

e-Science and the Semantic Web: a Symbiotic Relationship

Carole Goble¹, Oscar Corcho¹, Pinar Alper¹ and David De Roure²

¹ School of Computer Science, University of Manchester,
Manchester M13 9PL, UK

{carole,ocorcho,penpecip}@cs.man.ac.uk

² School of Electronics and Computer Science, University of Southampton,
Southampton SO17 1BJ UK
dder@ecs.soton.ac.uk

Abstract. 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 reusing knowledge and resources generated by others in perhaps unintended ways. It is about harvesting and harnessing the “collective intelligence” of the scientific community. The Semantic Web is an extension of the current Web in which information is given well-defined meaning to facilitate sharing and reuse, better enabling computers and people to work in cooperation. Applying the Semantic Web paradigm to e-Science has the potential to bring significant benefits to scientific discovery. We identify the benefits of lightweight and heavyweight approaches, based on our experiences in the Life Sciences.

1 Introduction

The term e-Science is normally used to describe computationally intensive science that is carried out collaboratively in highly distributed network environments [1]. Typically, a feature of such collaborative scientific enterprises is that they require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists. Brain neuroscientists remotely control and collect data from the world’s largest and most powerful transmission electron microscope in Japan (telescience.ucsd.edu). Astronomers steer telescopes from their offices, collect the data using remote archive repositories, and process it by exploiting the availability of machines of other institutions. The International Virtual Observatory Alliance (www.ivoa.net) makes available vast digital sky archives to all astronomers not just a lucky few. ^{my}Grid-Taverna (www.mygrid.org.uk) allows biologists to assemble personalised exploratory *in silico* experiments that interoperate between remotely and locally available applications, code-bases and databases to identify new genes [2]. The Human Genome effort is an

example of e-Science – over 500 datasets and tools are available on the web for bioinformaticians to piece together our understanding of life [3, 4].

The e-Science infrastructure supports and enhances the scientific process by enabling scientists to generate, analyse, share and discuss their insights, experiments and results in a more effective manner, particularly in the context of the deluge of data resulting from new experimental practices [5, 6].

Scientific progress increasingly depends on pooling resources, know-how and results; making connections between ideas, people, and data; and finding and interpreting knowledge generated by others, in ways that may not have been anticipated when it was created. 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 computing power or expensive instrumentation. It is about making connections between decoupled resources and people in the broadest context of diverse scientific activity, outside the bounds of localised experiments and closed projects, and enabling scientific endeavour “in the wild”.

The Semantic Web is defined as an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation [7]. Applying Semantic Web to Science [8] has attracted great interest particularly in the Life Sciences [9-13] which has been proposed as a “nursery” for incubating the required technological developments [14]. We take a perspective in which the Semantic Web is seen foremost as an infrastructure for gathering and exploiting collective intelligence; i.e. the capacity of human communities to evolve towards higher order complexity and integration through collaboration and innovation.

Section 2 introduces the Semantic Web and distinguishes between lightweight and heavyweight approaches. In Sections 3 to 5 we present three aspects of the Semantic Web – annotation, integration and inference – and sketch how the methods, techniques and tools used for each of them could provide benefits to scientists. In Section 6 we reflect on the symbiosis between e-Science and the Semantic Web.

2 The Semantic Web

Annotation is the process of associating metadata with an object. Metadata, defined as structured data about an object that supports functions associated with it, can be generated for any entity for which contextual data can be recorded [15]. We can annotate any object, be it a document, dataset, publication, codes, notebooks, and so on, within the scientific process – even a person or scientific instrument. Metadata can be expressed in a wide range of languages (from natural to formal ones) and with a wide range of vocabularies (from simple ones, based on a set of agreed keywords, to complex ones, with agreed taxonomies and formal axioms). It can be available in different formats: electronically or even physically (written down in the margins of a textbook). It can be created and maintained using different types of tools (from text editors to metadata generation tools), either manually or automatically.

