EarthCube

Roadmap for Creating the Semantic/Ontologic Infrastructure for the Geosciences

Prepared by Semantic/Ontology Community Group

Roadmap for Creating the Semantic/Ontologic Infrastructure for the Geosciences

Content

Submitted by Krishna Sinha, PI for NSF award 1238438,
Drawing the roadmap for the semantic/ontology based infrastructure for Geosciences

Report dedicated to the memory of Robert Raskin (1957-2012)

Summary and Overview

A. Krishna Sinha
Department of Geoscience
Virginia Tech

An integrative view of the earth, based on multi-disciplinary data, has become one of the most compelling reasons for research and education in the geosciences. It is necessary to create a modern infrastructure that can support the transformation of data to information to knowledge. Such an infrastructure for geosciences constitutes the vision of EarthCube. It is now possible to conduct web based "smart searches" that deliver data of interest to the user, as well as visualization and computational capabilities that promote a better understanding of the science associated with earth processes. Such capabilities lie at the foundation of EarthCube whose ultimate goal is to facilitate the use of complex, multidisciplinary data in seeking solutions to geoscience based societal challenges, and a deeper understanding of the earth as a system.

Semantics and ontologies play a key role in enabling the vision of EarthCube, and was the primary focus of a workshop held in Arlington, VA. Video, reports and presentation materials related to the workshop are available http://earthcube.ning.com/group/semantics-and-ontologies. The role of semantics and ontologies in both research and teaching is clearly evidenced by the recognized need to share, access, discover, integrate and model data (SADIM) towards new knowledge for the geoscience community (Figure 1). It is significant that geoscientists around the world are working towards the goal of discovering new knowledge through a better understanding of the fundamental principles behind complex and heterogeneous data products: a foundation for why the data values are, what they are, or an indication as to how observations would change over time through physical, chemical and biological processes. The vision of EarthCube will support a new generation of computational thinking that will enable knowledge discovery, and markedly broaden our understanding of geoscience and allied sciences for solving challenging and complex problems previously not even imagined.

The path defined by SADIM requires collaborative work between geoscientists and semantics/ontology experts. Making sharing of data a simple task for a data provider through innovative use of metadata as well as semantic tags for data would enable many thousands of geoscientists to act as data nodes. Similarly community based development of ontology enabled tools and services (APPS as used by https://explore.data.gov/catalog/apps/) could be automatically linked to users choice of datasets, and thus facilitate reuse of tools and models. These data and service ontologies have to be robust and should utilize semantically enabled registration technologies. However, significant challenges will have to be addressed in developing new semantically enabled techniques to evaluate data quality, provenance as well as management of legacy data. Ontologies that facilitate access and discovery of all data types will require many semantic technologies, such as ontology and vocabulary mapping, and development of data ontologies at various levels of granularity. Making existing and emerging ontologies available for reuse and modification can only be achieved through the establishment of a research oriented ontology repository capable of developing and sharing new semantic technologies e.g. ontology alignment and metadata extraction. As geoscientists foresee open access to data and services, integration of data will require new semantically enabled software engines that remove structural, syntactic and semantic heterogeneities, and enable the application of appropriate computational tools through recommender services.

For the Geoscience community wanting to engage in building, developing and monitoring the semantically enabled infrastructure, it will require self organization, as well as a significant outreach to the both the geoscience and computer science communities. It is critical that adoption of semantic technologies be encouraged to meet all the goals envisioned by EarthCube.

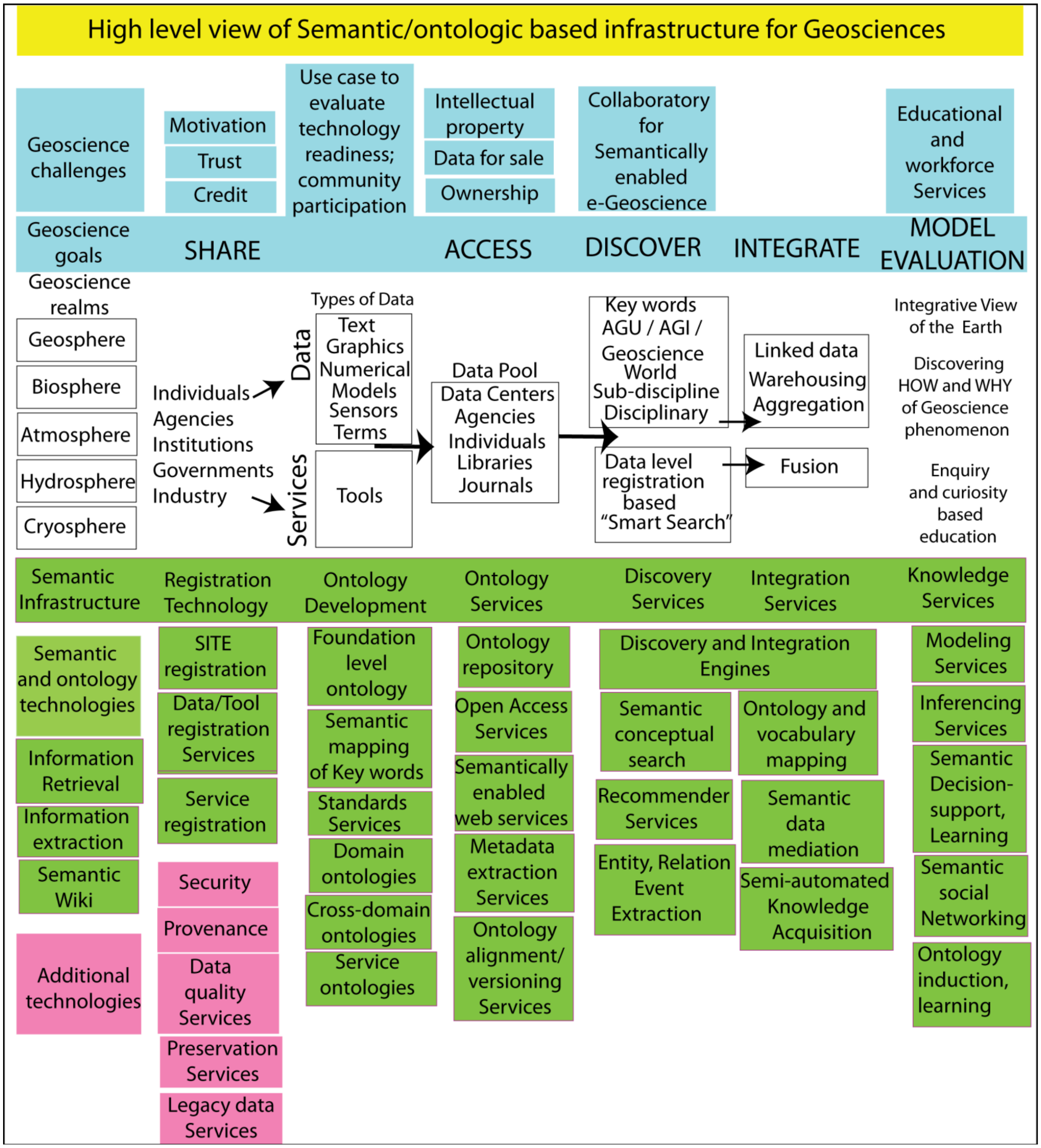


Figure 1. A generalized RoadMap for developing the semantic/ontologic infrastructure for the Geosciences. Facilitating a range of required activities for geoscientists that range from sharing data and tools to model evaluation leading to knowledge discovery requires a strong partnership with ontology/semantics expert community. Providing technologies to enhance enquiry and curiosity based education will prepare the workforce for the future. All technology requirements identified in the accompanying report are represented in this diagram, as well as others that may become significant as the infrastructure evolves. Some of the semantic technology capabilities presented in the figure were provided by Leo Obrst, Anne Thessen and Nancy Wiegand

RoadMap for semantic/ontologic infrastructure for the Geosciences

This document is a reformatted and edited version of the initial panel reports prepared at the Semantics/Ontology community group workshop during which each panel addressed one of the ten roadmap topics provided by NSF. The workshop took place in Arlington on April 30th-May 1, 2012. Panelists who led the workshop discussions and person(s) who submitted the initial written report are identified for each topic. Editing of the preliminary workshop report was done by Krishna Sinha and Nancy Wiegand. Reports by the Technology and Joint committees are presented as written by the committees.
Original unedited version of the panel reports, workshop agenda and presentation materials are available at http://earthcube.ning.com/group/semantics-and-ontologies/page/workshops.
Complete video recordings of the workshop proceedings are available at
[*https://vimeo.com/groups/140684*](https://vimeo.com/groups/140684)

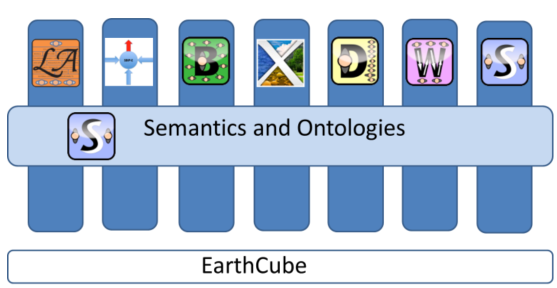
1. Purpose: *Introduction, including community(ies) to be served, technical area(s) of the roadmap, and brief discussion what improvements in the present state-of-the-art in geoscience data discovery, management, access, or utilization it will enable. Also include examples of how the outcomes from your effort will enable the community to be more productive and capable.*

Panelists: Peter Fox (initial report), Boyan Brodaric, Naijun Zhou, Calvin Barnes

Introduction: The purpose of this community group is to create a roadmap for EarthCube for developing and deploying a semantically enabled infrastructure for the Geoscience community and citizen scientists. To meet this objective, existing and new technologies will be deployed for modeling and ontology development, knowledge and data services, as well as tools to integrate and analyze data.

Semantics and ontologies cover a number of research areas such as the development and use of ontologies to standardize the meaning of terms, support the use of standards, resolve heterogeneous terms as well resolve terms across subdomains (bridging). Ontologies are also needed in understanding metadata and for provenance-aware services. Smaller ontologies represented as Ontology Design Patterns (ODPs) are also useful technologies. An ontology repository is needed to provide management for versioning services for all geoscience related ontologies. The repository will facilitate new methods to find, access, and align ontologies.

Semantics and ontologies fit well with other EarthCube groups for interoperability and brokering. In fact, semantics is cross-cutting across all the groups as shown in the graphics below.



Communities: Communities to be served are: Geosciences (atmosphere, ocean, earth sciences), plus geography and bordering communities, such as biology, ecology, environmental science, computer science, and social science. Additional discussion is required for including application communities as well as citizen scientists, educators, industry, and policy makers.

Improvements: One of the most significant improvements will be greater accessibility to geoscience resources (e.g., data, models, tools, workflow templates). However, a geoscience community challenge exists in identifying incentives for scientists to contribute data, as community support is required for the success of the semantic infrastructure. Making more and relevant data available to earth-scientists readily and easily for the research they are performing, without them knowing anything about the underlying semantics, will result in increased productivity of research results. That is, they will have access to data they never knew was there.

Technical Areas: Semantic/ontology enabled technical areas are related to search and discovery, interoperability, and use of complex applications through tools and services such as knowledge modeling and reasoning. It is clear that **semantically enabled** improved data discovery can be a trigger to community participation (long tail of science) for data and tool sharing.

Use cases: Use cases must be powerful enough to get buy in by geoscience disciplines. Additionally, it is important for use cases to capture sufficient context of the data for it to be utilized by other communities as well. If EarthCube is implemented through utilization of use cases, then such cases must be carefully defined in an open environment. Carefully composed, cross-cutting use cases (described with use-case templates for uniformity) can drive development of applications. At least some degree of cross-cutting generality should be the criterion for the selection of suitable use cases.

Outcomes: A **shorter term goal** of the cyberinfrastructure is **greater data availability** to scientists. The **longer term goal** is **knowledge availability**. Improvements to the technical areas of search and discovery, more complex applications, interoperability and quality assessment, through the use of semantics, will enable the community to be more productive and capable in supporting the use of geoscience knowledge in meeting societal challenges. It also is well recognized that semantics is the key to quality assessment of data. Enabling such quality assessment and to propagate the knowledge needed across brokering and other technology layers is an important component of EarthCube capabilities.

Timeline: A roadmap is useful to develop a realistic timeline (immediate opportunities leading to longer term projects). This timeline would be revisited and revised often to reflect developments. Again, there is general consensus that short term goals should emphasize greater data availability, while longer term goals should support knowledge availability. Although not identified, questions were raised as to metrics to characterize success based on a timeline. Quantifying success along the time line would be very important for different community sectors.

2. Communications: *Description of a communications plan with end users, developers, and sponsors, as well as links to and feedback from other EarthCube community groups and EarthCube concept projects to promote systems integration and accelerate development.*

Include a discussion of needed interactions with allied fields, agencies, and other related activities (present and desired).

Panelists: Anne Thessen, Krishna Sinha (initial report), Tim Finin, Bob Arko

Liaisons: The need to communicate the application and development of semantic technologies must be centered on developing an aggressive outreach plan, including identifying community members as liaisons with other EarthCube community groups and with other groups (e.g., ESIP, Ontolog, etc.). In this regard, an Outreach Committee was established (see Appendix 1). Two additional committees were also established (to develop a potential use case and to document the status of current technology, see Appendix 1). It should be noted that the Semantics and Ontologies group itself already has a strong mix of domain and technical people, who, furthermore, have extensive contacts with other relevant people nationwide.

Workshops, Publications, Repositories, and Demonstrations: To engage the broader geoscience community, workshops in conjunction with established geoscience and other conferences should be organized. Publication of special volumes on semantics for geosciences would also become a useful outreach and educational objective. Such challenges could also be more readily met through the publication of a high level article clarifying the goals of semantics in geoscience research and education. New communication strategies to enable open participation would require an easily accessible repository of vocabularies and higher level ontologies. Additionally, to reach the broad spectrum of geoscientists (long tail of science) demonstration projects documenting success stories would be necessary.

