

The Design and Realisation of the *my*Experiment Virtual Research Environment for Social Sharing of Workflows

David De Roure^a Carole Goble^b Robert Stevens^b

^a*University of Southampton, UK*

^b*The University of Manchester, UK*

Abstract

In this paper we suggest that the full scientific potential of workflows will be achieved through mechanisms for sharing and collaboration, empowering scientists to spread their experimental protocols and to benefit from those of others. To facilitate this process we have designed and built the *my*Experiment Virtual Research Environment for collaboration and sharing of workflows and experiments. In contrast to systems which simply make workflows available, *my*Experiment provides mechanisms to support the sharing of workflows within and across multiple communities. It achieves this by adopting a social web approach which is tailored to the particular needs of the scientist. We present the motivation, design and realisation of *my*Experiment.

Key words: Scientific Workflow, Workflow Management, Virtual Research Environment, Collaborative Computing, Taverna Workflow Workbench

1 Introduction

Scientific workflows are attracting considerable attention in the community. Increasingly they support scientists in advancing research through *in silico* experimentation, while the workflow systems themselves are the subject of ongoing research and development (1). The National Science Foundation Workshop on the Challenges of Scientific Workflows identified the potential for scientific advance as workflow systems address more sophisticated requirements and as workflows are created through collaborative design processes involving many scientists across disciplines (2). Rather than looking at the application or machinery of workflow systems, it is this dimension of collaboration and sharing that is the focus of this paper.

Understanding the whole lifecycle of workflow design, prototyping, production, management, publication and discovery is fundamental to developing systems that support the scientist’s work and not just the workflow’s execution. Supporting that lifecycle can be the factor that means a workflow approach is adopted or not. Workflow design is challenging and labour-intensive, and reusing a body of prior designs through registries or catalogues is highly desirable (3). Reuse is a particular challenge when scientists are outside a predefined Virtual Organisation or enterprise. These are individuals or small groups, decoupled from each other and acting independently, who are seeking workflows that cover processes outside their expertise from a common pool of components. This latter point arises when workflows are shared across discipline boundaries and when inexperienced scientists need to leverage the expertise of others.

In this paper we present the motivation, design approach and realisation of the *my*Experiment Virtual Research Environment for collaboration and sharing of experiments (4), which aims to provide a ‘workflow bazaar’ for any workflow management system. While individual workflow systems may provide workflow repository mechanisms, *my*Experiment is distinctive in that it facilitates the sharing of workflows and these may come from multiple systems. In Section 2 we consider the use of workflows for science, the power of workflows as shared entities, and the requirements for sharing. This is followed in Section 3 by a discussion of the *my*Experiment design. Section 4 describes how *my*Experiment has been realised as a service for the community. We conclude in Section 5.

2 Sharing Scientific Workflows

There are many workflow systems available – we found over 75 after conducting an informal search. These systems vary in many respects: e.g. who uses them, what resources they operate over, whether the systems are open or closed, how workflows are expressed (e.g. how control flow is handled), how interactive they are, when and how tasks are allocated to resources, and how exceptions are handled; see (1) for a comprehensive discussion. Our focus here is on scientific workflows which are near the application level rather than those further down in the infrastructure; i.e. we are interested in composing scientific applications and components using workflows, over a service oriented infrastructure (which may include Grid services). These are the workflows which are close to the scientist, or indeed the researcher in any domain. We also note the distinction between workflow templates and workflow instances: the former describes the steps and order of the process without identifying particular end points of services (or codes), while the workflow instance binds in the concrete executions (3). In this paper, workflow refers to both.

One immediate attraction of workflows which encourages their uptake is the easing of the burden of repetitive manual work. However, we suggest that the key feature for scientific advancement is reuse. Workflow descriptions are not simply digital data objects like many other assets of e-Science, but rather they capture pieces of scientific process – they are valuable knowledge assets in their own right, capturing valuable know-how that is otherwise often tacit (5). Reuse is effective at multiple levels: the scientist reuses a workflow with different parameters and data, and may modify the workflow, as part of the routine of their daily scientific work; workflows can be shared with other scientists conducting similar work, so they provide a means of codifying, sharing and thus spreading the workflow designer’s practice; and workflows, workflow fragments and workflow patterns can be reused to support science outside their initial application.