The foundational pillars of the Semantic Web are the tagging of entities with machine-processable metadata which asserts facts using terms, and the associated languages which define these terms and their relationships [16]. These languages extend existing markup languages like HTML or XML in order to represent knowledge – they enable *Semantic Annotation*. The publishing and sharing of the annotations and languages is crucial to realising the benefits of this approach.

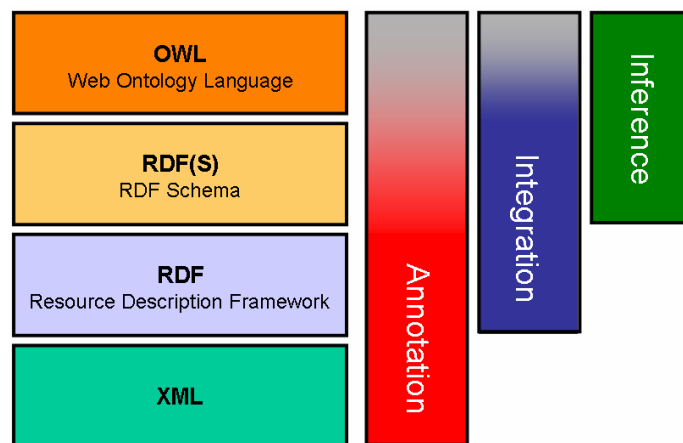


Fig. 1. Markup languages and the Semantic Web of Annotation, Integration and Inference

Figure 1 shows the mark-up languages that have been developed in recent years. Each language is based on the ones that are under it (i.e., OWL is based on RDF Schema, RDF and XML). The right-hand part of the figure shows that this stack of languages can be used for different purposes, distinguishing three functions that the Semantic Web can provide:

- **Annotation.** The Semantic Web can be used to provide machine-processable descriptions of resources in any of the languages in the stack, asserted using RDF statements. Anything can be annotated, and annotations can be shared.
- **Data Integration.** The Semantic Web can be used as a means of integrating information from diverse sources, connecting through shared terms (in RDF(S) or OWL) and shared instances (in RDF) and preserving context and provenance.
- **Inference.** The Semantic Web can be used as a powerful tool to infer new knowledge or detect inconsistencies in the knowledge already described in it, provided in RDF Schema and OWL. Each of these languages has different expressiveness, and different reasoning mechanisms that can be applied to its descriptions, so the inferences that can be achieved are different in each case.

These three different but complementary views of the Semantic Web are discussed in the following sections in the context of e-Science.

3 The Semantic Web as the Annotation Web

3.1 Annotation in e-Science

The annotations associated with Web resources form their own overlaid and intertwined Web – the Annotation Web. This is potentially a powerful tool which enables e-Science to be carried out collaboratively in highly distributed network environments. Not only does it provide a means for describing the resources being dealt with by e-Scientists (new findings, experiment results and provenance, etc.), but it also allows content and people to be connected, hence allowing the harvesting and harnessing of the collective intelligence of the scientific community. Let us see some examples:

- The ^{my}Grid-Taverna workflow environment [17] is a significant example of a new platform for scientific discovery [2]. Services are described according to a service ontology that specifies the model to be used to describe the service inputs and outputs. Following this approach, annotated services can easily be reused for the design and execution of scientific workflows. The annotation of services is done collectively using annotation tools like Pedro [18] and maintained by a curator in order to ensure their quality.
- The CombeChem project [19] (www.combechem.org) focuses on the notion of “publication at source”, capturing comprehensive annotations in order to facilitate interpretation and reuse of results. Using RDF to interlink information in multiple datastores, both internally and externally, CombeChem has established a *Semantic DataGrid* containing tens of millions of RDF triples [20]. The annotation commences in the laboratory [21]. The project also captured scientific discourse as part of the provenance record, through provision of tools to annotate meetings [22].
- The Friend-Of-A-Friend (FOAF) initiative (www.foaf.org) is being adopted for science and scientific publications. Scientific FOAF (www.urbigene.com/foaf/) is an example of how so-called “Web 2.0 technologies” [15] can be used to relate content and people in order to improve e-Science. For each article in the NCBI Pubmed bibliographic database (www.ncbi.nlm.nih.gov/entrez/), users are asked if a specific person is unambiguously one of the authors and whether they know any of the co-authors (which may not be necessarily the case when there are a large number of collaborators). Authors' interests are defined using the MeSH (Medical Subject Headings) terms of their papers. All this information is used to generate the FOAF profile of scientists.
- Recording the provenance of scientific results is important in facilitating interpretation and reuse of data, as well as for regulatory purposes. Semantic Web technologies are used to provide a flexible and extensible record of the experimental process in the *in silico* experimentation of ^{my}Grid [23] and commencing in the laboratory in CombeChem [24].

