

MASC Bioinformatics- 2007 Subcommittee Report

Prepared by Chris Town (Co-chair, cdtown@tigr.org) and Heiko Schoof (Co-chair, schoof@mpiz-koeln.mpg.de)

The main activities in the last year have been a continuation of the Web Services project that was initiated early in 2006 with funding from the DFG and NSF, details of which can be found at the project web site (<http://bioinfo.mpiz-koeln.mpg.de/araws>). Two developers' workshops were held in the spring of 2006, one at MPIZ-Cologne and one at TIGR. During the workshops, an increasingly detailed set of notes on how to set up the appropriate IT environment and how to implement web services was developed and made available on the project web site. Three of the participants at the Cologne workshop went on to develop their own services once back at their home institution. At the TIGR workshop, several simple web services were set up and deployed from the participants' home institutions during the course of the workshop. One additional service was deployed later. In addition, by remote consulting, a new web service for SeedGenes was deployed without any face-to-face interaction between the project personnel and the SeedGenes informaticians. A "show-and-tell" workshop was held at the International Conference on *Arabidopsis* Research in Madison, 2006 to demonstrate and promote web services. Following the meeting, work continued at MPIZ to advance the concept of a "one stop shop" by developing aggregator pages that will simultaneously query multiple locations for certain types of data. Their essential feature is that relevant web services are discovered automatically: once a data provider makes a BioMoby web service public, that data will be included in the aggregator automatically. Currently, the two best examples of these are tAIGa and LitRep (<http://bioinfo.mpiz-koeln.mpg.de/araws/searchtools>). tAIGa is short for 'the Arabidopsis Image Gallery' and acts as aggregator for services which take an AGI locus code as input and return one or more image(s). Currently, tAIGa queries seven services giving access to image data from SeedGenes, ABRC, RAPID, ATIDB, AtNoPDB, ProtLocDB and Arex. Like tAIGa, litRep requires the input of an AGI locus code and returns publications as PubMed or PubMedCentral identifiers. Currently four services are integrated, most notably the manually curated collections of bibliographic references of the project partners TAIR and Aramemnon. It will be important in the future both to enrich these pages by encouraging the deployment of additional web services for these data types as well as to begin to aggregate new types of data. Other activities during the project have been a significant effort by the German partners to update the code and thus re-activate many web services developed and hosted by NASC. A list of all services implemented and/or updated can be found at <http://bioinfo.mpiz-koeln.mpg.de/araws/web-services/public-ws>.

Lessons Learned and Challenges Faced

The project has (two) long-term goals:

- 1a. To educate data providers in the value of web services for data dissemination and to provide the instruction and tools necessary to accomplish this in a relatively straightforward fashion and, ultimately, without the need for a high level of technical expertise;
- 1b. To encourage the use of web services by an increasing number of data providers;
2. To acquaint and educate users in the potential benefits of web services, initially through the use of aggregator-type "one stop shop" interfaces, but also progressively by introducing them to the concept of workflows through the use of tools such as Taverna.

There are several challenges, but the most significant is identifying a group of data providers who will be proactive both in developing web services *and in maintaining them*. Based upon our experience to date, we believe that the combination of an instructional workshop that provides an overview of Biomoby, its implementation in Perl and Java, its deployment in a test environment and the use of Taverna to develop workflows, segueing into a hackathon where first services are implemented will be the most productive for future efforts.

Instructional media combined with remote consulting is a workable model but the "flying geek" approach has not been tested. Compared with pure remote consulting, workshops provide a stimulating and synergistic environment. Future workshops should be structured around groups of providers that host datasets that are complementary and potentially synergistic. It is important to deploy services during the workshop, since people tend to lose focus on web services very quickly after a workshop is finished. We think this is due in large part to insufficient priority granted to this method of data sharing by the PIs which in turn reflects the lack of community interest or pressure to provide data in this form. To address this, we will strive for better visibility of the achievements of the project as well as to provide novel integration of services that can currently only be performed manually by visiting multiple web sites. We intend to create links to the clients from prominent sites like TAIR, MIPS and TIGR. Also, a paper summarizing our results and experiences is being considered. In the remaining time of the project, we will try to motivate partners to set up more services and enlist new partners. A list of planned services is available on the project home page.