Project Description

The science and engineering communities are producing very large data sets that are also increasingly complex and diverse. These data sets are very well suited for particular narrowly-defined, discipline-specific purposes. In principle, these data sets could be used for solving more broadly-defined scientific problems such as understanding whole organisms, ecosystems and human populations. However, incorporating multiple data types from multiple sources to solve these problems remains a significant challenge. For example, a testable macroscopic biological hypothesis might involve the effect of environmental or climatic change on the genomic makeup of a given organism. As another example, a macroeconomic hypothesis concerning the most efficient use of resources to improve the quality of life in a region will depend on cultural and environmental knowledge as well as economic statistics.

While the data sets that are currently being developed typically engender the greatest level of enthusiasm by the communities that are creating them, data sets created in the past can have equal importance for related communities. Biodiversity is a case in point. The painstaking observations by generations of biologists over centuries represent an important resource for modern ecology and biodiversity studies, but those observations are locked in old textbooks and monographs that are not easily accessed by modern computing technology. The problem is not just the differences in recording media (paper versus disks) but also the enormous changes in terminology over time. Current data sets run the risk of an even more rapid obsolescence as the meaning of the data fields is forgotten even by the individuals who introduced them.

We believe in the promise of semantic technologies based on logic, databases and the Semantic Web as a means of addressing the problems of meaningful access to and integration of data over decades and centuries. Such technologies enable distinguishable, computable, reusable, and sharable meaning of information artifacts, including data sets, documents and services. We also believe that making this vision a reality requires additional supporting resources, and that these resources should be open, extensible, and provide common services over the ontologies. Our belief in this vision is based not only on current experience but also on the deep philosophical foundations that underly modern ontological engineering.

We propose to develop an open ontology repository (OOR) of controlled vocabularies and knowledge models that have been encoded in RDF, OWL, and other knowledge representation languages. More specifically, we propose to develop an open repository for the metadata and data sets of the following communities:

- Biology, especially the genomics, proteomics and other "omics" communities. This will be based on the highly successful BioPortal repository.

- Biodiversity, especially the species pages in the Encyclopedia of Life.

- Climate and environmental communities (including both natural environments and built environments).

- Human culture and sociology.

The data sets of these communities share a number of characteristics that make them well suited to the proposed OOR:

- They represent large data sets that are of considerable importance to their respective communities.

- Some of the data sets, especially those of the biodiversity community, represent data that is very old, sometimes centuries old, yet still of considerable value.

1

- The integration of these data sets opens up exciting research opportunities not only for the natural sciences but also for environmental and social sciences.

- The data sets have complex semantics, and there is no clear distinction between data and metadata. As a result, modern relational database technology is poorly suited for modeling the data sets.

While these data sets provide a compelling case for the proposed OOR, the prospect of broader impacts is even more compelling. As an integral part of the proposed project, we intend to foster a vigorous educational outreach program to bring other data-intensive research communities into the OOR initiative. Since the OOR will be an open, federated architecture and infrastructure, it is intended to be utilized by communities to host their own ontologies as well as allowing the communities to adapt previously established ontologies for their own purposes.

To address the issue of long-term sustainability, we propose to develop a new paradigm for maintaining semantic linkages available through the Internet. Specifically, we will develop a federated knowledge repository that can collectively correct for multiple points of failure and can foster collaborative stewardship of scientific knowledge. Particular emphasis will be given to the development of technological solutions that build on existing, proven architectures for maintaining biological (e.g., BioPortal, OBO Foundry and the International Nucleotide Sequence Data Consortium) and abiotic data (e.g., the National Climatic Data Center), as well as standards for metadata and services (e.g., ISO XMDR, WSDL and UDDI).

1 The OOR Statement of Purpose

The purpose of an OOR is to provide an architecture and an infrastructure that supports a) the creation, sharing, searching, and management of ontologies, and b) linkage to database and XML Schema structured data and documents. Complementary goals include fostering the ontology community, the identification and promotion of best practices, and the provision of services relevant to ontologies and instance stores. Examples of anticipated services include automated semantic interpretation of content expressed in knowledge representation languages, the creation and maintenance of mappings among disparate ontologies and content, and inference over this content. We believe that the OOR will ultimately support a broad range of semantic services and applications of interest to enterprises and communities.

Achieving these goals will help reduce semantic ambiguity whenever and wherever information is shared, thereby allowing information to be located, searched, categorized, and exchanged with a more precise expression of its content and meaning. The artifacts of the repository will provide a semantic grounding for diverse formats and domains, ranging from the conceptual domains and specific disciplines of communities to technical schemas such as WSDL, UDDI, RSS, and XML schemas, and of course expressed in standard ontology languages such as RDF, OWL, Common Logic, and others. Perhaps most importantly, the repository will enable wide-scale knowledge re-use and reduce the need to re-invent the wheel when defining concepts and relationships that are already understood.