3.2 Lightweight and heavyweight annotation

We can make a distinction between what we call *lightweight* and *heavyweight* annotation, by considering the vocabularies used to create metadata. Both forms of

annotation are relevant to scientific tasks such as those above but they have significantly different characteristics. Our classification is analogous to that of lightweight and heavyweight ontologies, introduced by Studer [25].

Lightweight annotation

We define lightweight annotation as the process of associating metadata with a resource, where the metadata does not necessarily refer to an existing ontology but consists of tags defined by the person(s) in charge of the annotation, which express the meaning of the information source by using terms instead of logical expressions. Furthermore, if the annotation is based on existing ontologies, it normally identifies instances of the ontology concepts, but not the relationships between those instances.

For example, should we have a Web document that describes a gene and the processes in which the gene is involved, a lightweight annotation would consist of tags assigned to different parts of the document. These tags may be: “Gene”, “BRCA2”, “Breast Cancer”, “disease”, etc. The CombeChem and SciFOAF examples above also illustrate lightweight annotation.

The advantage of this type of annotation is that it eases the collection of the most important terms and relations of a community, and can be a good starting point to achieve agreement in a specific area. Lightweight annotation has appeared in the context of the Web 2.0 initiative, which refers to a second generation of tools and services on the Web that lets people collaborate and share information online [26]. While lightweight annotation does not focus on relating annotations to existing common vocabularies, vocabularies can be created out of them in what we call *folksonomies* [27].

Heavyweight annotation

We define heavyweight annotation as the process of associating metadata with a resource, where the metadata refers to an existing ontology that is implemented in a formal language (e.g., RDF Schema, OWL, etc.); and the annotation does not only identify instances of ontology concepts but also relationships between those instances, which are compliant with the underlying ontology.

For example, should we have a Web document that describes a gene and the processes in which the gene is involved, a heavyweight annotation of this document would comprise metadata based on an ontology about genes, such as the Gene Ontology (www.geneontology.org). The metadata consists of instances of concepts that are defined in the Gene Ontology, plus instances of relations between those concepts, including genes, the processes where they are involved, etc. The ^{my}Grid-Taverna service discovery framework uses heavyweight annotation [5].

Heavyweight annotations normally give precise information about the contents of a document. This information is not necessarily exhaustive and is normally costly, since it is usually authored by domain experts. The ontologies in which those annotations are based are normally shared in a community.

3.3 Tools for annotation

Tools for creating more lightweight annotations are based on the Web 2.0 philosophy and emerging tools. SemanticMediaWiki allows adding semantic information to a Web document while creating it; the semantic information added consists of typed links and typed attributes (wiki.ontoworld.org/index.php/Semantic_MediaWiki). del.icio.us (<http://del.icio.us/>) is a social tagging tool that allows annotating Web resources by adding tags that are not specifically connected to an existing ontology, but that can be used to derive folksonomies. Flickr (www.flickr.com) is a social tagging tool that allows annotating images in a similar way to del.icio.us.

