Linking Earth and Climate Science: Semantic Search Supporting Investigation of Climate Change

Line C. Pouchard,[a1] Marcia L. Branstetter,[a] Robert B. Cook,[b] Ranjeet Devarakonda,[b] Jim Green,[b] Giri Palanisamy,[b] Paul Alexander,[c] Natalya F. Noy [c]

[a] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6290*
[b] *Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6290*
[c] *Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA 94305*

Abstract. Linked Science is the practice of integrating and aggregating structured data and information in physical, chemical, biological, sociological, and other traditional fields of scientific study. Much of this data does not live in the cloud or on the Web, but rather in multi-institutional data centers that provide tools and add value through quality assurance, validation, curation, dissemination, and analysis of the data. In this paper, we focus on the data in Earth and Climate Sciences and on the use of ontologies to facilitate search and integration of this data. Mercury, developed at Oak Ridge National Laboratory, is a tool for distributed metadata harvesting, search and retrieval. Mercury currently provides uniform access to more than 100,000 metadata records; 30,000 scientists use it each month. We augmented search in Mercury with ontologies, such as the ontologies in the Semantic Web for Earth and Environmental Terminology (SWEET) collection. We use BioPortal, developed at Stanford University, as an infrastructure to store and access ontologies. We use BioPortal REST services to enable faceted search based on the structure of the ontologies, and to improve the accuracy of user queries.

Keywords: Linked Science ontologies, BioPortal, semantic search, climate change

[1] Corresponding author. E-mail: pouchardlc@ornl.gov

1. Data Requirements for Linked Science

The ways in which scientists conduct research in physics, chemistry, biology, geography, ecology, sociology, and other scientific fields is changing significantly. Often, the most challenging research questions require them to understand and use the data from many scientific disciplines. These changes produce several key trends in the role of data, which are often referred to as *Linked Science*. Linked Science is the practice of integrating and aggregating structured data and information derived from physical, chemical, biological, sociological, and other traditional fields of scientific study.

In the modern cycle of scientific discovery, researchers use observations and detailed studies of natural processes to develop simulation models. Results from experimental studies (e.g. impacts of diverting precipitation or warming the soils from an ecosystem) extend the range of conditions under which the simulation models are valid. The computer models and simulation experiments test theoretical concepts and scientific hypotheses. Researchers validate the computer-modeling results with additional observations and experimental studies, leading, in turn, to the refinement of the theoretical concepts, the design of new experiments, and new computer models. This process, naturally, produces large amounts of scientific data. Scientists store and disseminate this data through archives and data centers, which are supported by organizations in government, academia, and industry. With computational modeling and simulation playing such a critical role in the scientific methodology, scientists increase the emphasis on validated datasets, value-added data products, and traceable information. Data quality and reproducible transformation of processes ensure trust in the credibility of scientific results. This trust is particularly essential in the study of climate change, where these results influence national and international policy. In other scientific domains, valid results translate into industry breakthroughs.

Thus, Linked Science must ensure meaningful collection, organization, classification, storage, discovery, access, transport, distribution, sub-setting, aggregation, dissemination, and visualization of large, diverse types of data. Data centers must include experimental, observational, and computer-generated data. This data varies in *scale* and *complexity*. Some domains operate on large individual datasets, indeed, so large that one file may not fit in memory. Others operate on large numbers of small files. In some cases, the complexity and size of datasets prevent visualization unless scientists first reduce the dimensions of datasets (i.e. the number of variables or degrees of freedom) by performing specialized analysis based on the features that they are interested in.

Access to the data poses another challenge. On the Web, with Linked Open Data, every resource has a unique identifier. In Linked Science, such uniform access is not always available because the data does not live in the cloud or on the Web, but in multi-institutional data centers. These data centers provide tools and add value through quality assurance, validation, curation, dissemination, and analysis of the data. This data also typically cannot be consumed by a browser, an audio or video reader, and usually require specialized applications that these data centers also provide.

Critically, each dataset must provide metadata in order to enable its meaningful use. Specifically, metadata must describe the way that the data was generated, potential errors, uncertainty or variability in the calculations and measurements. For instance, data produced by a given simulation run is not reproducible unless one has the input data and the values of input parameters set in the input script. For simulation runs on parallel machines and high performance systems, metadata should also include system configuration.

As the discussion above demonstrates, *metadata* about various aspects of the datasets is a critical component of the distributed and linked science today. Ontologies and semantic descriptions of the scientific data and processes provide the necessary entities supporting the production of new knowledge by allowing interoperability of the processes, shared annotations and integration of the data.

This paper makes the following contributions:

- We describe the motivation for semantic annotations in a large scientific domain, Earth and Climate Sciences.
- We present a tool that uses ontologies to improve the quality of search across heterogeneous big-data resources.
- We analyze the domain coverage that several existing ontologies in Earth Sciences provide for the large collection of datasets.