The latter point illustrates the tremendous potential for new scientific advance. An example of this is a workflow used to help identify genes involved in tolerance to Trypanosomiasis in east African cattle (6). The workflow enabled a large volume of data to be processed systematically, leading to a new scientific result which arose within part of the data that would have remained unexamined had the work been conducted manually. Then the same workflow was reused successfully by another scientist over a new dataset, to identify the biological pathways implicated in the ability for mice to expel the *Trichuris muris* parasite. This reuse was made easier by the explicit, high-level nature of the workflow that described the analytical protocol.

Workflows bring challenges too – they require skill to design, so they can be difficult and expensive to develop. Consequently, workflow designers need design assistance, and prefer not to start from scratch. Furthermore it is easy for the reuse of a workflow to be confined to the project in which it was conceived. In the Trypanosomiasis example, the barrier to reuse was how the knowledge about the workflow could be spread to the scientists with the potential need. In this case it was word of mouth within one institution; this barrier needs to be overcome. So, we have a situation of workflows as reusable knowledge commodities, but with potential barriers to the exchange and propagation of those scientific ideas that are captured as workflows (7).

Significantly, there is more to a workflow than a declaration of a process. An individual workflow description may take the form of an XML file (for example), but this does not sit in isolation. We identify a range of properties that are factors in guiding workflow discovery and reuse, including: descriptions of its function and purpose; documentation about the services with which it has been used, with example input and output data and design explanations;

workflow provenance, including version history and origins; reputation and use within the community; ownership and permissions constraints; quality, whether it is reviewed and still works; and dependencies on other workflows, components and data types. Workflows also enable us to record the provenance of the data resulting from their enactment, and logs of service invocations from workflow runs can inform later decisions about workflow and service use.

By binding workflows with this kind of information we assist discovery, reuse and repurposing, and we provide a basis for workflows to be trusted and interpreted unambiguously. But like the workflows themselves, the associated information is currently often confined to the system from which it originated and thus is not reusable as a useful commodity in its own right.

2.2 *Sharing workflows*

It is apparent then that we can view workflows as potential commodities, as valuable first class assets in their own right, to be pooled and shared, traded and reused, within communities and across communities, to propagate like memes. Workflows themselves can be the subject of peer review. Furthermore we can conceive of packs of workflows for certain topics, and of workflow pattern books – new structures above the level of the individual workflow. We call this perspective of the interacting data, services, workflow and their metadata within a scientific environment the *workflow ecosystem* and we suggest that by understanding and enabling this we can unlock the broader scientific potential of workflow systems.

Workflow management systems already make workflows available for sharing, through repository stores for workflows developed as part of projects or communities. For example, the Kepler Actor Repository is an LDAP-based directory for the remote storage, query and retrieval of actors (processes) and other workflow components (8) and the SCEC/CME workflow system has component and workflow libraries annotated with ontologies (9). These follow the tradition of cataloguing scripting libraries and codes. Inforsense’s online Customer Hub (10) is a central repository for their users to share best practices and leverage community knowledge potentially across projects, particularly aimed at an enterprise setting.

We go beyond these, focusing not only on making workflows available for sharing but actually facilitating and encouraging that sharing across broad and decoupled communities using a variety of workflow systems. So we are taking a more social approach, and we believe that the key to sharing is to recognise the use of workflows by a community of scientists. This acknowledges a central fact, sometimes neglected, that the lifecycle of the workflows is coupled with

the process of science – that the human system of workflow use is coupled to the digital system of workflows.

Intuitively, the sharing we propose is about scientists giving away their know-how. Why would a scientist release such valuable commodities to the wider community – why would scientists share? However, this is the nature of the established scholarly knowledge cycle. The efficient unfolding of new knowledge in science rests on a set of idealised institutional norms, one of which is the sharing of knowledge among scientists (11). The citing of published material is a form of reuse. Citing a scientist’s paper is almost as valuable as the publication itself. By sharing or publishing a workflow, with the appropriate attribution, a scientist can allow their work to be reused with the concomitant spread of their scientific reputation – their workflow is, in effect, being cited.

It is clear then that a scientist must be allowed entry at any point in the experimental or scholarly lifecycle, over and above *de novo* workflow construction. Furthermore there are two sets of social issues to be addressed:

The individual: Attribution of scholarly work – if scientists are to share intellectual property then the commodity needs to carry appropriate attribution. This is the means by which reputation is propagated through the community.

