11th International Conference on Open Repositories
Trinity College, Dublin, Ireland
13th - 16th June, 2016

Increasing the Availability and Reuse of Earth Science Ontologies Using a Repository

Line C. Pouchard, Purdue University, pouchard@purdue.edu
Michael N. Huhns, University of South Carolina, huhns@sc.edu
Thomas W. Narock, Marymount University, tnarock@marymount.edu

Session Type: Presentation

Abstract: Earth Science depends on a diverse and complex data and metadata ecosystem that includes satellite and remote sensing observations, experimental results, model data, and more. The Federation of Earth Science Information Partners (ESIP) is an open community that brings together science, data, and information technology practitioners. One goal is to facilitate the description and exchange of Earth Science data and metadata. Ontologies have become a popular means of achieving this goal; yet, ontologies can be difficult to find . They are also difficult to reuse as their purpose influences the semantic modeling of the domain they intend to cover. We have designed a technical and organizational approach to help address these challenges: 1) we have deployed and seeded an instance of the BioPortal ontology repository, 2) we are working through ESIP to promote the repository to Earth Science communities, and 3) we have created mappings between ontology entities to help with ontology reuse. We achieved our initial objectives: the repository is operating and populated, supports mappings, and is used by the ESIP community. Our approach also reveals that a sustainable development and maintenance plan for this repository is necessary to continue the availability and currency of Earth Science ontologies.

Conference Themes

- Supporting open scholarship, open data, and open science
- Repositories of high volume and/or complex data and collections

Keywords: Ontologies, ontology repository, ontology mappings, ESIP federation

Audience: This presentation is of interest to repository managers, developers, data producers, librarians, and anyone with an interest in ontologies and the Semantic Web. The intended audience also includes those interested in deploying and supporting a repository from a mature research prototype within the context of a community-run, non profit federation.

Background: Our submission addresses conference themes and OR2016 overarching topic.

1) We discuss our experience around an ontology repository for Earth science, which involves very complex, heterogeneous data from multiple disciplines and sources.
2) We discuss the deployment of a repository instance that was developed for a different domain.
3) We discuss how we make an important technology in Earth Science – ontologies – more discoverable and open to users.
4) We discuss our additional features to improve reuse of the technology.
5) We show the role of a community-based, nonprofit organization in providing support, outreach, and feedback for the deployment, maintenance and use of the ontology repository.

Presentation content

Introduction and motivation: Earth Science represents a diverse and complex data and metadata ecosystem that includes satellite observations and remote sensing, observations, experimental results, model data, and more. These data and accompanying metadata and descriptions are held in data centers that focus on types of data, such as ocean or ice data, hydrology data, climate model and simulation data. In order to perform analyses of specific phenomena, for instance predicting flash floods or visualizing a major perturbation like a hurricane, scientists need to integrate multiple datasets with different parameters at various resolution levels from different sources. In order to do this accurately, they need detailed descriptions and annotations of the chosen data sets, such as those provided by ontologies. Ontologies provide machine-processable entities organized in a graph that can describe these datasets so that datasets are more discoverable. Ontologies also provide labels and relationships between the entities so that ontology-annotated datasets can be discovered in multiple ways, by following relationships between entities, or by searching on related labels.

Problems we are trying to solve: Problems we are seeking to address include improving the discovery and reuse of ontologies as they in turn support discovery and reuse of datasets. In Earth Science, most data and software are open source and available on the web. What makes these ontologies hard to find is the sheer volume of data, metadata, and annotations available from numerous sources. They often live on individual project web sites or in some cases in multi-purpose repositories such as Github. Increasing availability helps improve quality through feedback and reuse.

Ontologies are intended to be reused. Reuse improves the quality of annotations, since basic entities and relationships already exist and have been vetted elsewhere, for instance in the form of ontology design patterns. Yet, reusing entire ontologies can be problematic, because the purpose for which they were designed influences the semantic modeling of the covered domain.

11th International Conference on Open Repositories
Trinity College, Dublin, Ireland
13th - 16th June, 2016

Finding mappings between ontology entities and linking ontologies through individual mappings allow ontology creators to reuse single concepts, rather than an entire ontology. Mappings also enable data from independently constructed datasets to be related and combined.