Summary: It was agreed that communication plans must provide a strategic and practical approach that would bring a fundamental understanding of semantics and ontologies (S/O), and its application in EarthCube, to the forefront in the minds of NSF EarthCube leaders, Geoscientists and other functional users. This will be enabled by:

- Sharing information and creating awareness of the S/O and their benefits, importance, and priority through timely, open and two way communications and knowledge sharing. Demonstration of success stories would be significant part of this goal.
- Sustaining interest in S/O development throughout the life of the roadmap through communicating updates regularly (right information at the right time) and through both virtual and face to face workshop meetings etc.
- Development of teaching material for conveying foundations, methods, technologies, and tools, targeted at Geoscientists user community.

3. *Challenges*: *Description of major drivers, trends, and shifts impacting or that could impact the focus of a working group, including but not limited to changing technology, adoption culture, and community engagement.*

Panelists: Ruth Duerr, Pedro Szekely, Xiang Li (initial report), Cyndy Chandler

Diverse areas: The most significant challenge faced in EarthCube is most likely due to the diverse research and application areas that a semantically enabled infrastructure will support, including data discovery, access and integration, as well as knowledge extraction and representation. Development of foundation ontologies and related domain based ontologies (developed by either top-down or bottom-up approaches), and access to tools and services required for this infrastructure are the keys to the success of EarthCube. As the geosciences cover a range of science domains with processes in these domains being often inter-related, development of ontologies requires involvement of disciplinary communities as concepts and terminologies need to be clarified and agreed upon for these domains.

Further, development of ontologies for knowledge representation in these domains and their integrations requires collaboration and coordination of many people and resources. Because of this diversity, ontology design patterns and a modular approach to ontology modeling may be a good approach.

Ontologies: Existing ontologies related to various aspects of the semantic infrastructure need to be organized while new ontologies at all levels of granularity are needed to describe data products which are required for data discovery and integration . It is well known that various data formats have been used for data products in geosciences domains. Therefore, data ontologies need to accommodate as broad a spectrum of formats as possible. Similarly, many data services have been and will be developed to serve the use of data products. As a result, service ontologies developed need to be generic and flexible enough to support services in all these domains and be capable of evolving to handle future developments. To support the vast scope that EarthCube encompasses, the developed ontologies must be accurate to reflect the science domains, but be flexible enough to adjust or adapt to advances in these areas. As there are many different standards to specify metadata for data products in these science domains, development of ontologies will require considerations of domain standards, as well as use specifications provided by standards from FGDC , OGC and others. An infrastructure challenge is that today's geo-web and semantic web are largely incompatible. More work is required on 'semantic enablement' of spatial data infrastructures.

Sociological aspect: Building the semantic infrastructure must be community-based and considered as a long term and evolutionary process. How to motivate scientists and organizations such as data centers, as well as other community members to contribute data and expertise will be another challenge as the benefits of this semantic based framework may happen incrementally. Clearly deployment of techniques to help innovate the process of community conversations, engagement, motivation, creativity, and how meeting/ workshop processes can be improved for knowledge intensive teams is critical for the establishment of the semantic infrastructure.

Technology: It is recognized that there may be limitations of semantic technologies which may not fully support the vision of the semantic infrastructure. For instance, current semantic applications are built using OWL technology to describe the semantics and for knowledge inference and extraction. However, OWL has limitations in semantic capabilities that are needed, such as addressing issues of fuzzy ideas and computations. Further, limitations of tool support for these semantic applications may also pose a challenge.

4. Requirements: *Process (es) to be used to get the necessary technical, conceptual, and/or community (i.e., end-user) requirements at the outset and during the life of the activity, including approaches to achieving community/end-user consensus.*

Panelists: Mark Schildhauer, Xiang Li (initial report), Pedro Szekely

Joint domain/semantic workshops: Since EarthCube intends to better serve the broad areas of geoscience, the first and key step towards the success is to better understand what geoscience communities need, and what the major obstacles that scientists face are in their daily research. This information will help technology experts understand the needs of domain scientists, and offer solutions to enable them to discover and integrate heterogeneous data. Participating in such sessions will motivate scientists to be more involved in the development of EarthCube. Narrowing the knowledge gaps between domain and technology experts is required for the success of EarthCube. Therefore, it was suggested that there is critical need for thematic workshops involving domain scientists facilitated by semantics experts. Development of use cases is very helpful to technology experts in understanding domain requirements. Additionally, it was suggested that there is need for semantics/ontology workshops involving KR (knowledge representation) engineers and those involved in semantic web technologies to enable timely communication about technology development, and their strengths and limitations. It was suggested that development and promotion of ontology design patterns may help enhance interoperability. The semantics group should support development of standard mechanisms for annotations, as well as identify best of class, highly useable tools to assist in semantically annotating data. Also, "advertising" foundational ontologic framework among EarthCube participants would enable them to "standardize" their technologies. Therefore, close communication and involvement from other working groups in EarthCube is required so that there is a consistent common vision for all working groups. Finally, there is need to identify incentives to keep participants enthusiastic and involved.

Other Projects: It was also suggested that the semantics group capitalize on the success of other semantic enabled projects such as ontology development and integration for biosciences, and related efforts such as DataONE, NeOn, Data Conservancy, NCAR, OOI, IOOS, etc.

Websites: It was recommended that Websites/portals need to be developed to facilitate communications among various groups during development of EarthCube, including domain and technology experts, data service centers, organizations for standards, and individuals who are interested in geosciences applications. Use cases, requirements and technology advancement can be posted to websites for comments and information sharing.

As semantics is considered to be the key component for developing EarthCube, any development undertaken in semantics should fully support the vision and progress of other working groups.

5. Status: *Description of the state of the art within the topical area of your roadmap. This should include approaches and technologies from geoscience, cyberinfrastructure, and other fields, the public or commercial sector, etc. that have the potential to benefit the EarthCube enterprise.*

Panelists: Anne Thessen, Peter Fox (initial report), Boyan Brodaric, Clinton Smyth

Long tail: Currently, a significant amount of Earth Science data sits on individual scientists' personal computers. These data are not available to others and eventually these data are lost as scientists retire or move to other positions. To capture and support an individual scientist's data there is need for long-term stewardship. An outstanding question to be faced is: How is the young scientist incentivized to share data? Handling this is an important but long-term goal which might be started on by a social process that institutionalizes social rewards for sharing data as well as expertise.

It is of significance that the vast majority of data creation is through the efforts of individual scientists (also referred to as the long tail of science). How can such data be shared/integrated/discovered worldwide? Discoveries are made by the individual scientist, and getting into the shared data space is a significant challenge. What is needed includes the development of specific formats, ontologies in different fields, semantic similarity measures and matching, and mapping within linked data. There could also be more communication among investigators and more integrated collaborative projects being funded.

Computational sciences: Advances in Web scale information retrieval, big data management, cloud computing, massively parallel computing, information extraction from text, machine learning, and semantic web technologies are enabling many disciplines, including the geosciences, to obtain and analyze heterogeneous data.

Informatics: 'State of the art' is the informatics approach involving small teams of domain experts, data curators, and computer science professionals working collaboratively to develop the infrastructure that meets the requirements identified through use case development.

Relation to other EarthCube topical areas: Emerging workflows also need to be semantically driven. Machine processes for experiments can be semantically driven by instructing a robot to re-do an experiment based on the interpretation of results. New ideas can be generated.

Data: Data currently are in DB tables (or Excel spreadsheets), but there is the Linked Open Data (LOD) initiative. Research is now being done on how to enable databases to appear as linked open data. It is estimated that 25% of Linked Open Data is geospatial (geographic?) data (or at least contains geographic reference, e.g., places or locations).

Tools: It was recognized that there was a need for tools that make semantics interesting and easy to use for the domain scientists in managing small projects that are populated with existing vocabularies (e.g. GeoSciML). Tools such as CMaps (cognitive maps) and TreeList Editor for editing taxonomies (available at www.georeferenceonline.com/TLE), as well as informal link and node mechanisms need to be advertised and exploited.

Ontologies: Existing vocabularies tend to be more terminological systems than ontologies with rich semantics.

Scaling: Scalability is forefront in many other disciplines, e.g. Life Sciences. Current ontologies were produced by humans with process knowledge to create semantic structure and nodes. Getting the knowledge into a formal structure, in a semantically consistent way, is a huge bottleneck. Scalability – there areTB, even PB of data. Studies with user interactions show that slower application response times tend to be less likely to be used. Bigger data systems (e.g., Facebook, Twitter) must be scaled horizontally rather than vertically to handle increasing user demands. Vertical scaling is not a cost effective long-term solution. Tweaking and simplifying semantic models and inference rules as a way to increase performance is not a cost effective solution and may be considered a form of vertical scaling. Horizontal scaling may include sharding and distributing queries and rule engines across a distributed triple store. The cloud is now widely used, Hadoop etc., and provides a better ability to handle larger databases. The number of semantic tools that are available is growing and freely available (see Appendix), and provides evidence of substantial progress in creating the infrastructure.

Language Options: Using controlled or restricted English, domain experts can express input that then becomes knowledge formalized by logic in the system. Ontology Extraction, going beyond language extraction, goes into relations and events, looks for entities and descriptions and how they relate taxonomically, sometimes can be easier to obtain than arbitrary events. Curation of ontologies is required. A documented workflow is required with a review process where people are responsible for curating the data and the semantics that describe the data. Current status is a wide variety of levels of language encoding needs for geoscience.

Assessing Benefits: A matrix of evaluation requirements may be beneficial. For example, some applications are real-time that may need "dynamic" ontology; thus a successful solution using predefined ontology may not work well here.

Metadata: We need to raise the bar. Currently, the minimum bar is of the type required in FGDC. With the capabilities enabled by web services infused with semantics and ontologies, we can raise the bar to include information sufficient to assess data quality.

Other Disciplines: The Life Sciences depend heavily on names, names management, digitization, mining, and semantically modeling species morphology. The major challenges for the Life Sciences revolve around liberation of data from text and the long tail of small providers. The state-of-the art for the former is TaxonFinder and Neti for finding taxon names and CharaParser for extracting morphology information. The latter will prove to be much more difficult as it requires buy-in from the community. Annotators and data repositories exist, such as Dryad (which works with publishers) to try to capture long tail data. We look to GenBank as an example of a data repository that has excellent community support. The state-of-the-art in Life Sciences' semantics includes projects like TaxonConcept, which represents species on LoD and Phenoscape, which, as its goal, performs inference to find evolutionarily important genes.

Life Sciences are unique in that taxonomic names are an important and near universal piece of metadata. The state-of-the-art in names management is represented by the Global Names Project and the Encyclopedia of Life demonstrates the beginnings of what can be accomplished by managing data around names.

6. ***Solution****: Process for the identification and comparison (pros and cons) of approaches and technology solutions that will contribute to the EarthCube goal of satisfying current and future research needs of the geoscience end-user.*

Panelists: Amit Sheth (initial report), Pedro Szekely, Ruth Duerr

The discussion of this topic first separates users from developers regarding solutions. The scope of technology solutions is then presented, **although a more comprehensive report by the Technology Committee is found in Appendix 4**.

Users: The process of identifying and developing solutions calls for realizing that we have two communities to serve – the community of users and the community of developers. The community of users will generally not be interested, nor should be interested, in specific technologies (e.g., Semantic Web technologies) underlying their applications and solutions. They will likely be exposed, however, to the ontologies that describe the concepts they will use in posting queries or questions to a system that will help them get information and insights they need. Even if the users are themselves not interested in details of semantic technologies, it would be valuable for them to be aware of "semantics-empowered" (analogy is "Intel inside") message, so that gradually they are a participant in making the solutions more powerful by being better users of semantics and by providing better semantics in the form of domain knowledge that can make systems incrementally more powerful (e.g., through ontology evolution).

Developers: The developer community should be aware of, and whenever possible use, relevant standards and community developed/adopted specifications that have good tooling support, such as those from W3C's Semantic Web initiative and other community efforts in Semantic Web (RDF/RDFS, OWL, SPARQL, Linked Open Data-LOD), W3C's Semantic Sensor Networking, (Semantic) Web Services (SAWSDL, SA-REST, WADL), OGC's Sensor Web Enablement, and ontologies such as SWEET. They should also learn from, and adopt with appropriate changes, the architecture of successful systems in other domains that support complex analysis, discovery, problem solving, and decision making. Such architectures may already encompass one or more of the following: ontology development and evolution; semantic annotation for a broad variety of data (text, images, video, sensor data of various modalities, etc.) and models; semantic search/browsing/filtering/querying; advanced semantic processing; and semantic reasoning (path and pattern finding, inferencing). Reasoners and ontology alignment systems are crucial for a highly heterogeneous and interdisciplinary setting.

The EarthCube community should support development and public sharing of open source vocabularies and ontologies, semantically annotated data, workflows and tools, use cases and best practices, challenges to evaluate diverse solutions to a common problem, and demonstration of successful applications and systems serving end user (scientist) needs. The latter can play a key role in adoption. Support for a semantic infrastructure or resource for the EarthCube community to host a registry and provide common services, such as one patterned after the National Center for Biomedical Ontologies for the biomedical domain, can be considered. Quality of data and annotations, including support for provenance, should be a key capability of such a resource. Linked Open Data (LOD) has already become a key semantic approach for data sharing, and is anticipated to also play a major role in the EarthCube community.

7. **Process:** *Process (es) to develop community standards, protocols, test data, use cases, etc. that are necessary to mature the functionality of the topical area and promote interoperability and integration between elements of EarthCube.*

Panelists: Calvin Barnes, Cyndy Chandler, Naicong Li, Philip Murphy

It was emphasized that communication within the community, providing technologies for use case capture and development and data discovery were priority themes. This section discusses the processes needed to create the solutions for these themes. Again, our shorter term goal is data availability with the longer term goal of reasoning and knowledge discovery.