The community: Recommendation of workflow, services, etc. is a vital part of enabling sharing through discovery by other scientists; the ability to review and comment is an inherent part of recommendation; communication of know-how about running or using an experiment is part of establishing and disseminating best practice.

The rise of harnessing the Collective Intelligence of the Web, the so-called Socio-Web, has dramatically reminded us that it is people who generate and share knowledge and resources, and people who create network effects in communities. Blogs and wikis, shared tagging services, instant messaging, social networks and semantic descriptions of data relationships are flourishing. Within the Scientific community we have examples: OpenWetWare, Connotea, and groups and applications on Facebook (see corresponding .org Web Sites and facebook.com).

By sharing an increasing body of workflows with an increasing number of users we have the potential for substantial benefits to scientists. In addition to contributing workflows, the greater the number of users, and the more workflow downloads and invocations, then the more evidence there is to assist in selecting a workflow. By mining the sharing behaviour between users within such a community we can provide recommendations for use, and by using the structure and interactions between users and workflow tools we can identify what is considered to be of greater value to users. Enactment generates service

usage and data provenance information which help track down workflows as well as data and services.

2.3 *Workflow Systems and Communities*

Scientific workflow systems with significant deployment include the Taverna workflow workbench (12), Kepler (8), Triana (1) and Pegasus (3). Taverna is the exemplar workflow system in this paper. Developed by the *myGrid* project (13) within the UK e-Science programme (14), Taverna is used extensively across a range of Life Science problems: gene and protein annotation; proteomics, phylogeny and phenotypical studies; microarray data analysis and medical image analysis; high throughput screening of chemical compounds and clinical statistical analysis.

Importantly, Taverna has been designed to operate in the ‘open wild world’ of bioinformatics. Rather than large scale, closed collaborations which own resources, Taverna is used to enable individual scientists to access the many open resources available on the Web and not necessarily within their enterprise. Many of the services are expected to be owned by parties other than those using them in a workflow. In practice they are volatile, weakly described and there is no contract in place to ensure quality of service; they have not been designed to work together, and they adhere to no common type system. Consequently, they are highly heterogeneous. By compensating for these demands (5), Taverna has made, at the time of writing, over 3500 bioinformatics orientated operations available to its users. This has been a major incentive to adoption. This openness also means that Taverna is not tied exclusively to the bioinformatics domain – any services can be incorporated into its workflows.

By way of comparison, the lifecycle of workflows in the Pegasus system has also been the subject of study (3). Pegasus has more of a computational and Grid emphasis. It maps from workflow instances to executable workflows, automatically identifying physical locations for workflow components and data and finding appropriate resources to execute the components; it reuses existing data products where applicable. Pegasus is used within large scale collaborations which are perhaps more typical of many grid and larger e-Science projects, while Taverna gives an interesting insight into another part of the scientific workflow ecosystem – it is being used by many scientists on their personal projects, constituting a distributed, disconnected community of users who are also the developers of the workflows. While e-Science has often focused on large scale collaborative projects with highly coordinated infrastructure, Taverna is used by the long tail of researchers doing everyday science.

Taverna provides a significant case study for our work because its distributed,

decoupled community is already beginning to find new mechanisms for sharing workflows, including for example web sites and Wikis. Through Taverna we can address the decoupled community of one workflow system, and from this learn how to work with other workflow systems and their communities; for example, we are doing this with Triana and Kepler. Ultimately we envisage that scientists will transcend individual systems, finding whatever workflow is appropriate for their applications and able to utilise multiple systems – what we call the *multiworkflow* environment, in which workflows can interwork by exchanging data or calling other workflows as services.

3 The design of *my*Experiment

To address and explore these issues, and to support a growing and distributed user base of workflow users and developers, we have designed and developed a Virtual Research Environment to support scientists using workflows. We call this *my*Experiment (4) and we envisage: a ‘gossip shop’ to share and discuss workflows and their related scientific objects, regardless of the workflow system; a bazaar for sharing, reusing and repurposing workflows; a gateway to other established environments, for example depositing into data repositories and journals; and a platform to launch workflows, whatever their system. *my*Experiment aims to take a step beyond existing workflow repositories by crossing project, community and product boundaries, emphasising social networking around the workflows, providing gateways to other environments and forming the foundation of a personal or laboratory workbench.