Tools for creating heavyweight annotations (ontology-based annotators) are primarily designed to allow inserting and maintaining ontology-based mark-ups in Web documents [28] [29]. First conceived as tools that could be used to alleviate the burden of including ontology-based annotations manually into Web resources, many annotators have evolved into more complete environments that use Information Extraction, Machine Learning and Natural Language techniques to propose semi-automatic annotations for any type of Web resources. In general, the more automated the annotation process is, the lower the quality of the annotations obtained.

Other tools are aimed at annotating data available in data sets by establishing mappings between the data set model and a set of terms coming from ontologies [30]. This annotation improves data discovery and integration, among others.

4 The Semantic Web as the Integration Web

Scientific discovery involves bringing together information from diverse sources, many of which may be available in the form of databases rather than Web pages. The set of these Web resources, also known as the “Deep Web” or “Data Web”, represents more than 80% of the total amount of data available on the Web [31].

Many of these databases contain overlapping or complementary information about the same individuals. For instance, the same protein might appear in UniProt and PDB, or two proteins might share the same Gene Ontology code as part of their annotation. One of the main challenges for the *Integration Web* is how to align the data that is available in heterogeneous distributed databases so that the appropriate connections are made among the pieces of information described in different disconnected databases.

Semantic technologies can be applied to this vital task of integrating information from multiple sources [13] [32]. They can describe the content of a set of databases according to a shared model, or according to a set of semantic models that can be aligned using ontology alignment or merging techniques (www.ontologymatching.org). Using these approaches we can leverage the information available in the Deep Web to information whose meaning is clearly described according to a common understood and/or agreed model of the domain.

In e-Science the information available in documents (e.g., journal publications) represents only a small percentage of the total information available. According to [33], “traditional biological journals will become just one part of various biological

data resources as the scientific knowledge in published papers is stored and used more like a database". One of the examples that support this vision is that of reading a description of an active site of biological molecule in a paper and being able to access immediately the atomic coordinates specifically for that active site, and then using a tool to explore the intricate set of hydrogen-bonding interactions described in the paper. A similar illustration is provided by the "Crystal EPrints" interface in CombeChem [20].

There are many examples in the literature that show how heterogeneous data sources can be integrated, supporting users by providing a common view of them and shortening the time needed to find and relate information. Here are two approaches: a data web and a wikipedia.

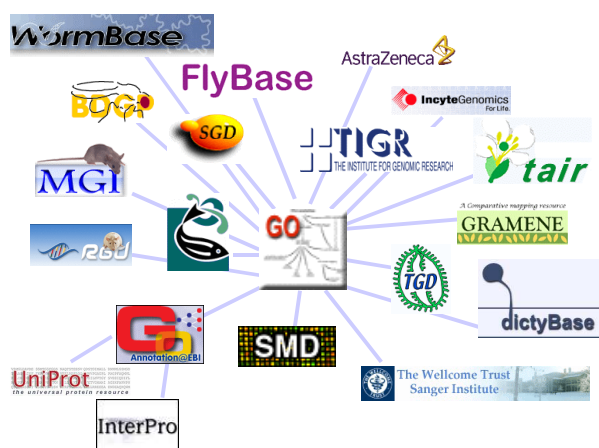


Fig. 2. Resources cross-linked through using the Gene Ontology

A **Data Web** approach encourages databases to export their data self-described in RDF, breaking down the barriers of different schemas. If the same instance – a protein, say – appears in many databases we can integrate the various RDF descriptions. The data entries are commonly annotated within the databases using controlled terms such as GO ids from the Gene Ontology (Fig 2). These “within record” annotations are exposed on the Data Web, and we can use shared terms to link between different database entries, and between the database records and other resources, such as documents, that are annotated using the Gene Ontology [34]. YeastHub is an RDF data warehouse that integrates different types of yeast genome data provided by different resources in different formats [35]. BioDASH takes advantage of UniProt’s RDF export facility and the BioPAX ontology to associate disease, compounds, drug progression stages, molecular biology, and pathway knowledge for a team of users. CombeChem used RDF to annotate entities but also to hold chemical data in RDF format [36], recognising the flexibility of this approach. The Collaboratory for Multi-scale Chemical Science (CMCS) [37] has developed an open source toolkit addressing provenance tracking and lightweight federation of data and application resources into cross-scale information flows, and provides another case study in the chemical data arena.