Use cases: In the short to midterm we need to develop a small number of use case scenarios that bridge disciplines and that are fully transportable to other EarthCube working groups, intellectual liaisons, and other organizations active in the area. **A summary of criteria to choose and evaluate use cases is given in Appendix 2.** An example for volcanism as a use case is also found in Appendix 2.

The following text has discussions from the workshop panel on use cases. It was emphasized that we need to create a process for gathering use cases, extracting core use cases for communication and reuse leading to gap analysis. The process for generating use cases can start with user meetings. If there are multiple use cases and not enough time to pursue all of them, then criteria from Appendix 2 will be used to select the best candidates. It will be important to address governance issues such as how to arrive at consensus about a use case within the community. An assessment of which realms of geosciences can be used to develop reusable ontologies and which ones will require domain-specific ontologies is an important step in developing the semantic infrastructure.

It was suggested that it was effective to work with early adopters to develop use cases to engage members of the research community. Use cases must have characteristics that recognize data heterogeneity as well as improving our thinking about models.

One approach for EarthCube might be to create a registry to store use cases. It was emphasized that discovery of data-related use cases would be highlighted, as scientists' ability to find new data relevant to their research generates immediate interest.

The current EarthCube survey is a good source of information for use cases (200+), but most are open ended and difficult to organize, which leads to suggestions that use case "templates" should be developed to help with organization. Use case templates should include a short form with several components including: summary statement, step-by-step description of the normal flow of the system being developed, an activity diagram as a pictorial representation of the steps and a concept map that would eventually develop into an information model. The words used in the use case and the concept map set the stage for semantic analysis and lead to new technology development. Through identifying and curating use cases, we provide direction to areas of content development: which ontologies need to be developed for which domains.

Creating ontologies: Once one or more use cases are identified, a workshop that includes domain scientists, knowledge engineers, and facilitators could be held to fully understand the use case, identify what kinds of ontologies are needed, and then start to populate/create the ontologies.

Further, the Semantics and Ontologies group needs to encourage and guide deeper use of semantics and ontology throughout the EarthCube project through identifying:

-- Sub-disciplines where aligning/coordinating emerging ontologies makes sense

-- Other areas with a lot of commonality – enabling the generation of parent ontology

-- Other areas with radically different ontologies -- then provide tools and guidance to connect them.

Currently many vocabularies are just terminologies. This represents lightweight semantics and is a good starting point. But, a multi-disciplinary use case such as the one presented by Sinha (http://earthcube.ning.com/group/semantics-and-ontologies/page/krishna-sinha-introduction-to-semantic-workshop-and-use-case) demonstrates the idea of reasoning and using semantics to discover new knowledge and make implicit facts explicit. This allows having a formal axiomatization, reasoning, and deep semantics that scientists want from an infrastructure.

Data discovery: Under midterm goals, in addition to use case capture and development, it was emphasized that technologies for data discovery was another priority theme. To achieve this goal, it was recommended that addressing data discovery challenges should be coupled with providing services that capture processes which occur in nature. Concern about buy-in from a diverse community would require introduction of semantics in data discovery and integration in a straightforward way.

Process of engagement: It was suggested that the group provide an information bulletin (posted on the semantics group website) for geoscientists which explains semantics and its relevance in lay terms that anyone can understand. This process of engagement is necessary as more end users start to be directed to the EarthCube site. Also, discussions on the utilization of use cases with semantically enabled solutions could be shown at national conferences, such as AGU, [ESIP Fed[1]], or GSA. This would enable more positive response to semantics from end users.

Sharing Data: There were discussions on how to develop a process that makes it easy to be a 'good citizen' wanting to share data. It was suggested that it was important to develop guidelines identifying the underlying technologies that can be useful, and provide pathways for data sharing. Participants asked if there is an active DataNet project for the geosciences because partnering EarthCube with the DataNet program to develop a cyberinfrastructure and semantics in support of GeoScience would have strong synergy. Anne Thessen responded that the Data Conservancy has some earth sciences data and is focused on being interdisciplinary, especially between bio and geo.

Introducing semantics: Semantics is not necessarily embraced by the entire science community. Therefore, there was agreement that forming a working group that focuses on semantic infusion would help to bridge the gap between the science and technology communities. Infusion processes would include assessing what are the science drivers, current capabilities, pain points, and wish lists.

[1] http://wiki.esipfed.org/index.php/Semantic_Web

8. *Timeline:* *Timeline for the project and all related sub-projects, including prioritization of activities and measurable milestones/major achievements and total resources (human and financial) required to achieve roadmap goals over a period of the next three to five years.*

Panelists: Krishna Sinha (initial report), Mark Schildhauer

In order to meet Roadmap guidelines, initial critical timelines identified by the semantics group were related primarily to short term activities. Initial concluded activities included:

1. Staffing of temporary committees (Outreach, Technology, and Joint (infusion) Committees). These committees fill a vital role in communication with other groups and agencies, provide an updated status of existing and emerging technologies and identify use case(s) that showcase semantically enabled discovery and integration of distributed and heterogeneous data.

2. Preparation and adoption of a summary document highlighting discussions of individual elements of the Roadmap. The document was the focus of a WebEx virtual meeting on Monday, May 7th at 2 pm (EST) and again on Thursday, May 17th at 2 pm (EST). The document was updated based on discussions of Monday, May 7th and then made available on the EarthCube Semantics Group site for community evaluation. The report was re-evaluated at the May 17th virtual meeting. This level of outreach and transparency will encourage continued support of EarthCube by scientists within the geoscience and semantics communities. The final report, i.e., this document, including the community's vision of the roadmap would be submitted to NSF by May 31st.

3. As part of the work for the Roadmap, two other documents were created, one on a potential use case and one on the status of semantic technologies. These are referenced in the Appendices.

The following goals although not explicitly discussed at the workshop, are possible suggested tasks, and should not be considered a final comprehensive list (presented by Nancy Wiegand to the semantics community group through group mailing, and further revised by Krishna Sinha):

Further short term goals (6 months to 1 year):

- Continue and increase communication with other EarthCube groups, e.g., brokering, workflows, interoperability, layered architecture, data mining that also need semantics as part of their solutions
- Organize Semantics and Ontology workshops to identify geoscientists and semantics experts who would be charged with the task of starting the process of cataloging existing geoscience related ontologies, and their application in demonstration projects developed around the use case already identified in this report
- Conduct a gap analysis for semantics technologies as applied to geoscience use case
- Support discussions through workshops with geoscientists to identify additional use cases
- Start creating initial foundational, domain and service ontologies, including Ontology Design Patterns for use in demonstration projects. Ontologies to be jointly created with domain experts and semantic/ontology scientists/engineers.

- Develop models for semantically enabled engines for registration of data, services and data nodes
- Develop models for semantically enabled engines for discovery and analysis of data

Midterm goals (2-4 years):
- Continue with the short term goals
- Refine further use cases
- Enhance initial ontologies for data and services
- Develop mapping across common vocabularies and ontologies
- Develop demonstration projects utilizing geoscience community endorsed use cases
- Work on provenance and ontologies
- Establish and populate ontology repository; identify process of sustainability of all resources
- Begin putting geoscience data into linked format (RDF)
- Work with the brokering, workflows, interoperability, layered architecture, and other EarthCube groups to include semantics and semantic components
- Develop tools and other software related to semantic capabilities
- Begin to show increased data discovery, access, and interoperability
- Increase community engagement, including the long tail
- Develop and disseminate educational materials
- Engineer semantically enabled data, service and discovery engines to handle use cases

Long term goal:
- Reasoning and knowledge discovery (knowledge availability (versus data availability) as per section 1)
- Complex analytics with information interoperable across all geoscience communities
- Improved data discovery technologies triggering increased community participation (long tail of science) for data and tool sharing

9. Management: *Management/governance/coordination plan and decision-making processes necessary to successfully establish standing committee(s) and subcommittees (if warranted), including a plan to identify and respond to shifts in technologies and changing needs at the end-point of use. Include discussion of approaches to educating end-users and achieving community consensus on advancing the capability/technological solution.*

Panelists: Krishna Sinha (initial report), Mark Schildhauer

To facilitate the development of semantics-enabled infrastructure, the workshop participants agreed to collectively author a manifesto explaining the importance of semantics in discovery and, integration of data and services leading to new knowledge for the Geoscience community. Use case and project registries would readily provide the information enabling evaluation of existing semantics technologies, including vocabularies. It was agreed that use case(s) be identified to showcase these capabilities by providing a semantic template that can provide a working solution for the use case. It was also agreed that solved use cases resolve communication with the broader community as they are able to see the impact of semantic technologies in the conduct of their science.

To make rapid progress towards these goals, working committees were formed immediately (committee members identified during WebEx meeting of May 7, 2012, see Appendix 1) and include:

- **Outreach committee** : charged with sharing the vision and capabilities of the semantics/ontology group with other communities (e.g. ESIP/federal agencies; geoscience societies)
- **Technology committee:** charged with assessment of current and future developments in semantic technologies
- **Joint (geoscience and technology) committee:** charged with developing use case(s) and to identify selected use cases with a goal of delivering semantically enabled solutions to the community. This working group is also charged with identifying where infused semantic solutions directly address gaps in the science communities.

These temporary standing committees and current memberships are likely to change through increased participation from the broader geoscience and semantics/ontology communities. Based on community needs additional committees maybe established in the future.

10. Risks: *Identification of risks and additional challenges to the successful establishment of any working group, and any unique risks associated with a working group associated with your topical area. With respect to identified risks, an approach to risk mitigation should be addressed.*

Panelists: Krysztof Janowicz (initial report), Naicong Li, Philip Murphy, Isabel Cruz

Two types of risk can be identified. Risks with respect to the semantics research roadmap and the risk of having a dedicated semantics and ontologies group in EarthCube.

With respect to the latter risk, two aspects were discussed. First, semantics plays a role in several of the other EarthCube groups, e.g., interoperability and workflows group. Therefore, one may consider addressing semantics in the specific groups rather than within a dedicated group. The participants agreed that this is not reasonable as it would create island solutions instead of a joint infrastructure and foster redundancy. Secondly, several aspects can only be addressed by a semantics group and which are a prerequisite for other groups. Examples include the semantic annotation and publication of scientific results as Linked Data, a set of shared core ontologies for the EarthCube community, alignment and matching of local ontologies, and information retrieval beyond simple keyword matching. The semantics group could use its expertise to form task groups that support other parts of EarthCube in terms of ontological modeling and reasoning needs.

In terms of the semantics research roadmap, the following risks have been identified:

First, domain experts may be scared away by the technical language in which ontologies are discussed and specified. While it is possible to hide these ontologies from the users and run them in the background to improve information retrieval and interchange, this is not feasible for the semantic annotation of datasets, integrity constraints, or using deep semantics for knowledge extraction and hypotheses generation. In such cases, domain experts require a detailed understanding of how knowledge engineers modeled domain facts to judge whether the resulting ontologies reflect their initial conceptualization. Consequently, ontologies should be implemented and documented via a transparent community process.

Second, the purpose of the ontologies has to be defined based on scenarios and application areas to ensure that they are not over or under engineered. Different degrees of detail and formalization may be required ranging from lightweight ontologies used for simple annotation to heavyweight ontologies that exploit the power of Semantic Web reasoning. Too generic ontologies may fail to restrict meaning to a degree where they cannot support semantic interoperability; too restrictive ontologies may hinder the semantic diversity inherent in a multidisciplinary science. The risk of the so-called knowledge engineering bottleneck can be minimized using ontology engineering methodologies, a modular, layered framework, and common modeling patterns.

Third, the wider EarthCube community may not adopt the proposed solutions, e.g., arguing that they are not suitable for certain tasks, do not match specific domain facts, or introduce ontological commitments that hinder their flexible reuse. Early and frequent usability testing, modularization, conformity testing, and semantic negotiation were identified as ways to mitigate this risk. Adoption is expected to be more likely if local and application-driven perspectives are supported by introducing a lattice of ontologies, i.e., semantic heterogeneity is not understood as a burden to be resolved. The fitness for purpose of the developed ontologies can be measured based on the amount of annotated data and the degree of interlinkage to other sources (especially in case of creating Linked Data).

Fourth, semantic aging and ontology evolution have been identified as risk. This risk can be mitigated by actively maintaining and curating knowledge infrastructures and ontologies. Additionally, risk mitigation would include understanding how other semantics communities, such as the biomedical community, address these challenges as those communities have been using ontologies for several years.

Finally, there is a risk of investing in a certain technology. While the group agreed that the methods and technologies should not be reduced to the Semantic Web layer cake alone, it was acknowledged that the Semantic Web offers the required standards, support from academia and industry, and the software required to implement an EarthCube knowledge infrastructure. While specific technologies may change and software tools may be discontinued, a high degree of standardization in conjunction with open and free source code mitigates the involved risks.

APPENDIX I

Temporary Standing Committees with Initial Members

This list of Committee members is in response to discussions related to Management needs (element #9 of the NSF EarthCube roadmap guideline) at the workshop.

Outreach committee: charged with sharing the vision and capabilities of the semantics/ontology group with other communities (e.g. ESIP/federal agencies; geoscience societies)
Liaison to EarthCube initiatives:
Data Discovery and data Mining Group: Tim Finnin; Workflow: Naicong Li
INTEROP: Nancy Wiegand, Philip, Karen Stock
Data Access/ Brokering/Layered architecture: Janet Fredericks
Weekly newsletter content to NSF: Nancy Wiegand, Krishna Sinha

Other communities:
ESIP, NASA Interoperability Working Group/Semantic Technologies Working Group: Hook Hua
Federal Agencies/ESIP: Peter Fox;
Geo-societies and International partners: Krishna Sinha/Peter Fox
Semantic Web: Krysztof Janowicz; SIGMOD: Nancy Wiegand; MMI: Karen Stocks
Geoscience Education: David Mogk
Ontolog Community and the Open Ontology Repository (OOR): Leo Obrst
International Association for Ontology and its Applications: Leo Obrst

Technology committee: charged with assessment of current and future developments in semantic technologies

The current group members are: Krysztof Janowicz, Leo Obrst, Amit Sheth, Gary Berg Cross, Pascal Hitzler, and Tim Finin.