3.1 *Web 2.0 perspective*

Our design methodology was closely based on our experience with Taverna, from which we identified principles for designing software for adoption by scientists and the associated mechanisms for user-engagement to support this process (15). However for *my*Experiment we were drawing inspiration from the Web 2.0 approach, from systems such as Facebook, MySpace and Amazon (see the corresponding .com sites), so we considered our principles in the light of the Web 2.0 design patterns (16). Here we summarise the application of each of the design patterns in the context of *my*Experiment:

- *The Long Tail* – as science becomes increasingly digital, more and more researchers are able to participate in sharing scientific assets with ease. A large number of scientists conducting the routine processes of science on a daily basis are set to make up the bulk of users. Hence *my*Experiment explicitly targets the decoupled community of everyday researchers.

- *Data is the Next ‘Intel Inside’* – our users are focused on data, and workflows themselves are the unique data of *myExperiment*. Hence we focus on supporting workflows through their lifecycle, including enactment – this gives us workflows, metadata, data inputs and products, the associated provenance information for both workflows and data, versioning and the collection of related digital assets into ‘experiments’.
- *Users Add Value* – to achieve the user-generated content that underlies the value of *myExperiment*, we make it as straightforward as possible for users to add workflows and other scientific assets to the pool. Additionally, developers are able to add value to the *myExperiment* software itself by developing over the API or enhancing the codebase. We have found that many of our science users will not be willing to contribute unless attribution mechanisms are in place and there is fine control over ownership and sharing.
- *Network Effects by Default* – many *myExperiment* users will simply wish to find workflows, not contribute them. To achieve network effects from these users we aggregate data as a side-effect of their use of *myExperiment*, for example the numbers of times workflows and services are used; this enables the community to benefit from usage without explicitly uploading new content. We also encourage tagging and reviewing of existing workflows.
- *Some Rights Reserved* – users require protection as well as sharing, but the environment must be designed for maximum ease of sharing to achieve collective benefits. Initiatives such as Science Commons provide a useful context for this. Meeting the particular needs of scientists in terms of ownership, attribution and licensing is one of the things that distinguishes *myExperiment* from other social web sites.
- *The Perpetual Beta* – Web applications are delivered to their users as services rather than software. *myExperiment* is building a collection of online services and aims to meet emerging user requirements in an agile manner. Hence it is designed and built from a service delivery viewpoint, and the users are effectively beta testers; the service is instrumented so that usage can be evaluated. Note, however, that we aim to provide greater stability in the API.
- *Cooperate, Don’t Control* – *myExperiment* exists in a network of cooperating data services with simple interfaces which make it easy to work with content. Hence *myExperiment* sets out to both provide services to others and reuse the services of others. It makes it as easy as possible for people to build ‘functionality mashups’ making use of *myExperiment* – to couple services together to meet the needs of scientists.
- *Software Above the Level of a Single Device* – There are many ways of accessing *myExperiment* other than from a desktop PC, ranging from mobile phones to tablet PCs in laboratories. Hence we have designed and tested *myExperiment* on a variety of devices from the outset, and we make it as easy as possible for people to build new interfaces.

Based on this motivation and approach, *myExperiment* was designed through a series of design and scoping workshops with end-users. A growing set of use cases drive the design, and one of these – illustrating individual use – is shown in Figure 1. Firstly the scientist uses *myExperiment* to discover a workflow from those publicly available or shared with the social network, using a variety of techniques including search and browsing, and assisted by tags, reviews and recommendations. Secondly the workflow is downloaded to be enacted and edited in the desktop Taverna environment. Finally, a new version of the workflow is uploaded to *myExperiment*, and the user has the option of keeping it private or perhaps sharing with a group on the site. As an alternative to enacting the workflow within the full interactive capabilities of Taverna, it can also be executed directly through the web-based interface of *myExperiment*. Other use cases involve multiple users; for example, a teacher might collect together workflows with example data and then make these available to students; the students then do their homework and submit their work back to the teacher, who then grades them using *myExperiment*.

