The Evolution of myExperiment

David De Roure[1,3], Carole Goble[2], Sergejs Aleksejevs[2], Sean Bechhofer[2], Jiten Bhagat[2], Don Cruickshank[3], Paul Fisher[2], Nandkumar Kollara[3], Danius Michaelides[3], Paolo Missier[2], David Newman[3], Marcus Ramsden[3], Marco Roos[4], Katy Wolstencroft[2], Ed Zaluska[3] and Jun Zhao[5]

[1] Oxford e-Research Centre
University of Oxford
Oxford, UK
dder@oerc.ox.ac.uk

[2] School of Computer Science
The University of Manchester
Manchester, UK

[3] Electronics and Computer Science
University of Southampton
Southampton, UK

[4] Leiden University Medical Centre and University of Amsterdam
The Netherlands

[5] Department of Zoology
University of Oxford
Oxford, UK

Abstract—**The myExperiment social website for sharing scientific workflows, designed according to Web 2.0 principles, has grown to be the largest public repository of its kind. It is distinctive for its focus on sharing methods, its researcher-centric design and its facility to aggregate content into sharable 'research objects'. This evolution of myExperiment has occurred hand in hand with its users. myExperiment now supports Linked Data as a step toward our vision of the future research environment, which we categorise here as'3rd generation e-Research.**

Keywords – scientific workflow, repository, Linked Data, scholarly communication, e-Research

I. INTRODUCTION

e-Science and e-Research are concerned with the future research environment. Scientific workflow systems [1] have emerged as a key part of this environment, supporting systematic data processing to handle a data deluge in a way that can be recorded, repeated, reproduced, reused and repurposed. The myExperiment social website was conceived to help researchers discover, share and publish workflows, addressing a gap in the scholarly knowledge cycle as researchers need to work with new forms of digital artifact that drop into the tooling of e-Research.

myExperiment has grown both in content and capability since its 2007 launch. It is in routine use by users and developers of workflows, particularly in bioinformatics [2] and increasingly in other disciplines from chemistry to social science and digital humanities. It has gained in the volume and diversity of its content and with over 1000 workflows it now represents the largest public repository of its kind.

To set out the ambitions for myExperiment we define three generations of the future research environment or "e-laboratory":

- **1st Generation**. The current practices of early adoptors of software tools. Characterised by researchers using tools within their particular problem area, with some reuse of tools, data and methods within the discipline. Traditional publishing is supplemented by publication of some digital artefacts like workflows and links to data. Provenance is recorded but not shared or reused. Science is accelerated and practice beginning to shift to emphasise *in silico* work.
- **2nd Generation**. The emerging e-Research practice. The key characteristic is reuse of the increasing pool of tools, data and methods across areas/disciplines. We see some freestanding, recombinant, reproducible 'research objects' and provenance analytics plays a role. New scientific practices are established and opportunities arise for completely new scientific investigations.
- **3rd Generation** The solutions we are developing now, characterised by global reuse of tools, data and methods across any discipline, and surfacing the right levels of complexity for the researcher. Radical sharing is key**.** Research is significantly data driven and we see increasing automation and decision-support for the researcher as the environment becomes assistive. Provenance assists design, and curation is both social and automated.

Early workflow systems were first generation, while myExperiment exemplifies the second generation and is evolving to the third. This evolution is the subject of this paper. It is not entirely in the hands of the technology: we fundamentally view the research environment as a socio-technical system, so this is a process of co-evolution with our users and is sympathetic with the design patterns of Web 2.0. myExperiment is therefore itself an experiment in creating an environment to support e-Research, with due attention to the social aspects of research practice, and how people use it is an important insight into future practice.

Successful uptake of new functionality in the research environment requires ease of use and return on investment: with the appropriate tooling we have an "intellectual access ramp" which helps researchers to engage in a graduated way as befitting their needs. We also need ease of assembly or configuration of the environment itself; i.e. an access ramp for the developers and research technologists who support the researchers. Both aspects are considered in this paper.

This paper updates our earlier presentation on the design of myExperiment [3]. It focuses on the co-evolution towards the third generation and it reflects on the design principles so that others may benefit from our experience. We present a user perspective in Section II, showing the progress in the use, content and functionality of the site. This is followed in Section III by a discussion of the move to Linked Data as part of realising our 3rd generation vision. We discuss the design principles in Section IV.

myExperiment is funded by JISC and the Microsoft Technical Computing Initiative, and is part of the myGrid and e-Research South consortia supported by the UK Engineering and Physical Sciences Research Council (EPSRC).