These goals cannot be achieved all at once, and must track the evolution of best practices as well as technology itself. It is also good system development practice to bound complexity by defining a system in terms of a series of short-term, achievable objectives. For this reason, as for other such initiatives, it is envisioned that the OOR will be developed in a series of phases, proceeding from the simple to the complex, with achievable goals that capitalize on previous experience and the emergence of technology over time. It is important to note that in any given phase, planning and prototyping for subsequent phases is always in progress.

2 State of the Art

The purpose of this section is to set out the major design decisions and the technology choices which are important to the creation of ontology repositories.

Ontology repositories support the storage, search, retrieval and interoperation of multiple ontologies.

Ontology repositories support macro-level storage, query and retrieval (across the collection of ontologies) and micro-level operations (within individual ontologies). At each level we would like to support both text search, and semantic search (variously faceted search, SPARQL, ontology and ontology language literate search). Some ontology repositories have used the same technologies for both macro-level and micro-level operations.

A key decision is the choice of a representation of the ontologies. Current practice includes: text, frames (e.g., OBO), graphs (e.g., RDF), and various types of logic, e.g., description logics (e.g., OWL-DL), first order logic (e.g., Common Logic), sorted logics, possibly higher order logic (HOL). Other possibilities include the use of UML (e.g., in the OMG Ontology Definition Metamodel).

Ontologies have been stored in long narrow relations, e.g., "triple stores" of RDF triples (subject, relationship, object), relational databases, customized data stores. Increasingly implementers are using "quad stores" in order to support Named Graphs. "Column stores" such as MonetDB and Vertica have also been used to store ontologies.

For the purposes of ontology interoperation it helps to have all of the ontologies in the repository encoded in a common representation. However, this requires the sometimes difficult and lossy translation of ontologies among various representations into the common representation. Some ontology repositories store ontologies in their native representation, with metadata to identify the representation language.

We also need some way to support ontology interoperation by specifying the mappings among entities, e.g., via relationships such as "same as," "is a," and "part of." Other mapping relationships include: "see also," "similar to." Some ontology mapping consistency checking tools check that mappings between partially ordered ontologies, e.g., taxonomies, preserve the partial orders.

Many ontology repositories which support partially ordered ontologies (taxonomies and partonomies) may decide to materialize the transitive closure of the partial order relation. This provides faster query evaluation at the expense of additional ingestion costs, storage, and maintenance.

Provenance of definitions in ontologies is important to the credibility, scientific attribution, and regulatory compliance of ontologies. In particular, many definitions are embodied in legislation, administrative regulations, court decisions, professional society standards.

Provenance and other metadata are distinguishing features of recent ontology repositories. Such metadata ranges from authorship, and creation date, version information, to evaluation and usage reports. Other metadata may include intended use (context).

Modularization support is useful for large ontologies, and for facilitating the reuse and mapping of portions of ontologies.

In a distributed setting, ontology repository developers increasingly are adopting Service Oriented Architectures (SOA), providing access, search, and other capabilities via web services. Two major approaches to SOA are REST and SOAP. REST is built on HTTP, with a small set of operators (GET, PUT, POST, DELETE) and the use of URL (or URI) addresses for all objects of interest. SOAP is based on XML RPCs. REST is much simpler to implement and should be adequate for typical ontology repository functions. SOAP is supported by a wide variety of software tools. Both SOA approaches are currently being used.

Finally, an ontology repository typically facilitates access to a variety of ontology related tools: creation, editors, pretty printers, visualization tools, differencing tools, modularization tools, import / export, version management, access control, inference engines, explanation, summarization.

3 Quality and Gatekeeping

We distinguish between gatekeeping and quality control. Gatekeeping criteria are a set of minimal requirements that any ontology within the OOR has to meet. The latter are intended to enable the users of the OOR to find quickly ontologies that fit their needs; the criteria are not supposed to ensure the quality of the ontologies.

3.1 Gatekeeping Criteria

The ontologies in the OOR have to meet the following criteria:

1. The ontology is submitted in a publicly described language and format.

2. The ontology is read accessible.

3. The ontology is expressed in a formal language with a well-defined syntax.

4. The authors of the ontology provide the required metadata as specified under section V.

5. The ontology has a clearly specified and clearly delineated scope.

6. Successive versions of the ontology are clearly identified.

7. The ontology is appropriately named.

It is particularly important that the required metadata include information about the process that is employed to create and maintain the ontology. (Is the ontology maintained in a cooperative and transparent process? Can anybody participate in this process?) Further, the metadata has to include information about the license under which the ontology is submitted.