Below we summarise three key design decisions which illustrate our design approach in action; information about other design decisions can be found in (17). Each of these presents a distinctive aspects of *myExperiment*:

Federation *myExperiment* provides an open source codebase, and individuals and laboratories are free to install their own *myExperiment* instances. They can then link them up using a federation model if and as they wish. This is achieved through the repository interoperability protocols of the Open Archives Initiative (OAI) (18), building on the experience and publishing ethos of the CombeChem (19) eBank-UK (20) projects. Additionally we provide a public site (www.myexperiment.org) so that scientists gain immediate benefit from available workflows. This federation approach provides a mechanism for scaling *myExperiment* deployment across multiple communities and providing appropriate community-specific customisations. It also facilitates the integration of *myExperiment* with existing repositories. We note that this open approach is in line with the Web 2.0 principles but it is not a characteristic of some of the major social networking sites.

Interface The public site provides a service for those who do not already have sharing and collaboration mechanisms in place. In addition to bringing the user to *myExperiment*, we undertake to bring *myExperiment* to the user – we break out *myExperiment* functionality through simple APIs so that it can be accessed through existing interfaces, including Wikis and web pages; for example, the interface might simply provide a control panel for executing pre-installed workflows, in which case the user does not see the workflow itself, or it could be a new interactive tool or a mashup. We also enable

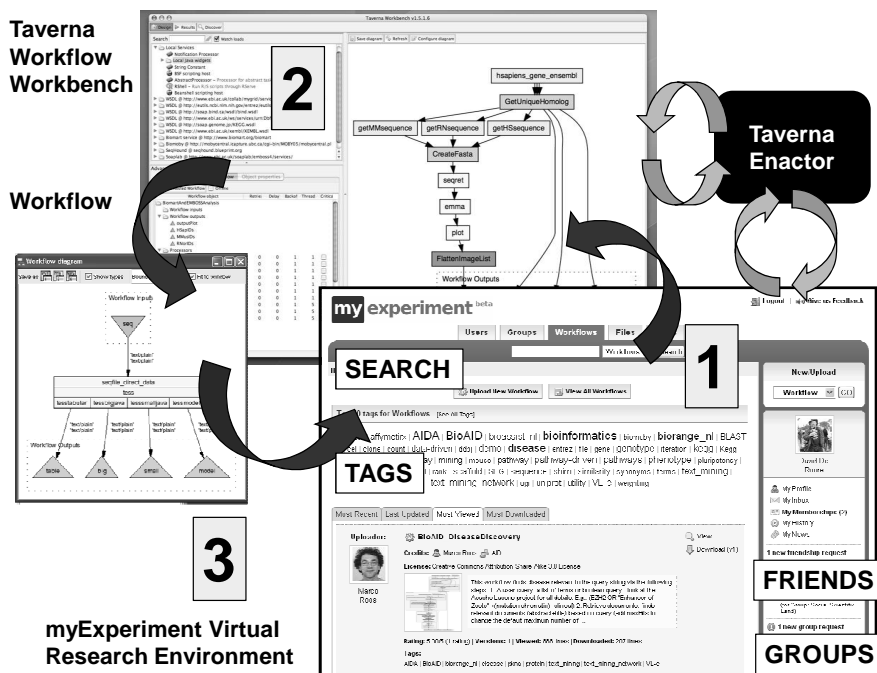


Fig. 1. The scientist (1) discovers a workflow in *myExperiment*, (2) executes and edits it in Taverna, and (3) uploads a new version to *myExperiment*.

the creation of Google Gadgets and *myExperiment* add-ons for sites such as Facebook. The role of *myExperiment* as provider of functionality rather than Web site *per se* is illustrated in Figure 2, which exemplifies the ‘cooperate don’t control’ design pattern.

Research Objects We have focused on workflows, but we know there is an immediate need to associate workflows with other information and also to work with collections of workflows. More generally, the *myExperiment* concept is about sharing digital objects which include data, results, provenance information, tags, associated documentation, etc. These individual items can be collected together to form *research objects*, for example to record an experiment. To address this we have designed a simple way of composing dispersed items into an *Encapsulated myExperiment Object* (EMO). For compatibility with open repositories, these aggregations are exported using the OAI *Object Reuse and Exchange* representation, which is based on named RDF graphs; we also adopt linked data practices. Thus the *myExperiment* environment is becoming a way of working with aggregations of scientific assets and of providing associated services for others to use. Our support for workflows and these compound research objects is in line with the data-centric approach of Web 2.0 and is a distinctive feature of *myExperiment*.