The paper is organized as follows. In the next section, we describe the domain application and provide a scientific scenario from the domain of

Earth and Climate Sciences where scientists must integrate heterogeneous data sources in order to perform a scientific investigation of a climate change scenario for river water transport. We also characterize the various types of datasets available in this domain. We describe the Mercury search engine, a tool for distributed metadata harvesting, search, and retrieval in Section 3. In Section 4, we motivate the use of ontologies in this scenario and describe a prototype tool to facilitate their use. We present the results that we obtained with the prototype and evaluate the ontology coverage in Section 5. Section 6 discusses related work and we analyze our results in Section 7.

2. Use case scenario and earth and climate science datasets

Consider the following scenario (Fig. 1). A hydrologist focuses on validating model simulation trends for nutrient transport within a river channel. The climate model simulation of the earth system used to investigate climate change has four components: land, sea ice, ocean and atmosphere [8].

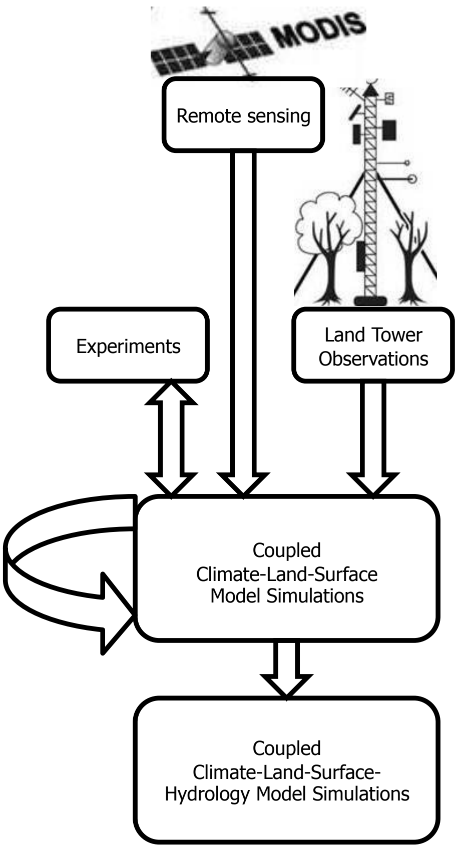

Fig. 1. Different types of data in the use case scenario.

The land component currently includes simulations of river flow. Future models of the earth system will contain biogeochemical species such as nitrogen, carbon, and phosphorus compounds (e.g., those contained in fertilizers). Changes in the chemistry of rivers from two different causes are relevant to climate change. First, biogeochemical species are washed from the soil, carried from water streams into larger rivers, and eventually end up in coastal oceans. Second, deforestation from biomass burning also causes changes to the chemical composition of the water that flows into rivers. The transport of biogeochemical species, particularly riverine nitrogen, may have an even larger effect: these species cause hypoxia (reduction in the oxygen concentration in water) and fish mortality in the coastal oceans [7]. In order to characterize these effects realistically, the hydrologist will need access to two types of data, which are generally available to earth scientists: (1) **computational** data that record the results of computer modeling and simulation; and (2) **observational** data that contain results of specific measurements. In our use case, the computational data will include models of river flow and transport of biogeochemical species; the observational data will describe stream flow, water quality, precipitation, air and water temperature, sediment data, biogeochemical species, and soil moisture.

For computational-model data, our hydrologist can turn to the Earth System Grid Federation (ESGF) gateway at the National Center for Atmospheric Research [1]. At the time of this writing, it contains 3,384 datasets of computational data totaling about 1.3 Petabytes of data and representing 368 variables. She will need to know, however, that file names in this source attempt to reflect variable name, such as "qchanr" (river flow), or "qchocnr" (river discharge into the ocean).

For observational data, the hydrologist can get data from the Gravity Recovery and Climate Mission [19] and the Tropical Rainfall Measuring Mission [20] from the National Aeronautics and Space Administration (NASA) to validate the outputs of the climate model simulation. These datasets contain remote sensing imagery for tropical precipitation and storage. Ground stream flow data is available from the United States Geological Survey (USGS). Fertilizer input and water-quality measurements may come from the Environmental Protection Agency and the US Department of Agriculture. The NASA-sponsored Distributed Active Archive Center at the Oak Ridge National Laboratory (ORNL DAAC) for biogeochemical data holds about 1,000 datasets (2

Terabytes) relevant to biogeochemical dynamics, ecological data, and environmental processes, as well as 60 TB from the MODIS Terra satellite land product subsets (measurements of surface radiance, reflectance, emissivity, and temperature).

A scientific user may typically be familiar with computational climate datasets, such as those found in ESG, or with observational earth and ecological science datasets such as those found in the ORNL DAAC, but not both. Both currently present their data in facetted searches along dimensions such as Project, Model, Experiment, Product, Variable Name, and Ensemble for ESG, and Parameter, Sensor, Topic, Project, Keywords in the ORNL DAAC. Note that in computational data the facet "Experiment" denotes experiments "in silico." In the observational data, one also finds "Models," a term typically reserved for simulations, where datasets are used in assessments and policy studies and simulate ecological systems: observational data can also be the result of simulations.

