned in the introduction, this effort was intended to be relatively authoritative, given the absence of other broadly derived controlled vocabularies. The related requirements for consistency, generality, and usability of the resulting ontology proved difficult to achieve in the two month window for the initial effort. More time must be allowed to create even a relatively straightforward ontology that is useful and authoritative.

3. APPROACH

The team began by using an existing vocabulary as a starting point for discussion, and working out appropriate names, relationships, and definitions for the terms in the ontology. (This process is similar to that used in the development of many other domain ontologies (Bermudez, Graybeal et al. 2005), Our team had the advantage of using the SWEET (Raskin and Pan 2003) earth science ontologies, and its author was a participant, so we received excellent advice and assistance with the SWEET ontology.

The work also benefited from the vocabularies used by participants—Roy Lowry provided a particularly strong set of terms from the British Oceanographic Data Center and the Common Data Index SeaSearch project (see http://www.sea-search.net)—and a number of on-line tools, including WordNet (Fellbaum) and Wikipedia.

The usual difficulties arose during our discussions. Ontology development is a balanced application of domain expertise, logical organization and transparency, assessment of common language usage, and semantic refinement. It involves trade-offs when not all those considerations can be ideally addressed. Ontologies often require repeated iterative development to become useful for their target audience(s).

Since our ontology was intended for a variety of uses, by many target audiences, making decisions required even more judgment as to the overall best approach. Hierarchies like the one shown in Figure 1 are difficult to develop, and they change quickly in the early stages.

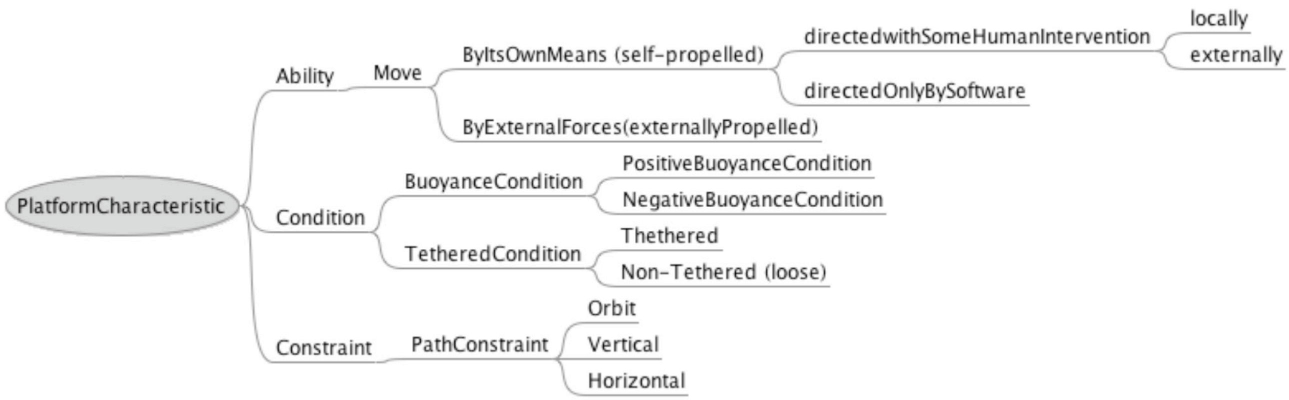

Figure 1 Hierarchy of Platform Characteristics (Version 0.2)

To take one example, we needed a term for the bodies of water that cover the earth's surface: oceans, seas, rivers, lakes, and even ponds, as some scientific platforms are deployed on all those bodies. Based on our backgrounds, we first selected the SWEET term *hydrosphere*, as it seemed to capture the appropriate concept. But the term 'hydrosphere' has a different meaning in the hydrologic community, where it also includes water vapor in the atmosphere. This 'collision of hydrospheres' suggests that either a new term may be necessary in SWEET—which would then be used in our ontology—or the existing SWEET terms needs to be defined so that all communities will understand how SWEET uses this word. As an initial approach we created our own term, which at least will not confuse people from different communities.

This example reflects the reality of working with words in a computational context. It represents a pervasive challenge when constructing and using ontologies, and in creating and using the semantic web itself. The lesson learned by the author was to allow time to find the most practical solutions, since ideal solutions often do not exist and many tradeoffs must be weighed.

4. THE DETAILS

One of the decision areas when developing an ontology is the rules to follow when creating terms. Some questions are relatively trivial: How do we capitalize terms? Others can be surprisingly tricky.

For example, if the major noun of a term precedes the other words, with any adjectives coming last, it is easy to see the organizational basis for the ontology just by looking at the terms. But any ability to match normal English usage will be lost: the unnatural word order of 'Profiler_Vertical' will frustrating for many users, and any transformation of 'autonomous underwater vehicle' would be equally awkward. The nominal English order was therefore chosen, because it is most useful to the potential users of this ontology: data managers, and the users of their systems. This approach also encouraged our focus on commonly used platform concepts, rather than exhaustively documenting the hierarchy of all possible concepts.