A **Wikipedia** is an example of how Web 2.0 technologies can be used to support information collection in e-Science. The Gene Wikipedia is proposed as a comprehensive knowledge collection about genes, gathering information from multiple data sources: GenBank, UniProt, Bind, Kegg, etc (www.bioinformatics.org/genewiki/wiki/). It is based on a common set of ontologies: NCBI taxonomy, Gene Symbol, Disease/Phenotype ontology, Protein Interaction databases, GenBank, Pathway and Gene Ontology allowing searches on the information based on these resources. The first wiki pages, and the curation of the available data, is based on the use of text mining tools, ensuring data quality and assurance.

The whole notion of “self-describing experiments” is taking hold as part of the W3C’s Health Care and Life Sciences Interest Group’s exploration of RDF, and initiatives such as King’s EXPO ontology of scientific experiments [38] and collection standards expressed in RDF [39]. Calls to the scientific publishers to annotate papers at publication with ontologies [40] and responses by publishers such as Nature to embrace collaborative tagging systems like Connotea (www.connotea.org) hold out the promise that we could build Integration Webs between data and publications.

5 The Semantic Web as the Inference Web

BioPax (the Biological Pathways Exchange) makes inferences over biological pathway data; an automated OWL reasoner is used to find overlapping and non-overlapping molecular interactions between two pathway datasets [41]. The ^{my}Grid-Taverna project provides the ability to search over scientist-centric descriptions in a subject specific way, using taxonomy information in the associated ontology [42]. These are both examples of inference: metadata that uses terms from OWL ontologies can be reasoned over using logic-based decision procedures, which means that new relationships can be inferred between statements that had not been explicitly made before.

There are several languages in the Semantic Web, with different expressivity and complexity. Depending on the formal language selected, different types of reasoning functions will be available. If we use RDF Schema to describe our ontologies (which means that we will use RDF to describe our annotations), we will be able to execute queries over the models created, using RDF query languages like SPARQL. We can do this to retrieve annotations, as well as during the data integration process once the data source to be queried has been identified and selected. The query engines that process these languages are able to perform basic taxonomic reasoning and consistency checking regarding the domain and range of properties. A more expressive language like OWL, based on the description logic formalism, enables us to perform more complex reasoning processes. For instance, we will be able to detect inconsistencies in the vocabulary that we are creating during the modelling process, such as finding missing terms in the Gene Ontology [43], and we will be able to derive concept taxonomies automatically from the descriptions that we have provided. Furthermore, we will be able to infer the concepts to which an individual belongs

given the information that we have from it, or we will be able to detect whether an individual description is inconsistent due to the constraints expressed in the ontology, for example finding new properties of proteins leading to possible drug discovery targets [39]. The complexity of modelling in OWL, however, should not be underestimated [44, 45].

Other logical formalisms, such as rules, allow inferring conditional relationships between individuals that were not possible with the previous formalisms – for example correspondences between data are not necessarily obvious to detect, requiring specific rules [46].

6 Symbiosis

The practice of science and e-Science can be enhanced by the technologies of the Semantic Web, enabling a dramatic reduction in the time needed to create new results and consequently the so-called “time to insight” and discovery [47]. Semantic Web-based applications promise help in: the development of controlled vocabularies, flexible metadata modelling, intelligent searching and document discovery, social and knowledge networking, advanced content syndication and publishing, data integration, aggregation and cross linking, application interoperability and knowledge mining [48]. A heavyweight approach (“pain today”) can be combined with a lightweight approach (“pain tomorrow”) to facilitate community participation. Sometimes a little semantics goes a long way [49].