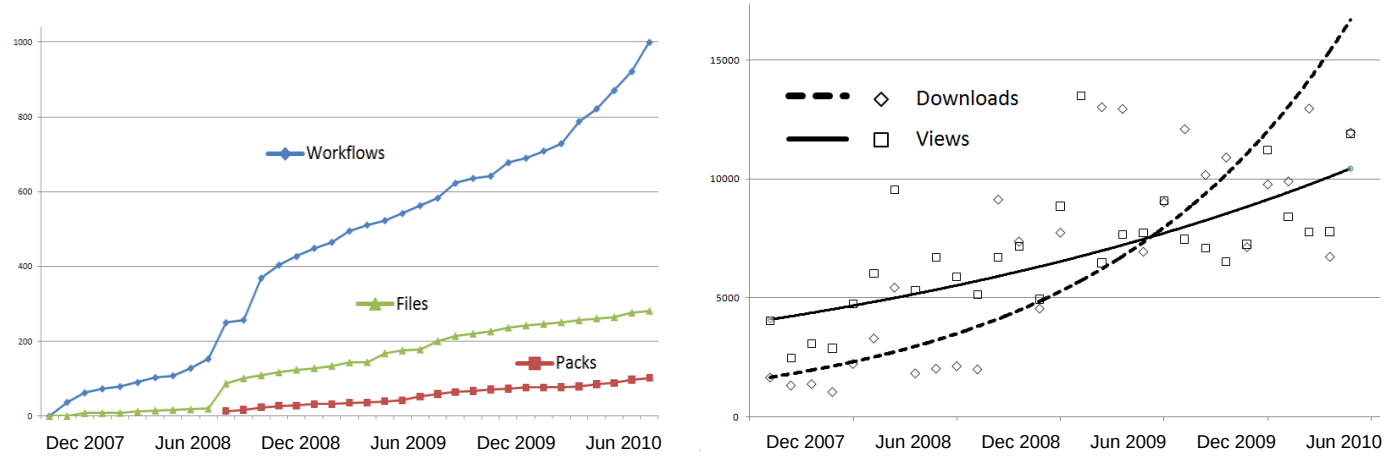

Figure 1. (a) (left) Growth in the content of the myExperiment repository; (b) (right) monthly views and downloads over a 30 month period to June 2010. The increasing number of downloads relative to views results from the growing use of alternative interfaces.

II. MYEXPERIMENT IN USE

With these ambitions in mind, we have addressed four important objectives in this phase of myExperiment:

1. Ease of discovery of workflows as content increases in scale and diversity. This is the key proposition for scientists – some may move on to use other features but the first reason for using the myExperiment "access ramp" is often workflow discovery;
2. Bringing myExperiment functionality into the researcher's work environment by supporting alternative interfaces. We facilitate adoption by making it easy to augment the current work environment rather than obliging the researcher to "come to us".
3. Integrating myExperiment with other tools and services in the emerging environment. This addresses the assembly challenge and is a step towards identifying the services that will underly the future research environment;
4. Exploring and anticipating emerging research practice, and thereby evolving myExperiment towards the 3rd generation e-Laboratory based on insights from its use and its users.

We set the scene by looking at the growth of content and then provide examples of addressing the first three of these challenges. The fourth is discussed in Section III where we consider Linked Data.

A. *Use of Content*

myExperiment has adopted the Web 2.0 approach of supporting one content type particularly well – like photos on Flickr or movies on YouTube. For this reason we focused on scientific workflows and, within that realm, we commenced by targeting particular workflow systems. In particular we recognised that Taverna [4] has a widely distributed user community with both need and incentive to share.

There are now nearly 30 different workflow types shared on myExperiment, ranging from Taverna, Project Trident [5], Meandre [6] and Bioclipse [7] to SPARQL queries, spreadsheets and makefiles. The extent of the custom support for a particular type ranges from automatic thumbnail generation to workflow enactment.

One of our original motivations for myExperiment was to bring workflows into the scholarly knowledge cycle. As a registry of workflows, myExperiment enables people to cite a persistent URI to refer to a particular workflow entry (a good example of this is [8] in which several workflows are cited in the references section of the paper). Another was to provide a basis for training, and this is evidenced through the collection of tutorial workflows. As well as research workflows there are benchmarks and test workflows used by workflow system developers.

In addition to workflows, myExperiment supports files and 'packs'. Packs were introduced because users wished to attach supplementary items to a workflow, such as example input and output data, papers and slides – they describe aggregations of content which could be inside or outside myExperiment, and they can be shared as first class objects. A typical pack might contain all the pieces associated with a given experiment or publication, or a workflow with example input and output data so that it can be tested. This secondary role as a registry of aggregations has become an important part of myExperiment's integration with other repositories, such as EPrints [9]. Packs are exported using the Object Reuse and Exchange (ORE) representation from the Open Archives Initiative [10].

The growth of the myExperiment content is shown in Figure 1, which depicts (a) user contributed content that is publicly available, and (b) monthly downloads of workflows (examined further in C below). The social network has also grown: the top 10 user networks (groups) have between 18 and 57 members. myExperiment also acts as a lens onto what people are sharing, and we note that the nature and functionality of the shared items has also evolved. The increasing use of workflows that make use of SPARQL and Linked Data are part of the motivation for the Linked Data support that we discuss in Section III.

Some other workflow systems support the idea of a repository, such as Kepler [11] which provides centralised

repository access from within the workflow system. Project Trident uses myExperiment as its community repository. Meandre provides a notion of repositories and is additionally supported within myExperiment to both share and execute workflows. The workflow collection in myExperiment has itself provided a basis for several studies; e.g. [12-14].