At least some of the team members might have argued about these details for a long time, but the team leader did the research needed to propose a fairly thorough set of rules to the team. Thanks to that starting point, the team quickly moved past those details and into the actual ontological and semantic issues. The important lesson here is to make sure someone has time and knowledge to lay the necessary foundation for the team's work, so time is focused on the key points in the ontology.

5. ORGANIZATIONAL FRAMEWORK

Not every ontology needs a single overarching framework to serve as the basis for its organization. It is often the case that ontological terms are related as a web of terms along different conceptual axes. Because of its scope, the semantic web will almost certainly have this broadly interconnected quality, with no particular categorization at the root level.

However, our team instinctively organized the platform ontology in a hierarchical framework, based at the highest level on the medium in which the platforms operate. In retrospect this decision was fortuitous, because it matched the way semantics have been organized to describe the platforms. Since the most important characteristic of observing platforms is where they operate, we could develop an ontology that was aligned with the common usage.

As the hierarchy of this ontology is descended, other characterizations distinguish key members of each group of platforms. Qualities included whether the platform was mobile, whether its motion was constrained (e.g., by a tether or directional limitations), and whether it had a crew or, lacking a crew, was autonomous or remotely operated.

Interestingly, although the platform quality concepts are hierarchical (Bermudez, Graybeal et al. 2006), the hierarchy of the ontology itself did not consistently follow these concepts. In each case, the developers considered the major categories of platform within that group, according to current practices in the field, and attempted to organize the ontology along those lines. Such a basis is prone to reinterpretation and revision, and it will be interesting to see whether this ontology persists in its current form or evolves to (or is replaced by) a significantly different framework.

In any case, the author concludes that the sensor ontology, envisioned as the team's next activity, can not be organized around an overarching conceptual framework. The applications, types, and categorizations of sensors are much more varied, and much harder to characterize in any credible way, than those of platforms. The team must anticipate the additional complexity and time required to create a non-hierarchical ontology.

6. OTHER CHALLENGES

When designing a multi-purpose ontology such as this, there can be surprises about the end uses. One of the participants found the distinction between 'research ship' and 'ship of opportunity' very important to characterize. The scientific value of this was unclear at first, but it turned out that operational and strategic planning goals required making this distinction, so that applications could assess available resources and their fitness for use. In similar ways, ontological frameworks can be extended and reorganized to meet an ever-growing number of applications, but at some point conflicts of purpose are likely to constrain the ability to create a "one size fits all" solution. The lesson for future similar projects is to discuss the broad priorities for the ontology at the beginning of the process, as this will guide the choices made throughout the ontology's development.

Another challenge is that of governance. This ontology was developed by a core group of interested individuals, and additional members joined the team during the process. The team certainly did not thoroughly represent the affected community, and changes will be ongoing. To establish the credibility of the platforms ontology, organizational resources will have to be dedicated to managing those changes, including evaluating each change for applicability, consistency, clarity, and completeness. For an ontology to be fully viable, the process and requirements that govern changes must be clearly specified and consistently followed, and the resulting changes must be fully traceable and versioned. Meeting this challenge will be required as part of the ongoing work on the platforms ontology, and should be included in the conception of future projects.

7. STATUS

The platform ontology is on-line at the MMI site and at SourceForge (http://sourceforge.net/projects/mmi/). It demands further support as noted above, but has already proven useful in the organization of several data systems and user interfaces. Interested community members are encouraged to contact the authors to help continue the development of the platform ontology and begin work on a sensors ontology.

8. CONCLUSIONS

The Marine Metadata Interoperability Project is committed to addressing sophisticated problems such as these ontologies, but also to providing solutions and guidance for the beginning data manager. The project's ontology work continues, and once platforms are sufficiently defined, will proceed to address data sources such as sensors and instruments, wiser for the lessons described here. Members of the community are welcome to join the MMI team working on this project, or on any of the many other projects MMI has initiated. Visit the site or send email to info@marinemetadata.org, or the author, for more information.

ACKNOWLEDGEMENTS

This activity was supported by the National Science Foundation through grant ATM-0447031. We would like to thank the other participants in this work: Robert Arko, Marilyn Drewry, Roy Lowry, Rob Raskin, and Kevin O'Neil.

REFERENCES

Bermudez, L. E., J. Graybeal, et al. (2006). "A Marine Platforms Ontology: Experiences and Lessons." Semantic Sensor Networks Workshop at the 5th International Semantic Web Conference (ISWC) 2006, Athens, GA.

Bermudez, L. E., J. Graybeal, et al. (2005). "Construction of Marine Vocabularies in the Marine Metadata Interoperability Project." 13th Ocean Sciences Meeting, Washington.

Fellbaum, Christiane, ed., (1998) WordNet: An Electronic Lexical Database, MIT Press, May 1998.

Graybeal, J., S. Watson, et al. (2006). "Marine Metadata Interoperability Workshop: Advancing Domain Vocabularies Workshop Report." Moss Landing, CA, MBARI. Access from http://marinemetadata.org/workshop05/results/workshopreport/.

Raskin, R. and M. Pan (2003). "Semantic Web for Earth and Environmental Terminology (SWEET)." Semantic Web Technologies for Searching and Retrieving Scientific Data (SCISW), Sanibel Island, Florida.