A Virtual Research Environment for Collaboration and Sharing of Experiments

Freestyle collaboration with my colleagues to share my e-Experiments in a way that suits Me

David De Roure, Jeremy Frey, Jonathan Essex, Simon Coles, Ed Zaluska

University of Southampton

Carole Goble, Robert Stevens, Steve Pettifer, Rob Procter

University of Manchester

Introduction

The ^{my}Grid (www.mygrid.org.uk) and CombeChem (www.combechem.org) e-Science pilot projects both constructed VREs for scientists. Both were motivated by reducing the "time to discovery" in the face of the data deluge which characterises much of e-Science. **^{my}Grid** focused on *in silico* experimentation by life scientists through the production, use, and exchange of scientific workflows that combine publicly available services (datasets, tools, applications etc). **CombeChem** focused on data acquisition in the chemistry laboratory and its processing and publication. Through experience with real users, both projects have observed the crucial need to support the scientists in the development of experiments – reducing "time to experiment".

These projects have taken scientists into a world where their experiments are described digitally – which means they can share and build experiments in ways simply not possible before. Experiments typically take the form of workflow descriptions and associated artefacts – already an established technique in the life sciences and generally applicable to any research which involves processing data to get results, running simulations, combining information sources etc. For example, the ^{my}Grid Taverna Workflow Workbench is an open-source desktop tool which provides a graphical interface to facilitate the easy building, running and editing of workflows. These workflows contribute to a know-how pool, as do the services that they orchestrate, so that they can be exchanged, traded, reused, and remixed.

myExperiment is a Virtual Research Environment which makes it easy for people to share experiments and discuss them. Scientists should be able to swap workflows and publications as easily as citizens can share documents, photos and videos on the Web. myExperiment owes far more to social networking websites such as MySpace (www.myspace.com) and YouTube (www.youtube.com) than to the traditional portals of VLEs and Grid computing, and is immediately familiar to the new generation of scientists[1]. The myExperiment VRE provides a personalised environment which enables users to share, re-use and repurpose experiments – reducing time-to-experiment.

The scoping and preliminary design work for myExperiment in the life science domain has already been conducted in ^{my}Grid, and the myExperiment concept is attracting the attention of the international community.[2] Building on this, and on CombeChem's experience embedding experiments in the scholarly knowledge cycle, the myExperiment VRE will be developed and deployed with life scientists and chemists, then extended to new user groups including social scientists and astronomers. We have already attracted a great deal of interest, anticipation and excitement, and we are confident that there is a very real need for this VRE.

The myExperiment project is a 24 month project commencing March 2007 in two phases – the development of the myExperiment VRE and deployment in life sciences and chemistry, and then the extension to new user communities such as the social sciences and astronomy. For all these users we will be investigating the benefits of sharing and collaboration enabled by the VRE.

myExperiment contributes to the VRE programme by bringing in the results and experience of the £5.8M e-Science investment in the ^{my}Grid and CombeChem pilot projects. Broadening first to two scientific domains and then into the wider community, myExperiment will extend the new research practices already established by early adopters in life sciences and stimulate change in research practice across further disciplines – both in the use of workflows and the sharing of experiments. The partners' combined experience of two large pilots and other VREs brings an understanding of interoperability, which is essential to provide the ease of use that users are already demanding. By building on solutions already in use, myExperiment will provide a significant example of VRE capabilities to the wider community.

---

1 See http://portalparty.blogspot.com/ for a partial list of such systems.

2 See New Scientist http://www.newscientisttech.com/article.ns?id=mg19225745.500; GRIDToday http://www.gridtoday.com/grid/963514.html; to appear in The Lancet Oncology;

Project Description

myExperiment is very much a development and deployment project and sets out to gain real experience in authentic user settings. The background R&D is already done through ^{my}Grid and CombeChem and, within the life science community, we have already conducted a scoping exercise, requirements capture and preliminary design work. myExperiment is not 'yet another portal' and doesn't try to do all things for all users – it will be developed through close interactions with the target user communities, evolving towards a customisable solution that can be configured to the requirements of specific communities. The keywords are useful, reliable and resilient.

In preparation for this proposal, over 20 life science users met in a workshop facilitated by NCeSS to identify a set of requirements (www.mygrid.org.uk/wiki/Portal). The 26 topics identified fall into 4 categories: social networking, artefact management, execution environment and gateways. Mock-ups of myExperiment web sites have been used as a basis for discussion and ideas (myexperiment.org/mockup/). A 'lightweight' repository of workflows is already in use on the ^{my}Grid WIKI, providing a common place for researchers using Taverna to publish their workflows and results (myexperiment.org). This design work continues in conjunction with the National Centre for eSocial Science in the run-up to this proposed project commencing in March 2007.

Based on this preliminary scoping and design work, here we describe the functionality of the myExperiment VRE through four scenarios, which correspond to our proposed pilot studies.