Significantly we have taken a holistic view of the scientific process. Rather than looking at specific, self-contained projects, we observe that scientific discovery occurs “in the wild” using diverse resources. The Semantic Web permits this decoupling between the people, the content of information sources, the metadata about the content, and the time when this content and metadata is created. The person who creates the data is not necessarily the one who digests it. Furthermore, data and metadata can be used for objectives that were not expected when they were created, and at any point in time since its creation.

The natural match between Semantic Web and Life Sciences has been recognised by W3C in their focus on Life Sciences [50]. The Semantic Web benefits from e-Science, since its emerging technologies are being tested on a large-scale thanks to their use in e-Science applications [14] and the piloting of Semantic Webs for Life Science such as Sealife (www.biotec.tu-dresden.de/sealife/) and SWAN [51]. Realising the benefits of bringing Semantic Web and Science together depends on establishing awareness and understanding across those communities [52].

In fact the Semantic Web technologies also have a role inside the e-Science infrastructure. The Semantic Grid [53] is characterised as an open system in which people, services and computational resources (all owned by different stakeholders) are brought together flexibly and dynamically. While Semantic Web technologies can be used *on* the Grid, the Semantic Grid emphasises applying the Semantic Web technologies *within* the e-Science infrastructure – the machine processable descriptions then enable the services and resources to be brought together automatically [54].

We suggest that there is in fact a deeper mutual benefit for e-Science and the Semantic Web. The Semantic Web will thrive in an environment where the annotations – heavyweight and lightweight – are produced and consumed through the routine use of the infrastructure. To reach that point there needs to be a mechanism for bootstrapping the semantic infrastructure, and this will occur where there is incentive and a low cost to entry. We have described the incentives which lead to new scientific discovery. The low entry cost emphasises the need for tools for e-Scientists and also highlights the value of the lightweight tagging approach to move the endeavour forward. The Web flourished through ease of use and incentive to publish as well as to consume content, and e-Science appears to provide similar circumstances for the Semantic Web.

Acknowledgements

The authors express their thanks to their many colleagues in the e-Science programme and beyond. We acknowledge the support of the EPSRC through the ^{my}Grid (GR/R67743/01, EP/C536444/1) and CombeChem e-Science projects (GR/R67729/01, EP/C008863/1). This work is also supported by the EU FP6 OntoGrid project (STREP 511513) funded under Grid-based Systems for solving complex problems, and by the Marie Curie fellowship RSSGRID (FP6-2002-Mobility-5-006668).

References

1. Hey, T. and A.E. Trefethen, *Cyberinfrastructure for e-Science*. Science, 2005. **308**(5723): p. 817-821.
2. Stevens, R., et al. *Exploring Williams-Beuren Syndrome Using myGrid*. in *12th International Conference on Intelligent Systems in Molecular Biology*. 2004. Glasgow, UK: Bioinformatics
3. NAR1, *Web Server issue*. Nucleic Acids Research 2006. **34**.
4. NAR2, *Database Issue* Nucleic Acids Research 2006. **34**.
5. Hey, A.J.G. and A.E. Trefethen, *The Data Deluge: An e-Science Perspective*, in *Grid Computing - Making the Global Infrastructure a Reality*, F. Berman, G.C. Fox, and A.J.G. Hey, Editors. 2003, Wiley and Sons. p. 809-824.
6. Szalay, A. and J. Gray, *2020 Computing: Science in an exponential world*. Nature, 2006. **440**: p. 413-414.
7. Berners-Lee, T., J. Hendler, and O. Lassila, *The Semantic Web*. Scientific American, 2001. **284**(5): p. 34-43.
8. Hendler, J., *Science and the Semantic Web*. Science 2003. **299**: p. 520-521.
9. Baker, C.J.O. and K.-H. Cheung, eds. *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*. 2006, In Press.
10. Schroeder, M. and E. Neumann, *Special Issue on Semantic Web in Life Sciences*. Journal of Web Semantics, 2006. **4**(3).
11. Stevens, R., O. Bodenreider, and Y.A. Lussier. *Semantic Webs for Life Sciences Session Introduction*. in *Pacific Symposium on Biocomputing*. 2006.
12. Kazic, T. *Putting Semantics into the Semantic Web: How Well Can It Capture Biology?* in *Pacific Symposium on Biocomputing*. 2006.
13. Neumann, E., *A Life Science Semantic Web: Are We There Yet?* Sci. STKE, 2005(283).