The data-processing tasks associated with the simultaneous use of observation and computational data, such as in our scenario, are daunting. Each data domain has its own portal, its own metadata formats, and its own query-building methods for obtaining datasets. The exact definition of variables and observational parameters may require substantial searches for unfamiliar topics. In order to advance investigation of climate change, scientists need access to formal descriptions of the multiple entities present in each activity and to the tools that permit seamless searches across all entities.

Thus, data solutions to the scientific question require the use of heterogeneous data. The hydrologist will also need to search for datasets from different data centers to discover useful data because each data center specializes in storing datasets relevant to their mission and focused on the needs of the sponsoring agency.

3. The Mercury tool: aggregating metadata

Mercury is a tool for distributed metadata harvesting, search, and retrieval. It was originally developed for NASA, and Mercury is currently used by projects funded by NASA, USGS, and U.S. Department of Energy (DOE) [6]. More than 30,000 scientists use Mercury each month.

Currently, Mercury provides a single portal to search quickly for data and information contained in disparate data-management systems. It collects metadata and key data from contributing project servers distributed around the world and builds a centralized index. It currently provides access to over 100,000 metadata records. The Mercury search interfaces then allow the users to perform simple,

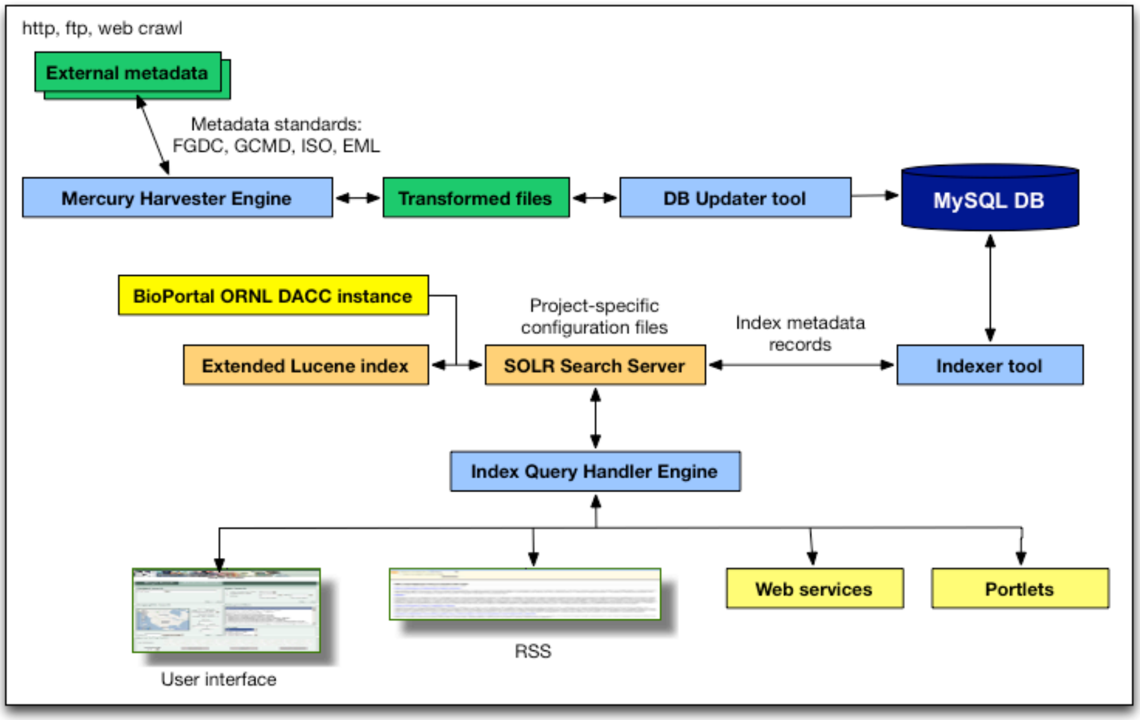

Fig. 2. Architecture of the Mercury Search Engine and its integration with BioPortal ORNL DACC instance. Blue boxes indicate reusable software components. Green boxes are metadata files. Yellow boxes are external services. The Mercury Search service calls BioPortal REST services to add ontology knowledge to the queries.