Scenario 1 – Life Sciences

A biologist decides to use myExperiment to facilitate her use of eScience tools in her data analysis. She visits the myExperiment.org portal and creates an account. She describes her area of scientific interest, including her biological interests (in this case, tolerance to trypanosomiasis in cattle), the experimental techniques such as microarrays and Quantitive Trait Locii, and the bioinformatics tools and techniques that are of interest to her. It gives her a list of potential collaborators who have similar interests, and is encouraged to nominate other potential myExperiment users to join in. She is shown a set of workflows and results that match her interests, with new and highly rated contributions highlighted. She sees three workflows that are of particular interest, all of which appear to be for analysing microarray results associated with QTL. She is told that one of these workflows does the analysis at the level of the gene, whereas the others perform the analysis at the level of the pathway. Comments from other users about these workflows tell her that analysing microarray data at the pathway level gives better results due to the larger granularity of analysis. Finally, the comments indicate that one of these workflows is more reliable in terms of completion than the other. She is able to choose the best workflow for her task, and either download it for local use, or execute it directly from the web page. In doing so she is linked in to the scientific community consisting of the workflow's creator and its other users. Having used the workflow, she is able to add her own comments about her experience, offering tips for using the workflow and selectively publishing her results.

Another scientist has created a new workflow for performing metabolomics experiments based on the genes present in defined locii. He wants to share this workflow with potential collaborators. As a user of myExperiment, he has found it an easy and rapid way to share his ideas. He logs on, and sees that he has new collaborators that are using his previously existing workflows and data. He looks at their comments, responding to some with comments of his own. From within the Taverna workbench environment, which he has previously configured with his myExperiment account details, he selects the 'publish' option, and his new in-silico protocol is automatically added to his portfolio of published workflows. He annotates his workflow, guided by the portal which highlights terms previously not encountered, and links up terms already in its vocabulary. Existing and potential collaborators are notified of this new resource according to their account preferences. One of these, because of her interest in QTL, is informed of this new experiment, and sees the opportunity for a new collaboration.

Scenario 2 – Chemistry

A chemist interested in materials for their electrochemical properties has synthesised a new compound which is electrochemically inactive. She decides to use an e-Science approach to determine whether this new compound might have any alternative applications and therefore visits a Virtual Research Environment (VRE). Using a chemical search engine she looks for related materials and finds that an analogue has been shown to have optical activity as a sensor that can operate at high temperatures. Unfamiliar with the field, she queries myExperiment for workflows to analyse structure and thermal property relationships and a number of results that are highly rated are returned. She is able to view comments from other users, as well as the creator and is able to assess the various different software codes employed. She elects to use a workflow that derives a melting point from a crystal structure. However, as she has no access to instrumentation or expertise to derive crystal structures, she queries an Experiment Resource Discovery service and finds the National Crystallography Service. She enrols with this experiment data collection service and submits a sample of her material by post for analysis. After some days she is alerted that her sample is on the instrument and data is being collected. Via the VRE she monitors and steers the experiment and then collaborates with the service in the crystallographic workup and analysis of the compound. Once a result is obtained she then subjects it to the melting point calculation workflow which is enacted on the Grid.

Scenario 3 – Social Statistics

A social scientist decides to turn to myExperiment for help in her studies of the impact of road pricing on traffic growth. She does this as managing the complex, iterative task of performing statistical analyses over distributed resources is proving a barrier to her work. She searches the myExperiment workflow repository to find instances of workflows which broadly match her needs. She finds a suitable workflow template which that she adapts, selecting her input datasets and data filters, and setting parameters for a series of simulation runs using the NCeSS MoSeS Node agent-based modelling service. She arranges for the outputs of the different simulation runs to be stored in her personal myExperiment workspace for later statistical analysis. Once her research is complete, she deposits her workflow and results back into her private area within the myExperiment repository. Once she has written up the results, she will publish them in the global repository.

Another social scientist wishes to do a comparative study of modelling assumptions used in studies of the impact of land use strategies on the environment. He searches the literature to find relevant papers and then cross references the resulting citations to see if any of these match publication recorded as being linked to workflows published in the myExperiment repository. He downloads the corresponding workflows and extracts the model parameters from the provenance metadata. In one of the workflows he also finds annotations added by another researcher which direct him to examine in more detail one of its model parameters. He decides to investigate this by re-executing this workflow with a range of values for this particular parameter. Once this is done, he emails the researcher who added the annotation to tell her about his findings.

Scenario 4 – Astronomy

An astronomer has just become interested in a particular area of night sky. He wants to gain access to all the data that is available for that area from the main providers of astronomical data (e.g. ESA, ESO, NASA). He uses myExperiment to search for data resources available for his key area. His search discovers workflows tagged with terms matching to some of his interests. Using myExperiment's social organisation support, he contacts the workflow author to discuss prior results and possible collaboration. He uses one of the workflows, modifying it to use one of his own locally developed analysis tools. He then publishes this workflow to myExperiment to share his experiment with the original author, giving her access to his results and the provenance of his workflow shows her contribution to his work. His new collaborator wants to try his workflow for herself and so he publishes details of his locally developed analysis tool to the myExperiment repository. He indicates access constraints in the registry as his institution has to impose these to prevent over demand on resources by keen hobbyists. He is able to publish the results of his analysis to the community through a respectable journal. As part of his paper, he references his workflow in the myExperiment repository and is able to share the experiment itself with the astronomical community as a whole.