Joint (geoscience and technology) committee: charged with developing use-case(s) and to identify selected use cases with a goal of delivering semantically enabled solutions to the community. The working group is also charged with identifying where infused semantic solutions directly address gaps in the science communities.

The current group members are: Calvin Barnes, Hassan Babaie, Anne Thessen, Pedro Szekely, and Mark Schildhauer. Janet Fredricks will cover Provenance in use case scenario.

APPENDIX 2
Process for Selection of Use Case(s) provided by Krishna Sinha

The process of identifying use case(s) must meet many criteria that enable the geoscience community to recognize the value of technology to be used in solving the use case scenario. The key aspects of any use case is (1) attract the long tail of science to participate (2) have a large spectrum of data types, and (3) provide an opportunity to find gaps in existing technologies that facilitate sharing, access, discovery, integration and modeling.
Important criteria include:

- availability of a rich vocabulary
- availability of data with different syntax and semantics
- availability of data in multiple formats
- availability of data in a globally distributed system on varying platforms
- availability of tools and models
- availability of high level ontologies and ready linkage to SWEET ontologies
- is societally relevant, engages citizen scientists and policy makers
- engage the long tail of science
- amenable to use of advanced computational methods
- provide a resource for teaching
- provide a resource for identification of ' technology gaps ' in meeting end user requirements
- provide integrative capabilities with other domains (ecology, geography, engineering, biology)
- Others

Based on these criteria, there was general consensus that volcanism (a phenomenon in SWEET ontology; http://sweet.jpl.nasa.gov/ontology/) could be considered an effective use case. It was also agreed that simple to complex queries could be readily created through the use of data and data products associated with this phenomenon for technology gap evaluation.

This phenomenon has hundreds of years of observational data (including sensors) in many formats, and is available in a globally distributed system (on different platforms). As many of the objectives of EarthCube require semantic/ontologic enhanced capabilities (data mining, workflows and others) to be placed in the hands of the scientist leading to integration across many disciplines, the use of volcanism as a use case platform would engage the broad community. Technologies that support sharing, accessing, discovering , integrating and modeling heterogeneous data across many disciplines in geosciences, as well as ,ecology, geography, engineering, biology would provide an integrative view of the earth at many scales and associated processes. This broadly defined use case also provides the opportunity to undertake a gap analysis of our existing semantic capabilities, as well as methods to assess quality of data, lineage, security, etc. Its societal impact provides a broad engagement for the "long tail of science", educators and policy makers.

APPENDIX 3

Presented by Joint (geoscience and technology) committee
Members: Calvin Barnes, Hassan Babaie, Anne Thessen, Pedro Szekely, Mark Schildhauer.

SEMANTICS AND VOLCANISM USE CASE

Introduction

The Semantics-based volcanism system will allow Earth scientists studying volcanic eruptions and their impact on Earth's major components (e.g., atmosphere, hydrosphere, biosphere), government and NGO workers providing help and care for the citizens affected by eruption, and students investigating different aspects of volcanism, to more efficiently integrate and query their data based on geological knowledge and Semantic Web technology.
There are few aspects of geosciences that encompass the entirety of the planet and, for that matter, the solar system. Volcanism is one. We have the opportunity to attract and involve a great diversity of geoscientists in making semantics work for the community.

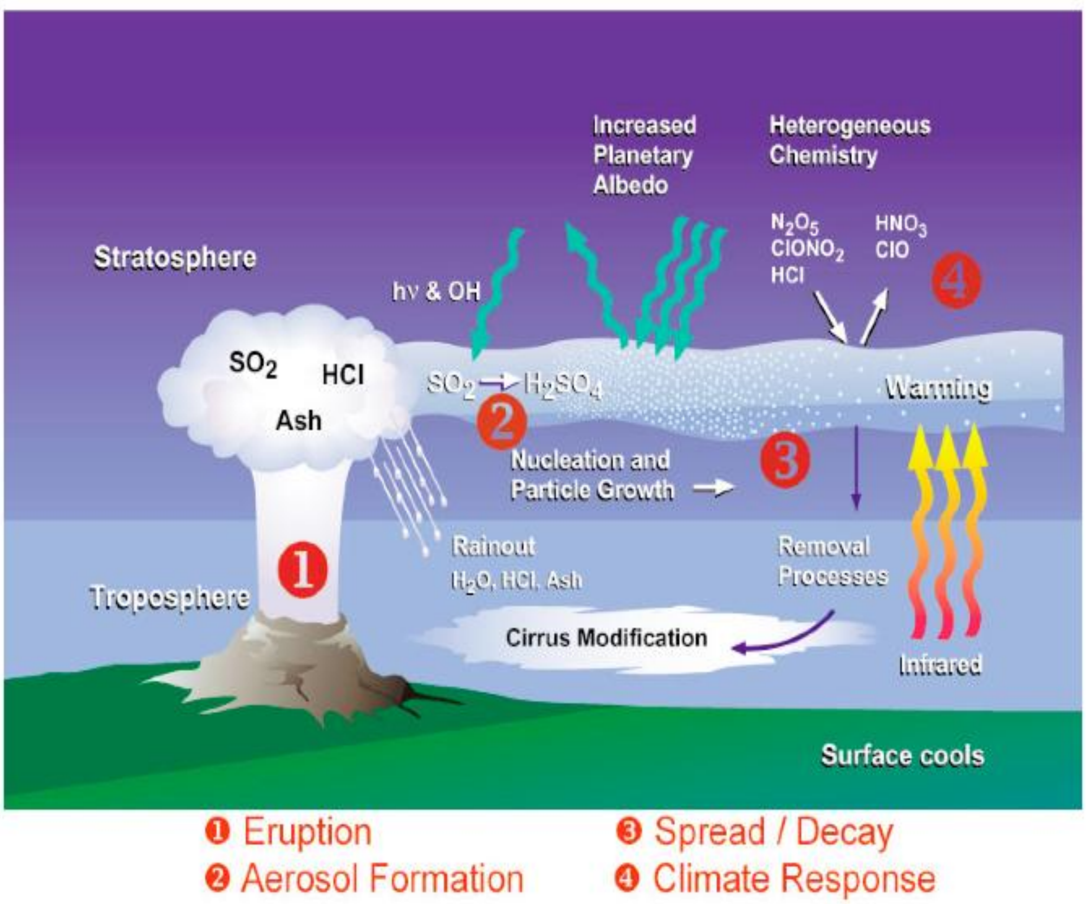


Figure 1. Interaction of different spatial and spatio-temporal entities during a volcanic eruption (Krishna Sinha Use case presentation, EarthCube Workshop, Arlington , April 30-May 1, 2012; Figure from Fox et al., 2007 , http://esto.nasa.gov/conferences/estc2008/papers/fox_a2p3.pdf)

Purpose

Opportunity

The knowledge-based system will allow automated reasoning by drawing new knowledge from existing, integrated data about interaction of Earth's components (atmosphere, geosphere, hydrosphere, biosphere, cryosphere) during volcanic eruptions, by applying the inference rules which are inherent in the ontology languages such as OWL.

Volcanism is one of the major geological processes that have shaped the earth and its atmosphere throughout geological time. It was responsible for the original introduction of water and oxygen into the atmosphere, and periodic global warming (excess CO_2) and cooling (aerosols blocking sunlight). A great number of people live near active volcanoes, and thus are vulnerable to its hazards which include violent eruption and explosion, ash fall, lahar, poison gas, and acidity in lakes due to CO_2 seepage. Volcanoes are mostly associated with plate boundaries and continental and oceanic hotspots. Thus, the relationship of volcanism to tectonism and deformation is of critical interest. Formation of many of our mineral resources is associated with volcanism and the magma that feeds it.

Designing, developing, and deploying ontologies to model the knowledge about the spatial and spatio-temporal aspects of volcanism and its related phenomena will help find solutions to important scientific questions (see below), and address many of our global, societal problems related to volcanic eruption.

Problem Statement

Volcanism directly encompasses major aspects of the geosciences. For example, petrologists study the solid products of volcanism (rocks and minerals) and are interested in their geometric relationships (field study), absolute and relative age relationships, and the chemical and isotopic compositions of the rocks and minerals. Physical volcanologists are interested in these types of data and also in the processes of eruption such as: the interplay between magma properties (composition, viscosity, temperature), volcano structure, plate tectonic setting, and explosivity. When combined with geophysics (seismology and potential-field data) and structural geology (extension, faulting, mid-ocean ridges, fissure eruption), models for locations and longevity of magma chambers, monitoring and prediction of eruptions and, communication of potential volcanic hazards to decision makers and the public may be developed.

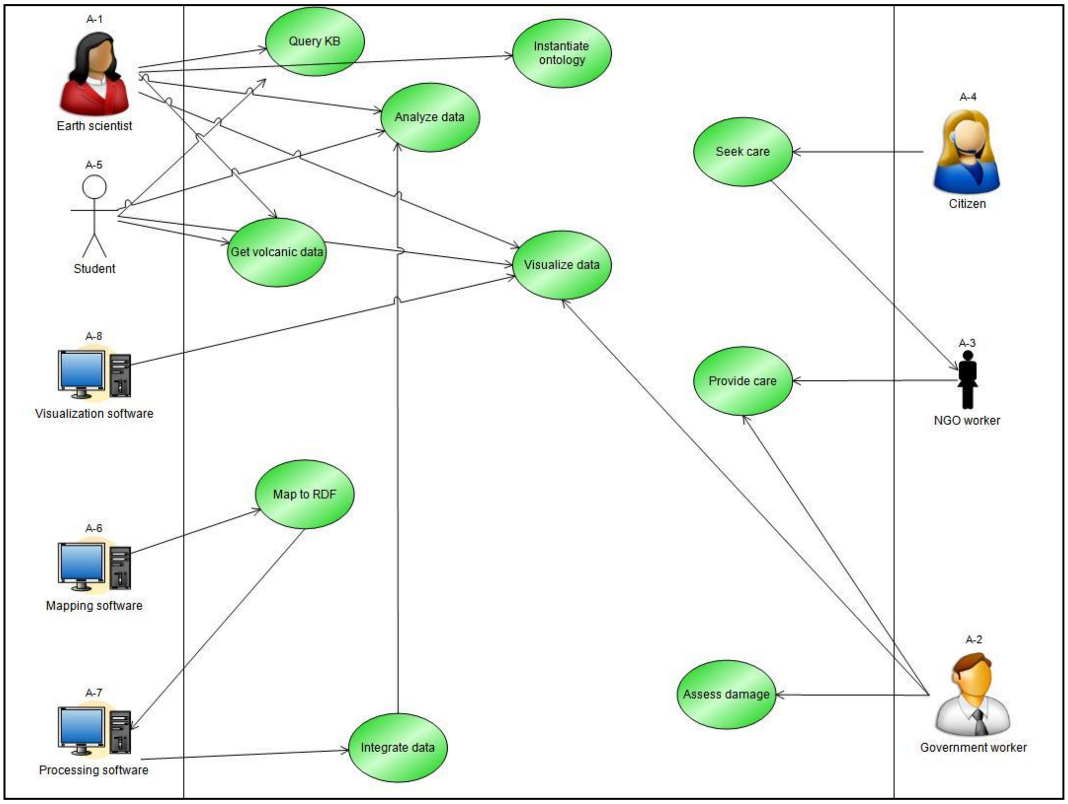
The problem is that the voluminous data collected by these scientists are heterogeneous in structure and are not integrated. They reside in globally distributed, relational databases and other types of files (e.g., Excel, text) that despite being related are not connected or integrated. It is very difficult and costly for software to interoperate with each other when dealing with the disparate data.

Market

Users of the system include Earth scientists such as geologist, oceanographer, and atmospheric scientist. Others include archeologists who combine regional wind and weather patterns with data from past eruptions to analyze and date ash beds associated with prehistoric remains, and climate scientists who need to assess the impact of volcanic eruptions on climate.

Graduate students who do their theses on ecosystem succession on volcanic fields after eruption. The student may need as much observational data about species occurrences in volcano-impacted zones as quickly as possible.

Other users include government worker who is tasked to develop short- and long-term predictive tools to identify population centers at risk from volcanic activity; Non-scientist users include NGO worker who helps the citizens impacted by the eruption, citizens, needing help and care from the government and NGO workers, and Software will use the system to convert all kinds of data into RDF, making integration of data possible, ontologies to model domain knowledge and allow automated queries, and Web services and tools will allow processing and visualization of the integrated data.



Product Position

The following functionalities differentiate the ontology-based system from the existing ones. Study of volcanic products (field and remote mapping, minerals, rocks, gas chemistry [in situ and remote]) is a main geological activity. Currently, data about these activities are stored in isolated databases and other forms of data files, in disparate data structures. Ontologies developed to model the knowledge about Earth materials will allow automated reasoning about volcanic products and processes, and help discovery of new knowledge through the language (RDF, RDFS, OWL) inference rules and uniform RDF data structure.

Study of volcanic activity, ancient and modern (as above, plus ground-based and satellite instruments; seismology; potential field geophysics). Ontologies, modeling the knowledge about physical and chemical processes and geological tasks and tools would help to more efficiently understand the inter-relationship between processes and their products through automation of knowledge discovery and formulation of knowledge-based queries.