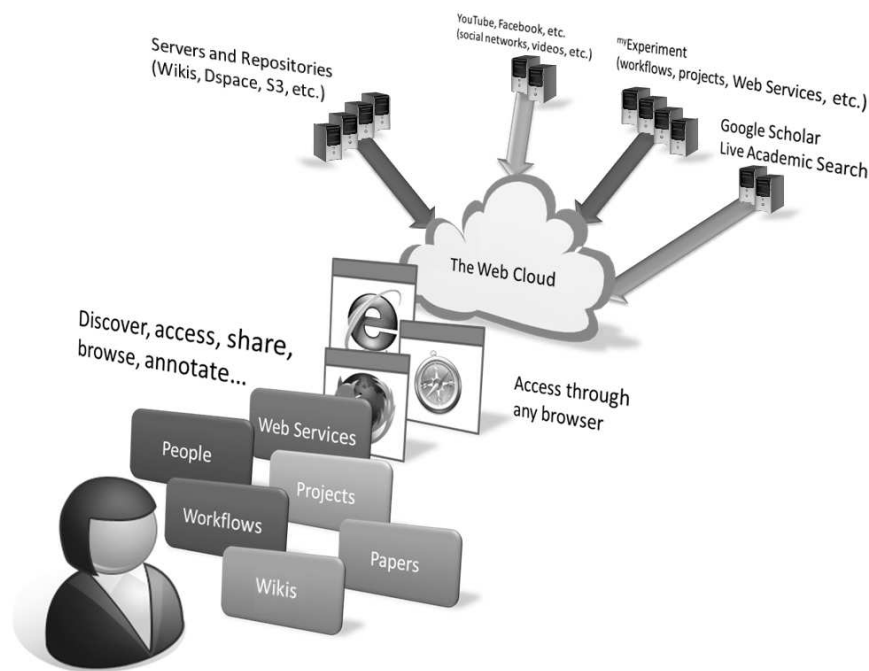


Fig. 2. *myExperiment* brings functionality to the user through their familiar interfaces and can be combined with other services.

4 Realisation

The realisation of *myExperiment* as a service for the community has adopted a highly user-centric approach, consistent with the notion of perpetual beta. Rather than a single progression of requirements capture, design, build and test, we delivered a prototype system which provided the mechanism for requirements capture, designed the *myExperiment* data model and architecture, and then delivered a service which has continuously evolved in response to user needs.

Figure 3 illustrates several of the architectural decisions that were made in the first phase of the project. The *myExperiment* core functionality is the social network, ownership, sharing and permissions and workflow support; these are all reflected in the data model. As far as possible it makes use of other services, so for example it can reference content in external repositories. The enactor is a separate component so that users can plug in their own – a standalone enactor has been developed for Taverna workflows, enabling users in a laboratory to execute workflows through the *myExperiment* interface or other interfaces built upon it, and Triana has also been integrated. The search engine is separate too, so that one engine can support multiple *myExperiment* instances. The API has been designed in conjunction with its users, i.e. those building interfaces over *myExperiment* or integrating *myExperiment* functionality elsewhere. *myExperiment* instances are federated through metadata exchange.

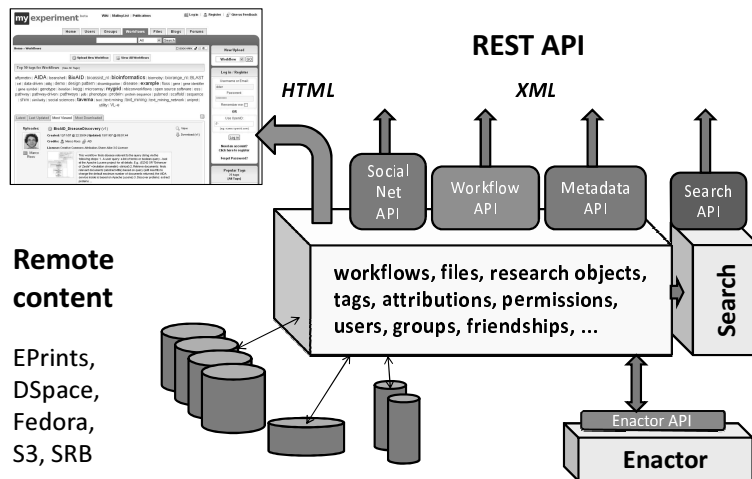


Fig. 3. Illustration of *myExperiment* architectural decisions.