Designing myExperiment

These scenarios illustrate the five major functionalities of the myExperiment VRE (Figure 1):

- **Social networking**. myExperiment is a place where digital artefacts are shared by members of a community. People create personal profiles and groups, and can tag digital artefacts for ease of discovery and re-use. It features a search capability;
- **Artefact management**. The digital artefacts associated with experiments – workflows, results, services, provenance records – are managed by the user, who can choose to keep them private, share them or make them public;
- **Metadata management**. The digital artefacts are annotated with semantic descriptions (for curated services and workflows) and/or loosely tagged, to support search, discovery and facilitate federation of VRE content. In particular, the shopping metaphor for "shopping for workflows" is a strong one, and already piloted by the ^{my}Grid team. The metadata needs seeding and curating;
- **Execution environment**. A user within the myExperiment environment can execute workflows directly, and a user of tools such as the Taverna workbench can seamlessly access artefacts within the myExperiment environment. The results of execution are published under the control of the user;
- **Gateways** to other services, web-based tools and desktop applications, for example publishing data to repositories, or letting others create tools which access a workflow collection. This is part of the underlying federation model.

We also note two key shifts in thinking in myExperiment compared to many e-Science and grid solutions:

1. A focus on making it easy to *publish* information, not just on discovering and using it. In myExperiment we see three degrees of publication: discovering and sharing experimental artefacts in an open environment, publishing results to standard community curated repositories (e.g. Uniprot) and publishing scholarly output. This draws in particular on the CombeChem publishing philosophy;
2. Use of *familiar social networking website techniques*, enabling informal interaction and progressive publishing with the user in control. This is in contrast to the more prescriptive, heavyweight and

specialist interfaces provided by many portal solutions which are focused around controlled access to content and services rather than publishing and sharing.

The immediate beneficiary of myExperiment is the researcher who wants to exchange experiments with others in the same community using the same tools – myExperiment enables them to establish best practice and accelerate their research. Additionally, users in the same community using different tools can share other artefacts of experiments, such as information about the execution of the workflows. On another axis, sharing may be between users within one community or across communities. Ultimately myExperiment enables new forms of collaboration between communities previously unable to share, enabling entirely new research. These relationships are exemplified by the user communities engaged in this project.

The myExperiment environment is a *federated system* involving *multiple myExperiment instances* talking open protocols in order to interwork, to make artefacts from elsewhere available in myExperiment and to make myExperiment artefacts available through other services and tools. This is the federation model established through the CombeChem and eBank projects, and it is enabled by metadata sharing between VREs. A direct parallel may be drawn with the Open Archives Initiative (www.openarchives.org) which enables access to Web-accessible material through interoperable repositories for metadata sharing, publishing and archiving. OAI arose out of the e-print community but now addresses other digital materials, promoting a low-barrier interoperability framework and associated standards.

Both ^{my}Grid and CombeChem placed an emphasis on recording the *provenance* of data. This is particularly important in an open environment where sharing of data occurs – data values themselves are meaningless without some understanding of what they represent and how they came about. Provenance records are generated by Taverna when workflows execute, and ^{my}Grid has developed a sophisticated means of storing these using four ontologies. CombeChem has developed the notion of "publication at source" in which the experimental plan is part of the context of the acquired data and it is possible to chase back from results to the source data and its context. This is illustrated by the ecrystals interface (ecrystals.chem.soton.ac.uk). myExperiment permits sharing of provenance records. Protocols for sharing are being developed through Provenance Challenge workshops (twiki.gridprovenance.org/bin/view/Challenge/WebHome) led by Luc Moreau in Southampton and Ian Foster at Argonne National Lab.

To demonstrate the use of myExperiment to share artefacts other than workflows, we will work with the Kepler (kepler-project.org) workflow team, with whom we have a long-standing collaboration through the myGrid Link-up project (www.mygrid.org.uk/linkup). The Kepler and Taverna workflow systems differ in design and execution. Transforming a workflow from one into the other is difficult, but being able to compare workflow results and experimental design between systems is valuable. myExperiment provides a common workflow discovery mechanism for Kepler and myGrid to enable comparisons between workflows and their results.

Preliminary discussions with Taverna users in the Chemoinformatics domain suggest that there will be a number of practical and cultural differences between our initial two communities. We note different behaviours with respect to sharing of data, and we anticipate that chemistry workflows are more likely to be tuned to local computing facilities. These are important issues that will be understood through the study in this project.

Once we have a workflow repository in place we will not only benefit from sharing but also the analysis of workflows. For example, we will be able to identify common patterns, and better understand the workflow discovery behaviour as numbers increase. This analysis will be a subject for future work.

