GRID TECHNOLOGY FOR COLLABORATIVE ONTOLOGY DEVELOPMENT

Stuart Aitken, Kemian Dang & Jonathan Bard

School of Informatics,
University of Edinburgh, UK

informatics

## http://www.biosphere-portal.org

ABSTRACT

In contrast with the centrally-organised curation of the Gene Ontology, many biological ontologies are developed by loosely-organised groups who develop their ontology remotely. These groups tend to be formed from scientists and bio-informaticians from research groups with a common interest, who want to create a resource that will be useful to the community, rather than being formally mandated. Until recently, technological support for bio-ontology development relied on stand-alone editors running on users' desktops for creating new ontology versions (e.g. OBO-Edit, COBrA and Protégé) and on private email, email lists and perhaps Wikis for the distribution of ontology files and discussions. Clearly, much better use could be made of the storage, versioning and visualisation techniques being developed by the database and e-Science communities. BioSphere is an online ontology editor supporting multiple users and is underpinned by a server that stores versions (in OWL-XML) and provides a discussion portal.
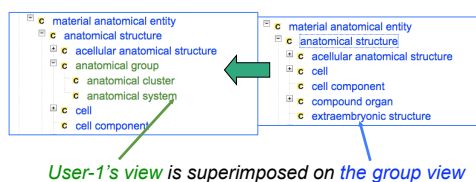
The BioSphere Portal
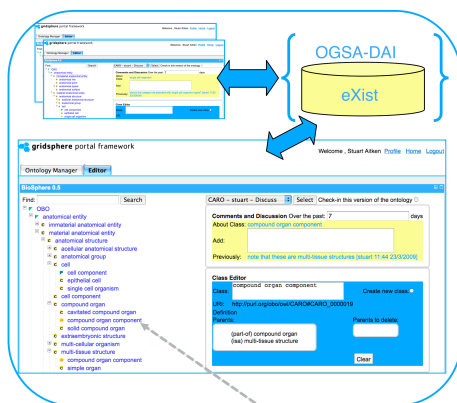
*The diagram shows several CARO users accessing BioSphere. The **bottom panel** shows a single user interface in more detail. **Yellow panel - comment & discussion**. One term has a note (yellow bubble) from an earlier discussion as shown in the blue text of the yellow panel. The current user can add their own text in the "Add" window. **Blue panel - editing**. Drag-and-drop facilities allow term parents to be added or removed and relationships to be changed.*

The BioSphere ontology portal provides a set of online tools for editing ontologies and for organising a developer community.

The system provides each user with their own view of the structure of the ontology (created on-the-fly) that is derived from the group's view – all comments are shared. The system allows an editor to reconcile the different views and provide a new consensus.

For each user, the system automatically manages the edits they make and supports ontology versioning. That is, while editing their view of the ontology, the modifications are saved in the central source document as updates. Users have editing or curation rights, according to the authority assigned to them by other members of the group.

User Views and Group Views

User-1's view is superimposed on the group view

Biosphere uses an XML database (eXist) and GridSphere, an open-source portal that allows tools and services to be provided securely through a standard web browser.

The XML DB allows searching across many ontology documents (stored in the XML syntax for the Web Ontology Language [OWL] version 1.1). Viewing ontologies as XML documents allows XML methods for versioning to be applied: essentially, each assertion in the ontology carries a timestamp annotation. The eXist XML database is accessed through an OGSA-DAI API.

The OWL XML API is currently in the GridSphere portlet, but we aim to relocate these data-access functions to the OGSA-DAI layer to make them more widely available as web and Grid services

While developing systems in the Java Enterprise style has its challenges (e.g. the multiplicity of layers), we believe that the potential for code reuse within the e-Science community is a key benefit beyond the immediate goal of providing a useful tool for a specific user group.

Future Work

These tools will be used in a number of case studies, the first (at the MRC HGU – Edinburgh) is re-structuring the mouse developmental anatomy.

http://www.biosphere-portal.org