Study of plate tectonics (global) and structural (regional) controls on volcanism. Ontologies about all aspects of plate tectonics, relating the spatial (e.g., subduction zone, backarc basin) and spatio-temporal entities (subduction, rifting), will help to correlate volcanism to tectonic processes.

he real advantage of the semantics-based volcanism system results from the various consequences of volcanism and the semantic links (i.e., meaningful, knowledge-based relations) that can be formed throughout the geosciences, and elsewhere, for example: terrestrial effects (lava, air-fall, pyroclastic flows, mudflows, sector collapse), marine effects (sector collapse leading to tsunamis; hydrothermal exchange of volcanic rocks and sea water), ecology of volcanic terrains, return of plant and animal communities to volcanically-disrupted regions, return of human communities to volcanically-disrupted regions, relationships of various types of volcanic terrains to land use, disruption of surface-water systems (drainage degradation, sedimentation, etc), disruption of ground-water systems, atmospheric effects of eruption (air traffic hazards; global climate effects), medical effects of volcanic aerosols, effects of volcanic substances on nearby food webs and thus the products humans rely on (fishing, agriculture).

Stakeholders

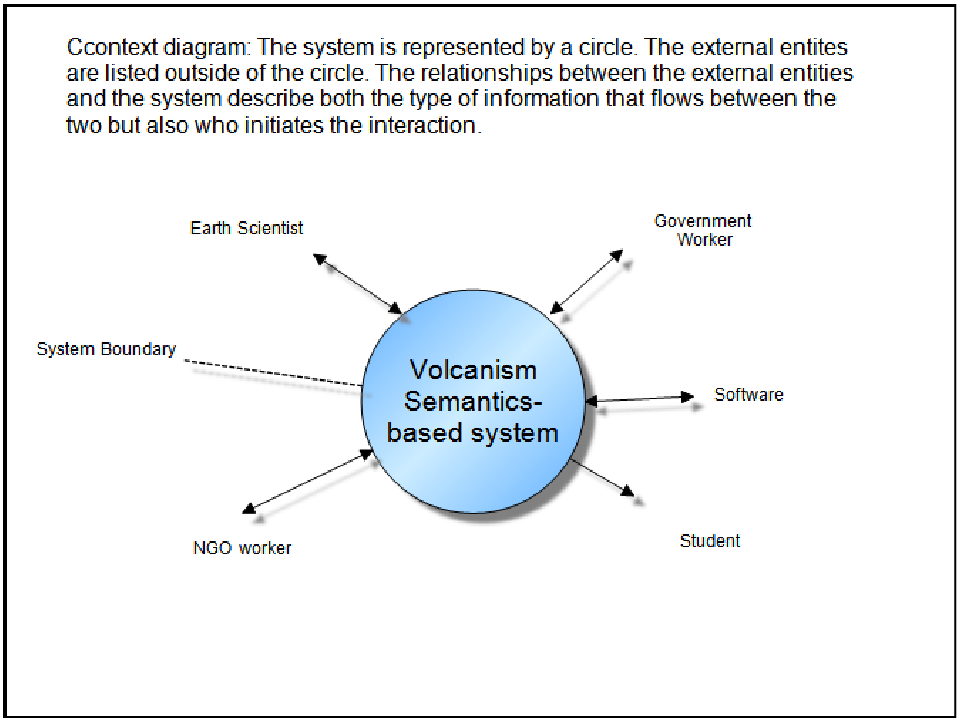
The main stakeholders are Earth scientists who collect, or have collected, a large set of data about present and past volcanic eruptions, and study their effect on the atmosphere, hydrosphere, geosphere, and biosphere. Other stakeholders include government and NGO workers, citizens, and students.

Stakeholder Goals

Ontologies will help us to integrate data (RDB, text, XML, XML, CSV, HTML) collected by government and professional organizations that deal with volcanism and its effects, such as NSF, NASA, NOAA, USGS, Canadian Geological Survey, Geoscience Australia, British Geological Survey, Geological Surveys of Japan, Indonesia, Chile, Argentina, and many others, American Geophysical Union, International Association of Volcanism and Chemistry of Earth's Interior, Geological Society of America, European Geophysical Union, and Geochemical Society. An ontology-based infrastructure will enable these communities to more easily integrate their data and make better decisions in relation to questions related to volcanism.

Product Overview

The knowledge-based volcanism infrastructure will be constructed by building a series of inter-related ontologies to model the knowledge about the fields in Earth sciences with direct interests in volcanic activity and its consequences, such as physical volcanology, petrology/mineralogy/geochemistry, geophysics, geochronology, structural geology, hydrology, climatology, atmospheric chemistry, economic geology, and oceanography.



Features and Benefits

The semantics-based volcanism infrastructure will help raise and answer questions that are currently not possible with existing systems. Generally, it would allow integration of heterogeneous, disparate, and globally distributed data; software interoperability; automated reasoning, and effective query.

Examples of questions which can be addressed with the ontology-based infrastructure:

- Resolve the signature of an eruption in the terrestrial lower atmosphere by "determining the statistical signatures of both volcanic and solar forcings on the height of the tropopause".
- Which species reinhabited disrupted volcanic terrain after the eruption of Mount St. Helens?
- What (people and other species) is at risk through varying types of volcanic eruptive products?
- What is the relationship between eruptive intensity, volume and gas content, and impact on global climate?
- What is the relationship between the amount of the CO_2 gas emitted out of a submarine volcano and carbonate deposition in the ocean and global warming?
- What will be the economic impact of an eruption of Taal volcano on the fishing economy of the Philippines?

External Requirements and Constraints

The system requires designing, developing, and employing a series of ontologies, at different granularity levels, for volcanism and all the fields that are related to it (see above). It also requires data from these communities to be available for translation (mapping) into the RDF data model based on these ontologies. The main constraint is to get access to the relational database schemas which must become available to the ontologists who convert the RDB data structure into the RDF triples. This practical problem needs to be resolved with the help of NSF and other funding agencies.

Some links to status of volcanology related ontologies and applications (provided by Krishna Sinha)

http://tw.rpi.edu/proj/portal.wiki/images/7/75/IN53B-1204_AGUFM07_SESDI_package.pdf

https://marinemetadata.org/references/sesdiontology

http://sesdi.hao.ucar.edu/cmaps/Atmospheric_Climate_III.jpg

http://ceur-ws.org/Vol-401/iswc2008pd_submission_71.pdf

APPENDIX 4

Presented by Technology Committee

Semantic Aspects of EarthCube

Version 1.0, May 22, 2012

By

Pascal Hitzler, Kno.e.sis Center, Wright State University
Krzysztof Janowicz, University of California, Santa Barbara
Gary Berg-Cross, SOCoP
Leo Obrst, MITRE
Amit Sheth, Kno.e.sis Center, Wright State University
Tim Finin, University of Maryland, Baltimore County
Isabel Cruz, University of Illinois at Chicago

— the Technology Subcommittee of the EarthCube *Semantics and Ontologies Group* —

with additional input by

Naicong Li, University of Redlands
Karen Stocks, University of California, San Diego.

Purpose

In this document, we give a high-level overview of selected Semantic (Web) technologies, methods, and other important considerations, that are relevant for the success of EarthCube. The goal of this initial document is to provide entry points and references for discussions between the Semantic Technologies experts and the domain experts within EarthCube. The selected topics are intended to ground the EarthCube roadmap in the state of the art in semantics research and ontology engineering.

We anticipate that this document will evolve as EarthCube progresses. Indeed, all EarthCube parties are asked to provide topics of importance that should be treated in future versions of this document.

Deep Versus Shallow Semantics

Ontology languages—and knowledge representation languages in general—differ in terms of expressivity, i.e., they differ with respect to the language primitives that they provide for modeling, and in the extent to which these language primitives are endowed with a formal semantics. Knowledge representation in this respect can be traced back to a history that is over two thousand years old (see, e.g., [HKR10, Chapter 1]), and the state of the art in ontology engineering and Semantic Web is reaping the rewards of this long-standing tradition.

It is in exactly this tradition that ontology languages, such as the W3C standards OWL [HKP+09] and RDF [MM04], are endowed with a so-called *formal semantics*, which is essentially based on the model-theoretic semantics of mathematical logic. In an ontology engineering context, this semantics can be understood as an *inferential* semantics, which, intuitively, determines how the joining of different pieces of knowledge *entails* new knowledge, in the sense of logical, deductive, inference.

OWL and RDF differ, e.g., in terms of language constructs, the meaning of which is captured by the respective formal semantics. For example, OWL provides the *owl:sameAs* language construct. The formal semantics of OWL essentially declares that it is used for identifying two resources (i.e., URLs), which refer to the same entity. Hence, whatever is said about the first, can be inferred to also hold for the second. While RDF does not forbid the use of owl:sameAs, its formal semantics does not capture this meaning, i.e., from an RDF perspective, owl:sameAs can only be used informally, and its meaning is up to the user or to the system that happens to encounter it.

OWL, in fact, provides a stronger (or deeper) semantics than RDF, in the sense that it has more language constructs with a formally defined semantics. OWL and RDF can thus be understood as being part of a spectrum of knowledge representation languages, which range from very shallow (or even formal semantics-free) languages, to languages with very strong formal semantics that significantly surpass OWL in terms of language constructs and engineering capabilities [Obr03]. For example, on the very shallow end of the spectrum is tagging with natural language terms (for which the semantics is not formally defined, but determined by our use of natural language and domain-specific scientific terms, i.e., informal or controlled, but relatively unstructured vocabularies), or the use of microdata (as, e.g., in schema.org). Taxonomies, thesauri, and class hierarchies also range at the shallower end. The OWL 2 tractable profiles [HKP+09] range between RDF and OWL 2 DL. At the deeper end, beyond OWL, are e.g., full first-order logic languages such as ISO Common Logic [CL], extensions of first- (or even higher-) order predicate logic by expressive means of the likes of uncertainty handling, commonsense reasoning, temporal modalities, to name just a few.

Both shallow and deep semantics approaches respective benefits and drawbacks. Shallow approaches, for example, are easier to get started with, but miss rigor as they suffer from a lack of power to restrict possible multiple interpretations, and thus make it more difficult to realize subsequent knowledge integration and interoperability. Deep approaches are harder to understand, and engineering is a serious effort, though with the advantage of making it easier to bridge heterogeneity gaps. In general, deeper approaches enable greater precision and accuracy in the applications which use them. It could perhaps be claimed that the recent success of ontology engineering and Semantic Web Technologies rests at least partially on the fact that RDF and OWL map out a *sweet spot*, a reasonable compromise between deep and shallow semantics.

For EarthCube, it is important to understand this issue, and to carefully navigate this spectrum while being aware of the trade-offs involved. In particular, it would be naive to expect that knowledge engineering could start on the shallow end of the spectrum, and then subsequently be refined or deepened as EarthCube progresses. A recent and rather prominent example for the fallacies in this approach is the use of the aforementioned owl:sameAs language construct in RDF-based Linked Data [BHB09]: While it is occurs in very substantial quantities, its usage is mostly informal and in particular is not aligned with the formal semantics that it should inherit from OWL [HHM+10]—a problem with, at hindsight, could have been avoided by taking deep semantics into consideration in the first place. With respect to EarthCube, a shallow semantics approach seems sufficient for tasks such as improved retrieval but is likely to fail for data integration.

Semantic Interoperability and Semantic Heterogeneity

While we still lack a formal definition of semantic interoperability, it is usually defined as the ability of services and systems to exchange data in a meaningful way [HKPBR99, GEFK99]. For example, from a systems perspective, semantic interoperability can be 'defined as the enablement of software systems to interoperate at a level in which the exchange of information is at the enterprise level. This means each system (or object of a system) can map from its own conceptual model to the conceptual model of other systems, thereby ensuring that the meaning of their information is transmitted, accepted, understood, and used across the enterprise.' [OWM99] In contrast, syntactic interoperability only focuses on the technical ability of systems to exchange data. To give a concrete example, a service may request data about wind speed and direction to compute the dispersion of a gas plume and request both values as floating point numbers. A service that can deliver such weather data is called syntactically interoperable. If, however, the first service expects a 'wind blows from' semantics while the second service offers data in a 'wind blows to' semantics, the results of the dispersion model will be wrong and potentially dangerous [PL04, K05]. While, strictly speaking, data cannot interoperate, the term is often used in a broader context. Semantic Interoperability is key to EarthCube and all other infrastructures in which data has to be published, reused, and integrated. The risk associated with a lack of semantic interoperability is that incompatible data is unwittingly combined or that unsuitable methods and models are applied to datasets. Many spectacular cases have been reported over the last years, the crash of the Mars Climate Orbiter due to a confusion between English and metric units being one of the most prominent examples [B06]. Geospatial ontologies [K05,E02], semantic annotation of geospatial data [FS02], matching/alignment of multiple, local ontologies, ontology-driven Web portals [MSSSS03, LRGJ12], query processing for heterogeneous geospatial sources [CX08], and geospatial ontology-driven analysis [ASRUAK06] have been proposed as one way to capture the body of knowledge of a specific domain and assist scientists in understanding methods and datasets.

Semantic technologies and ontologies are believed to be core components in establishing semantic interoperability as both help to restrict the interpretation of domain terminology towards their intended meaning and hence allow for more intelligent metadata. As ontologies are best thought of as constraint networks [K09], semantic interoperability can never be guaranteed in infrastructures that require the on-the-fly combination of data or service chaining. Hence, in addition to ontologies, reasoning services are required that support service matching [GH07] or translate between ontologies. Over the last 10 years, service interoperability has been addressed by several proposals and standards such as OWL-S [OWLS04], WSMO [RKL+05], WSDL-S [A05]/SAWSDL (a W3C recommendation) [VS07], SA-REST [SGL07] or USDL [USDL11]. While these approaches propose service ontologies as an essential part, additional work is required to ensure that different knowledge and service infrastructures can interoperate. For instance, the so-called Geo Web that is largely based on services defined by the Open Geospatial Consortium (OGC) cannot communicate with the Semantic Web which will be a major roadblock for EarthCube. A Semantic Enablement Layer [JSBKMS09] can transparently mediate between both infrastructures and, hence, allow Spatial Data Infrastructures (SDI) to access reasoning services, Linked Data, and ontology repositories from the Semantic Web as well as the other way around, e.g., enable Semantic Web applications to dive into the Geo Web. Such a layer needs to be transparent to ensure that no changes to existing and well standardized infrastructures are required. First implementations for the semantic enablement of several SDI components have recently been published [BMJNM11, SSOR09, JBSSEL11, MMR12, HPST09]. GeoSPARQL has recently been standardized and proposed as a common query language for the Geospatial Semantic Web; see [BKta] for an introduction.