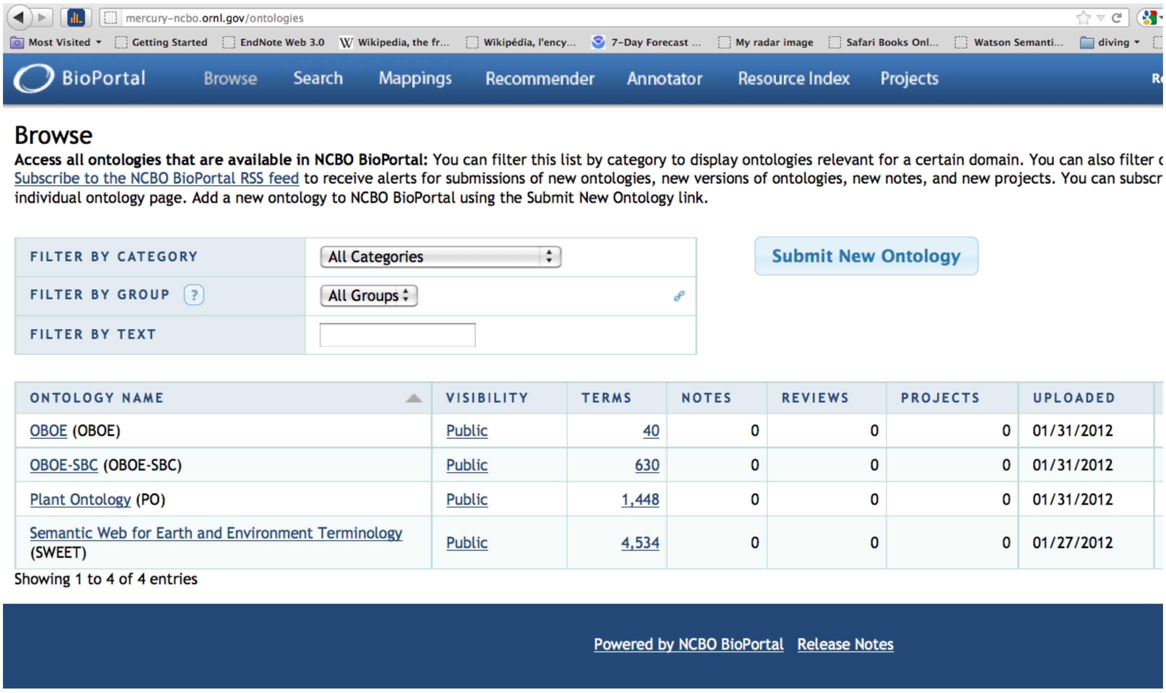

Fig. 3. BioPortal instance at ORNL DACC. The screenshot shows the ontologies that the repository currently contains.

attribute-based, spatial and temporal searches across these metadata sources. This centralized repository of metadata with distributed data sources provides extremely fast search results to the user, while allowing data providers to advertise the availability of their data and maintain complete control and ownership of that data. Fig. 2 shows a diagram of the Mercury architecture.

Mercury supports several widely used metadata standards and protocols such as the Federal Geographic Data Committee, Dublin Core, Darwin Core, the Ecological Metadata Language, the International Standards Organization's ISO-19115, XML, Library of Congress protocols Z39.50 and Search/Retrieve via URL (SRU), and Amazon subsidiary A9's OpenSearch.

The Mercury architecture includes a harvester, an indexing tool, and a user interface. Mercury's harvester typically harvests metadata records from publically available external servers. Data providers and principal investigators create metadata for their datasets and place these metadata in a publically accessible place such as a web directory or FTP directory. Mercury then harvests these metadata and builds a centralized index and makes it available for the Mercury search user interface. Mercury also harvests metadata records from external catalogs using the Open Archives Initiative Protocol for Metadata Harvest (OAI-PMH) [5] and other web-based harvesting techniques.

Mercury's query engine is built using a service-oriented architecture, which includes a rich user interface. This interface allows users to perform various types of search capabilities, including 1) simple search, which performs a full text search, 2) advanced search, which allows users to search against controlled-vocabulary keywords, time period, spatial extent and data provider information, and 3) web browser tree search, which enables a drill-down through the metadata facets using a hierarchical keyword tree.

4. Adding semantics to Mercury

With the breadth of sciences represented within the Mercury metadata records, scientists can potentially address some key interdisciplinary scientific challenges related to climate change and its environmental and ecological impacts, including carbon sequestration, advance of seasons, as well as questions related to the mitigation of these effects. However, the wealth of data and metadata also makes it difficult to pinpoint the datasets relevant to particular scientific inquiries.

We have applied semantic technologies—ontologies, in particular—to improve the relevance of search results. There are several reasons for using this approach. First, simply using popularity of datasets to determine their relevance typically is not useful in the case of scientific data queries. Each scientific inquiry tends to be unique, and what is relevant for one inquiry is not relevant for another. Thus, we must be able to rank search results based on the *meaning* of the data descriptions. Second, scientific queries are unlike everyday queries because they return specific datasets, which themselves have numerous parameters that may or may not be exposed to the search. For example, the Earth System Grid Federation (ESGF) gateway exposes 368 variables to search. These are deep-web queries. Third, each domain science has its own terminology, more or less curated and consensual, and with various degrees of standardization. The same terminology covers different concepts across domains (the semantic plurality problem), and different terms mean the same thing (the synonymy problem). Interdisciplinary research is arduous because a scientist who is already an expert in a domain must become fluent in the language of another, just to find the relevant datasets to start addressing a question. For all these reasons, we decided to use scientific ontologies because they can provide a context for search results, in a way that string-based keywords never will.