B. *Discovery and curation*

This increasing volume and breadth of content means that greater support is required for workflow discovery. We have introduced filters to refine the workflow display. For familiarity the design is inspired firstly by shopping sites – we are, after all, supporting people shopping for workflows – as well as other interfaces familiar to this community such as online library interfaces. We support filtering on workflow types, tags, authors and curation categories. By making authors' names a facet in this interface we tie into the myExperiment social network: this also serves to boost visibility of workflow contributors, thus contributing to our reward and incentive structure.

Significantly, the content also exhibits a wide spectrum in the quality and reusability of the contributions. Best practice is demonstrated by popular workflows and we have also created a set of reference workflows. While popular content 'floats to the surface', we found it necessary to deal with incomplete content, such as people creating test content when trying the site for the first time and making experimental use of features.

We have addressed this through support for curation. Our approach acknowledges the role of the expert curator whilst respecting the "wisdom of the crowd": we provide additional support for users with curator status to add curation tags (e.g. Requires example, Requires description, Runnable, Obsolete, Test workflow, Example data, Not runnable, Tutorial / example, Whole solution, Component).

From the outset we only obliged users to enter minimal structured metadata about contributions to myExperiment, partly because we did not wish to impose barriers that would deter contributors and also because, given the diverse and non-prescriptive use of the site, we did not have common structures and vocabularies for all kinds of contributed object. The downside of this is of course a deficiency in categorised metadata to facilitate discovery and reuse.

Ease of contribution versus quality of metadata is an important equilibrium and our approach to this problem is threefold: introduction of templates and controlled vocabularies for contribution types, provision of feedback mechanisms to encourage users to provide comprehensive metadata, and greater automated assistance in recommending metadata. We also note that there are multiple opportunities to assign metadata during the lifecycle of the object, not just at upload time and not just in the myExperiment interface.

C. *Alternative Interfaces*

In order to facilitate developers in creating alternative interfaces, and integrating myExperiment functionality into other environments, the REST API was provided early in the evolution of the site, complete with interactive documentation, examples and a test server. Several interfaces have been built including Google gadgets, two facebook applications, a Silverlight interface and an Android interface, as well as integration with Windows 7 and with twitter.

The most widely used alternative interface is the plugin to Taverna Workflow Workbench, shown in Figure 2, which enables the user to access myExperiment content without leaving the Taverna environment, providing tabs for MyStuff, Tag Browser, Search, History and access to the "starter pack" of Taverna workflows on myExperiment. This interface now comes prepackaged in a new Taverna installation.

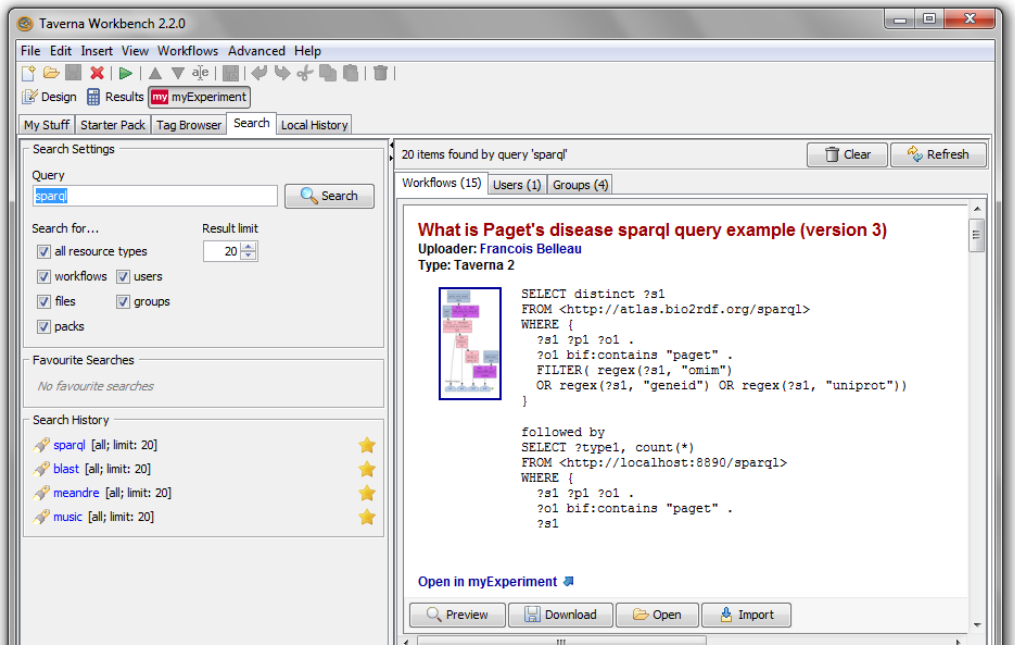

Figure 2. myExperiment functionality in the Taverna T2 workflow workbench, integrated using the REST API. The workflow shown is a Taverna workflow that uses a SPARQL query.