3.2 Quality Control

It is not sufficient for the OOR just to store ontologies, but that it needs to enable the evaluation of the ontologies within it. The OOR will offer functionalities like those on social networking sites which would allow users to comment on ontologies and rank them. Further, the OOR will enable selective views of the repository using tags provided by subcommunities that characterize ontologies with respect to their chosen criteria. For example, such a view might select for ontologies for specific fields of research or industries, or for ontologies satisfying specific quality criteria or levels of organizational approval.

4 Metadata for Ontologies

To support the sharing and reuse of ontologies within the repository the OOR will store both ontologies and metadata for ontologies.

The metadata will allow users to:

- determine whether an ontology is suitable for a user purpose;

- capture the design rationales that underlie the ontology;

- find information about author, author credentials, and source of ontology reference material

- retrieve ontologies for use in domain applications;

- retrieve ontologies to be integrated with other ontologies;

- retrieve ontologies that will be extended to create new ontologies;

- determine whether or not an ontology can be integrated with given ontologies;

- determine whether a set of ontologies retrieved from the repository can be used together;

- determine whether an ontology in the repository can be partially shared.

There will be policies for creation and modification of metadata and documentation of ontologies and the management of the persistence and sustainability of ontologies.

Users (including end-users, ontology and repository developers, subject matter experts, stakeholders) should participate in the collaborative ontology development life cycle and in decisions regarding what metadata are suitable for ontologies in the repository.

The metadata will include both logical metadata (logical properties of the ontology independent of any implementation or engineering artifact) and engineering metadata (properties of the ontology considered as an engineering artifact).

4.1 Logical Metadata

4.1.1 Language

The first logical property is to identify the language used to specify the ontology.

The report "Evaluating Reasoning Systems" contains a classification of formal languages used to specify ontologies. A formal language has a syntax (logical symbols together with a formally specified grammar) and a model theory (which specifies the conditions under which expressions in the language can be given particular truth assignments).

A formalizable language has a syntax, although it does not have a model theory. Examples of such approaches include Topic Maps and folksonomies (which are written in XML) and ISO 15926 (which is written in EXPRESS).

Finally, some ontologies are only specified in natural language, including Wordnet, taxonomies, and thesauri.

4.1.2 Modularity

A second property of ontologies is based on modularity – is a particular ontology a monolithic set of axioms, or is it composed of a set of smaller modules? Furthermore, is each module considered to be a separate ontology within the repository? If not, what are the relationships between the modules and which modules of an ontology can be used separately?

For example, the Process Specification Language (PSL) consists of a set of modules which are extensions of a common core theory PSL-Core. Metadata for each module specifies which other modules must also be included when using the module.

4.1.3 Relationships between ontologies

We can also specify various logical relationships between ontologies within the repository, including mutual consistency, extension, and entailment, and semantic mappings.

4.2 Engineering Metadata

In addition to the logical metadata for ontologies, we need to specify metadata for ontologies as considered as engineering artifacts. This includes

- provenance

- versioning

- existing applications of the ontology (e.g. interoperability, search, decision support)

- domain-specificity (e.g. biology, supply chain management, manufacturing

4.3 Conclusions regarding Metadata

The Ontology Metadata Vocabulary (OMV), Dublin Core, ISO 11179, ISO 19763, and other existing approaches to provenance and versioning metadata are all candidates for aspects of the metadata for ontologies in the OOR.

We will use an empirical approach to the identification and evaluation of ontology metadata. Proposals for ontology metadata already exist, and we will evaluate them using use-case scenarios. These scenarios both motivate the use of the metadata and help establish best practices.

5 Repository Architecture

The Architecture of a repository for enabling wide-scale searching and sharing of ontologies will be open and extensible. The Architecture design will be modular in nature and provide for ontology storing, sharing, searching, governance, and management of the repository infrastructure and content.

5.1 Architecture Approach

The core approach for the Open Ontology Repository is a federated, service oriented architecture. This approach provides for distributed ontology storage, repository management and service support. Metadata will be provided for every ontology in the repository. The repository will also provide connections for logical services, inference engines etc.

Those who engage in the federation must include required metadata. This metadata must include any access constraints.

Over the repository will be an ontology that is inclusive of both the metadata of ontologies and the information required for operational use.