The Semantic Web for Earth and Environmental Terminology (SWEET) [18] is a mature foundational ontology developed at the NASA Jet Propulsion Laboratory. SWEET currently contains over 6,000 concepts organized in 200 OWL ontologies classifying 9 top-level concepts. For SWEET 2.3 these top-level concepts are:

- Representation (math, space, science, time, data)
- Realm (ocean, land surface, terrestrial hydroshere, atmosphere, heliosphere, cryosphere, geosphere)
- Phenomena (macro-scale ecological and physical)
- Processes (micro-scale physical, biological, chemical, and mathematical)
- Matter (living thing, material thing, material thing)
- Human Activities (decision, commerce, jurisdiction, environmental, research),
- Property (binary, categorical, ordinal, quantity)
- Role (physical, biological, space, chemical),
- Relation (human, physical, space, time, chemical).

We used the SWEET ontologies to improve the accuracy of the Mercury search interface. The ontologies provide context by linking individual keywords to a scientific realm and suggest additional keywords for searches.

In order to incorporate the SWEET ontologies and other ontologies that are relevant to earth sciences into the Mercury architecture, we chose BioPortal as an ontology repository. BioPortal is a community-based ontology repository developed by the National Center for Biomedical Ontology (NCBO) [16, 25]. While the instance of BioPortal that runs at NCBO is a repository of biomedical ontologies—with more than 300 of them at the time of this writing—the BioPortal software is domain-independent and there are several BioPortal installations that run ontology repositories in other domains. The BioPortal at ORNL DAAC is one such installation (Fig. 3). The portal allows users to browse ontologies and to search for specific ontologies that have terms that are relevant for their work. The mappings between ontologies in BioPortal not only allow users to compare the use of related terms in different ontologies, but also allow analysis of how whole ontologies compare with one another. BioPortal provides access to the ontologies through a REST interface, thus enabling easy integration with Mercury (Fig. 2).

In order to provide access to ontology entities in the ORNL DACC BioPortal instance, we designed an ontology service that allows integration of ontology entities into search results. The Mercury search system passes its search parameter to BioPortal, which returns one or several entities (classes, properties, terms) through the REST interface. The user can choose any of these entities as additional search parameters for Mercury, or directly display the results indicated by the ontology sub-class terms.

5. Results

As Mercury user interface already uses a facetted search approach, we can present the ontology results to the user in the same user interface (Fig. 4). In this figure, the five top boxes ("Filter by") show the facetted results without semantic search. The bottom four boxes ("Ontology") present the results of the semantic search. Unlike a facetted search that highlights attributes within a set of results but cannot enlarge the set, the semantic solution can implement both restrictions (improving precision) and expansions (improving recall) of the initial set of results.