The increase in downloads relative to views in Figure 1(b) is a result of increasing use of interfaces that lead directly to a download. This emphasises the difficulties in collecting statistics when a variety of interfaces are in use which may cache views and workflows in different ways, and although we eliminate 'bots' it is difficult to identify 'genuine downloads' reliably and consistently. This is an important problem because usage figures provide important feedback and help build reputation and incentive.

D. BioCatalogue integration

BioCatalogue [15] is a sister project to myExperiment that provides a registry of Web Services in the Life Sciences (www.biocatalogue.org). It is built to the same design principles and draws closely on the myExperiment experience, with community curation of content. It brings together service providers, service consumers, expert curators and tool developers, encouraging annotation and curation by all. Web Services (and their various operations, endpoints, inputs and outputs) are described in detail and are constantly monitored for availability and changes to their programmatic interface.

These are powerful tools in combination and we are working towards a rich symbiosis between myExperiment and BioCatalogue:

- The metadata about Web Services and their service status information available from BioCatalogue can be made available to myExperiment users to assist in workflow selection;
- The collection of workflows on myExperiment provides information about services that are used and the interconnections between services.

To achieve the first part, we have introduced a Web Services tab to myExperiment (with Latest Services, Updated Services, etc.) which links through to BioCatalogue, a mechanism to harvest service descriptions and support for searching services. Workflow descriptions also link through to the BioCatalogue website, and we show "similar workflows" based on services. The index of Web Services harvested into myExperiment is illustrated in Figure 3.

The "similar workflows" functionality is an example of a range of recommendation features which are under development. These make use of similarity measures based on the descriptions and tags of the contributions, using Latent Semantic Analysis, as well as the social network. We also plan to recommend workflows and services based on the types of input and output data as the integration evolves.

III. LINKED DATA

The workflows and packs on myExperiment give an important insight into future research practice and to the combinations of external resources that researchers are using. Our final objective is the co-evolution towards the third generation environment, characterised as radical reuse and exemplified here by myExperiment's support for Linked Data.

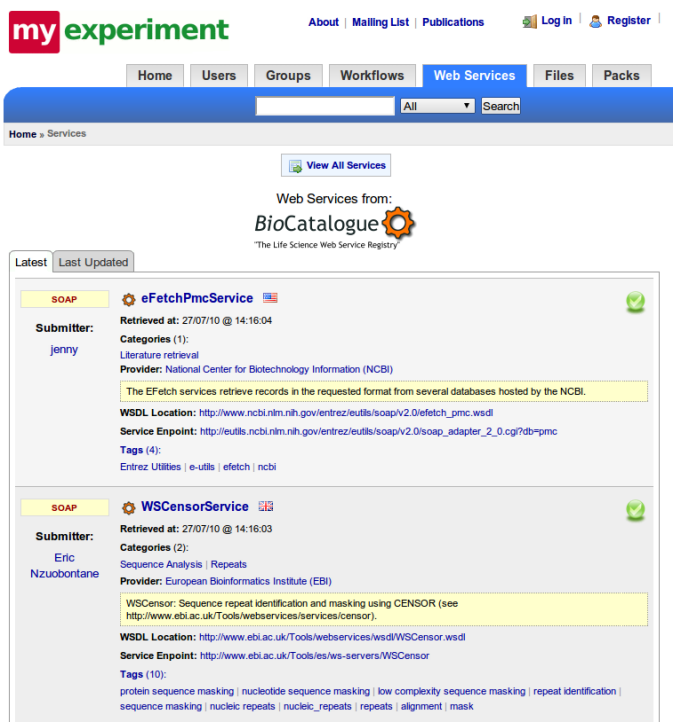

Figure 3. The services tab in myExperiment showing service descriptions from the BioCatalogue.

A. Use cases

The Linked Data movement [16] is gaining significant traction in research, with important data sources increasingly available in this format. Early adopter domains include life sciences, social sciences (notably through the Open Government Data initiatives) and digital humanities.

We increasingly see workflows in myExperiment which work at this level of abstraction. Workflows that use SPARQL endpoints as data sources and which populate triplestores on the fly have been shared for over a year (notably those of Francois Belleau, one of which appears in Figure 2). This applies not just to the data resources but to the services in the research environment, such as repositories.

Hence we have explored how myExperiment itself fits into the Linked Data environment that these researchers are using. We illustrate this with two use cases which use myExperiment's Linked Data support to answer research questions in two different domains: computational musicology and bioinformatics.

In our first example, Page et al [17] have developed an operational proof-of-concept system in the music information retrieval domain that demonstrates the utility of linked data for enhancing the application of workflows. It integrates:

1. An Audio File Repository which serves digital audio 'signal' and publishes a small RDF sub-graph describing each locally stored audio file as linked data;

2. A Collection Builder that enables a researcher to select a set of signals described by linked data services then publish the collection as RDF;
3. A Meandre workflow, stored and executed on myExperiment, for music genre classification. The workflow accepts RDF published by the Collection Builder, dereferences resources from the Audio File Repository and runs the classification algorithms;
4. A Results Repository which publishes the analysis output as linked data.