14. Goble, C. *Using the Semantic Web for e-Science: inspiration, incubation, irritation*. in *4th International Semantic Web Conference (ISWC2005)*. 2005. Galway, Ireland: Springer.
15. Greenberg, J., *Metadata and the World Wide Web*. The encyclopedia of library and information science, 2002. **72**: p. 244-261.
16. Antoniou, G. and F. van Harmelen, *A Semantic Web Primer*. 2004: MIT Press
17. Oinn, T., et al., *Taverna: A tool for the composition and enactment of bioinformatics workflows*. *Bioinformatics Journal*, 2004. **20**(17): p. 3045-3054.
18. Garwood, K., et al. *Pedro ontology services: A framework for rapid ontology markup*. in *2nd European Semantic Web Conference*. 2005. Heraklion, Greece: Springer-Verlag.
19. Frey, J.G., D. De Roure, and L.A. Carr. *Publication At Source: Scientific Communication from a Publication Web to a Data Grid*. in *Euroweb 2002 Conference, The Web and the GRID: from e-science to e-business*. 2002. Oxford, UK: BCS.
20. Taylor, K., et al. *A Semantic Datagrid for Combinatorial Chemistry*. in *6th IEEE/ACM International Workshop on Grid Computing*. 2005. Seattle.
21. Hughes, G., et al., *The Semantic Smart Laboratory: A system for supporting the chemical e-Scientist*. *Organic & Biomolecular Chemistry*., 2004. **2**(22): p. 3284-3293.
22. Bachler, M., et al. *Collaborative Tools in the Semantic Grid*. in *GGF11 - The Eleventh Global Grid Forum*. 2004. Honolulu, Hawaii, USA: Global Grid Forum.
23. Zhao, J., et al. *Using Semantic Web Technologies for Representing e-Science Provenance*. in *3rd International Semantic Web Conference ISWC2004*. 2004. Hiroshima, Japan: Springer.
24. Frey, J., et al. *CombeChem: A Case Study in Provenance and Annotation using the Semantic Web*. in *International Provenance and Annotation Workshop (IPAW'06)*. 2006. Chicago, USA: Springer.
25. Corcho, O., M. Fernández-López, and A. Gómez-Pérez, *Methodologies, tools and languages for building ontologies. Where is their meeting point?* *Data & Knowledge Engineering*, 2003. **46**(1): p. 41-64.
26. O'Reilly, T. *What Is Web 2.0*. 2005 [cited July 2006]; Available from: www.oreillynet.com/go/web2.
27. Gruber, T. *Ontology of Folksonomies: A Mash-up of Apples and Oranges*. 2005 [cited July 2006]; Available from: <http://tomgruber.org/writing/ontology-of-folksonomy.htm>.
28. Corcho, O., *Ontology-based document annotation: trends and open research problems*. *International Journal of Metadata, Semantics and Ontologies*, 2006. **1**(1): p. 47-57.
29. Uren, V., et al., *Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art*. *Journal of Web Semantics*, 2006. **4**(1).
30. Bizer, C. *D2R MAP – A DB to RDF Mapping Language*. in *12th International World Wide Web Conference (WWW2003)*. 2003. Budapest, Hungary.
31. Bergman, M.K., *The deep Web: surfacing hidden information*. The Journal of Electronic Publishing, 2001. **7**(1).
32. Stephens, S., et al., *Aggregation of bioinformatics data using Semantic Web technology*. *Journal of Web Semantics*, 2006. **4**(3).
33. Bourne, P., *Will a Biological Database Be Different from a Biological Journal?* *PLoS Comput Biol*, 2005. **1**(3): p. e34.
34. Bechhofer, S., R. Stevens, and P. Lord. *Ontology Driven Dynamic Linking of Biology Resources*. in *Pacific Symposium on Biocomputing*. 2005. Hawaii.
35. Cheung, K.H., et al., *YeastHub: a semantic web use case for integrating data in the life sciences domain*. *Bioinformatics* 2005. **21**(1): i85-i96).
36. Taylor, K.R., et al., *Bringing Chemical Data onto the Semantic Web*. *J. Chem. Inf. Model.*, 2006. **46**: p. 939-952

