myExperiment: A Web 2.0 Virtual Research Environment for Research using Computation and Services

David De Roure

Intelligence Agents Multimedia Group,
School of Electronics and Computer Science,
University of Southampton, UK

dder@ecs.soton.ac.uk

Carole Goble

Information Management Group,
School of Computer Science,
The University of Manchester, UK

carole.goble@manchester.ac.uk

ABSTRACT myExperiment is a social web site for the born-digital items arising in contemporary research practice, in particular scientific workflows and *Research Objects*. myExperiment can be seen from many perspectives: as a Virtual Research Environment which majors on social sharing, as "Facebook for scientists" but without the implicit openness which is actually a deterrent to scientists, as a second generation digital library which combines a repository with a place for conducting *in silico* research, or as the foundation of the future e-Laboratory.

1. Background

In 2004 the UK's JISC Committee for Support of Research introduced the notion of the "Virtual Research Environment" (VRE), a term which arose by analogy with the "Virtual Learning Environment". The VRE Working Group stated that "a VRE should support the processes of conducting research, including marshalling of resources, scholarly discourse and publication, and the creation and maintenance of collaborations, across disciplines, institutions and countries, including support for meetings and organisational processes" [1]. The programme has been successful in delivering case studies and deployed environments, and now focuses on institutional embedding.

myExperiment, a VRE from phase 2 of the programme, distinctively adopted a Web 2.0 approach in delivering a "social web site for scientists". Recognising a need to support sharing of the new artefacts of digital research that are poorly supported in the traditional scholarly knowledge lifecycle, and following the Web 2.0 principle by which sites major on support for one content type (e.g. movies, photos or slides), myExperiment focused initially on sharing scientific workflows.

Workflows are a means of composing available data and computational services in order to conduct "in silico" research, in a repeatable and sharable way [2]. They bring an immediate benefit to the researcher by providing automation to handle a deluge of data systematically, and a longer term benefit in that they enable the sharing of methods in order to facilitate new research outcomes. In a world of data-intensive science we observe not only a data deluge but a method deluge, and this focus on methods has become myExperiment's trademark [3].

2. myExperiment

While it shares many characteristics with other Web 2.0 sites, the distinctive features of myExperiment to meet the needs of its research user base are support for credit, attributions and licensing, fine control over privacy, a federation model and the ability to execute workflows [4]. A prototype was launched in July 2007 to gather requirements, and the current www.myexperiment.org site went live after a further 5 months of development and has evolved substantially in "perpetual beta". It now has 1800 registered users, with thousands downloading public content, and some 600 workflows across multiple workflow systems. See wiki.myexperiment.org for more information about the project.

Significantly, we have recognised that researchers do not work with just one content type, and have developed support for "packs" – collections of items, both inside and outside myExperiment, which can be shared as one; for example, a pack might contains workflows, example input and output data, results, logs, PDFs of papers and slides. Packs can be exported using the Object Reuse and Exchange representation which is gaining increasing adoption in the open repositories community.

The system is implemented in a web application framework, Ruby on Rails, and the user interface is designed to be familiar to the next generation of researchers. myExperiment also has a RESTFUL API and a SPARQL interface, facilitating programmatic access. Developers have built a variety of alternative interfaces including Google Gadgets and Windows 7 integration, demonstrating that it is possible to bring myExperiment's functionality into the researchers' existing environment rather than obliging them to go to myExperiment.

In addition to www.myexperiment.org there are now multiple instances of myExperiment supporting different contribution types and different communities. The next phase of myExperiment emphasises integration with institutional repositories: EPrints in Southampton and Fedora in Manchester. It will also bring support for expert curators, controlled vocabularies, faceted browsing and new contribution types including scripts and workflows for the Meandre and Kepler systems. We are analysing usage and taking steps to encourage community curation of content – for example to combat the challenging problem of "workflow decay".

A sister project, "BioCatalogue" (www.biocatalogue.org), provides a registry of Web Services in the life sciences domain. Biocatalogue enjoys a symbiotic relationship with myExperiment, with myExperiment providing service usage information to BioCatalogue, and BioCatalogue providing service availability information to myExperiment. Significantly, BioCatalogue adopts myExperiment's curation philosophy, with service metadata maintained by the service provider, experts and community, assisted by automated tools.

3. e-Laboratories

myExperiment has been adopted in various ways in a number of projects which broadly fall under the banner of "e-laboratories", across domains from bioinformatics and chemistry to social statistics and music. These e-Laboratories support the research lifecycle through resource discovery and acquisition, computation, publication and curation. While scientific workflow systems gave us the first generation of e-Laboratories, myExperiment helps define the 2nd generation. As myExperiment evolves, our goal is to be able to assemble e-Laboratories with ease, through common components and services and a shared notion of "Research Object" – an elaboration of packs, carrying information about the relationships between elements of a pack and also between packs. Research Objects are replayable, repeatable, reproducible, reusable, repurposeable, replicatable and reliable, and their routine use characterises the next generation of e-Laboratory.

1st Generation	**2nd Generation**	**3rd Generation**
Current practice of early adoptors of e-Labs tools such as Taverna, ELNs, LIMS. Characterised by researchers using tools within their particular problem area, with some re-use of tools, data and methods within the discipline. Traditional publishing is supplemented by publication of some digital items like workflows and links to data. Provenance is recorded but not shared and re-used. Science is accelerated and practice beginning to shift to emphasise *in silico* work.	Designing and delivering now, based on experience with Taverna, myExperiment and Lablogs. Key characteristic is re-use – of the increasing pool of tools, data and methods, across areas & disciplines. Contain some freestanding, recombinant, reproducible Research Objects. Provenance analytics plays a role. Expert curation supplemented by community curation. New scientific practices are established and opportunities arise for completely new scientific investigations.	e-Labs we'll be delivering in 5 years – illustrated by open science and open source science. Characterised by global reuse of tools, data and methods across any discipline, and surfacing the right levels of complexity for the researcher. Key characteristic is radical sharing. Research is significantly data driven – plundering the backlog of data, results and methods. Research Objects supersede papers. Increasing automation and decision-support for the researcher – the e-Laboratory becomes assistive. Provenance assists design. Curation is autonomic and social. New research outcomes are obtained.

References

[1] Allan, R., Allden, A., Boyd, D., Crouchley, R., Harris, N., Lyon, L., Robiette, A., De Roure, D. and Wilson, S. Roadmap for a UK Virtual Research Environment: Report of the JCSR VRE Working Group, JISC, 2004. http://www.jisc.ac.uk/uploaded_documents/VRE%20roadmap%20v4.pdf

[2] Gil, Y., Deelman, E., Ellisman, M. et al. "Examining the Challenges of Scientific Workflows". IEEE Computer 40(12): 24-32. 2007.

[3] Goble, C. and De Roure, D. "Curating Scientific Web Services and Workflows". Educause Review, 43 (5). EDUCAUSE Review, vol. 43, no. 5, September/October 2008.

[4] De Roure, D., Goble, C., Bhagat, J., Cruickshank, D., Goderis, A., Michaelides, D. and Newman, D. "myExperiment: Defining the Social Virtual Research Environment". In 4th IEEE International Conference on e-Science, 7-12 December 2008, Indianapolis, Indiana, USA.