The demonstrator (known as "How country is my country?") is shown in Figure 4, which also illustrates that the Linked Data tooling is hidden behind the scenes of the researcher's interactive interface. Further information can be found on www.nema.ecs.soton.ac.uk.

In a second example, Roos and colleagues have conducted a "proof of principle" in the bioinformatics domain using multiple resources [18]. As well as illustrating the needs of a real investigation, this example shows the complexity of the method that needs to be captured for reuse and reproducibility. They integrate:

1. Taverna provenance records exposed as RDF;
2. A myExperiment RDF document for a protein discovery workflow;
3. A mocked-up BioCatalogue document using myExperiment RDF data as example;
4. Provisional RDF documents obtained from the ConceptWiki (conceptwiki.org) development server;
5. An RDF document for an example protein, obtained from the RDF interface of the UniProt web site.

These use cases show research occurring at a new level of abstraction over the tooling, in which myExperiment is an integrated part of the Linked Data research environment. They illustrate 3rd generation behaviour, assembling resources into the environment and supporting automation.

B. *Supporting Linked Data*

myExperiment's Linked Data support comes at two levels: a SPARQL service endpoint allowing querying to data hosted by myExperiment, and a Linked Data interface. The SPARQL endpoint is a web service that implements the standard RDF query protocol, and through the Linked Data support users can retrieve RDF descriptions about every type of myExperiment entity, be it a workflow, a pack, a user or a group. The SPARQL endpoint (http://rdf.myexperiment.org/sparql) was released in 2009 and immediately attracted usage in the myExperiment user and developer community, the Semantic Web community and indeed in the broader myExperiment team where it has become an essential utility in site maintenance and reporting. SPARQL queries are shared on myExperiment itself.

The significant point about the Linked Data support is that it provides a common interface over multiple repositories, enabling the same tooling to be applied without enforcing any prior agreement between those sites. For example, we have produced specialist code to integrate myExperiment and EPrints, but the Linked Data approach can provide this integration and with a wide variety of other repositories too. This was the basis of a presentation at Open Repositories 2010 [19] in which we demonstrated use of a Linked Data browser to view myExperiment content and navigate to other repositories.

This "human in the loop" approach makes a point but is not the end goal: we anticipate greater use of Linked Data tooling as the Linked Data community shifts focus from production to use. Already we are seeing benefits of making metadata available in this way, as other Linked Data services are now aware of myExperiment. For example, myExperiment can be discovered through public voiD (Vocabulary of InterLinked Datasets) stores [20] used by any Linked Data query federation engines, and querying a voiD store would identify multiple instances of myExperiment and related servers – these could include annotation servers, perhaps based on the Open Annotation Collaboration (with which myExperiment's own annotation model is consistent) [21].

Supporting Linked Data involves implementation of a consistent URI scheme with content negotiation, publication of data as RDF and preferably a SPARQL endpoint query interface [22]. The myExperiment SPARQL endpoint provided the latter two capabilities first. This was achieved by creating a separate server with its RDF data synchronised to the public content of the myExperiment server, and published in RDF according to the myExperiment ontology which draws as far as possible on existing ontologies including Dublin-core, Friend of a Friend (FOAF) and Semantically Interlinked Online Communities (SIOC) [23].

myExperiment has always had persistent URIs of the form http://www.myexperiment.org/workflows/15 – it is these which appear in publications and emails. Linked Data recommends data publishers to indicate the representations of a resource in its URIs, for example, using URIs like http://www.myexperiment.org/workflows/data/15 to indicate information about workflow 15 represented in RDF format. In order to be backward compatible with existing myExperiment URIs, we choose the following scheme in myExperiment:

- *http://www.myexperiment.org/workflows/{identifier}* to identify a workflow, a non-information resource;
- *http://www.myexperiment.org/workflows/{identifier}.html* to identify information about a workflow represented in HTML format;
- *http://www.myexperiment.org/workflows/{identifier}.rdf* to identify information about a workflow represented in RDF/XML format, that can be consumed by a Linked Data browser or a query engine.

The multiple representations of the data about a workflow can be retrieved through HTTP content-negotiation.

The benefits of supporting Linked Data have been discussed above, but here we see two of the costs. The first is a usability concern – if people bookmark or exchange the URI in the browser then this refers to the HTML representation and not the non-information resource. We have addressed this by