Pilots

We are piloting with 4 user communities at different levels of maturity with respect to the use of workflows. Nominated *scientific community workers*, resourced through additional funds, will work with each community to seed each test-bed with content, curate content and metadata, and build the community.

Life Scientists. ^{my}Grid was targeted at skilled bioinformaticians who learned to use the tools and techniques and can make workflows. myExperiment broadens the user base to those who will not make workflows but can discover them. This community includes a significant number of the PhD and postdoctoral biologists in the UK, numbering thousands. The Manchester Centre for Systems Biology headed by Doug Kell, the Bioinformatics Group at Manchester led by Andy Brass and Robert Stevens, Anil Wipat's Bioinformatics Group at the University of Newcastle, Dawn Field, Head of the Molecular Evolution and Bioinformatics and Director of the NERC Environmental Bioinformatics Centre, and Misha Kapushesky from the Microarray Informatics Team, European Bioinformatics Institute will drive the testbed. In addition we have support from international myExperiment enthusiasts including Vrije University Amsterdam and the ICapture Laboratory at British Columbia.

Chemists: The CombeChem, eBank and R4L projects target the experimental process and research and information discovery lifecycle. The myExperiment proposal provides a mechanism by which these test-beds can be linked up and hence provide a provenance trail from discovery all the way back to the start of the data generation process in the laboratory, and to facilitate the creation of these workflows. Much of chemistry is dependent on structure and its relation to physical properties or reactivity and hence the ability to make these links, get back to source data and reuse workflows will affect a vast proportion of the academic research community, both in the UK (some 30-40 departments, thousands of PhD and PDRA researchers) and worldwide. This test-bed will be driven by Jeremy Frey and Simon Coles at the University of Southampton, with international support from the University of Indiana.

Social Statisticians. The use of complex modeling techniques are found in many university social, medical and bio science departments where they commonly work collaboratively in interdisciplinary teams. They are also found in most government departments and local authorities, working in support of policy analysts. This scattered community amounts to well over a thousand, leaving aside all the social science PhD students for whom social statistics is a required part of their training. This test-bed will be driven by Rob Proctor, Director of the National Centre for e-Social Science, in collaboration with David De Roure who is a member of the Southampton Statistical Sciences Research Institute (www.s3ri.soton.ac.uk).

Astronomers: The concept of workflows and data/analysis sharing is not new to the astronomy community; and the astronomy community is mature in this respect. A number of Virtual Observatories exist and the International Virtual Observatory Organisation is working to standardise interfaces to allow easier sharing of resources within the community. myExperiment offers the social networking infrastructure to support this community. The number of PhD students in Astronomy in the UK alone numbers in the hundreds. This test-bed with be driven by Nick Walton of the AstroGrid project (www.astrogrid.org).

Work plan

The project is committed to the VRE development model, and the partners have direct experience of this model through their engagement with OMII-UK. Based directly on the 'figure of 8', the project brings together:

- a technical development team, responsible for construction of the VRE;
- scientific community workers who, with the developers, are responsible for the pilots;
- a collection of volunteer community advocates (drawn from the OMII-UK advocates programme and elsewhere), responsible for stakeholder engagement;
- our user communities.

Technical Developers (2 RAs full time 12 months; 1 RA full time for 6 months) will: (a) deliver the software to OMII-UK quality assured standard, (b) deploy, host and tend the pilot artefact repositories; (c) build exemplar gateways to tooling over the myExperiment artefacts by (i) desktop tools, specifically a Taverna workflow execution capability and (ii) web-based tools, specifically a scientific publication capability.

- Developer 1 (D1), David Withers, based at Manchester has responsibility for the workflow and service artefact repository and special accessibility tooling (WP2) and the Taverna workflow execution capability (WP4).
- Developer 2 (D2), Don Cruickshank, based at Southampton, has responsibility for the scientific publication capability (WP5) and metadata federation.
- Both Developers work together to assemble, build and deploy the myExperiment framework (WP1); the artefact accessibility capability and management and developers interoperability interface (WP1) and the social networking capability (WP2).

Scientific community workers (SCWs) (2 x 0.5 RAs funded partly through external contributions for 18 months) have a guardianship and facilitating role for community building, and a responsibility to populate and curate the test-beds. They will work with the scientific communities, and in particular the community advocates, to: (a) articulate requirements; (b) seed the pilot myExperiment instances, or stimulate the seeding process; (c) nurture the myExperiment trails by stimulating the community interaction with the myExperiment instances and (d) curate, or stimulate the curation, of core artefacts contained within myExperiment instances. We will use and leverage the metadata annotations already available in the current ^{my}Grid service repository, using the myGrid service ontology, and incorporate the ^{my}Grid Feta service discovery tool (WP6).

Life Science:	*Social Statistics:*	*Chemistry:*	*Astronomy:*
Duncan Hull;	Danius Michaelides;	Simon Coles;	AstroGrd Nominee.