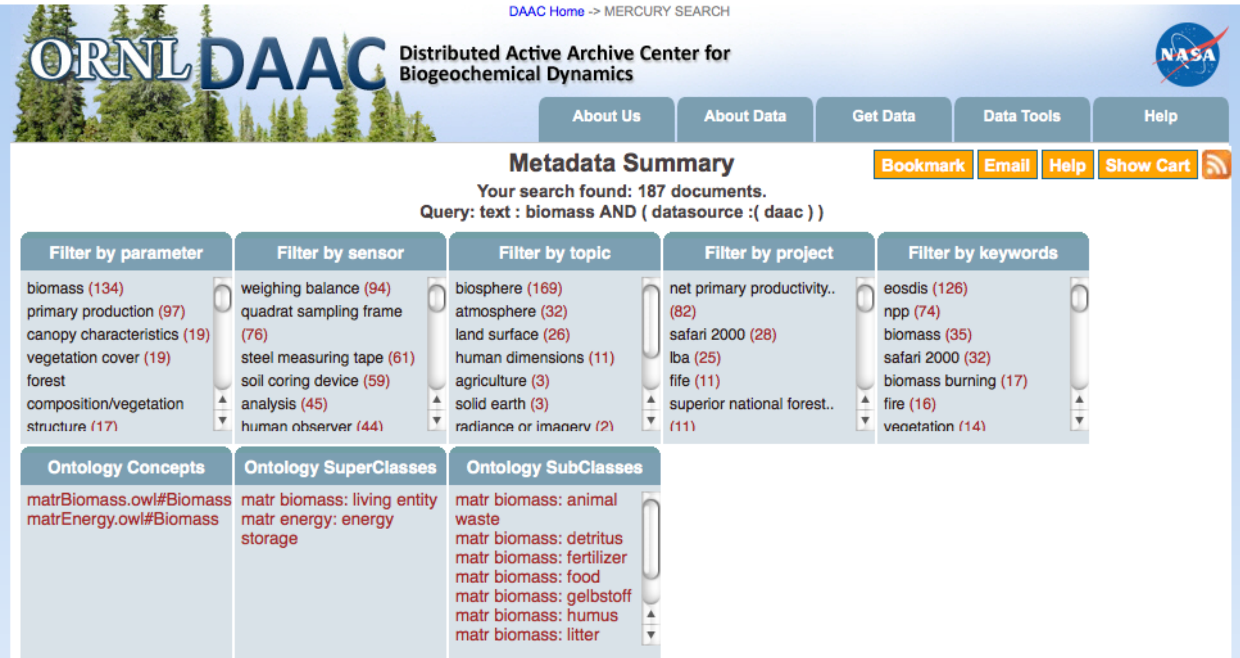

Fig. 4. User interface for the semantic search in ORNL DACC. The user has searched for "biomass" and the interface suggest additional related terms based on the ontology search.

Specifically, there are four facets powered by ontologies: Ontology Concepts, Ontology Super-classes, Ontology Sub-classes and Filter by keywords and all sub-classes. *Ontology concepts* present each search term within the ontological hierarchy. *Ontology Super-classes* shows the hierarchical level one level up and *Ontology sub-classes*—one level down. The facet "Filter by keywords and all sub-classes" the ontology service sends the sub-class terms to Solr, which returns links to datasets of interest (not shown in the figure).

The ontology service provides domain context, parameter attribute, and entity annotations to the Mercury search system.

5.1. Using ontologies to improve recall

Recall the scenario that we described in Section 2. Our hydrologist will need to search for datasets annotated with "biomass" because she wants to analyze the transport of biochemical species in the river flow. She will search of the ORNL DAAC for datasets containing the term "biomass." A Mercury search using controlled vocabulary keywords returns 35 datasets, a full-text search returns 187 datasets. A search for "biomass OR humus" (a type of biomass) returns 192 datasets, indicating that 5 potentially relevant datasets are not included in the search on biomass.

Querying the SWEET ontologies through BioPortal's REST API, the ontology service exposes "humus" as an additional search term for Mercury in the first discovery session about "biomass." Humus is a sub-class of biomass in SWEET. Thus, the semantic search returns the five additional datasets without the user having to know about specific types of biomass. "Biomass" also acquires scientific context when the ontology service exposes that it can be a form of Energy Storage and a Living Entity.

5.2. Using ontologies to improve precision

"Carbon" is another popular search term in Mercury, since the increase in the concentration of carbon dioxide in the atmosphere is considered a potential factor of climate change. A Mercury search for "carbon" returns 264 datasets from the ORNL DAAC. With the ontology service integrating the results of an ontology search into the facetted search, "carbon" acquires a scientific context and additional query terms that can be used to improve the precision of the original search. For example, the individual in one of the ontologies, "stateTimeGeologic2:Carboniferous," links results to datasets relevant to geological times (paleo-environmental science), while the sub-class "carbon offset" links to datasets relevant to "human environmental control" and "human activity." In addition, "offset" is not a facet offered by the Mercury search system but the ontology search suggests this sub-class to reduce the result set further. Limiting the search to both "carbon" and "offset" produces only two results.

5.3. Analyzing the coverage of ontologies

The BioPortal instance at ORNL currently contains 4 ontologies, with 6,652 classes among them (Fig. 3). In addition to SWEET 2.3, the collection includes the Plant Ontology, which describes structure and developmental stages of a plant [2], and the Extensible Observational Ontology (OBOE) for representing scientific observations and measurements [13], and its extension to represent ecological and environmental data.

We evaluated how well the terms in these ontologies cover the top 100 controlled-vocabulary keywords indexing datasets for the ORNL DAAC in Mercury. "Biomass" is the top keyword currently indexing 138 datasets.

Fig. 5 shows the results of this evaluation. 21 of the tops 100 keywords do not appear in the ontologies. Thus, 79% of the top 100 keywords in the ORNL DAAC have at least one match in the selected ontologies. At the long tail of the distribution one keyword (water) has 38 matches, and two (air, carbon) have 28 matches.

6. Related work

Researchers in the Semantic Web community have studied *semantic search* and a variety of approaches to it. A recent survey [21] presented a general model for semantic search and identified different types of semantic search. In general, there are two key approaches. In one, the (linked) data is represented in RDF or OWL and the search engine provides access to a collection of such data, either through keyword search or through SPARQL (e.g., SWSE [9], or Sindice [22]). Uren and colleagues provide a survey of this type of semantic-search engines [23]. The second class of semantic-search applications are document-retrieval applications that use semantics to

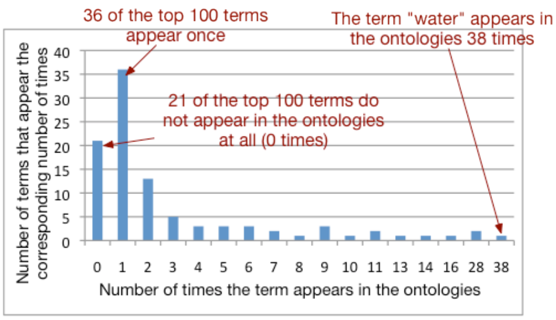

Fig. 5. Ontology coverage of the top 100 controlled-vocabulary keywords.

expand or constrain the user query (e.g., [3, 4]). The application that we describe here does not fall into either of these two categories, however. On the one hand, it provides access to heterogeneous collections of structured data, but this data is not represented in Semantic Web formats. At the same time, it uses semantics on the "front-end", augmenting the user query, but we use this query expansion to access structured data and not a set of documents. Thus, to the best of our knowledge, the application that we have described is unlike many semantic-search applications because it uses semantics on the query side but provides access to structured data, but not in RDF and OWL format.