Instead of continuing the millennia old search for the universal ontology, different types of ontologies have been proposed in computer science. The classification of ontologies based on their granularity and thematic scope into top-level, domain, task, and application ontologies was first introduced by Guarino [G98]. An alternative classification into global and local ontologies has been proposed by Uschold [U00], while others distinguish between domain-independent and domain-specific ontologies. Several global, top-level ontologies such as DOLCE, SUMO, BFO, GFO, and Upper Cyc have been proposed as well as domain ontologies for the Earth sciences such as SWEET [RP05]. Initially, it was assumed that each scientific discipline could agree on a domain-level ontology and that these ontologies could all refer back to one common foundational ontology. Lower level ontologies, e.g., application ontologies, were thought of as mere specializations of these ontologies. It turns out, however, that even within very specific domains it is difficult to get scientists to agree on a common definition for their domain vocabulary and especially to align these definitions with the very abstract and loaded classes from top-level ontologies. For instance, lenticular clouds can be classified as events or physical objects at the same time [G04] while these two classes are often defined as core distinctions in top-level ontologies. More recently, Sinha and Mark [SM10] demonstrated that feature types such as Hill can be specified as physical objects, features, or amount of matter, while these three classes are among the core distinctions proposed by the DOLCE foundational ontology for physical endurants. In other terms, many types are multi-aspect phenomena [G04] to a degree where even top-level distinctions cannot be utilized without reference to context. However, there are approaches that attempt to address linkages between these notions via bridge axioms, including BFO notions of SNAP and SPAN, i.e., temporal snapshots vs. temporal spans [GS04]. Similarly, the theory of granular partitions [BS03, BSM07] does take into consideration context, and tries to resolve problems related to parts of objects. The relation between objects and events and their ontological distinction has been an active area of research for many years, see, e.g., [GM09]. Consequently, taking heterogeneity as reality, the increasingly popular Linked Data approach does not follow the idea of a few authoritative ontologies but proposes to define local and application centric ontologies to suit the needs of specific data sets and repositories.

This paradigm shift is accompanied by a changing focus towards ontology matching, alignment, semantic translation [SE08,JHSVY10], and multi-ontology query processing [MKSI96] that allow to directly interact between different ontologies without the need to agree on one common reference first. Ontology design patterns, a (partial) analogy to the successful software engineering design patterns, have been proposed to support the development of a multitude of ontologies [G05]. Examples of such patterns that have been applied to semantics-based systems in the geosciences include the Semantic Sensor Network ontology [W3CSSN12]. In a highly interdisciplinary setting, semantic heterogeneity should not be misunderstood as a burden but is a consequence of diverse models, methods, and viewpoints brought in by different scientific disciplines and ongoing debates with domains [J10]. Semantic Web technologies and knowledge engineering frameworks should assist domain experts in making their conceptualizations explicit, and hence foster data sharing and reuse by supporting semantic interoperability without giving up on diversity [J12]. Similarly, there is no need to agree on one common representation framework. To support a knowledge infrastructure and community such as EarthCube, ontologies based on description logics (e.g., OWL) have to go hand in hand with numerical and statistical models [SRT05]. Currently, Semantic Web research is investigating how machine learning can assist in extracting knowledge from data and in reducing the burden of ontology engineering [SKW08, TMBS08, LH10, FDF12, RLTDF12, J12]. Such a data-driven perspective is also gaining ground in the area of geospatial semantics [BMT08, SM10b]. Bottom-up approaches, however, cannot replace top-down engineering–both have to work hand in hand. EarthCube will require a lattice of theories that fosters interoperability and at the same time allows for multiple perspectives. Such a lattice of theories will consist of top-level and domain-level ontologies, local and application-centric micro-ontologies, as well as bottom-up learned fragments. Semantic Web reasoning systems will enable integration within this lattice.

Limits of the Ontological Approach

Like every knowledge representation language, the Web Ontology Language OWL [HKP+09] has advantages and disadvantages. The main advantage of using OWL as a basis for EarthCube is, that in doing so, EarthCube is aligning with the current mainstream, which will make it easiest to import new methods, tools and existing ontologies and data.

Concerning some of the known drawbacks of OWL, it is important to notice that state of the art research is actively addressing them, and while newest developments take time before being incorporated in a standard, some methods are mature enough to be used in conjunction with

OWL. Other capabilities may in fact remain out of scope for EarthCube, or should be incorporated only in a very careful manner in cases where they are unavoidable. The following are some of the important "known drawbacks" of OWL.

Use of rules paradigms

The Rule Interchange Format, RIF [KB10], is a W3C standard for expressing rules on the Web. Rules and OWL were for a long time thought to be very complementary, with radically different design rationales. However, it is important to notice that many (monotonic, Datalog-style) rules can already be represented in OWL 2 DL [KMH11] (but were not representable in OWL 1 DL). Furthermore, recent research is significantly closing the gap. For example, a new construct called nominal schemas [KMKH11], which is a very light-weight extension to OWL [CKH12], makes it possible to represent arbitrary Datalog rules, even without restriction on the arity of the predicates. It thus completely captures, for example, DL-safe SWRL [HPB+04] and RIF Core [BHK+10]. Another light extension of OWL, so-called conjunctive roles [CH12], makes it possible to capture most Datalog rules under a first-order logic semantics. Providing tool support for these rules-extensions of OWL could easily be realized within EarthCube by extending existing tools for OWL (e.g., the Protege OWL editor). It can be expected that such "light" extensions, which cover Datalog-style monotonoic rules, would cover many of the rules-modeling requirements within EarthCube.

Non-monotonicity, i.e., local world closure and defaults

Capabilities of non-monotonic logics, such as default reasoning or the (local) closed-world assumption are not present in OWL. However, there is a significant body of knowledge on extending OWL with such capabilities (some of which is closely related to the OWL and Rules integration discussed in the previous paragraph). For an overview, see the related work sections in [KMH11, KSH11] and [KHM12]. However, since the research discussion is still very much in flux, it may be advisable to incorporate such capabilities only in a very careful manner, until the foundations have been solidified. *Simple* adaptations for modeling local closure or defaults, may be possible, and the concrete methodological approach will have to be determined based on use case requirements.

Modeling of uncertainty and probabilities

OWL in its current form does not provide for the modeling of uncertainty or probabilistic knowledge. Extensions have been developed, indeed there is a significant body of work on this, but it is still rather unclear which of the proposals would constitute "preferred" paradigms. Since the research discussion is still very much in flux, it may be advisable to incorporate such capabilities only in a very careful manner, until the foundations have been solidified. *Simple* adaptations for modeling uncertainty may be possible, but the concrete methodological approach will have to be determined based on use case requirements.

Spatial Reasoning Ontology Standards

A large amount of Earth Science linked data has an inherent spatial context. Without spatial reasoning, however, the value of this spatial context is limited to Earth Scientists. Over the past decade several vocabularies and query languages with varying levels of support for fundamental geospatial concepts have been attempted to exploit this knowledge and enable spatial reasoning. [ASRUAK06] gives examples of and strategies for computing geospatial relationships such as topological relations, cardinal direction, and proximity relations. Recently OGC sponsored a new standard called GeoSPARQL [OGC11, BKta] that attempts to unify data access for the geospatial Semantic Web. GeoSPARQL promised to be a standard vocabulary for many current data sets. Since the standard started with spatial comparisons, some feature add-ons are possible in the next two to four years. These should be of value to the Earth Science community. An example would be the addition of different coordinates reference systems used in earth sciences to supplement the current geographic coordinate reference systems.

Knowledge Acquisition (including Extraction from text), CMaps (cognitive maps) and Domain Expert Support

Developing domain ontology/knowledge bases remains a major bottleneck and risk since it is a resource intensive and time-consuming task. Quality knowledge acquisition has had a high barrier and requires the cooperation of earth science domain specialists, who provide concepts, and ontologists/knowledge engineers to faithfully structure and represent these in processable forms. It is generally impractical for the average earth science subject-matter expert to learn knowledge engineering and proper structuring of formal ontologies. (Initiatives within the Social Semantic Web recognize this problem, and have begun to identify mechanisms to collective motivate users to volunteer time and resources to participate in the semantic content creation process [SS10].) Due to time constraints it is also generally impractical for a large group of knowledge engineers and ontologists to master the concepts, terminology and principles of one or more earth science domain and/or independently extract domain knowledge from documents. Collaborative methods and the use of ontology patterns are one way the problem is being addressed, but several technologies and complementary tools are also important [Cue05]. These include:

- Conceptual modeling (Cmap) tools that are easy enough to learn and simple enough to use to allow domain experts to capture their background and domain knowledge along with reasoning approaches in intermediate, expressive forms. CMAP tools present a simple graphical representation in which instances and classes are presented as nodes, and relationships between them are shown as arcs. Resulting concept maps can be used to enable discussion and to later generate formal domain ontologies and supporting background knowledge. Tools of this type include various concept map tools such as CMAP, OntoEdit & Mind2Onto, MAP2OWL, and COE.
- Tools and technology that support conversion of conceptual models into a formal representation including if-then rules (COE).

- Tools to extract background and domain knowledge from text including Web documents and represent it formally (e.g., Text2Onto [CV05], TEXCOMON, Text to Knowledge Mapping [OE06], YAGO [SKW08].).
- A class of tools that provides enhanced metadata descriptions to text (e,g, YAGO-NAGA tool and approach).
- Tools to allow domain expertise to express their knowledge in a controlled form of natural language to manage both the syntactic and semantic ambiguities of ordinary language by enforcing a single definition for every term. Controlled forms can then be converted to some formal representation such as OWL (e.g. Attempto Controlled English (ACE), Rabbit & Roo [HJD08], Peng-D [Sch05], ClearTalk [Sku03] and Gino (guided input natural language ontology editor*)*.
- Tools that visualize formal languages in a simple form and allow easy editing. Examples include GrOwl [KWV07].

Semantic Mediators and Intelligent Brokers

Agent Brokering employs central mechanisms to help resolve such things as disparate vocabularies, support data distribution requests, enforce translatable standards and to enable uniformity of search and access in heterogeneous operating environments. Broker architectures have an important role in addressing interoperability and data integration issues in federated data and agent-based systems. Brokering is currently employed widely by Spatial Data Infrastructures (SDIs), current examples of which include the USA National Spatial Data Infrastructure in the USA, and INSPIRE in Europe. Examples in the geosciences include GEON, CUAHSI, OneGeology and the Semantic Mediator of the Marine Metadata Interoperability (MMI) Project. The MMI mediator allows registering a vocabulary or service and issuing a semantic query on these. Projects that have used the MMI semantic mediator include the International Coastal Atlas Network (ICAN), OOSTethys (OGC Ocean Science nteroperability Experiment), Oceans Innovation Demo 2008 and Q2O (QARTOD 2 OGC). Based on such experience they have been proposed as part of the CI approach within the EarthCube Interop group.

This makes sense given that broker approaches and their implementation are rapidly maturing with embedded capabilities that now include new technologies such as terminology and semantic meditation. In the Biomedical realm, for example, standardized terminological services have been developed as an insertable module to refine user queries. They have also been used for mapping the user's terms to appropriate medical vocabularies. Mediating broker have been developed to help with composition of services and information on the Semantic Web. In such efforts compositional knowledge is used to help automate Web service flow generation. This includes operational (syntactic), semantic and pragmatic knowledge. Operational knowledge helps assure that correct output and input types for possible service composition, while the semantic component uses domain-specific expert knowledge to shape the Web service compositionality.

Knowledge Sifter [KCD+04] is an example of a scalable agent-based system that supports access to heterogeneous information sources such as the Web, open-source repositories, XML-databases and the emerging Semantic Web. The Knowledge Sifter architecture consists of layers of specialized agents reside that perform well-defined functions to supports interactive query specification and refinement, query decomposition, query processing, integration, as well as result ranking and presentation.

EuroGEOSS is an Earth Science Brokering framework (i.e., a family of brokers including semantic mediators) employed to bind various heterogeneous resources and adapt them to different community tools. In collaboration with FP7 GENESIS project (http://www.genesis-fp7.eu/), EuroGEOSS prototyped a Semantic Discovery Broker extending functionality of the existing Discovery Broker capacity. It implements a "third-party discovery augmentation approach": enhancing discovery capabilities of infrastructures by developing new components that leverage on existing systems and resources to automatically enrich available geospatial resource description with semantic meta-information. Currently, the EuroGEOSS DAC is able to use existing discovery (e.g. catalogs and discovery brokers) and semantic services (e.g. controlled vocabularies, ontologies, and gazetteers) in order to provide users with semantics enabled query capabilities, helping to bridge a critical gap that hinders multidisciplinary infrastructures.

Brokering frameworks, such as EurpGEOSS, follow several simplifying principles to help manage risks and enable improved semantic brokers to be added to the architecture:

- Use *Autonomy and Modularity* to keep the existing capacities as independent as possible by interconnecting and mediating standard and non-standard capacities;
- *Enhance and Supplement*, but do not supplant, system mandates and prior governance arrangements;
- Provide *Low Entry Barriers* for both resource users and data producers;
- Support *Flexibility and Extensibility* to accommodate existing and future information systems and information technologies; and
- *Incrementally Build On* existing cyberinfrastructures and incorporate heterogeneous resources by introducing distribution and mediation functionalities to interconnect heterogeneous resources.

Ontology Repositories and Management

Earth Science and affiliated fields like climatology increasingly has a need to assemble, integrate and analyze large datasets. This remains a challenge because archived data reflects heterogeneous data models and independent conceptualizations, which makes meaningful data sharing difficult. Metadata to annotate the meaning of data is a central feature of information sharing infrastructure to provide such capabilities as:

- data and service discovery,
- facilitating interoperability and
- linking.

