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.

1 Vision and Rationale

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.

1.1 Opportunity and Impact

The OOR has the potential for having a major impact on data set interoperability in general, but to ensure that this potential is fully realized, the project will include a vigorous education and outreach program. This program will seek 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 to allow the communities to adapt previously established ontologies for their own purposes.

Yet another opportunity for a broader impact is to use ontologies as a tool for education at all levels, even at the elementary school level. Since ontologies formalize the language of a community, they can be the basis for education and training for their discipline, provided that the ontologies are properly annotated. Kenneth Baclawski is currently engaged in an NIH-sponsored research project to develop ontology-based automated health counseling tools. These tools use "relational agents" that converse with patients. Relational agents have also been developed for special-purpose educational tools. Ontologies can offer the potential for scaling up these special-purpose educational tools to apply to virtually any domain, provided that the ontology is properly annotated. The research problem is to determine what annotations are required and to minimize the cost of providing the annotations, and then to build educational tools that are based on these annotated ontologies.

1.2 Sustainable Economic and Technology Models

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).

In disciplines where opinions and facts can change over time, there is a need to retain versioning information. In biodiversity, for example, the concept of a species may be adjusted as more information (e.g., molecular) about a particular species is attained. Taxonomic hierarchies may thus change over time; however, an investigator may wish to still compile what assertions were made with respect to an "older" species concept. These types of retrospective studies can be highly informative, especially in the context of understanding how biodiversity knowledge changes over time.

The OOR and the individual repositories of the communities being served will be developed in a series of phases which will include a series of milestones to assess progress toward achieving both the technical and sustainability goals.

- **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.

The OOR team currently meets weekly using a virtual collaborative environment. This environment is managed by Peter Yim. In addition to the weekly meetings, subteams will have their own meetings, on a weekly or biweekly schedule.

2 Activities

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:

- Biomedicine, including ontologies for genomics, proteomics, diseases, anatomy, model organisms, and other resources served by the highly successful BioPortal repository.

- Biodiversity, such as the species pages in the Encyclopedia of Life.
- Climate and environmental communities (including both natural environments and built environments).
- Human culture and sociology.

2.1 Support the full data preservation and access lifecycle

To truly support discovery, innovation and learning well into the future the OOR collaboration will manage the full data life cycle by providing 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. The OOR will support a broad range of semantic services and applications of interest to enterprises and communities.

The OOR will develop 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] The OOR will design and implementation service-oriented architectures and services, including automated and semi-automated service orchestration and parallel optimization to support the repository. [40, 52, 68]

The following are the core requirements for the OOR:

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.

2.1.1 Data deposition/acquisition/ingest

The OOR will develop 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. Logical relationships between ontologies will be supported within the repository, including mutual consistency, extension, and entailment, and semantic mappings. The use and linking of metadata, controlled vocabularies, and ontologies, for intelligent search and decision support. [14, 42, 47, 63, 65, 64]

The OOR will 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 and access control. While the emphasis is on the metadata level, ontologies also include instance data. For knowledge-rich domains such as the targeted sectors, the ontology includes all of the data as well as the metadata. For other domains, the data will be managed by special-purpose applications, and the ontology will be only part of the database, playing the roles that are most appropriate, such as encoding access policies and procedures, enabling discovery and interoperability, and ensuring that data remains accessible and understandable over timelines of decades or more.

2.1.2 Data curation and metadata management

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. The ontologies in the OOR must 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; (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 especially 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.

In addition to the logical metadata for ontologies, the OOR will include metadata for ontologies considered as engineering artifacts. This includes provenance, versioning, existing applications of the ontology (e.g. interoperability, search, decision support) and domain-specificity (e.g. biology, supply chain management, manufacturing).

The Ontology Metadata Vocabulary (OMV), Dublin Core, ISO 11179, ISO 19763, and other existing approaches to provenance and versioning metadata will be used as the basis for the metadata for ontologies in the OOR. An empirical approach will be used to identify and evaluate ontology metadata. Proposals for ontology metadata already exist, and they will be evaluated using use-case scenarios. These scenarios both motivate the use of the metadata and help establish best practices.

2.1.3 Data protection

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. The core mechanisms to be provided in the first version include enforcement of policies for access, submission, governance, change management, and control over user and administrative access. The later version of the OOR will provide capabilities to: create usage reports, validate syntax, check logical consistency and automatically categorize a submission.

2.1.4 Data discovery, access, use, and dissemination

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.

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 criteria: domain, author/creator/source, version, language, terminology and controlled vocabularies, quality, mapping, and inference. The interfaces for discovery will be suitable for both specialist and non-specialist users, using GUIs, web services, and language-based APIs.

2.1.5 Data interoperability, standards, and integration

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).

2.1.6 Data evaluation, analysis, and visualization

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.