37. Myers, J.D., et al. *A Collaborative Informatics Infrastructure for Multi-scale Science*. in *Second International Workshop on Challenges of Large Applications in Distributed Environments*. 2004. Honolulu, Hawaii.
38. Newscientist.com news service and *Translator lets computers "understand" experiments*. 2006 [cited July 2006]; Available from: <http://www.newscientist.com/article/dn9288-translator-lets-computers-understand-experiments-.html>.
39. Wang, X., R. Gorlitsky, and J.S. Almeida, *From XML to RDF: How Semantic Web Technologies Will Change the Design of 'Omic' Standards*. *Nature Biotechnology*, 2005. **23**(9): p. 1099-103.
40. Blake, J., *Bio-ontologies—fast and furious*. *Nature Biotechnology* 2004. **22**: p. 773-774.
41. Cheung, K.-H., et al., *A semantic web approach to biological pathway data reasoning and integration*. *Journal of Web Semantics*, 2006. **4**(3).
42. Lord, P., et al. *Feta: A light-weight architecture for user oriented semantic service discovery*. in *2nd European Semantic Web Conference*. 2005. Heraklion, Greece: Springer-Verlag.
43. Wroe, C.J., et al. *Methodology To Migrate The Gene Ontology To A Description Logic Environment Using DAML+OIL*. in *8th Pacific Symposium on Biocomputing (PSB)*. 2003. Hawaii.
44. Stevens, R., et al., *Managing OWL's Limitations in Modelling Biomedical Knowledge*. *Intl Journal of Human Computer Systems*, Accepted for publication.
45. Zhang, S., O. Bodenreider, and C. Golbreich. *Experience in Reasoning with the Foundational Model of Anatomy in OWL DL*. in *Pacific Symposium on Biocomputing*. 2006.
46. Neumann, E.K. and D. Quan. *Biodash: A Semantic Web Dashboard for Drug Development*. in *Pacific Symposium on Biocomputing*. 2006.
47. Gates, W. *Opening Keynote*. in *SuperComputing 2005 (SC05)*. 2005. Seattle, Washington.
48. Goble, C., R. Stevens, and S. Bechhofer, *The Semantic Web and Knowledge Grids*. *Drug Discovery Today: Technologies*, 2005. **2**(3): p. 193-302.
49. Wolstencroft, K., et al. *A little semantics goes a long way in Biology*. in *4th International Semantic Web Conference (ISWC2005)*. 2005. Galway, Ireland.
50. W3C, *Workshop on Semantic Web for Life Sciences*. 2004: Cambridge, Massachusetts USA.
51. Gao, Y., et al., *SWAN: A distributed knowledge infrastructure for Alzheimer disease research*. *Journal of Web Semantics*, 2006. **4**(3).
52. Goble, C.A. and D. De Roure. *The Semantic Grid: Building Bridges and Busting Myths*. in *16th European Conference in Artificial Intelligence (ECAI 2004)*. 2004. Valencia, Spain.
53. De Roure, D., N.R. Jennings, and N.R. Shadbolt, *The Semantic Grid: Past, Present, and Future*. *Proceedings of the IEEE*, 2005. **93**(3): p. 669-681.
54. Goble, C.A., et al., *Enhancing Services and Applications with Knowledge and Semantics*, in *The Grid 2: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Editors. 2004, Morgan-Kaufmann. p. 431-458.