5.2 Core Requirements

The requirements presented are important to the enablement of wide-scale knowledge re-use.

1. The repository architecture shall be scalable.

2. The architecture shall be optimized for sharing, collaboration and reuse.

3. The repository shall be capable of supporting ontologies in multiple formats and levels of formalism.

4. The repository architecture shall support distributed repositories.

5. The repository architecture shall support explicit machine usable/accessible formal semantics for the meta-model of the repository.

6. The repository shall provide a mechanism to address intellectual property and related legal issues/problems.

7. The repository architecture shall include a core set of services, such as support for adding, searching and mapping across ontologies and data related to the stored ontologies.

8. The repository architecture shall support additional services both directly within the province of the repository and as external services.

9. The repository should support all phases of the ontology lifecycle.

5.3 Repository Management

An ontology repository requires mechanisms for effective management. The understanding is that as a repository and its infrastructure evolve, more management support mechanisms will be included.

Required mechanisms will provide the capabilities to:

1. enforce access policies

2. enforce submission policies

3. enforce governance policies

4. enforce change management policies

5. control user and administrator access

Highly recommended mechanisms will provide the capabilities to:

1. create usage reports

2. validate syntax

3. check logical consistency

4. automatically categorize a submission

5.4 Service and Application Support

OOR interfaces should support internal and external services and applications including:

- Ontology creation tools

- Ontology editors

- Ontology differencing tools

- Ontology modularization tools (clustering, etc.)

- Ontology export

- Ontology visualization (e.g., graph visualization)

- Version management

- Access control

5.5 Discovery Support

To facilitate knowledge discovery the repository shall provide metadata capabilities to support search capabilities, governance process, and management. The repository will support discovery by at least the following:

- domain

- author/creator/source

- version

- language

- terminology and controlled vocabularies

- quality

- mapping

- inference

6   Key Personnel

Kenneth Baclawski is the PI. He is responsible for overall coordination of the project and directing of outreach to other communities, especially those represented by the DataNet Partners. This will specifically provide for active community input and participation in all phases and all aspects of DataNet Partner activities. In his capacity as a faculty member at Northeastern University, he will develop new tools and capabilities for learning that integrate research and education at all levels.

Leo Obrst will lead a subcontract at MITRE. The MITRE sub-team will address a number of aspects relevant to an OOR which will include a repository of both ontologies, instance data, rules, and potentially raw or source data linked to instances and ontologies, and services to support public access, integration, analysis, adaptation, and preservation of knowledge rich information for science. These aspects are the following:

1. Ontology Evaluation. The development of methods, practices, services, and artifacts to support automated and human reviewed evaluation and comparison of ontologies stored in the repository. [35, 45, 44]

2. Ontology Architectures, Modularization, and Alignment/Mapping. The development of requisite ontology-based architectures, including ontology lifecycle management, theories and implementations of ontology modularity, upper and middle ontologies, and research and software development of methods for automatically and semi-automatically aligning and mapping ontologies. The use and linking of metadata, controlled vocabularies, and ontologies, for intelligent search and decision support. [14, 42, 47, 63, 65, 64]

3. Ontology, Instance, and Rule Reasoning. The development and implementation of efficient logic programming-based reasoning methods that amalgamate Semantic Web-based ontologies and rules with extended Prolog and Answer Set Programming, to be used for reasoning over the ontologies, instances, and rules of the repository. [56, 55, 67]

4. Service Orchestration and Optimization to Support OOR Artifacts. Design and implementation of service-oriented architectures and services, including automated and semi-automated service orchestration and parallel optimization to support the repository. [40, 52, 68]

5. Outreach to communities in bioinformatics, national command and control, and intelligence. [37, 36, 34, 48, 46, 50, 63, 66, 67, 68]

Mark Musen will lead a subcontract at the NCBO. The NCBO sub-team will be responsible for enhancing the BioPortal project to serve as the web server and database engine for the OOR. While BioPortal is currently used as a centralized repository for biomedical ontologies, it was actually designed as a general purpose ontology repository. So it is an ideal foundation for the OOR. The NCBO sub-team will work with the LBNL sub-team to integrate the XMDR capabilities into BioPortal.

Neil Sarkar will lead a subcontract at MBL. The MBL sub-team is responsible for coordination with the Biodiversity communities, especially those communities involved in the Encyclopedia of Life, the natural environmental communities, and the climate communities. The sub-team will engage in research on a variety of topics especially methods for dealing with documents and other artifacts as

terminology and ontologies evolve over time, and methods for linking terminology in related, but distinct, communities.