An alternative approach would have been to adopt established portal technologies. However, the user interaction experience with *myExperiment* is quite different to most portals, and part of our goal is to deliver an interface that is familiar to the next generation of scientists. In fact the relationship of *myExperiment* and portals is symbiotic, because they solve different problems: a portal can make use of *myExperiment* functionality behind the scenes, and *myExperiment* can be used as a ‘skin’ for existing portals.

The system is implemented using Ruby on Rails (21), an open source Web application framework that has served us well in enabling the developers to spend less time writing code and more time working with users; this strategy was reinforced by the two developers being embedded within the scientists’ work environment at the outset of the project. The closed beta system was released three months from the project start and provided basic social networking facilities, serving an important role in capturing both community attention and requirements. User feedback on the closed beta was positive and brought to the fore the different requirements of a social web site supporting scientists.

While this was in use, the team designed the data model and a new codebase with a RESTful API which released at six months and immediately attracted developers (soon outnumbering the *myExperiment* development team) and the creation of Google Gadgets. With particular emphasis on support of ownership, attribution and sharing, the public beta was made available eight months into the project after user trials, in November 2007. Over the period January–April 2008 the site gained over 750 registered users, with many others accessing public content without needing an account; visits came from 78 countries. The site was well received by the community, by then carrying around a third of

all publicly available Taverna workflows.

Predominantly used by life scientists at the outset, several other communities became involved, notably chemistry, social statistics and music information retrieval. A combination of methods is in use for ongoing requirements gathering and evaluation. These have been selected for their complementary strengths and include: heuristic evaluation (22) for early and rapid detection of basic usability problems; questionnaires to profile the user population, solicit their requirements and feedback; semi-structured interviews to probe more deeply user attitudes towards, and experiences of, *my*Experiment; and ethnographic studies (23) to observe *my*Experiment in everyday use to reveal whether and how people actually find a role for it in their work.

5 Conclusions

We have made the case for a mechanism for sharing workflows in order to realise their scientific potential, and have identified the issues in achieving this. Enabling incentive models for sharing within a community of practice and supporting an emergent model of sharing is a challenge. The Virtual Organisations of Grid computing often attempt to achieve a similar objective, although they are typically centred on a common technically defined problem and do not focus on social aspects that might involve different incentive structures.

To rise to this challenge we have created *my*Experiment, which enables sharing of workflows in multiple workflow systems and approaches this from a social perspective, adopting Web 2.0 techniques. Fundamentally it is the simplicity of Web 2.0 for users and developers which is attractive. The benefits of community participation are becoming apparent, particularly with respect to community curation; for example, where workflows naturally ‘decay’ as services change or disappear, they can now be maintained or deprecated by the community. Hence *my*Experiment is benefiting not only the design of workflows but also the *quality* of workflows, in ways that were not available before.

Through creating *my*Experiment we are effectively conducting our own experiment: we are exploring whether communities of scientists will share to the extent that they benefit from the network effects from a social web approach. Through doing this we are gaining insights into the sharing behaviours of research communities across multiple domains. Uptake of the site confirms that, in our initial domains, *my*Experiment is already bringing benefit to both individuals and communities.

Acknowledgements

This work is funded by JISC VRE2 (04/06) with additional support through EPSRC grants EP/D044324/1, EP/C536444/1 and EP/C008863/1 and Microsoft. The National Centre for e-Social Science has provided invaluable assistance with user evaluation. Figure 2 is due to Savas Parastatidis. Thanks especially to the *my*Experiment, *my*Grid and CombeChem teams, and to the Taverna and *my*Experiment user communities (especially Douglas Kell, Paul Fisher, Duncan Hull, Marco Roos and Simon Coles). We are grateful to Ian Taylor, Bertram Ludascher and Ewa Deelman for providing links with Triana, Kepler and Pegasus respectively.