The OOR will develop methods, practices, services, and artifacts to support automated and human reviewed evaluation and comparison of ontologies stored in the repository. [35, 45, 44]

2.2 Engage at the frontiers of science and engineering research and education

Ontology engineering is rapidly developing research area. There are both computational and storage challenges as the size of a knowledge base grows since the computational complexity of logical inference is much greater than the complexity of traditional database query processing. The OOR team is closely connected both with the ontology engineering community and with the communities specifically targeted by this proposal. As shown in the biosketches, the team members have substantial ontology research experience. Ontology engineering research will be an integral part of the OOR effort.

2.3 Education and training

As an integral part of the proposed project, the OOR will support 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.

The Ontolog Forum has been engaging in educational and outreach activities for 6 years, reaching over 40 distinct communities. Examples include communities in bioinformatics, national command and control, and intelligence. [37, 36, 34, 48, 46, 50, 63, 66, 67, 68]

2.4 Community and user input and assessment

Community and user input is an integral part of the process of ontology development. Assessing the effectiveness of the OOR in meeting its responsibilities is an important component of the project. Katherine Goodier has project assessment and evaluation as her primary responsibility.

2.5 International Participants

The OOR was introduced by participants in the Ontolog Forum. The Ontolog Forum is an open international community dedicated to the advancement of ontological engineering. The forum sponsors numerous virtual meetings and collaborations which include international participation. The OOR will make use of virtual meetings and environments as a means of establishing international collaborations with new communities. Kenneth Baclawski has a team of graduate students that are working with Peter Yim to extend and enhance the collaboration tools to have semantic capabilities based on ontologies.

3 Organizational Structure

3.1 Leadership and Management

3.2 Comprehensive expertise and infrastructure capacity/capabilities

The OOR team possesses considerable expertise in library and archival sciences (Mark Musen, Indra Neil Sarkar); computer, computational and information sciences (all team members); cyberinfrastructure (all team members); and domain sciences (discussed in Section 3.3 below). The team has been collaborating at virtual meetings for over a year, as well as at a face-to-face meeting at NIST in May of this year. The team has developed an effective organizational structure that enables shared responsibility, close coordination and cooperation, and catalyzes the rapid exchange of ideas. The team has access to considerable computational, storage, network access, dissemination, interaction and communication resources. Peter Yim is responsible for managing this infrastructure.

3.3 Diverse, multi-sector participation

The initial sectors that will be served by the OOR include the following: (1) **Biology**, (2) **Biodiversity**, (3) **Climate and Environment**, (4) **Human culture and sociology**.

BioPortal is a centralized repository for biomedical ontologies. It primarily serves the biomedical research sector, including the genomics, proteomics, diseases, anatomy, and model organism communities. However, BioPortal is a general purpose ontology repository, and it will be the foundation for the OOR. Mark Musen and his team at the NCBO will provide the expertise for this sector.

Indra Neil Sarkar and his team at MBL will provide the expertise for serving the biodiversity community. The focus will be on the evolutionary history and taxonomic communities that strive to create dynamic syntheses of information such as the species pages in the Encyclopedia of Life, He will also be providing expertise in the natural environment sector.

John Graybeal, Luis Bermudez and their team at the Monterey Bay Aquarium Research Institute will be collaborating with the OOR team to integrate The Marine Metadata Interoperability (MMI) project with the OOR. MMI has the key mission objective of developing a broad community presence to address marine metadata issues. This puts us in a unique position to understand community needs, address those needs in open and broadly applicable ways, and obtain community engagement and buy-in for open solutions.

The man-made environment is a diverse sector that includes architecture and engineering communities. Katherine Goodier will provide expertise for serving this sector. Finally, Michelle Raymond and her colleague Thomas Lyndon Wheeler will be providing expertise for serving the human culture and sociology sector.

For each of these sectors, the OOR collaboration will be providing: (1) Guidance for data providers, metadata providers, and ontology providers; (2) Organized references on all facets of metadata needs and solutions; (3) Services targeting semantic interoperability in marine and related domains, including vocabulary lists, ontology repository and associated services, and vocabulary creation and maintenance tools, services, and guidance; (4) Community collaboration environment (files, emails, web pages either visible or secure); (5) Access to work in progress on metadata tasks and projects. In addition, the OOR collaboration can provide purposeful capabilities: (1) Targeted identification and evaluation of resources (vocabularies, standards, tools, services); (2) Identification and engagement of key community participants (projects or individuals) in metadata initiatives; (3) Training and workshops in metadata technologies and techniques, particularly dealing with semantic tools and services, including vocabulary and ontology development, metadata standards and their application, as well as metadata-enlightened architectural development and analysis; (4) Community environment(s) to advance particular topics or discussions. The references section of the proposal includes links to resources that are already available in the targeted sectors.

3.4 Data Network

The primary purpose of ontologies is to achieve interoperability. The OOR will initially be focused on collaboration and coordination with its target communities. As the project continues, it will collaborate and coordinate closely with other DataNet Partners and digital preservation/access organizations both nationally and internationally. The ultimate goal is to allow for seamless, single entry point discovery, access, and use of data from any source. The Ontolog Forum in general, and the OOR team in particular, are already engaged in developing and disseminating best practices and principles. The Ontolog Forum has a substantial, multi-year track record for shared governance and coordination that will be leveraged for the proposed 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 51^{st} 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.