Bruce Bargmeyer will lead a subcontract at LBNL. The LBNL sub-team will contribute the results of the eXtended Metadata Registry (XMDR) project to date and extend the work as needed for the OOR project. The XMDR project is concerned with the development of improved standards and technology for storing and retrieving the semantics of data elements, terminologies, and concept systems (thesauri, taxonomies, ontologies, etc.) in metadata registries. Existing metadata registry standards include the ISO/IEC 11179 family of Metadata Registry standards (e.g., ISO/IEC 11179). The XMDR project proposes extensions of the ISO/IEC 11179 family of metadata registry standards to support more diverse types of metadata and enhanced capabilities for semantics specification and queries. The LANL sub-team has already created and tested a prototype extended metadata registry (See xmdr.org). The primary responsibility of the LBNL sub-team will be to work together with the NCBO sub-team to integrate capabilities of XMDR and BioPortal. We would work on standards to make it feasible for wide deployment of the OOR.

Katherine Goodier is a co-PI and consultant. She will assist the PI in project coordination and outreach activities as detailed in the responsibilities of Prof. Baclawski. She is specifically responsible for directing a vigorous and comprehensive assessment and evaluation during all phases of the project. She is responsible for the human culture ontology and data set, and will engage in research to integrate this ontology with the biodiversity, environmental and climatic ontologies.

Peter Yim is a consultant and co-PI. He is responsible for managing team collaboration efforts. Although the team is geographically dispersed, it has been engaging in virtual meetings several times each week. He will not only be providing the collaboration infrastructure for the team, but he will also be engaging in research and development of new methods for improving the collaboration experience. This will be especially important for outreach efforts to DataNet Partners, which we envision to be mainly virtual meetings in a rich virtual environment.

Mike Dean is a consultant and co-PI. He will be responsible for developing a set of modular software interfaces (loosely modeled after the Apache Server) allowing OOR instantiations to choose and configure the OOR capabilities, languages, and policies they want to support. He is also responsible for the data management life cycle, including integration, release packaging, and testing. Finally, he will be engaging in research on federation among OOR and non-OOR registries, repositories, and collaborative development environments (including Semantic MediaWiki).

John Graybeal is a consultant. He will be responsible for outreach to the oceanographic and marine science communities. He is currently engaged in the Marine Metadata Interoperability (MMI) project which is bringing semantic interoperability to marine science through a combination of semantic framework description, ontology registry development, vocabulary discovery, user-centered use cases, workshop hosting, and other community engagement. He will continue this effort in the context of DataNet and the OOR where he will also be engaging more general environmental and climatic communities. In addition, he will engage in research on vocabulary creation, harmonization and term mapping.

Thomas Lyndon Wheeler is a consultant. He will be working with Katherine Goodier.

7 Activities

The OOR and the individual repositories of the communities being served will be developed in a series of phases.

- **Year 1**: Gathering of requirements from the initial communities being targeted by the OOR. Development of APIs and planning for adaptation and integration of the existing ontology repository platforms. Education and outreach programs are limited to selected subcommunities.

- **Year 2**: Completion of the first release of the OOR platform. Initial delivery of OOR services to the targeted communities. Feedback from the communities will be gathered for use in the next phase. Planning begins for the second release. Education and outreach programs expand but remain within the targeted communities.

- **Year 3**: Development of the second release begins, based on lessons learned. Education and outreach programs expand beyond the targeted communities to other DataNet partners. Delivery of OOR services to additional research communities.

- **Year 4**: Second release of the OOR platform. Education and outreach expands to include commercial and industrial communities. Transition to self-sustaining OOR begins with planning for the organization structure.

- **Year 5**: Transition to fully self-sustaining mode. Education and outreach are now the primary role played by the project.

References

[1] K. Baclawski. Search system and method based on multiple ontologies, July 23 2002. United States Patent No. 6,424,973. Assigned to Jarg Corporation, Waltham, MA.

[2] K. Baclawski. *Introduction to Probability with R*. CRC Press, January 23 2008.

[3] K. Baclawski, M. Kokar, P. Kogut, L. Hart, J. Smith, J. Letkowski, and P. Emery. Extending the Unified Modeling Language for ontology development. *Software and System Modeling*, 1(2):142–156, 2002.

[4] K. Baclawski, M.M. Kokar, R. Waldinger, and P. Kogut. Consistency checking of Semantic Web ontologies. In *Proc. of the International Semantic Web Conference*, volume 2342, pages 454–459. Springer-Verlag, Berlin, June 2002.

[5] K. Baclawski, C. Matheus, M. Kokar, and J. Letkowski. Toward a symptom ontology for Semantic Web applications. In *ISWC'04*, volume 3298, pages 650–667. Springer-Verlag, Berlin, 2004.

