# Accelerating Time to Experiment – the myExperiment approach to Open Science

David De Roure, University of Southampton, UK
Carole Goble, University of Manchester, UK
Jiten Bhagat, University of Manchester, UK

*myExperiment has set out to provide the social software and services to support the scientific process, focusing on the 'time to experiment' phase of the scholarly knowledge cycle rather than the data deluge from new experimental techniques. In the context of open science this phase is also experiencing a deluge of scientific objects – not just data but protocols, methods and the new artefacts of digital science such as workflows, provenance records and ontologies. myExperiment has already demonstrated the role of a social web site in addressing this challenge, with significant usage in the first few months of the service. In this paper we consider the second way in which myExperiment undertakes to reduce time-to-experiment – by integrating with the everyday work practices of scientists.*

## Introduction

e-Science projects and infrastructure have often focused on the phase of the scientific process *after* collecting experimental data, with an emphasis on the application of emerging computational capability to handle the deluge of data from new experimental techniques such as DNA microarrays, combinatorial chemistry or sensor networks. However, 'time to discovery' also depends significantly on the time taken to get to the stage of conducting the experiment in the first place. While the computer science community has focused on accelerating processing, if we are to accelerate time to discovery then our research effort needs to look at 'accelerating' the human part of the cycle. This is precisely the goal of the myExperiment social web site for scientists – to reduce time-to-experiment and avoid reinvention [1].

The key to this acceleration, we believe, is to facilitate the sharing of data and methods – making it easy for scientists to progress effectively and efficiently to the experiment, and to make shared items as reusable as possible. Scientific advance has always relied on a social process in which scientists share ideas, methods and data, but traditionally this discourse has been mediated by the scholarly publishing process and established structures such as conferences. Scientists are increasingly turning to blogs, wikis and social networks to facilitate this process, a phenomenon sometimes characterised as Science 2.0 [2]. With this we also see a movement to open science where large scale, open distributed collaboration is enabled by making data, methods and results freely available on the Web.

Hence there is a tremendous potential in providing social tools to support the scientific process. This is made possible because increasingly the various research objects that we use – data, methods, publications, workflows – are available digitally. However there is also a tremendous challenge – we now have a data and method deluge which demands new techniques, especially in the context of open science. The new instrument that we bring to bear on this challenge is provided by society itself – it is the scale of community participation and the network effects that this brings. This instrument offers new ways of tackling difficult challenges; for example, the 'decay' over time of research objects can be addressed by community curation.

myExperiment set out to tackle the 'time to experiment' challenge using Web 2.0 techniques, working with scientists to build a social web site for sharing scientific workflows and their associated research objects. At the outset there was considerable scepticism as to whether scientists would be prepared to use a social web site and doubt that critical mass could be achieved in order to enjoy the network effects of Web 2.0. We now have an evidence base of several months of usage data which provides evidence that, for our communities at least, scientists do share and that their social networks develop [3]. Our analysis also shows that an enthusiastic core is willing to share quality workflows but expects credit for doing so, acting as provider to the wider community [4]. We are working with our social science colleagues to analyse the sharing behaviours in different communities.

## Supporting Open Science

However, myExperiment is more than a social web site that supports the research objects of scientists. It is also an open and extensible development environment and community, which follows the 'cooperate don't control' maxim of Web 2.0 so that myExperiment makes its own functionality easily available for reuse by others and draws on other services as much as possible. We do not oblige the scientist to come to myExperiment, but rather we make it as easy as possible to bring myExperiment to the scientist. We believe that this is essential for adoption [5] – which in turn is essential to build community and catalyse the community effects that we seek.

Hence myExperiment has paid as much attention to its developer community as it has to designing the user interface:

- By creating tools to manage its RESTful API, the exposed functionality is highly customisable in response to requirements. The API has enabled new interfaces to be built and existing interfaces to incorporate myExperiment functionality, such as a wiki or workflow systems. Developers without specialist training can build solutions to meet the particular needs of science users focused on particular tasks – functionality mashups over myExperiment.
- In turn myExperiment can reach out to the content and services needed to support the scientist in their research – it makes their research objects accessible and *actionable* beyond the core repository. In particular it is designed to be part of the scholarly knowledge cycle and is compatible with Open Archives Initiative protocols and the emerging tooling and practice of the Semantic Web.
- In contrast to social web sites like Facebook and mySpace, developers can download, reuse and repurpose myExperiment itself. The codebase is evolving as it is used across multiple projects.