Approach: We have designed a technical and organizational approach to address these problems and we will discuss each step in the presentation. First we discuss the advantages of deploying and populating an ontology repository for discovery. We have deployed an instance of the BioPortal Ontology Repository (the parent system) to host ontologies in Earth Science[1]. This is reuse of a technology framework developed by the National Center for Biomedical Ontologies at Stanford University to host medical and biological ontologies. The framework contains many features including ontology validation, visualization, metadata, search and browse features, and a recommender. We are re-purposing the repository for Earth Science ontologies and adapting it for this domain. One advantage of our re-purposed repository is to provide a single site where Earth Science ontologies can be found, where researchers can share their ontologies, and build additional constructs on each other's foundation. We have seeded the repository with some ontologies of interest in Earth Science. Our ontology collection includes the Semantic Web Earth and Environmental Taxonomy (SWEET) and the Environment Ontology (ENVO). ENVO contains 1911 classes. SWEET contains over 4,000 classes and covers many foundational sub-domains of interest to Earth Science, including basic physics and mathematical processes.

Second, facilitating the reuse of ontologies includes an evaluation of their scope. One possible method for performing this evaluation is to provide mappings between ontology elements in different ontologies. Obtaining mappings involves locating terms with similar semantics between two ontologies. We used the equivalence relationship to discover similarities between concepts. Agreement Maker Light, an algorithm from the Ontology Alignment Evaluation Initiative provided the mappings. We uncovered 841 mappings between SWEET and ENVO. These mappings show a 13% overlap between classes in the two ontologies. The implications for reuse are that 1) these ontologies do not duplicate each other, which supports having them both in the repository; 2) there exists some common entities that can be used to relate concepts in the two ontologies; 3) the mappings broaden the scope of the collections contained in our repository since users can avail themselves of additional relationships; and 4) semantic mappings between ontology entities can provide the basis for ontology design patterns that can re-used.

[1] http://semanticportal.esipfed.org

Third, we will discuss the organizational aspect of our project and involvement with the ESIP federation. We are working with the ESIP federation to support the repository deployment, gather feedback from members, and perform outreach to the community of Earth scientists. The Federation of Earth Science Information Partners (ESIP) founded in 1998 by NASA, is an open, networked community that brings together science, data and information technology practitioners that handle end-to-end Earth science data. The ESIP community includes those responsible for all elements of Earth observation, including observation data, research, and ultimately, application and education. Scientists and members of ESIP have been developing ontologies for the Earth sciences domains for many years. ESIP is organized in committees and clusters. Anyone interested can join the monthly teleconferences and semi-annual meetings. We have been working within the Products and Services Committee to obtain minimal funding for prototyping the repository. We have also worked with the Semantic Technologies Committee to support outreach, entice users to deposit their ontologies, and gather user feedback. Deploying the Semantic repository under the auspices of ESIP ensured that we reached the intended community and had a wide audience of Earth scientists who integrate data from many sources and would thus need ontologies. However, as our repository and ontology collection mature, it is no longer feasible to support the instance, install new releases, and ensure outreach with scant resources. A long-term plan for the maintenance and sustainability of the repository is needed. These discussions are just getting started in ESIP.

Interest to the intended audience: An audience interested in ontologies, the Semantic Web, and Earth Science, will want to hear this presentation as it describes our efforts to re-purpose an ontology repository developed for another domain and improve discovery and reuse of ontologies. Discussions around the workings of a non-profit, community-based federation to deploy, maintain and improve community infrastructure will arise out of our involvement with the ESIP federation.

Conclusion: In this project. we provided a solution to the problem of hard-to-discover Earth Science ontologies by integrating many pieces developed independently, and maximizing reuse. We have provided a technology (the ontologies) and an infrastructure to a new community (Earth Science) for access and re-use beyond their original purpose, thus opening our collections to new users. We made the ontologies available to a larger audience by harnessing the resources of a non-profit organization. We also suggest a path for discovering ontology design patterns.