[6] K. Baclawski and T. Niu. *Ontologies for Bioinformatics*. MIT Press, Cambridge, MA, October 2005. ISBN 0-262-02591-4.

[7] K. Baclawski, T. Niu, and E. Neumann. Using DAML format for representing complex gene networks: implications in novel drug discovery. *Amer. J. Human Genetics*, 69 (Suppl)(4):457, 2001. Presented at the 51st Annual Meeting of the Amer. Soc. Human Genetics in San Diego, CA on October 12–16, 2001.

[8] B. Bargmeyer. Common sense, common logic. *ISO Focus*, 3(3):2006, March 2006. ISSN 1729-8709.

[9] B. Bargmeyer, editor. *Special Edition of International Journal of Metadata, Semantics and Ontologies*. Inderscience Publishers, 2008. to apper.

[10] B. Bargmeyer and S. Chance. Environmental data: Edge issues and the path forward. Technical Report 58869, LBNL, September 20 2005.

[11] L. Bermudez, P. Bogden, E. Bridger, G. Creager, D. Forrest, , and J. Graybeal. Toward an ocean observing system of systems. In *Proceedings of the Oceans'06 MTS/IEEEBoston*, September 18-21 2006.

[12] L. Bermudez, J. Graybeal, , and R. Arko. A marine platforms ontology: Experiences and lessons. In *Proceedings of the Semantic Sensor Networks Workshop at the 5th International Semantic Web Conference ISWC 2006*, Athens, GA, November 5-9 2006.

[13] L. Bermudez, J. Graybeal, A. Isenor, R. Lowry, , and D. Wright. Construction of marine vocabularies in the marine metadata interoperability project. In *Proceedings of the 13th Ocean Sciences Meeting*, Washington, D.C., September 19-23 2005.

[14] M. Daconta, L. Obrst, and K. Smith. *The Semantic Web: The Future of XML, Web Services, and Knowledge Management.* John Wiley, June 2003.

[15] V. Duggar and K. Baclawski. Integration of decision analysis in process life-cycle models. In *International Workshop on Living with Uncertainties*, Atlanta, Georgia, USA, November 5 2007.

[16] J. Graybeal, , and L. Bermudez. When hydrospheres collide: Lessons in practical environmental ontologies. In *Proceedings of the 7th International Conference on Hydroscience and Engineering (ICHE 2006)*, April 2007. ISBN: 0977447405.

[17] J. Graybeal. Esperanto, klingon, or other: Metadata implications for global marine observations. In *Proceedings of the Oceans'06 MTS/IEEE-Boston*, Boston, MA, September 18-21, 2006. IEEE Press.

[18] J. Graybeal. Marine Metadata Interoperability Community Vocabulary Use Cases, 2008. `http://marinemetadata.org/usecasesvocs2008`.

[19] J. Graybeal. Marine Metadata Interoperability Ontology Registry, 2008. `http://marinemetadata.org/community/teams/ont/mmirepository`.

[20] J. Graybeal. Marine Metadata Interoperability Ontology Registry and Repository, 2008. `http://mmisw.org/or`.

[21] J. Graybeal. Marine Metadata Interoperability Semantic Framework, 2008. `http://marinemetadata.org/semanticframeworkconcept`.

[22] J. Graybeal. Marine Metadata Interoperability Semantic Interoperability Workshop Series, 2008. `http://marinemetadata.org/workshops/oosinterop2008`.

[23] J. Graybeal. Marine Metadata Interoperability VINE Tool, 2008. `http://marinemetadata.org/vine`.

[24] J. Graybeal. Marine Metadata Interoperability Voc2RDF Service, 2008. `http://marinemetadata.org/voc2rdf`.

[25] J. Graybeal. Marine Metadata Interoperability vocabulary references, 2008. `http://marinemetadata.org/ont`.

[26] J. Graybeal, J. Bellingham, and F. Chavez. Data systems for ocean observation programs. *Sea Technology Magazine*, 46(9), September 2005.

[27] J. Graybeal, L. Bermudez, P. Bogden, S. Miller, and S. Watson. Marine metadata interoperability project: Leading to collaboration. In *Proceedings of the Global Data Interoperability Workshop*, Sardinia, Italy, June 2005.

[28] J. Graybeal, L. Bermudez, and K. Brink. Improving ocean research (meta)data management. *ORION Newsletter*, 2(1), Summer 2005.