References

- [1] Ian J. Taylor, Ewa Deelman, Dennis B. Gannon, and Matthew Shields. *Workflows for e-Science: Scientific Workflows for Grids*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [2] E. Deelman and Y. Gil, editors. *Final report of NSF Workshop on the Challenges of Scientific Workflows*. National Science Foundation, May 2006.
- [3] Ewa Deelman and Yolanda Gil. Managing large-scale scientific workflows in distributed environments: Experiences and challenges. In *E-Science '06: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing*, page 144, Washington, DC, USA, 2006. IEEE Computer Society.
- [4] Carole Goble and David De Roure. *my*Experiment: social networking for workflow-using e-scientists. In *WORKS '07: Proceedings of the 2nd workshop on Workflows in support of large-scale science*, pages 1–2, New York, NY, USA, 2007. ACM.
- [5] Chris Wroe, Carole Goble, Antoon Goderis, Phillip Lord, Simon Miles, Juri Papay, Pinar Alper, and Luc Moreau. Recycling workflows and services through discovery and reuse: Research articles. *Concurr. Comput. : Pract. Exper.*, 19(2):181–194, 2007.
- [6] P. Fisher, C. Hedeler, K. Wolstencroft, H. Hulme, H. Noyes, S. Kemp, R. Stevens, and A. Brass. A systematic strategy for large-scale analysis of genotype phenotype correlations: identification of candidate genes involved in african trypanosomiasis. *Nucleic Acids Res.*, 35(16):5625–33, August 2007.
- [7] Antoon Goderis, Ulrike Sattler, Phillip W. Lord, and Carole A. Goble. Seven bottlenecks to workflow reuse and repurposing. In *International Semantic Web Conference*, pages 323–337, 2005.
- [8] Kepler Web Site. <http://kepler-project.org>

- [9] Philip Maechling, Hans Chalupsky, Maureen Dougherty, Ewa Deelman, Yolanda Gil, Sridhar Gullapalli, Vipin Gupta, Carl Kesselman, Jihic Kim, Gaurang Mehta, Brian Mendenhall, Thomas Russ, Gurmeet Singh, Marc Spraragen, Garrick Staples, and Karan Vahi. Simplifying construction of complex workflows for non-expert users of the southern california earthquake center community modeling environment. *SIGMOD Rec.*, 34(3):24–30, 2005.
- [10] Inforsense Web Site. <http://www.chub.inforsense.com>
- [11] J.M. Ziman. *Public Knowledge. An Essay Concerning the Social Dimension of Science*. Cambridge University Press, Cambridge, England, 1968.
- [12] Tom Oinn, Mark Greenwood, Matthew Addis, M. Nedin Alpdemir, Justin Ferris, Kevin Glover, Carole Goble, Antoon Goderis, Duncan Hull, Darren Marvin, Peter Li, Phillip Lord, Matthew R. Pocock, Martin Senger, Robert Stevens, Anil Wipat, and Chris Wroe. Taverna: lessons in creating a workflow environment for the life sciences: Research articles. *Concurr. Comput. : Pract. Exper.*, 18(10):1067–1100, 2006.
- [13] myGrid Web Site. <http://www.mygrid.org.uk>
- [14] Tony Hey and Anne E. Trefethen. The UK e-Science core programme and the grid. *Future Gener. Comput. Syst.*, 18(8):1017–1031, 2002.
- [15] David De Roure and Carole Goble. Six principles of software design to empower scientists. *IEEE Software (to appear)*, 2008.
- [16] T. O’Reilly. What is Web 2.0 – design patterns and business models for the next generation of software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [17] David De Roure and Carole Goble. myExperiment – a web 2.0 virtual research environment. In *International Workshop on Virtual Research Environments and Collaborative Work Environments*, 2007.
- [18] Open Archives Initiative Web Site. <http://www.openarchives.org>
- [19] K. Taylor, R. Gledhill, J. W. Essex, J. G. Frey, S. W. Harris, and D. De Roure. A semantic datagrid for combinatorial chemistry. In *GRID ’05: Proceedings of the 6th IEEE/ACM International Workshop on Grid Computing*, pages 148–155, Washington, DC, USA, 2005. IEEE Computer Society.
- [20] Monica Duke, Michael Day, Rachel Heery, Leslie A. Carr, and Simon J. Coles. Enhancing access to research data: the challenge of crystallography. In *JCDL ’05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, pages 46–55, New York, NY, USA, 2005. ACM.
- [21] Ruby on Rails Web Site. <http://www.rubyonrails.org>
- [22] Jakob Nielsen. *Heuristic evaluation*. John Wiley & Sons, Inc., New York, NY, USA, 1994.
- [23] R. J. Anderson. Representations and requirements: The value of ethnography in system design. *Human-Computer Interaction*, 9(2):151–182, 1994.