Community advocates (CAs) (two from each community) are nurtured volunteers who have an elevated role above the community membership as a whole, acting as scouts and sentinels of their community's myExperiment. They are hand-picked voices of the myExperiment community for their community. They will work closely with the community worker to support the stimulation, seeding, and curation, as well as observing the pilots. They will work with the development team to propose improvements, capabilities and identify partners for third party added-value capabilities. The Life Sciences and chemistry already have these advocates in place through the OMII-UK User Advocates programme (www.omii.ac.uk). Others will be drawn from the AstroGrid project and on advice from the NCeSS.

Work Packages

WP	Description (deliverable milestone months)	
1	**myExperiment Framework and basic capability** T1: Confirm technology choices and assess GridSphere flexibility, interop & scalability (1)	D1, D2

	T2: Assemble, build and deploy the myExperiment framework. (4, 8) T3: Host and support 4 myExperiment pilot installations. (4, 8,12, 18) T4: Accessibility capability and developers interoperability interface for desktop and web-based tools (5) T5: Package myExperiment to be deployable by third parties. (16)	
2	**Social networking capabilities** T1: Social networking framework – federating blogs, wiki, chatrooms, RSS feeds, messaging, shared lists, shared artefacts, cross cutting security. T2: Tagging and recommender capability	D1, D2 SCWs
3	**Artefact management** T1: Workflow repository, based on prototype developed by myGrid, T2: Incorporation of the myGrid Service repository and Feta discovery tool. T3: Special accessibility tooling based on tagging and shopping metaphors based on WP1-T4 capability T4: Transparent seamless access to local private resources (view workflow results or private papers, etc) using WP1-T4(7)	D1, D2 SCWs
4	**Workflow desktop tool interoperability gateway** T1: Interoperability with Taverna workbench (an acquired workflow/service is executed) uses WP1-T4 (4, 8) T2: Kepler workflow execution (14)	D1 SCWs
5	**Publication web-tool interoperability gateway** T1: Scientific publication interoperability capability (an acquired paper is obtained or published) (9).	D2 SCWs
6	**Community building** T1: Seek, marshal, articulate and prioritise requirements (1-12) T2: Seed myExperiment pilots, and stimulate the seeding process, with workflows, discussions, workflow packs, service packs etc (4/5, 8/9) T3: Stimulating the community interaction such as discussion forums, and build incentives for sharing such as ratings (4 onwards) T4 Curate, and stimulate the curation, of core artefacts with metadata (4 onwards) T5: Convene and run the User Advocate workshops (6, 12) T6: Promote the myExperiment pilots to the wider community (throughout) T7: Evaluation of Trails report (18)	SCWs CAs.

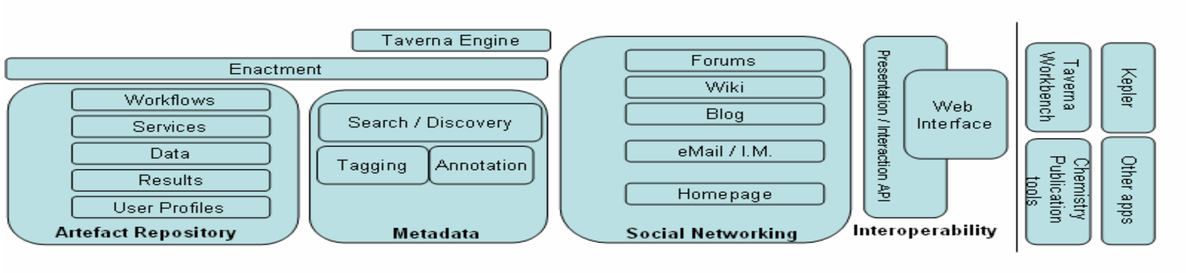

Figure 1: The conceptual architecture.

Phases and Pilots

We propose to run two **pilots** with each of the four pilot communities during the first year, progressively adding capability two each pilot. In each case the developers and the community workers will spend a month working closely with the users with assistance from the other developer. The pilots are staggered so to manage the launching of each pilot myExperiment instantiation.

	Month		Capability
Phase1 : Develop	4	Life Science 1	Basic framework, social networking capability, simple repository support, simple Taverna workflow launching; limited content; current myGrid service metadata tagging
	5	Chemistry 1	Basic framework, social networking capability, simple repository support, simple publication interoperability; first phase metadata tagging;
	8	Life Science 2	Extended social networking capability; full repository support; tagged discovery; extensive content; extended Taverna execution; private resources accessibility; simple publication interoperability; extended metadata tagging;

	9	Chemistry 2	Extended social networking capability; full repository support; tagged discovery; extensive content; extended publication interoperability; extended metadata tagging;
Phase2:Trial	12	Astronomy	Full capability
	12	Social Statistics	Full capability
	12-18	All	Incremental developments and sustainability practice deployed
	16	All	Evaluation of trials and review

