

## Information Integration Issues in Scientific Applications

Deborah L. McGuinness

Acting Director, Knowledge Systems, Artificial Intelligence Laboratory  
Stanford University

As research moves into its next phase as a distributed inter-connected multi-disciplinary global community, many issues in information access and integration become more challenging and simultaneously more important. We are investigating a number of aspects of semantic information integration ranging from underlying representation languages and infrastructure, to tools for analysis, evolution and maintenance, to applications ranging from immunology to solar-terrestrial physics.

We take a multi-dimensional approach:

1 – provide infrastructure for tracking and explaining where information came from, how it was manipulated, and why recipients might decide to trust information. Our major research thrust is on Inference Web [McP04] – an infrastructure to support explanations for answers from question answering systems. As part of this effort, we have designed the proof markup language [PMF06] – a representation Interlingua for encoding knowledge provenance. This effort is funded largely by DARPA for use in explaining cognitive assistants, and in particular task processing [MPGW06] (in the DARPA PAL program), for use in explaining integrated learners (in the DARPA integrated learning program), and for use in explaining analyst tools [WMP+05] (in the DTO novel intelligence for massive data program).

2 – explore semantic technology-based infrastructural approaches to access and integration. In particular, we are engaging in an NSF-funded effort to design and implement a virtual observatory for solar terrestrial physics [FMM+06]. We have also begun a NASA-funded effort to explore semantically-enabled scientific data integration with the initial domain areas of volcanoes and climate [FMRS06].

3 – explore information privacy issues, initially focusing on usage of information (as opposed to collection of information). We are engaged in an NSF-funded cybertrust program effort to provide transparent accountable data mining systems [WAB06].

At this workshop, I would be interested in describing the proof markup language and discussing how it can be and is being used to encode and explain data provenance. I would also be interested in discussing needs from scientific applications for information integration and semantic approaches for supporting cyberinfrastructure. This latter topic may be broken into thrusts for semantic integration and access, as well as ontology evolution and maintenance in a collaborative distributed environment, such as the web.

### References:

[FMM+06] Peter Fox, Deborah L. McGuinness, Don Middleton, Luca Cinquini, J. Anthony Darnell, Jose Garcia, Patrick West, James Benedict, and Stan Solomon. Semantically-Enabled Large-Scale Science Data Repositories. To appear in the Proceedings of the Fifth International Semantic Web Conference, Athens, Ga, November 5-9, 2006.

[FMRS06] Peter Fox, Deborah L. McGuinness, Rob Raskin, A. Krishna Sinha. Semantically-Enabled Scientific Data Integration. Proceedings of the Geoinformatics Conference, Reston, Virginia, May 10-12, 2006.

- [McG05] Deborah L. McGuinness. Why Should You Trust Answers from the Web? In Proceedings of the Joint Conference on Information Sciences, Web Intelligence and Security Track, Salt Lake City, Utah, July 2005.
- [MFRW00] Deborah L. McGuinness, Richard Fikes, James Rice, and Steve Wilder. *An Environment for Merging and Testing Large Ontologies*. In the Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR2000), Breckenridge, Colorado, USA. April 12-15, 2000.
- [MPGW06] McGuinness, D.L.; Pinheiro da Silva, P.; Glass, A.; Wolverson, M. Explaining Task Processing in Cognitive Assistants. 2006. Stanford Knowledge Systems Laboratory Technical Report KSL-06-06.
- [McP04] Deborah L. McGuinness and Paulo Pinheiro da Silva. Explaining Answers from the Semantic Web: The Inference Web Approach. Web Semantics: Science, Services and Agents on the World Wide Web Special issue: International Semantic Web Conference 2003 - Edited by K. Sycara and J. Mylopoulis. Volume 1, Issue 4. Journal published Fall, 2004.
- [MZP+06] Deborah L. McGuinness, Honglei Zeng, Paulo Pinheiro da Silva, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal. Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study. WWW2006 Workshop on the Models of Trust for the Web (MTW'06), Edinburgh, Scotland, May 22, 2006.
- [MMP+06] J. William Murdock, Deborah L. McGuinness, Paulo Pinheiro da Silva, Chris Welty, and David Ferrucci. Explaining Conclusions from Diverse Knowledge Sources. To appear in the Proceedings of the Fifth International Semantic Web Conference, Athens, Ga, November 5-9, 2006.
- [PMF06] Paulo Pinheiro da Silva, Deborah L. McGuinness and Richard Fikes. A Proof Markup Language for Semantic Web Services. Information Systems, Volume 31, Issues 4-5, June-July 2006, Pages 381-395. Previous version, technical report, Knowledge Systems Laboratory, Stanford University,
- [SPB+05] Muthuraman Sathiamurthy, Bjoern Peters, Huynh-Hoa Bui, John Sidney, John Mokili, Stephen S Wilson, Ward Fleri, Deborah L McGuinness, Philip E Bourne and Alessandro Sette. An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. Immunome Research 2005, 1:2 (20 Sep 2005).
- [WAB+06] Daniel J. Weitzner, Hal Abelson, Tim Berners-Lee, Chris P. Hanson, Jim Hendler, Lalana Kagal, Deborah L. McGuinness, Gerald J. Sussman, K. Krasnow Waterman. Transparent Accountable Inferencing for Privacy Risk Management. Proceedings of AAAI Spring Symposium on The Semantic Web meets eGovernment. AAAI Press, Stanford University, USA 2006. Also available as MIT CSAIL Technical Report-2006-007 and Stanford KSL Technical Report KSL-06-03.
- [WMP+05] Christopher Welty, J. William Murdock, Paulo Pinheiro da Silva, Deborah L. McGuinness, David Ferrucci, Richard Fikes. Tracking Information Extraction from Intelligence Documents. In Proceedings of the 2005 International Conference on Intelligence Analysis (IA 2005), McLean, VA, USA, 2-6 May, 2005.
- [ZPM05] Ilya Zaihrayeu, Paulo Pinheiro da Silva and Deborah L. McGuinness. IWTrust: Improving User Trust in Answers from the Web. Proceedings of 3rd International Conference on Trust Management (iTrust2005), Springer, Rocquencourt, France, 2005.