## Examples

*Developing new Interfaces.* We have two exercises in building entirely new user interfaces to myExperiment's functionality. Firstly we are using Silverlight to build a rich similarity search and socially-driven workspace mashup that uses the myExperiment API together with other common data sources like Google Search, Google Scholar, CiteULike, Connotea, PubMed and so on. Figure 1 shows how a user can initiate a search from the "Top Tags" cloud built from live myExperiment data. Secondly we have built Google Gadgets for myExperiment. Two of these are shown in Figure 2: the Tag Cloud gadget, and the Workflows Feed Reader gadget which shows the latest myExperiment workflows.

*Bringing myExperiment to existing interfaces.* We have integrated with the Taverna workflow workbench by building a Taverna plugin for myExperiment, so that Taverna users can access the myExperiment capabilities from within the Taverna environment (figure 3). We are currently integrating with Microsoft's Trident Scientific Workflow Workbench [6], and for this we have developed preliminary support in myExperiment for sharing Windows Workflow Foundation (WWF) workflows. Finally we are working in conjunction with our open science colleagues in chemistry to bring myExperiment together with work on Electronic Lab Notebooks and 'blogging the lab' [7].

*Scholarly publishing.* myExperiment can refer to external repository content and we are developing mechanisms to import metadata directly from the repositories, focusing first on EPrints integration. myExperiment supports a general notion of research object, which captures aggregations of objects and also encompasses the other forms of data in myExperiment – for example members, groups, tags and the social network. We call these objects EMOs (Encapsulated myExperiment Objects). EMOs are represented in RDF and we have developed a myExperiment ontology which uses Dublin Core metadata for research objects and FOAF for the social network information. To interwork with repositories we have adopted the Object Reuse and Exchange representation from the Open Archives initiative, which is based on named RDF graphs. We envisage that the scholarly publishing process will evolve to support this more general notion of scientific research object, which will facilitate reusable and reproducible research.

## Conclusion

Accelerating time to experiment requires *social infrastructure* for open science. myExperiment demonstrates the value of the social web site approach but also of the open development practices which enable the power of the social infrastructure to be brought into play at the point of use by scientists.


[1] De Roure, D.; Goble, C.; Stevens, R., The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows, Future Generation Computer Systems, July 2008. http://eprints.ecs.soton.ac.uk/15709/

[2] M. Mitchell Waldrop, "Science 2.0: Great New Tool, or Great Risk?", Scientific American, Published online January 9, 2008 on http://www.sciam.com/article.cfm?id=science-2-point-0-great-new-tool-or-great-risk

[3] De Roure, D., Goble, C., Bhagat, J., Cruickshank, D., Goderis, A., Michaelides, D. and Newman, D. myExperiment: Defining the Social Virtual Research Environment. Submitted to 4th IEEE International Conference on e-Science, 7-12 December 2008, Indianapolis, Indiana, USA. http://eprints.ecs.soton.ac.uk/16560/

[4] Goderis, A., De Roure, D., Goble, C., Bhagat, J., Cruickshank, D., Fisher, P., Michaelides, D. and Tanoh, F. Discovering Scientific Workflows: The myExperiment Benchmarks. Submitted to IEEE Transactions on Automation Science and Engineering . http://eprints.ecs.soton.ac.uk/15662/

[5] De Roure, D. and Goble, C. Six Principles of Software Design to Empower Scientists. IEEE Software. In Press, 2009. http://eprints.ecs.soton.ac.uk/15032/

[6] See http://www.microsoft.com/mscorp/tc/trident.mspx

[7] Neylon, C. Openwetware blog. See http://blog.openwetware.org/scienceintheopen/