Existing tools and proposed technologies

myExperiment inspiration comes from social networking websites such as MySpace rather than conventional Grid portals. The emphasis is on flexibility and self-determination for the scientist; "mash-up" accessibility for third party tool application providers and the leveraging of the increasing available and maturing mainstream and proven tooling for such sites provided by the open source community and large commercial vendors. By aligning ourselves with this "Web 2.0" wave of activity, we transfer mainstream technology for the mainstream citizen to the scientific community. We want to be able to rapidly develop the base of myExperiment. Consequently we apply existing Web 2.0 technologies, such as AJAX with the Google Web Toolkit, using the standard J2EE platform, using proven open source technology such as the Spring framework, Hibernate and Apache Tomcat. See oreillynet.com/lpt/a/6228 for a description of Web 2.0, offering some comparisons between the past and the present/future. A recent survey (ajaxian.com/archives/ajaxiancom-2006-survey-results) identifies that most widely used Java based AJAX toolkit is DWR (getahead.ltd.uk/dwr). For Wiki pages we propose to use Xwiki. Open source tagging and blogging tooling is also available. For the workflow and service repository (Figure 1) we will adopt the tools already available through myGrid, using a WebDAV backend, and the GRIMOIRES registry available from OMII-UK. Unlike the workflows and services repository, which will be held centrally, workflow results will be held locally at the users' site, using the myGrid Baclava and KAVE repository infrastructure, or in local repositories that will be configured to be accessible to myExperiment through the interoperability interface. Metadata annotation and tagging will use the already deployed myGrid Service Ontology. For the workflow execution environment we will use Taverna, interworking with R (building on CombeChem experience) for handling social statistics. Taverna is available open source from OMII-UK. R is available as GNU GPL.

We have experience in conventional portal frameworks such as UPortal, Sakai and GridSphere. By following established portlet standards such as JSR168, myExperiment modules, such as running or discovering a workflow, could be integrated with existing portal software. However, we have serious concerns about this as the primary interface due to lack of flexibility in interaction models, which are forced on the users. We propose, therefore, to undertake stress trials and to attempt a rapid prototype with GridSphere in the first month of the project (WP1). The myExperiment VRE aims to interwork with these solutions, as well as repositories such as ePrints as pursued in the JISC eBank project (www.ukoln.ac.uk/projects/ebank-uk).

The partners already have a collaboration agreement in place for OMII-UK, which addresses IPR and was designed to be consistent with JISC practice.

Management

The project will be managed on a day to day basis by June Finch, who is the project manager for the myGrid team of OMII-UK and a member of the operations team. This represents a considerable institutional contribution. The project will have a weekly meeting of all its developers and scientific community workers. These will be face to face at the sites and virtual between the sites. Every three weeks it will have a bigger virtual meeting of the community. These meetings will dovetail with the myGrid meetings that are already in place under the OMII-UK operational practice. We will leverage the technical management and User Advocate Programme already developed and being rolled out by the OMII-UK (reference). This ensures a quality of engagement and sustainability

There will be at least two workshops bringing everyone together, at month 6 to reflect upon the first pilots and at month 12 to reflect on the second pilot phase. The development strategy will follow the OMII-UK software engineering and quality assurance procedures to ensure quality of software and make it fit for purpose.

Risk register

Risk	*Likelihood L/M/H*	*Impact L/M/H*	*Containment Strategy*
Tooling unavailable	L	H	Tooling already used worldwide
Recruitment problems	L	M	Large teams at both sites
Fail to engage community	L	M	Advocates, content and social incentives, community workers

Unable to clarify requirements	L	M	Already achieved in Life Sciences
Sustainability of software	M	H	Standards-based solution
Sustainability of pilots	M	H	Refocus resource
Availability of content	L	M	Already proven
Failure of principle	L	H	Already proven

Impact and Sustainability

We are committed to open source development and already work in conjunction with JISC OSS-watch. The myExperiment software will be made available through the OMII-UK. The development practices will follow the OMII-UK QA process, and the manager of the project is on the operations team of OMII-UK. The myExperiment pilots' lifetimes depend to some extent on their take-up in the community. The Life Science pilot will be sustained by ESNW, in partnership with NIBHI for 12 months post the project, during which time we will seek funds to sustain it through the BBSRC's BBR programme and other mechanisms. We plan in due course that an organisation such as the EBI will host a national myExperiment for Life Sciences. The Social Science pilot will be sustained, developed and hosted by NCeSS through its e-infrastructure and services for social sciences project. There are already mechanisms in place for disseminating and supporting Taverna through OMII-UK. The myExperiment software will be additional to the Taverna offering and will follow the same dissemination practices.

Benefits to JISC Community

In addition to the benefits articulated to the VRE programme, the two most significant benefits of myExperiment to the JISC community are:

- The provision of a workflow solution which is easy to use by non-specialist researchers. Workflow is a key technology to bring together JISC services within the research environment. For example, we are already engaged in discussions with MIMAS and AHDS, and these providers are interested in lightweight service access via workflow as well as grid-enabled access;
- Generic technology solutions (Taverna and myExperiment) which have real utility across a broad range of disciplines and are highly customisable.

Our engagement with JISC through a variety of projects enables us to conduct this work embedded closely within the JISC context, including for example information environment activities.