[29] J. Graybeal, K. Gomes, M. McCann, B. Schlining, R. Schramm, and D. Wilkin. MBARI's SSDS: Operational, extensible data management for ocean observatories. In *The 3rd International Workshop on Scientific Use of Submarine Cables and Related Technologies*, Tokyo, 2003.

[30] J. Graybeal, S. Watson, A. Isenor, L. Bermudez, M. Howard, et al. Sensor metadata interoperability workshop report. Technical report, MBARI, July 2007.

[31] K. Johnson. Lizards rapidly evolve after introduction to island, April 21 2008. http://news.nationalgeographic.com/news/2008/04/080421-lizard-evolution.html.

[32] D. Koning, I.N. Sarkar, and T.D. Moritz. TaxonGrab: Extracting taxon names from text. *Journal of Biodiversity Informatics*, 2:79–82, 2005.

[33] P.R. Leary, D.P. Remsen, C.N. Norton, D.J. Patterson, and I.N. Sarkar. uBioRSS: Tracking taxonomic literature using RSS. *Bioinformatics*, 23(11):1434–1436, 2007.

[34] J. Luciano. A pathway to pathways. In *Genome Technology*, pages 15–16, June 2006.

[35] J. Luciano and L. Obrst. Ontology evaluation methods and metrics, 2008. Proposed MITRE internal research.

[36] J. Luciano and R. Stevens. e-Science and biological pathway semantics. *BMC Bioinformatics*, 8(Suppl 3), 2007.

[37] J. Luciano and R. Stevens. OWL: PAX of mind or the AX? Experiences of using OWL in the development of BioPAX. In *OWL: Experiences and Directions*, Gaithersburg, MD, USA, April 1-2 2008.

[38] C. Matheus, K. Baclawski, and M.M. Kokar. BaseVISor: A forward-chaining inference engine optimized for RDF/OWL triples. In *International Semantic Web Conference*, 2006.

[39] C. Matheus, K. Baclawski, and M.M. Kokar. BaseVISor: A triples-based inference engine outfitted to process RuleML and R-Entailment rules. In *Proc. Intern. Conf. on Rules and Rule Languages for the Semantic Web*, Athens, GA, November 2006.

[40] D. McCandless, L. Obrst, and S. Hawthorne. Dynamic web service assembly using OWL. In *W3C Workshop on Semantic Web in Energy Industries; Part I: Oil and Gas*, Houston, TX, Dec. 9-10 2008.

[41] T. Niu, K. Baclawski, Y. Feng, and H. Wang. Database schema for management and storage of single nucleotide polymorphism. *Am. J. Human Genetics*, 73 (Suppl):A467, 2003.

[42] L. Obrst. Ontological architectures. In M. Healy, A. Kameas, and R. Poli, editors, *TAO Theory and Applications of Ontology, Volume 2: The Information-science Stance*. Springer, 2008.

[43] L. Obrst, P. Cassidy, S. Ray, B. Smith, D. Soergel, M. West, and P. Yim. The 2006 upper ontology summit joint communiqué. *J. Appl. Formal Ontology*, 1(2), 2006.

[44] L. Obrst, W. Ceusters, I. Mani, S. Ray, and B. Smith. The evaluation of ontologies: Toward improved semantic interoperability. In C. Baker and K.-H. Cheung, editors, *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*. Springer, 2007.

[45] L. Obrst, T. Hughes, and S. Ray. Prospects and possibilities for ontology evaluation: The view from the national center for ontological research (ncor). In *Workshop on Evaluation of Ontologies for the Web (EON2006)*, Edinburgh, UK, May 22 2006.

[46] L. Obrst, T. Janssen, and W. Ceusters, editors. *Ontologies for Intelligence Analysis.* IOS Press, 2008. forthcoming.

[47] L. Obrst, H. Liu, and R. Wray. Ontologies for corporate web applications. *Artificial Intelligence Magazine, special issue on Ontologies*, pages 49–62, Fall 2003. http://portal.acm.org/citation.cfm?id=958676.

[48] L. Obrst, D. McCandless, S. Stoutenburg, K. Fox, D. Nichols, M. Prausa, and R. Sward. Evolving use of distributed semantics to achieve net-centricity. In *Regarding the "Intelligence" in Distributed Intelligent Systems*, Arlington, VA, Nov. 8-11 2007.

[49] J. Owen. Heat triggers sex change in lizards by "turning off" key gene, April 19 2007. http://news.nationalgeographic.com/news/2007/04/070419-sex-lizards.html.

[50] M. Parmelee, D. Nichols, and L. Obrst. A net-centric metadata framework for service oriented environments. *International Journal of Metadata, Semantics and Ontologies (IJMSO)*, 2008. forthcoming.