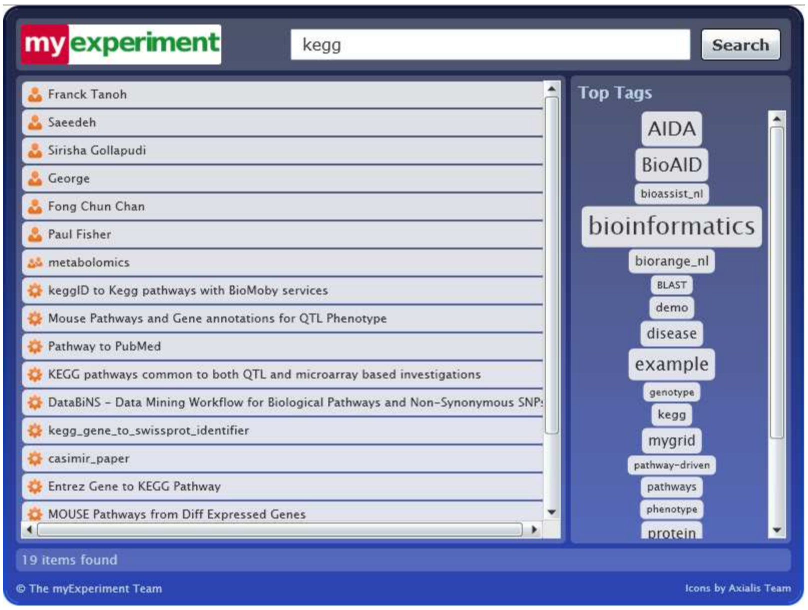


**Figure 1.** Silverlight functionality mashup over myExperiment using live data.

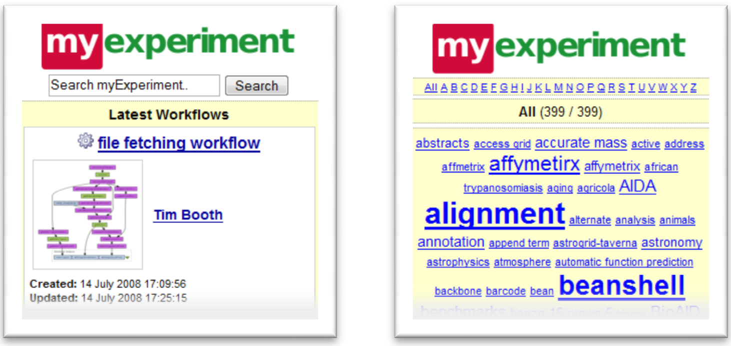


**Figure 2.** Google Gadgets for the myExperiment workflow feed and tag cloud.

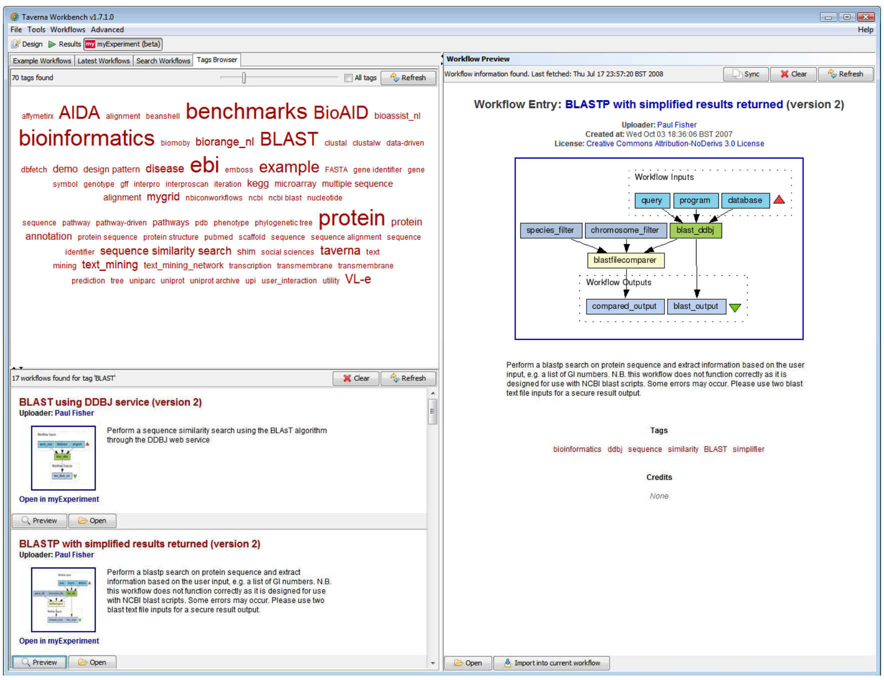


**Figure 3.** Integrating myExperiment with the Taverna Workbench, which is widely used in the life sciences. We are currently integrating with Microsoft's Trident Scientific Workflow Workbench.