Budget

Directly incurred costs	Soton	Manchester	Soton	Manchester	Soton	Manchester	
	Yr1	Yr1	Yr1	Yr1	Yr2	Yr2	FEC £
Personnel	[redacted]						
Staff							
Consultant							
Equipment							
Travel & subsistence							
Other							
Total incurred							

Directly allocated costs							
	Yr1	Yr1	Yr1	Yr1	Yr2	Yr2	
Personnel	[redacted]						
Institutional estate							
Total allocated							
Indirect costs							
General services							
Total FEC							

Expected contribution from Partners	
JISC contribution	[redacted]
Institutional contribution	
% JISC contribution	
% Institutional contribution	

Southampton: 1 developer (Cruickshank) full time for year 1 and 50% time for year 2, 20% of chemist community (Coles) 12 months and 20% social stats community (Michaelides) 12 months. **Manchester**: 1

developer (Withers) full time for year 1 and 50% time for year 2, life science community worker (Hull) and astronomy community worker (to be nominated by Astrogrid) supported by 6 month contribution from ESNW and support for collaboration using the myGrid platform and the OMII-UK link with Astrogrid. There is also a contribution from NCeSS prior to project start. Co-investigators bring skills necessary to successfully fulfil the "figure-of-8"; model and time calculations are based on weekly meetings plus additional specialist input in appropriate project phases. **Equipment** costs are based on developer PCs plus running the VRE servers within the pilots. **Travel** and **consumables** costings are based on project management meetings, workshops at months 6, 12, facilitated user engagement meetings and conference attendance (e.g. All Hands, NCeSS conference, OGF, NeSC workshops). We have based these figures on actual costs incurred in community engagement activities in other e-Science projects, in particular OMII-UK.

Consortium

No sites in the UK have any greater experience of e-Science pilot projects. Four of the 6 projects funded at the beginning of the e-Science programme involved Manchester (^{my}Grid, RealityGrid, Geodise) or Southampton (CombeChem, Geodise, ^{my}Grid). Manchester is also the home of the ERSC-funded National Centre for e-Social Science Hub. Its goal is to help social scientists to make the best use of new and emerging e-science technologies in their research. Of the many collaborative activities between the sites, the most notable is perhaps that Southampton and Manchester, with Edinburgh, form the Open Middleware Infrastructure Institute UK (OMII-UK). Both sites have also been involved in the VRE phase 1 programme, through the MEMETIC, SAGE and CORE projects. Other activities include MIMAS at Manchester and ePrints in Southampton. The Semantic Grid initiative, which influenced both ^{my}Grid and CombeChem, is driven primarily out of Southampton and Manchester. Manchester also hosts the e-Science North West regional centre (www.esnw.ac.uk) who are making a significant financial contribution this project.

Key Personnel

Developers

Don Cruickshank is a Research Fellow in the Pervasive Systems Centre in the School of Electronics and Computer Science, University of Southampton where he is currently employed on the CombeChem platform grant. Don's expertise is in distributed data-driven visual programming environments. He previously worked on the Equator e-Science project *Grid-based Medical Devices for Everyday Health* in conjunction with Oxford, in which he developed an interactive portal for multiple types of user engaged with the system. Don's PhD was on the Visual Programmability of Desktop Interfaces. Don has also been involved with the DTI-funded FloodNet project which collects data from deployed sensor nodes and performs flood prediction using Grid-based simulation. *Developer. 18 months funding by this proposal.*

David Withers is a Software Engineer in the School of Computer Science, University of Manchester where he is currently employed on the myGrid project. David's expertise is in working closely with users to design and implement graphical interfaces. He previously worked on 'expressik', a European Space Agency funded project, where he developed an open source toolkit for information model analysis and data validation. David has also worked on the EPSRC funded project 'Model Based Process Management for Electronic Design and Manufacture', which developed web-based user interfaces for managing the process of design and manufacturing of electronics products. *Developer. 18 months funding by this proposal.*

Community workers

Danius Michaelides is a Senior Research Fellow in the Pervasive Systems Centre in the School of Electronics and Computer Science, University of Southampton. He has been responsible for Southampton's role in the MEMETIC VRE project. Prior to that was employed by the e-Science project CoAKTinG (Collaborative Advanced Knowledge Technologies in the Grid) on which MEMETIC was based, and conducted trials of the CoAKTinG tools in conjunction with NASA. In the Equator Interdisciplinary Research Collaboration he was responsible for the design, development and deployment in the Chawton House project, also involving extensive user trials. Danius's PhD was joint between Social Statistics and Computer Science, funded under the ESRC Analysis of Large and Complex Datasets programme. *Scientific Community Worker for Social Statistics.*