Kauppinen and colleagues frame the challenges of linked science in the form of an "executable paper" [10], with publication of validated and well-sourced data as one of the key requirements. Contributions to the recent First Linked Science workshop [11] investigated several issues related to Linked Science and Linked Data but did not focus on semantic searches for structured datasets in dedicated archives. Researchers discussed the requirements for Linked Science in the geo-physical sciences [14]; the use of rules for interactively mapping data sources in databases to ontology and generating RDF triples [12]; the need for trust in the data sources with an emphasis on formally describing the relationship between data and sources in bibliographic resources [15]; challenges in the bioinformatics [24] and astronomy domains. Thus, researchers are actively addressing the trends in Linked Science and our effort is complementary to the approaches described in these papers.

7. Discussion

Our approach to the investigation of climate change has led to the programmatic integration of search capabilities and the development of a semantic service for discovering multi-disciplinary datasets in Earth and Environmental sciences. Scientists can use our semantic service to discover both the new datasets that were not included in the original search results (improving recall) and additional features for a search that they can use to restrict the number of results (improving precision).

We used a BioPortal instance as a source for ontologies rather than a triple-store or an OWL API to process the ontologies for several reasons. First, the REST service interface that BioPortal provided was easy to integrate into the Mercury architecture. Second, ontology authors sometimes use idiosyncratic approaches to representing some features of their ontologies, such as preferred names or synonyms for terms. These lexical features are key to user searches but ontologies use different properties to represent them. Even though SKOS, a W3C Recommendation for representing vocabularies on the Web provides standard properties for these features, our experience shows that ontology authors do not yet follow that recommendation. BioPortal uses ontology metadata to extract these properties and provides its users with a single service call to access this information across all ontologies in a repository. Finally, BioPortal enables scientists to submit new ontologies through its web interface and these ontologies become available to the semantic search in Mercury. Thus, if a scientist discovers a new ontology that covers her domain of interest, she can simply add it to her set of ontologies to expand the meaningful results from her semantic search..

We set up the ORNL DAAC instance of BioPortal because this user community needs a stable ontology repository that covers the Earth and Environmental Sciences domains. This instance of BioPortal is accessible to ORNL DAAC users with all the functionality provided by BioPortal, including annotations, ontology extensions, and term mappings. New community additions to the ontologies made through this instance are directly accessible to the semantic service.

However, our approach has several limitations. First, the facetted display becomes crowded very quickly and a more dynamic presentation of search results may be beneficial. Another, more serious, limitation is that the quality of the newly discovered metadata is contingent on the quality of the ontologies used in our implementation. BioPortal curates the ontologies by enforcing compliance to ontology language standards and resolving relationships and axioms to detect potential conflicts, but it cannot check for coverage or correctness in terms of domain expertise. Search terms and thesaurus keywords in the ORNL DAAC may be absent from current ontologies, or the ontology classification may not bring additional information that is not already presented by the facetted terms. However, as semantic technologies mature, more substantial ontologies become available in many scientific domains. And the mapping features in BioPortal will allow for new ontology entities to be related.

8. Conclusion

The solution that we presented in this paper leverages the federated search capabilities in Mercury that collect metadata records from several scientific domains, and the storage, access and curation functionality of BioPortal. With minimal additional development, this approach builds on two mature systems and enables finding relevant datasets for interdisciplinary inquiries. The paper thus indicates a direction for linking environmental, ecological and biological sciences.