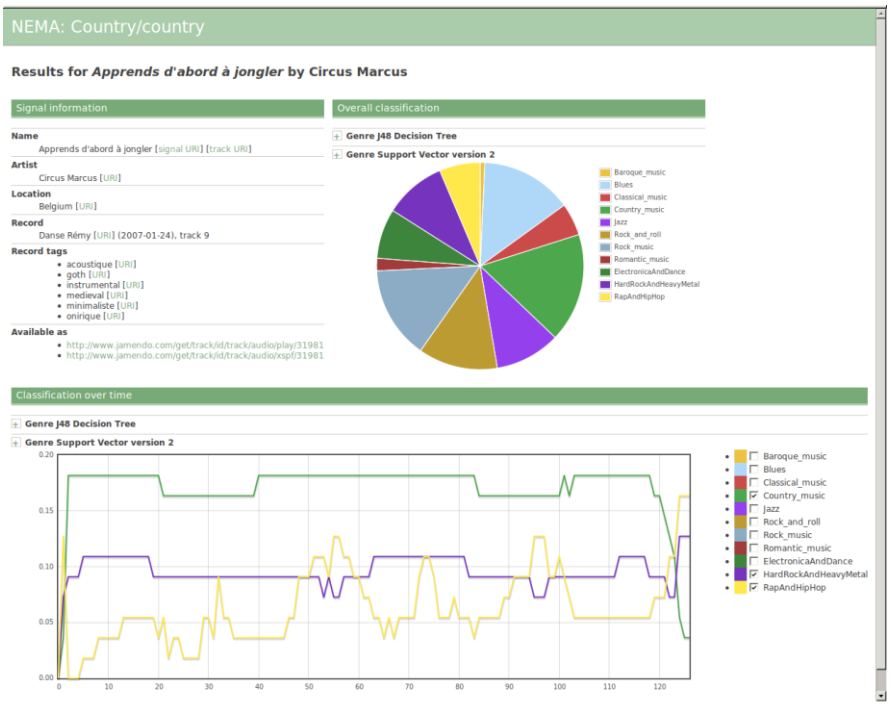

Figure 4. Integration of myExperiment into the research environment based on Linked Data. Here the Result Viewer web application shows analysis for a music collection (top) and music genre weightings over time for a specific Signal (bottom).

providing extra links in the page to access the various representations. Furthermore, we permit users to copy URIs from their web browser and get a useful response when they paste it into a Semantic Web application (by redirecting via the non-information resource when a .html URI gives rise to a content type mismatch).

Secondly, Content-negotiation involves an additional round-trip and in certain circumstances this could be a significant cost. By writing redirect rules as part of the Apache configuration, the myExperiment Rails codebase does not have to handle content negotiation. Performance tests have shown only very minor increases in response time compared to before.

In general we found that current Linked Data practice is very much focused on publishing and not so much on consuming. Although this is a logical order, it means publishing practice is not yet fully informed by cases of use.

Packs are published as Linked Data using ORE. At the moment, a Pack consists of a set of components with metadata describing how they are each related to the pack itself. There is a clear need to be able to express the relationships between individual items, and for the relationships to draw upon controlled vocabularies. We are developing the user interface for presenting and describing these relationships. We also have use cases for large and complex packs which will be created programmatically but require visualisation in the Web interface.

Another aspect of Linked Data is representation of provenance. Workflows and Packs themselves provide valuable assistance with understanding the provenance of results, assisting with interpretation trust and reuse. myExperiment also provides some socially-maintained provenance information for workflows themselves through credit and attribution. Publishing this information is a first step; we anticipate development of provenance analytics tools that consume it to support the researcher.

IV. REFLECTION

We have previously described myExperiment as a "Social Virtual Research Environment" [3] and as other environments and sites adopt a Web 2.0 approach we expect the principles illustrated in myExperiment to become more widespread. Here we reflect on our experience in myExperiment and some adjacent projects as we evolve from the 2^{hd} to 3rd generation: firstly on what we have built, and then how we have built it. We consider both the researcher and developer "ramps".

A. *What we have built*

Part of the "experiment that is myExperiment" was the question as to whether researchers would share sufficiently – the assumption is that successful Web 2.0 sites are predicated on this behaviour. Our usage shows that sharing does indeed occur (an analysis can be found in [24]). It also shows different sharing behaviours in different communities. The SysmoDB project, which builds on myExperiment to support sharing of data, models and experimental protocols in systems biology, is an excellent example of addressing data sharing from a "social VRE" approach (www.sysmo-db.org).

We focused first on workflows – on the specific before the generic – though the site could be used to share all sorts of objects. We have retained our focus on sharing methods and thereby sharing know-how and building capacity. Within the

context of e-Research this makes a powerful point, that there is pervasive data collection and attention to data curation but methods do not get the same attention [25]. Researchers are developing techniques to cope with a deluge of data and we believe that these should be shared and curated also. This is exemplified by the MethodBox project which builds on myExperiment to share statistical methods for epidemiology and public health research (www.methodbox.org).

Through our focus on how researchers work today and will work in the future, myExperiment has gone on to provoke discussions about not just how people share but *what* they will be sharing. Will our evolved Packs be the shared digital artefact of future research? While others approach this by looking at the evolution of the academic paper [26], we are coming at this from "what is the shared digital artefact?" [27].

myExperiment's move to Linked Data is very much part of this story. Instead of a repository which is inward-looking, myExperiment is contributing to the Linked Data web – not just content but functionality, as a community-curated registry of workflows and aggregations. Equally, other Linked Data tools and services (e.g. coreference resolution and open annotation) add value to myExperiment without any extra effort. The latest contributions to myExperiment demonstrate that myExperiment is providing methods for Linked Data – what we might call 'Linked Open Methods'.