[51] P.J. Planet and I.N. Sarkar. mILD: A tool for constructing and analyzing matrices of pairwise character incongruence tests. *Bioinformatics*, 21:4423–4424, 2005.

[52] M. Prausa. *Parallelized Dynamic Service Orchestration.* PhD thesis, Colorado Technical University, September 2008. Obrst: PhD advisor.

[53] S. Ray, P. Yim, F. Olken, K. Baclawski, D. Holmes, D. Bedford, and S. Turnbull. Lessons learned from virtual organizing for the ontology summit 2007, July 17 2007.

[54] J.A. Rosenfeld, I.N. Sarkar, P.J. Planet, D.H. Figurski, and R. DeSalle. ORFcurator: Molecular curation of genes and gene clusters in prokaryotic organisms. *Bioinformatics*, 20:3462–3465, 2004.

[55] K. Samuel and L. Obrst. Answer set programming: Final report on a comparison between ASP and Prolog for Semantic Web ontology and rule reasoning. Technical report, MITRE, October 2007. Result of MITRE Sponsored Research Innovation Grant experiment.

[56] K. Samuel, L. Obrst, S. Stoutenberg, K. Fox, P. Franklin, A. Johnson, K. Laskey, D. Nichols, S. Lopez, and J. Peterson. Applying prolog to semantic web ontologies and rules: Moving toward description logic programs. *The Journal of the Theory and Practice of Logic Programming (TPLP)*, 8(03):301–322, May 2008. http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=1853440.

[57] I.N. Sarkar. Biodiversity Informatics: Organizing and linking information across the spectrum of life. *Briefings in Bioinformatics*, 8(5):347–357, 2007.

[58] I.N. Sarkar. Grand challenges in biodiversity informatics. *Asia-Pacific Biotech News*, 11(1):15–18, 2007.

[59] I.N. Sarkar, M.G. Egan, G. Coruzzi, E.K. Lee, and R. DeSalle. Automated Simultaneous Analysis Phylogenetics (ASAP): An enabling tool for phylogenomics. *BMC Bioinformatics*, 9(1):103, 2008.

[60] I.N. Sarkar, P.J. Planet, and R. DeSalle. CAOS software for use in character based DNA barcoding. *Molecular Ecology Resources*, 2008. in press.

[61] I.N. Sarkar, R. Shenk, and C.N. Norton. Exploring historical trends using taxonomic name metadata. *BMC Evolutionary Biology*, 8(144), 2008.

[62] I.N. Sarkar, J. Thornton, P.J. Planet, B. Schierwater, and R. DeSalle. A systematic method for classification of novel homeoboxes. *Molecular Phylogenetics and Evolution*, Sep;24(3):388–399, 2002.

[63] S. Semy, M. Pulvermacher, and L. Obrst. Toward the use of an upper ontology for U.S. government and U.S. military domains: An evaluation. Technical Report MTR 04B0000063, MITRE, November 2005. http://www.mitre.org/work/tech_papers/tech_papers_05/04_1175/index.html.

[64] S. Stoutenburg. *Refining Ontology Alignment: New Methods for Relationship Acquisition and Advanced Web Applications.* PhD thesis, U. of CO, Colorado Springs, CO, 2009. Obrst: PhD committee member.

[65] S. Stoutenburg, J. Gray, and J. Kalita. Refining ontology mapping techniques: Acquiring properties beyond similarity and equivalence. In *ODBASE*, 2008. submitted.

[66] S. Stoutenburg, L. Obrst, D. McCandless, D. Nichols, P. Franklin, M. Prausa, and R. Sward. Ontologies for rapid integration of heterogeneous data for command, control, and intelligence. In *Ontologies for the Intelligence Community Conference*, Columbia, MD, Nov. 28-30 2007.

[67] S. Stoutenburg, L. Obrst, D. Nichols, P. Franklin, K. Samuel, and M. Prausa. Ontologies and rules for rapid enterprise integration and event aggregation. In *Vocabularies, Ontologies and Rules for the Enterprise (VORTE 07), EDOC 2007*, Annapolis, MD, Oct. 15-19 2007.

[68] S. Stoutenburg, L. Obrst, D. Nichols, K. Samuel, and P. Franklin. Applying semantic rules to achieve dynamic service oriented architectures. In *RuleML 2006: Rules and Rule Markup Languages for the Semantic Web*, volume 4294, pages 581–590, Heidelberg, November 10-11 2006. Springer-Verlag. http://www.mitre.org/work/tech_papers/tech_papers_06/06_0904/index.html.