Large catalogs or repositories of meta-data are now part of several cyberinfrastructures (CIs) engaged in overcoming the challenging of sharing primary data in the Earth Sciences. Examples include the Earth System Grid (ESG) [WAB+08], INSPIRE [NBR+09] and the HydroCatalog & the metadata services used by CUAHSI (http://semanticcommunity.info/@api/deki/files/13844/=056_Tarboton.pdf). ESG projects can register appropriate data characteristics (e.g., dataset title, variable names, spatial and temporal boundaries, etc.). INSPIRE includes a Discovery Service about web service capabilities but also to discover and get metadata for specific resources based on the resource unique IDs.

Such efforts are a useful first step, however, the establishment of community metadata standards, frames and creation of applications and standard formats to facilitate collection remains a challenge. For one thing there remain many meta-data formats which allow semantic mismatches. Most metadata lacks proper and systematic semantics to handle diverse data bases and schemas. Current metadata standards, including those specified by the FGDC for spatial data, were not designed for automated Web searching. More semantic languages, such as RDF(S) and OWL, along with tools for dealing with ontologies, can be used to provide better common metadata with useful knowledge structures. As part of the EarthCube CI, large repositories of Earth Science data should be converted into RDF and linked to the existing linked data cloud (http://linkeddata.org/).

To help handle semantic heterogeneity for querying and processing, background and local ontologies in proper semantic languages are needed to resolve individual data source terms or parameter identifiers within and between domains. Further, some integrated upper level suite of ontologies, driven by use case requirements combining particular aspects of domain ontologies would help integrate between different domains. These need to be stored in easily accessed repositories to enable wider use.

Ontology repositories are now part of the semantic thrust in other fields such Biomedicine. BioPortal (http://bioportal.bioontology.org/) provides a good example of a well-maintained virtual repository for ontologies and other knowledge sources, along with a number of services to improve reusability of ontologies, annotations and mappings. Other individual and inter-related ontology repositories are now being created, including several members of the EarthCube community, as part of the Open Ontology Repository (OOR, http://openontologyrepository.org/) [BS09]) effort. This is a volunteer effort to promote the global use and sharing of ontologies by:

- establishing a hosted registry-repository
- enabling and facilitating open, federated, collaborative ontology repositories
- federating independent registries to enable sharing of common vocabularies and ontologies
- community based annotation and mapping, along with search and other capabilities to promote sharing and reuse
- establishing best practices for expressing interoperable ontologies and taxonomy work in registry-repositories.

All work is in compliance with open standards and uses:

- open technology (open source)
- open knowledge (open content)
- open collaboration (transparent community process)
- open to integration with "non-open" repositories via an open interface

OORs are currently used to collect such things as geospatial ontologies as part of an NSF INTEROP project (www.socop.org). Another example is the Ontology Registry and Repository (ORR) developed by the Marine Metadata Interoperability program [RBF09]. ORR is a key enabler for the MMI mission to promote the exchange, integration and use of marine data through enhanced data publishing, discovery, documentation and accessibility [GIR12]. ORR leverages and integrates well-known libraries and open source technologies to provide the oceanographic community with easy-to-use tools for creation and maintenance of vocabularies and term mappings, as well as a central location for such artifacts to greatly facilitate discovery and sharing.

Such ontology repositories have capabilities to store, manage and share ontologies, map between ontology terms, and provide browsing and search for ontologies.

Once developed, ontologies for EarthCube could be stored in these existing repositories, to allow search and update. These should be distributed, or a dedicated EarthCube ontology repository. In either case, the EarthCube cyberinfrastructure needs to supplement the existing metadata catalog to access ontologies in repositories modularly designed to work with the CI architecture.

Semantically Driven Workflows

Semantics and ontologies can play an important role in scientific workflows composition for distributed scientific data analysis [PDS10, GGK+12]. Furthermore, semantics and ontologies can play an broader role in large scale spatial planning and decision support [LRGJ12], where one encounters different levels of workflows, with information associated with higher level workflows constituting semantic constraints for lower level workflows. Large scale land-based environmental planning and decision making problems typically involve collaborative research across earth science domains as well as social and information science domains. The process for solving such "Grand Challenge" problems (e.g. regional-scale assessment and planning process for reducing conservation conflicts between threatened and endangered species conservation and energy development projects) typically follow a high level workflow consisting of steps such as defining the planning goal/objectives and the decision problem, establishing evaluation criteria for desired state of the system, developing or adopting domain process models, developing data, assessing current states of the system, designing plan alternatives, performing impact analysis on design alternatives, evaluating design alternatives and selecting a plan, etc. Each of the steps could involve a series of sub steps, and some of them involve domain process modeling – creating conceptual models for domain processes (e.g. cause-effect models among anthropogenic activities, stressors, habitat resources and species), and creating corresponding computational workflows to be used in assessing the current or simulated state of the system. Such computational workflows would in turn include scientific workflows for data processing, modular analytical tasks, visualization, and so on. Just as we have a choice from deep and formal semantics to shallow, lighter weight and implicit semantics, we will have a choice of using enterprise class semantic web services along with semantic search, discovery, composition and orchestration vis-à-vis semantically annotated RESTful services and semantic mashups [SGL07, SBRSS08] .

Ontologies can be used to formalize planning process workflows, domain process workflows, and scientific workflows. Besides coding the various levels of workflow templates, ontologies can be used to semantically annotate the resources needed for instantiating a workflow template (data sets, models and tools), indicating their purpose or classification, for example. Semantic registration of these resources is essential for automatic resource discovery on CyberInfrastructure, which is essential for automatic workflow orchestration. Coupled with semantic reasoning, ontologies can further guide workflow template instantiation, or guide new workflow template composition. There are many ways that the attributes of a specific planning process can semantically inform lower level workflows. For example the type and characteristics of a specific planning problem (e.g. site search or selection) may provide guidance or constraints on the type of algorithms (e.g. optimization) to be used in the computational workflow during the solution alternative design step, or the presence of multiple participant types in the planning process may suggest the use of some specific type of algorithms for deriving a common set of evaluation criteria weights (e.g. the consensus convergence algorithm). When composing a scientific workflow under a domain process computational workflow, the semantic information on the “entity type” being considered and the bounding geographic area for the entity distribution in the domain process workflow can be used to specify the input data requirement for the scientific workflow. Some of these semantic constraints are propagated down from the planning process workflow (e.g. from the planning objectives and planning spatial extent) to the domain process model workflows. All this will furthermore affect the choice of software tools (which implement algorithms) to be used during a scientific workflow.

Effective collaborative research between the Semantics and Ontologies Working Group and the Workflow Working Group can benefit from working on a common Grand Challenge type problem use case. Large-scale environmental planning problems can provide such use cases since they involve applying computational workflows to process massive amounts of heterogeneous spatial data with ever increasing analytic complexity, work that cuts across different earth science domains to social and information sciences, and can provide end-to-end interoperability use cases for EarthCube initiatives.

Conclusions

The Web has been a great boon to scientists by making it easier for them to collaborate, share documents and develop common resources and tools. The Semantic Web technologies will enhance, deepen and accelerate their ability to collaborate by enabling scientists to share their data, scientific models and software services in ways that support automated discovery, interoperability, fusion and reuse. There are tremendous opportunities in the Geosciences for applying this approach to develop a cyber infrastructure that will help to advance the field with both short and long term payoff. This document has outlined several of the immediate steps that can be taken as well as identifying some of the longer term issues and goals. We welcome feedback and contributions from the Geoscience and Computing communities.

References

[A05] Akkiraju, et al. The Web Service Semantics – WSDL-S. W3C member submission, November 2005. http://www.w3.org/Submission/WSDL-S/

[ASRUAK06] Ismailcem Budak Arpinar, Amit P. Sheth, Cartic Ramakrishnan, E. Lynn Usery, Molly Azami, Mei-Po Kwan. Geospatial Ontology Development and Semantic Analytics. T. GIS 10(4): 551-575 (2006).

[B06] Yaser Bishr. Geospatial semantic web. In Sanjay Rana and Jayant Sharma, editors, Frontiers of Geographic Information Technology, pages 139–154. Springer, 2006. ISBN 978-3-540-31305-2.

[BHB09] Christian Bizer, Tom Heath and Tim Berners-Lee, Linked Data - The Story So Far, International Journal on Semantic Web and Information Systems 5(3), 2009.

[BKta] Robert Battle and Dave Kolas, Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. Semantic Web journal, to appear. Available from http://www.semantic-web-journal.net/

[BMT08] B. Bennett, D. Mallenby, and A. Third. An ontology for grounding vague geographic terms. In Formal Ontology in Information Systems – Proceedings of the Fifth International Conference (FOIS 2008), volume 183, pages 280–293. IOS Press, 2008.

[BHK+10] Harold Boley, Gary Hallmark, Michael Kifer, Adrian Paschke, Axel Polleres, Dave Reynolds, RIF Core Dialect. W3C Recommendation 22 June 2010, http://www.w3.org/TR/rif-core/

[BMJNM11] Bröring, A., Maue, P., Janowicz, K., Nüst, D., and Malewski, C. (2011): Semantically-Enabled Sensor Plug & Play for the Sensor Web. Sensors 2011, 11, 7568-7605.

[BS03] Bittner, Thomas, and Barry Smith. 2003. A Theory of Granular Partitions. In: Foundations of Geographic Information Science, M. Duckham, M. F. Goodchild and M. F. Worboys, eds., London: Taylor & Francis Books, 2003, 117-151.

[BS09] Baclawsk, Kenneth; Todd Schneider. 2009. The Open Ontology Repository Initiative: Requirements and Research Challenges. Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK 2009). ISWC 2009, Oct. 25-29, 2009, Chantilly, VA.

[BSM07] Bittner, Thomas; Barry Smith; Maureen Donnelly. 2007. The Logic of Systems of Granular Partitions. http://ontology.buffalo.edu/smith/articles/BittnerSmithDonnelly.pdf.

[CH12] David Carral Martinez, Pascal Hitzler, Extending Description Logic Rules. In: E. Simperl at al. (eds.), Proceedings ESWC2012, Lecture Notes in Computer Science Vol. 7295, Springer, Heidelberg, to appear.

[CKH12] David Carral Martinez, Adila Alfa Krisnadhi, Pascal Hitzler, Integrating OWL and Rules: A Syntax Proposal for Nominal Schemas. In: Proceedings OWLED 2012. To appear.

[CL] Common Logic. ISO/IEC 24707:2007 - Information technology — Common Logic (CL): a framework for a family of logic-based languages. http://iso-commonlogic.org/.

[CV05] Philipp Cimiano, Johanna Völker: Text2Onto. In: Andrés Montoyo, Rafael Muñoz, Elisabeth Métais (Eds.): Natural Language Processing and Information Systems, 10th International Conference on Applications of Natural Language to Information Systems, NLDB 2005, Alicante, Spain, June 15-17, 2005, Proceedings. Lecture Notes in Computer Science 3513 Springer 2005, pp. 227-238.

[GS08] I. Cruz and W. Sunna. Structural alignment methods with applications to geospatial ontologies. Transactions in GIS, special issue on Semantic Similarity Measurement and Geospatial Applications, 12(6):683–711, 2008.

[CX08] I. F. Cruz and H. Xiao, Data Integration for Querying Geospatial Sources. Geospatial Services and Applications for the Internet, eds. John Sample, Kevin Shaw, Shengru Tu, and Mahdi Abdelguerfi, Springer, pp.113-137, 2008.

[Cue05] Cuel R., Cristani M., A Survey on Ontology Creation Methodologies, Int. J. Semantic Web Inf. Syst. (2005), Vol. 1, no. 2, pp. 49-69.

[E02] Egenhofer, M.J. Toward the Semantic Geospatial Web. In *Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems,* McLean, Virginia, 2002.

[FDF12] Nicola Fanizzi, Claudia d'Amato, Floriana Esposito: Induction of robust classifiers for web ontologies through kernel machines. J. Web Sem. 11: 1-13 (2012)

[FS02] F. Fonseca and A. Sheth, The Geospatial Semantic Web, UCGIS White paper, 2002.

[G04] A. Galton. Fields and objects in space, time, and space-time. Spatial Cognition and Computation, 1:39–68, 2004.

[G05] A. Gangemi. Ontology design patterns for semantic web content. In Y. Gil, E. Motta, R. Benjamins, and M. Musen, editors, The Semantic Web – ISWC 2005, 4th International Semantic Web Conference, volume 3729 of Lecture Notes in Computer Science, pages 262–276. Springer, 2005.

[G98] N. Guarino. Formal Ontology and Information Systems. In N. Guarino, editor, International Conference on Formal Ontology in Information Systems (FOIS1998), pages 3–15. IOS Press, Trento, Italy, 1998.

[GEFK99] Goodchild, M.F., Egenhofer, M.J., Fegeas, R., and Kottman, C.A. (eds.) *Interoperating Geographic Information Systems*. New York, Kluwer, 1999.

[GGK+12] Gil, Y.; Gonzalez-Calero, P. A.; Kim, J.; Moody, J.; and Ratnakar, V. (2012) A Semantic Framework for Automatic Generation of Computational Workflows Using Distributed Data and Component Catalogs. *To appear in the Journal of Experimental and Theoretical Artificial Intelligence*.

[GIR12] Graybeal, J., Isenor, A., Rueda, C. (2012): Semantic Mediation of Vocabularies for Ocean Observing Systems. Computers & Geosciences. Volume 40, March 2012, Pages 120-131, ISSN 0098-3004.

[GH07] Stephan Grimm, Pascal Hitzler, Semantic Matchmaking of Web Resources with Local Closed-World Reasoning. International Journal of e-Commerce **12** (2), 89-126, Winter 2007-8.

[GM09] Antony Galton and Riichiro Mizoguchi. 2009. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Appl. Ontol.* 4, 2 (April 2009), 71-107.