Simon Coles is the Manager of the internationally renowned UK National Crystallography Service (EPSRC) and has led the NCS involvement in the UK e-Science program from the outset of the CombeChem EPSRC) project. SJC is a founder member of the CrystalGrid Collaboratory (e-Science Sisters project, EPSRC) and a co-investigator on the digital repository based projects eBank-UK and Repository for the Laboratory (JISC). SJC is crystallographic editor to two chemistry journals, on the editorial board of ChemistryCentral and coauthor of over 300 peer reviewed (primarily data-based) journal publications. SJC has recently been appointed to the advisory board of the Research Information Network (Engineering and Physical Sciences), the council of the British Crystallographic Association and is chairman of the Young Crystallographers committee. *Scientific Community Worker for Chemistry.*

Duncan Hull Duncan Hull is completing his PhD student in the BioHealth Informatics group in the University of Manchester. A biologist by training, Duncan has focused on turning myGrid Web service discovery into a feasible activity for life scientists. The work required close interaction with bioinformaticians and biologists. Before academia, he was an XML software engineer with CSW Informatics Ltd and built web applications for the National Health Service, BBC Monitoring and the Ford Motor Company. Duncan has a BSc in Plant Sciences and an MSc in Computer Science, both from the University of Manchester. *Scientific Community Worker for Life Sciences.*

Managers

June Finch is the Project Manager of the ^{my}Grid project at Manchester. As part of her role, she also sits on the OMII-UK Operations committee and is responsible for management of the Manchester branch of OMII-UK. Prior to employment at the University of Manchester, June worked as a Project Manager and Software Engineer in the industrial sector in the UK. She has over ten years experience of software delivery in the telecomms industry.

Ed Zaluska is a Senior Lecturer in the School of Electronics and Computer Science, based in the Intelligence Agents Multimedia Group. A member of the ECS team in CombeChem, in which he had a coordination role, his interests include advanced computer architectures and distributed computing systems. He has published in parallel and distributed computing and e-Science. Ed will continue his management role in this project, coordinating the liaison between the activities in ECS and Chemistry

Investigators

Rob Procter is Research Director of the ESRC-funded National Centre for e-Social Science (NCeSS). His role at NCeSS focuses on developing Centre research strategy, coordinating development of applications of e-infrastructure and services in social sciences, and investigating usability issues which may influence wider adoption. He is a member of the EPSRC e-Science Strategic Advisory Team, the JISC VRE and e-Infrastructure Programme Advisory Boards, the eSI Scientific Advisory Board, e-Science Usability Task Force, e-Science User Group, OMII-UK User Group and the AHRC ICT Programme Steering Committee.

Robert Stevens is a Senior Lecturer in Bioinformatics in the School of Computer Science at the University of Manchester. His principle interests are in the development and use of formal ontology to both describe and analyse biological data, and the use of eScience tools in the transformation of bioinformatics to an industrial scale. Robert is the PI at Manchester of the myTea project on which some of the ideas of myExperiment are based, and is a CI and key member of the ^{my}Grid team, driving the Life Science community activities. He is a CI of the OMII-UK project for ^{my}Grid.

Carole Goble is Professor of Computer Science at Manchester. She is the Director of the EPSRC ^{my}Grid project, the chair of OMII-UK and the co-director of the e-Science North West regional centre. She is also a CI on the myTea project. Her principle interests are in knowledge based systems for applications and middleware, and she has an international reputation in the Semantic Web, e-Science, Grid Computing and the Semantic Grid. She works within an application context, most notably in the Life Sciences. She sits on the BIRN-CC EAB.

David De Roure is a Professor of Computer Science at Southampton, where he is a Director of the Pervasive Systems Centre and PI of OMII-UK Southampton. He led the CS team in CombeChem and associated projects, and was Co-Director of Southampton Regional e-Science Centre. He has been involved in the EPSRC, ESRC and AHRC e-Science programmes and is a member of the Steering Group of the Open Grid Forum, where he also leads the Semantic Grid activities in conjunction with Carole Goble.

Jeremy Frey is a Reader in the Department of Chemistry. His research in physical chemistry is in the application of laser techniques to the investigation of molecular properties. He has been involved in a number of inter- and multi-disciplinary projects. He is the PI of the CombeChem EPSRC e-Science project on the application of e-Science and Grid technologies to chemical structure and property prediction and participates in associated JISC projects.

Jonathan Essex is a Reader in the School of Chemistry at the University of Southampton, having been a Royal Society University Research Fellow from 1994-2002. His research interests lie in the application of classical computer simulations to problems of organic & biochemical interest. He is co-investigator on a number of large-scale collaborative grants, including the BioSimGrid and IntBioSim e-science projects funded by the BBSRC, the comb-e-chem project supported by the EPSRC, and the “4 billion bases a day” Basic Technology project. He has been responsible for the JISC Schools Malaria project (e-Malaria) with Jeremy Frey.

Steve Pettifer is a lecturer in Computer Science and member of the Advanced Interfaces Group. His research interests include distributed systems, visualisation, computer graphics, virtual environments, human computer interaction and usability. He recently led the technical component of the EPSRC/DTI funded UTOPIA project, which has produced user friendly integrated analysis tools in the field of protein bioinformatics, and he is currently an investigator on the EMBRACE EU FP6 Network of Excellence which aims to provide coherent access to a suite of biological services around Europe.