B. How we built it – our design principles

The design principles of Taverna and myExperiment are presented in [28]. Here we review the myExperiment design against the Web 2.0 principles [29] in order to examine their relevance in the move to the third generation research environment. This is important because it is the first consideration of Linked Data in the context of this design framework.

1) The Long Tail

In myExperiment we see two aspects of the Long Tail. Firstly we are directly supporting the tail of the distribution of research practitioners and not just a few large players [30] – we support "long tail science" and also the communication between this and "Big Science". Our second long tail is in the distribution of web sites, since myExperiment Packs reach out to anywhere on the Web and not just a few large repositories. Linked Data emphasises the tail, as we are already witnessing a growing number of Linked Data resources.

2) Data is the Next Intel Inside™

myExperiment has demonstrated the value of doing one content type well and focusing first on the specific (workflows, starting with Taverna) rather that the generic (sharing arbitrary contributions). Making a small number of researchers happy first is more likely to lead to adoption and practice that can be translated to others – the myExperiment codebase could share any sort of contribution but we do not attempt to do that. Hence we become an authoritative Linked Data source.

3) Users Add Value

As a site of user-generated content, all the value is added by the users. However, research content is different to photos and movies: it has specialist application and there are not yet universal 'players' for our content. The challenge then is to make the content as reusable as possible. Our proactive curatorship model is part of this (corresponding loosely to the notion of editors on Wikipedia) as well as our social and assistive approaches to improve structured metadata.

4) Network Effects by Default

Capturing usage information adds value to the site by providing feedback and as a basis for recommendations. Although download figures have proven to be problematic because of the variety of clients in use and programmatic access, we see effects in our content due to its richly linked nature; for example, our similar workflows recommendations come from the content itself (workflows interlinked by services) and analysis has revealed a similar interconnectedness of content in packs. These intrinsic effects in the content will be enhanced by the BioCatalogue integration. Furthermore, Linked Data opens new scope for network effects as our content interlinks with the wider web.

5) Some Rights Reserved.

The site facilitates the use of creative commons licensing and makes it easy to make content publicly available, but it is an important principle that we do not mandate this: rather, researchers have full control over privacy and licensing. In this respect we differ from other open science sites like OpenWetWare (openwetware.org). This absolutely reflects our users, some of whom are deterred by the idea of Web 2.0 simply because they believe this implies everything is open. It is important to note our distinction between discovery and acquisition; e.g. Linked Data can help discover a workflow and then obtaining permission for use may follow.

6) The Perpetual Beta

Running an agile website is completely different to managing software releases that need to be installed at the client end, crucially because it enables a rapid cycle of co-design with a diverse base of users, both researchers and developers. Behind the scenes there are multiple virtualised myExperiment servers – for development, testing new features and providing a sandbox for programmatic use – so that the team can be very responsive to requests and move rapidly from test to deployment of new functionality.

7) Cooperate, Don't Control

This is the single most important principle in the myExperiment design. It absolutely underpins our alternative interfaces, integration with other services and the move to Linked Data. myExperiment makes itself as reusable as possible (e.g. through the REST API and SPARQL endpoint) and makes use of other services as much as possible. The BioCatalogue integration is a good example of symbiosis rather than reinvention. This principle underlies the research user ramp and the developer ramp.

8) Software Above the Level of a Single Device

This is consistent with our approach to alternative interfaces discussed above: our users often require bespoke, task-specific interfaces. With respect to devices, the Android interface was an excellent exercise in rethinking the myExperiment interface in the context of the different modes of use and interactive capability. Generally, by providing notifications we can also interact through twitter or RSS feeds.

V. CONCLUSION

The evolution we have discussed in this paper is effectively a co-evolution of myExperiment with its research users. The "experiment that is myExperiment" has led to a novel repository which acknowledges the primacy of method and a community social network of people and interlinked artefacts of digital research. It has demonstrated that researchers do share, and it has brought new digital artefacts into the scholarly knowledge lifecycle. It provides an "intellectual access ramp" both for research users and developers.

We have illustrated the relevance of the Web 2.0 design principles in the context of e-Research as we evolve to the third generation research environment. We have also observed a significant design synergy with Linked Data, which truly meets Web 2.0 in the "cooperate, don't control" paradigm: it is also inherently data-centric, leverages the long tail, benefits from open licensing for mashing and remixing, and enables network effects in the content to flourish.

ACKNOWLEDGMENT

The authors acknowledge the input of all the myExperiment users, especially the "friends and family" who help with design and testing, and our collaborators including Andrea Wiggins and Ravi Madduri. The characterisation of three generations of e-Laboratories is due to Iain Buchan. Thanks to the Taverna and e-Labs teams, to Hugh Glaser, Ian Millard, Chris Gutteridge and Les Carr for their many helpful discussions about Linked Data, and to Kevin Page, Ben Fields and Scott Marshall for helping provide the use cases.