[GS04] Grenon, P., and B. Smith. 2004. SNAP and SPAN: Towards dynamic geospatial ontology. Spatial Cognition and Computation, 4(1), 2004, pp. 69 -104.

[HHM+10] Harry Halpin, Patrick J. Hayes, James P. McCusker, Deborah L. McGuinness, Henry S. Thompson: When owl: sameAs Isn't the Same: An Analysis of Identity in Linked Data. In: Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, Jeff Z. Pan, Ian Horrocks, Birte Glimm (Eds.): The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I. Lecture Notes in Computer Science 6496 Springer 2010, pp. 305-320

[HJD08] Hart, G., Johnson, M., Dolbear, C.: Rabbit: Developing a Control Natural Language for Authoring Ontologies. 5th European Semantic Web Conference (ESWC'08), Tenerife, Spain (2008)

[HKPBR99] F. Harvey, W. Kuhn, H. Pundt, Y. Bisher, C. Riedemann. Semantic Interoperability A Central Issue for Sharing Geographic Information. in The Annals of Regional Science, Vol. 33 (1999), pp. 213-232.

[HKP+09] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, OWL 2 Web Ontology Language: Primer. W3C Recommendation, 27 October 2009. http://www.w3.org/TR/owl2-primer/

[HKR10] Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, Foundations of Semantic Web Technologies, CRC Press / Chapman & Hall, 2010.

[HPST09] Cory A. Henson, Josh K. Pschorr, Amit P. Sheth, and Krishnaprasad Thirunarayan. 2009. SemSOS: Semantic sensor Observation Service. In *Proceedings of the 2009 International Symposium on Collaborative Technologies and Systems* (CTS '09). IEEE Computer Society, Washington, DC, USA, 44-53.

[HPB+04] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, 21 May 2004, http://www.w3.org/Submission/SWRL/

[JHSVY10] Prateek Jain, Pascal Hitzler, Amit P. Sheth, Kunal Verma, Peter Z. Yeh: Ontology Alignment for Linked Open Data. International Semantic Web Conference (1) 2010: 402-417

[J10] Krzysztof Janowicz: The Role of Space and Time For Knowledge Organization on the Semantic Web. Semantic Web Journal. Volume 1, Number 1-2, pp. 25-32, IOS Press. 2010

[J12] Janowicz, K. (2012; forthcoming): Observation-Driven Geo-Ontology Engineering. Transactions in GIS.

[JBSSEL11] Janowicz, K., Bröring, A., Stasch, C., Schade, S., Everding, T., and Llaves, A. (2011): A RESTful Proxy and Data Model for Linked Sensor Data. International Journal of Digital Earth. DOI:10.1080/17538947.2011.614698, pp. 1-22.

[JSBKMS09] Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maue, P. and Stasch, C. (2010): Semantic Enablement for Spatial Data Infrastructures. Transactions in GIS 14(2), Blackwell Publishing, pp. 111-129.

[K05] W Kuhn - Journal on data semantics III. Geospatial semantics: why, of what, and how?. Lecture Notes in Computer Science, Volume 3534/2005, pp 1-24. 2005. URL http://www.opengeospatial.org/standards/requests/80. Document 11-052r3.

[K09] W. Kuhn. Semantic Engineering. In G. Navratil, editor, Research Trends in Geographic Information Science, pages 63–76. Springer, 2009.

[KCD+04] L. Kerschberg, M. Chowdhury, A. Damiano, et al, "Knowledge Sifter: Ontology-Driven Search over Heterogeneous Databases," *In Proc. 16th Int. Conf. Scientific and Statistical DB Management*, 2004.

[KHM12] Matthias Knorr, Pascal Hitzler, Frederick Maier, Reconciling OWL and Non-monotonic Rules for the Semantic Web. In: Proceedings ECAI2012. To appear.

[KR10] Michael Kifer, Harold Boley, RIF Overview. W3C Working Group Note, 22 June 2010. http://www.w3.org/TR/rif-overview/

[KMH11] Adila Alfa Krisnadhi, Frederick Maier, Pascal Hitzler, OWL and Rules. In: A. Polleres, C. d'Amato, M. Arenas, S. Handschuh, P. Kroner, S. Ossowski, P.F. Patel-Schneider (eds.), Reasoning Web. Semantic Technologies for the Web of Data. 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures. Lecture Notes in Computer Science Vol. 6848, Springer, Heidelberg, 2011, pp. 382-415.

[KWV07] Krivov S, Williams R, Villa F: GrOWL: A tool for visualization and editing of OWL ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web* 2007, 5(2):54-57.

[KMKH11] Markus Krötzsch, Frederick Maier, Adila Alfa Krisnadhi, Pascal Hitzler, A Better Uncle For OWL - Nominal Schemas for Integrating Rules and Ontologies. In: S. Sadagopan, Krithi Ramamritham, Arun Kumar, M.P. Ravindra, Elisa Bertino, Ravi Kumar (eds.), WWW '11 20th International World Wide Web Conference, Hyderabad, India, March/April 2011. ACM, New York, 2011, pp. 645-654.

[KSH11] Adila Krisnadhi, Kunal Sengupta, Pascal Hitzler, Local Closed World Semantics: Keep it simple, stupid! In: Riccardo Rosati, Sebastian Rudolph, Michael Zakharzaschev, Proceedings of the 2011 International Workshop on Description Logics (DL2011), Barcelona, Spain, July 2011. CEUR Workshop Proceedings, Vol. 745, pp. 532-542.

[LH10] Jens Lehmann, Pascal Hitzler, Concept Learning in Description Logics Using Refinement Operators. Machine Learning 78(1-2), 203-250, 2010.

[LRGJ12] Li, N., Raskin, R., Goodchild, M. and Janowicz, K. (2012; in print) An Ontology-Driven Framework and Web Portal for Spatial Decision Support. To appear in Transactions in GIS vol. 16 (s1).

[MKSI96] E. Mena and V. Kashyap and A. Sheth and A. Illarramendi. OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies}, 1996, 14--25, IEEE Computer Society Press.

[MM04] Frank Manola, Eric Miller, RDF Primer, W3C Recommendation 10 February 2004. http://www.w3.org/TR/rdf-primer/

[MMR12] Patrick Maué, Henry Michels and Marcell Roth. Injecting Semantic Annotations into (Geospatial) Web service descriptions. Semantic Web jorunal. DOI 10.3233/SW-2012-0061, 2012.

[MSSSS03] Maedche A, Staab S, Stojanovic N, Studer R and Sure Y 2003 SEmantic portAL: The SEAL Approach. In Fensel D, Hendler J A, Lieberman H, and Wahlster W (eds) Spinning the Semantic Web, MIT Press: 317-359

[NBR+09] Nogueras-Iso, J., Barrera, J., Rodriguez, A. F., Recio, R., Laborda, C., and Zarazaga-Soria, F. J. (2009). Development and deployment of a services catalog in compliance with the inspire metadata implementing rules. In van Loenen, B., Zevenbergen, J., and Besemer, J., editors, Spatial Data Infrastructure Convergence: Research, Emerging Trends, and Critical Assessment, volume 48 of Groene series, pages 21-34. The Netherlands Geodetic Commission (NGC), The Netherlands.

[Obr03] Obrst, L. 2003. Ontologies for Semantically Interoperable Systems. Proceedings of the Twelfth ACM International Conference on Information and Knowledge Management (CIKM 2003), Ophir Frieder, Joachim Hammer, Sajda Quershi, and Len Seligman, eds. New Orleans, LA, November 3-8, New York: ACM, pp. 366-369.

[Obr11] Obrst, Leo; Pat Cassidy. 2011. *The Need for Ontologies: Bridging the Barriers of Terminology and Data Structures.* Chapter in: GSA Memoir on Societal Challenges and Geoinformatics, Sinha, Krishna; Linda Gundersen; Ian Jackson; David Arctur, eds. Publication of a Memoir Volume. Geological Society of America (GSA). November 1, 2011.

[Obr10] Obrst, Leo. 2010. Ontological Architectures. Chapter 2, pp. 27-66 in the book: TAO – Theory and Applications of Ontology: Computer Applications, Roberto Poli, Johanna Seibt, Achilles Kameas, eds. September, 2010, Springer.

[OGC11] Open Geospatial Consortium. OGC GeoSPARQL - A Geographic Query Language for RDF Data. Open Geospatial Consortium, 2011.

[OE06] Ou, W. & Elsayed, A. (2006). A knowledge-based approach for semantic extraction. International Conferenceon Business Knowledge Management, Macao, 2006.

[OWLS04] http://www.w3.org/Submission/OWL-S/

[OWM99] Obrst, L., G. Whittaker, A. Meng. 1999. Semantic Context for Object Exchange, AAAI Workshop on Context in AI Applications, Orlando, FL, July 19, 1999.

[PDS10] C. Pedrinaci, . Domingue and A. Sheth, 'Semantic Web Services,' In Handbook: Semantic Web Technologies, Volume 2: Semantic Web Applications. J. Domingue;D. Fensel;J. Hendler, (Eds.). Springer, 2010.

[PL04] Probst, F. and M. Lutz. Giving Meaning to GI Web Service Descriptions. in 2nd International Workshop on Web Services: Modeling, Architecture and Infrastructure (WSMAI-2004). 2004. Porto, Portugal.

[RLTDF12] Achim Rettinger, Uta Lösch, Volker Tresp, Claudia d'Amato, Nicola Fanizzi: Mining the Semantic Web - Statistical learning for next generation knowledge bases. Data Min. Knowl. Discov. 24(3): 613-662 (2012)

[RKL+05] Dumitru Roman, Uwe Keller, Holger Lausen, Jos de Bruijn, Rubén Lara, Michael Stollberg, Axel Polleres, Cristina Feier, Christoph Bussler, and Dieter Fensel: Web Service Modeling Ontology, *Applied Ontology*, 1(1): 77 - 106, 2005.

[RP05] R. Raskin, M. Pan, Knowledge representation in the Semantic Web for Earth and environmental terminology (SWEET), Computers & Geosciences 31 (9), 1119-1125, 2005.

[RBF09] Rueda, C., Bermudez, L., Fredericks, J. (2009). The MMI Ontology Registry and Repository: A Portal for Marine Metadata Interoperability. OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges. Oct. 2009.

[SBRSS08] S. Sahoo, O. Bodenreider, J. Rutter, K. Skinner, A. Sheth. **An ontology-driven semantic mashup of gene and biological pathway information: Application to the domain of nicotine dependence.** Journal of Biomedical Informatics 41 (5), 2008, 752-765

[Sch05] Schwitter, R.: A Controlled Natural Language Layer for the Semantic Web. In: Zhang, S., Jarvis, R. (eds.) AI 2005. LNCS (LNAI), vol. 3809, pp. 425–434. Springer, Heidelberg (2005)

[SE08] P. Shvaiko and J. Euzenat. Ten Challenges for Ontology Matching. In R. Meersman and Z. Tari, editors, On the Move to Meaningful Internet Systems: OTM 2008, volume 5332 of LNCS, pages 1164–1182. Springer, 2008.

[SGL07] Amit P. Sheth, Karthik Gomadam, Jonathan Lathem. 'SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups,' IEEE Internet Computing, November - December 2007, 84-87. Also W3C submission: http://www.w3.org/Submission/SA-REST/

[Sku03] D. Skuce. 2003. A Controlled Language for Knowledge Formulation on the Semantic Web. http://www.site.uottawa.ca:4321/factguru2.pdf.

[SKW08] F.M. Suchanek, G. Kasneci, G. Weikum, YAGO: A large ontology from Wikipedia and WordNet. Journal of Web Semantics, 2008.

[SM10] G. Sinha and D. Mark. Toward a foundational ontology of the landscape. In Extended Abstracts of GIScience 2010, 2010.

[SM10b] G. Sinha and D. Mark. Cognition-based extraction and modelling of topographic eminences. Cartographica, 45(2):105–112, 2010

[SRT05] A. Sheth, C. Ramakrishnan, and C. Thomas. Semantics for the semantic web: the implicit, the formal and the powerful. International Journal on Semantic Web and Information Systems, 1:1–18, 2005.

[SS10] Katharina Siorpaes and Elena Simperl. Human intelligence in the process of semantic content creation. World Wide Web (WWW) Journal, 13(1), 2010.

[SSOR09] K. Stock, M. Small, Y. Ou, and F. Reitsma. OGC Discussion Paper 09-010 – OWL Application Profile of CSW. Technical report, Open Geospatial Consortium, 2009.

[TMBS08] Christopher Thomas, Pankaj Mehra, Roger Brooks, Amit P. Sheth: Growing Fields of Interest - Using an Expand and Reduce Strategy for Domain Model Extraction. Web Intelligence 2008: 496-502.

[USDL11] Kay Kadner and Daniel Oberle (editors) Unified Service Description Language XG Final Report. W3C Incubator Group Report 27 October 2011.

[U00] M. Uschold. Creating, integrating and maintaining local and global ontologies. In 14th European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany, August 2000.

[VS07] Kunal Verma, Amit P. Sheth, Semantically Annotating a Web Service, IEEE Internet Computing, March/April 2007, 11(2), 83-85. Also W3C recommendation: http://www.w3.org/TR/sawsdl/

[WAB+08] D N Williams, R Ananthakrishnan, D E Bernholdt, S Bharathi, D Brown, M Chen, A L Chervenak, L Cinquini, R Drach, I T Foster, P Fox, D Fraser, S Hankin, P Jones, C Kesselman, D E Middleton, J Schwidder, R Schweitzer, R Schuler, A Shoshani, F Siebenlist, A Sim, W G Strand, N. Wilhelmi, "The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data" Bulletin of the sAmerican Meteorological Society (BAMS), 2008

[W3CSSN12] Compton, M., Barnaghi, P., Bermudez, L., Garcia-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., and Taylor, K. (accepted): The SSN Ontology of the W3C Semantic Sensor Network Incubator Group. Journal of Web Semantics.