Putting Semantics into e-Science and Grids

Carole Goble

The School of Computer Science, The University of Manchester, Oxford Road, Manchester, M13 9PL carole@cs.man.ac.uk

Abstract

What is the Semantic Grid? How can e-Science benefit from the technologies of the Semantic Web? Can we build a Semantic Web for e-Science? Would that differ from a Semantic Grid? Given our past experiences with Scientists, Grid developers and Semantic web researchers, what are the prospects, and pitfalls, of putting semantics into e-Science applications and Grid infrastructure?

1. e-Science, Grids and Semantics

e-Science is scientific investigation performed through distributed global collaborations between scientists and their resources, and the computing infrastructure that enables this. Scientific progress increasingly depends on pooling know-how and results; making connections between ideas, people, and data; and finding and interpreting knowledge generated by strangers in new ways other than that intended at its time of collection. It is about harvesting and harnessing the *collective intelligence* of the scientific community. It has as much to do with intelligent information management as with sharing scarce resources like large scale compute power or expensive instrumentation.

The Semantic Web is an initiative to enable and operate a semantic infrastructure for gathering and exploiting the Web's collective intelligence, exploiting technologies primarily from artificial intelligence and data management computing. Applying the Semantic Web paradigm to e-Science seems like it might be a winner [1].

Grid computing is a component of the e-Science einfrastructure. Similar to an e-Scientist, a Grid depends on understanding the available resources, their capabilities, how to assemble them and how to best exploit them. Grid applications and the Grid middleware they utilize thrive on metadata and the knowledge to apply that metadata intelligently. However, Grid Services commonly deal with their semantic infrastructure in ad-hoc and hidden ways, providing poor mechanisms for sharing and openly processing knowledge. Thus, analogous to the Semantic Web, the Semantic Grid is an activity to systematically expose semantically rich information associated with resources to build more intelligent Grid services [2].

Although the Semantic Web and Semantic Grid initiatives are gathering pace, with many pioneering examples, there are barriers. Some are technical – the architectural or theoretical foundations, the mutual maturity of technologies, their appropriateness for the required tasks, and their scalability. Others are operational – the difficulty of gathering and maintaining the semantic content, for example, and an over-reliance on unavailable tooling. Many, and the most difficult to overcome, are sociological – the interplay between Scientists and Computing specialists, the Semantic and the Grid communities and the interfactional battles within those communities.

The talk presents the case for semantic web technologies as a delivery mechanism for the semantic fabric needed to support collective collaborative intelligence in e-Science and Grids. Technical, operational and sociological issues are addressed, and a research and development agenda sketched. Initiatives in the Semantic Web for Life Sciences, and personal experiences (good and bad) from the ^{my}Grid¹ Geodise, and OntoGrid projects, will provide food for thought.

2. References

[1] C Goble, Using the Semantic Web for e-Science: inspiration, incubation, irritation in 4th Intl Semantic Web Conference, Galway, Ireland, November 2005 to appear

[2] C Goble and D De Roure *The Semantic Grid: Building Bridges and Busting Myths*, 16th European conference on Artificial Intelligence, Valencia, Spain, August 2004.

¹ <u>http://www.mygrid.org.uk</u>, <u>http://www.geodise.org</u>, http://www.ontogrid.net