References

1. Bernholdt, D., Bharathi, S., *et al.*, The Earth System Grid: Supporting the next generation of climate modeling research. Proceedings of the IEEE, 2005. 93(3): p. 485-495.
2. Bruskiewich, R., Coe, E.H., *et al.*, The Plant Ontology (TM) Consortium and plant ontologies. Comparative and Functional Genomics, 2002. 3(2): p. 137-142.
3. Castells, P., Fernandez, M., and Vallet, D., An adaptation of the vector-space model for ontology-based information retrieval. Knowledge and Data Engineering, IEEE Transactions on, 2007. 19(2): p. 261-272.
4. Chu-Carroll, J., Prager, J., Czuba, K., Ferrucci, D., and Duboue, P. Semantic search via XML fragments: a high-precision approach to IR. In Proceedings. 2006: ACM. p. 445-452.
5. Devarakonda, R., Palanisamy, G., Green, J.M., and Wilson, B.E., Data sharing and retrieval using OAI-PMH. Earth Science Informatics, 2011. 4(1): p. 1-5.
6. Devarakonda, R., Palanisamy, G., Wilson, B.E., and Green, J.M., Mercury: reusable metadata management, data discovery and access system. Earth Science Informatics, 2010. 3(1-2): p. 87-94.
7. Doney, S.C., The Growing Human Footprint on Coastal and Open-Ocean Biogeochemistry. Science, 2010. 328(5985): p. 1512-1516.
8. Gent, P.R., Danabasoglu, G., *et al.*, The Community Climate System Model Version 4. Journal of Climate, 2011. 24(19): p. 4973-4991.
9. Hogan, A., Harth, A., Umbrich, J.r., Kinsella, S., Polleres, A., and Decker, S., Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine. Web Semantics:

Science, Services and Agents on the World Wide Web, 2011. 9(4): p. 365-401.
10. Kauppinen, T. and de Espindola, G. Linked open science? Communicating, sharing and evaluating data, methods and results for executable papers. In Proceedings of International Conference on Computational Science, ICCS 2011. 2011, 4. p. 726-731.
11. Kauppinen, T., Pouchard, L.C., and Kessler, C., eds. Proceedings of the First International Workshop on Linked Science (LISC 2011). Vol. CEUR Workshop Proceedings 783. 2011.
12. Knoblock, C.A., Szekely, P*., et al.* Interactively Mapping Data Sources into the Semantic Web. In Proceedings of First International Workshop on Linked Science (LISC2011) at ISWC 2011. 2011. Bonn, Germany, CEUR Workshop Proceedings 783.
13. Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., and Villa, F., An ontology for describing and synthesizing ecological observation data. Ecological Informatics, 2007. 2(3): p. 279-296.
14. Mäs, S., Müller, M., Henzen, C., and Bernard, L. Linking the Outcomes of Scientific Research: Requirements from the Perspective of Geosciences. In Proceedings of First International Workshop on Linked Science (LISC2011) at ISWC 2011. 2011. Bonn, Germany, CEUR Workshop Proceedings 783.
15. McCusker, J.P., Lebo, T., Ding, L., Chang, C., Pinheiro da Silva, P., and McGuinness, D.L. Where did you hear that? Information and the Sources They Come From. In Proceedings of First International Workshop on Linked Science (LISC2011) at ISWC 2011. 2011. Bonn, Germany, CEUR Workshop Proceedings 783.
16. Musen, M.A., Noy, N.F., Shah, N.H., Whetzel, P., Chute, C.G., Storey, M.A., and Smith, B., The National Center for Biomedical Ontology. Journal of American Medical Informatics Association (JAMIA), 2011.
17. Mwebaze, J., Boxhoorn, D., and E., V. Supporting Scientific Collaboration Through Class-Based Object Versioning. In Proceedings of First International Workshop on Linked Science (LISC2011) at ISWC 2011. 2011. Bonn, Germany, CEUR Workshop Proceedings 783.
18. Raskin, R.G. and Pan, M.J., Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Computers & Geosciences, 2005. 31(9): p. 1119-1125.
19. Tapley, B.D., Bettadpur, S., Watkins, M., and Reigber, C., The gravity recovery and climate experiment: Mission overview and early results. Geophysical Research Letters, 2004. 31(9).
20. Theon, J.S., The Tropical Rainfall Measuring Mission (Trmm). Remote Sensing of Earths Surface and Atmosphere, 1993. 14(3): p. 159-165.
21. Tran, T., Herzig, D.M., and Ladwig, G.n., SemSearchPro--Using semantics throughout the search process. Web Semantics: Science, Services and Agents on the World Wide Web, 2011. 9(4): p. 349-364.
22. Tummarello, G., Delbru, R., and Oren, E. Sindice.com: Weaving the Open Linked Data. In Proceedings of International Semantic Web Conference (ISWC-2007). 2007. Busan, Korea: Springer Berlin / Heidelberg, 4825. p. 552-565.
23. UREN, V., LEI, Y., LOPEZ, V., LIU, H., MOTTA, E., and GIORDANINO, M., The usability of semantic search tools: a review. The Knowledge Engineering Review, 2007. 22(04): p. 361-377.
24. Vision, T., Blake, J., Lapp, H., Mabee, P., and Westerfield, M. Similarity Between Semantic Description Sets: Addressing Needs Beyond Data Integration. In Proceedings of First International Workshop on Linked Science (LISC2011) at ISWC 2011. 2011. Bonn, Germany, CEUR Workshop Proceedings 783.
25. Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C.I., Tudorache, T., and Musen, M.A., BioPortal: Enhanced Functionality via New Web services from the National Center for Biomedical Ontology to Access and Use Ontologies in Software Applications. Nucleic Acids Research (NAR), 2011. 39(Web Server issue): p. W541-5.