REFERENCES

[1] Taylor, I.J., Deelman, E., Gannon, D.B., and Shields, M.: Workflows for e-Science (Springer, 2007. 2007)

[2] Goble, C.A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P., and De Roure, D.: myExperiment: a repository and social network for the sharing of bioinformatics workflows, Nucl. Acids Res., 2010.

[3] De Roure, D., Goble, C., Bhagat, J., Cruickshank, D., Goderis, A., Michaelides, D., and Newman, D.: myExperiment: Defining the Social Virtual Research Environment. Proc. IEEE Fourth International Conference on eScience, Indianapolis, 7-12 December 2008 pp. 182-189

[4] Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M.R., Li, P., and Oinn, T.: Taverna: a tool for building and running workflows of services, Nucl. Acids Res., 2006, 34, (suppl_2), pp. W729-732

[5] Barga, R., Jackson, J., Araujo, N., Guo, D., Gautam, N., and Simmhan, Y.: The Trident Scientific Workflow Workbench. Proc. Proceedings of the 2008 Fourth IEEE International Conference on eScience2008 pp. 317-318

[6] Llor, X., Ã • cs, B., Auvil, L.S., Capitanu, B., Welge, M.E., and Goldberg, D.E.: Meandre: Semantic-Driven Data-Intensive Flows in the Clouds. Proc. Proceedings of the 2008 Fourth IEEE International Conference on eScience2008 pp. 238-245

[7] Spjuth, O., Alvarsson, J., Berg, A., Eklund, M., Kuhn, S., Masak, C., Torrance, G., Wagener, J., Willighagen, E., Steinbeck, C., and Wikberg, J.: Bioclipse 2: A scriptable integration platform for the life sciences, BMC Bioinformatics, 2009, 10, (1), pp. 397

[8] Kuhn, T., Willighagen, E., Zielesny, A., and Steinbeck, C.: CDK-Taverna: an open workflow environment for cheminformatics, BMC Bioinformatics, 2010, 11, (1), pp. 159

[9] Brody, T., Carr, L.A., and Tarrant, D.: From the Desktop to the Cloud: Leveraging Hybrid Storage Architectures in Your Repository, in From the Desktop to the Cloud: Leveraging Hybrid Storage Architectures in Your Repository (Georgia Institute of Technology, 2009, edn.).

[10] Lagoze, C., Sompel, H.V.d., Nelson, M., Warner, S., Sanderson, R., and Johnston, P.: A Web-based resource model for scholarship 2.0: object reuse & exchange, Concurrency and Computation: Practice and Experience, DOI 10.1002/cpe.1594

[11] Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B., and Mock, S.: Kepler: An Extensible System for Design and Execution of Scientific Workflows. Proc. Proceedings of the 16th International Conference on Scientific and Statistical Database Management2004 pp. 423

[12] Stoyanovich, J., Taskar, B., and Davidson, S.: Exploring repositories of scientific workflows. Proc. Proceedings of the 1st International Workshop on Workflow Approaches to New Data-centric Science, Indianapolis, Indiana2010 pp. 1-10

[13] Groth, P., and Gil, Y.: Analyzing the Gap between Workflows and their Natural Language Descriptions. Proc. Proceedings of the 2009 Congress on Services - I2009 pp. 299-305

[14] Wassink, I., Vet, P.E.v.d., Wolstencroft, K., Neerincx, P.B.T., Roos, M., Rauwerda, H., and Breit, T.M.: Analysing Scientific Workflows: Why Workflows Not Only Connect Web Services. Proc. Proceedings of the 2009 Congress on Services - I2009 pp. 314-321

[15] Bhagat, J., Tanoh, F., Nzuobontane, E., Laurent, T., Orlowski, J., Roos, M., Wolstencroft, K., Aleksejevs, S., Stevens, R., Pettifer, S., Lopez, R., and Goble, C.A.: BioCatalogue: a universal catalogue of web services for the life sciences, Nucl. Acids Res., 2010, 38 (suppl_2), pp. W689-94

[16] Bizer, C., Heath, T., and Berners-Lee, T.: Linked Data - The Story So Far, International Journal on Semantic Web and Information Systems (IJSWIS), 2009, 5, (3), pp. 1-22

[17] Page, K.R., De Roure, D., ONeill, G., Nagel, B.J., Crawford, T., and Fields, B.: Semantics for music analysis through linked data: How country is my country?. IEEE e-Science, Brisbane, Australia 2010.

[18] Roos, M., Bechhofer, S., Zhao, J., Missier, P., Newman, D., De Roure, D., and Marshall, M.S.: A Linked Data Approach to Sharing Workflows and Workflow Results. Proc. ISoLA 2010 - Tools in Scientific Workflow Composition, Crete, October 2010

[19] De Roure, D.: Repositories and Linked Open Data: the view from myExperiment. Proc. Open Repositories, Madrid, Spain, July 6 2010

[20] Alexander, K., Cyganiak, R., Hausenblas, M. And Zhao, J. voiD Guide - Using the Vocabulary of Interlinked Datasets. http://rdfs.org/ns/void-guide

[21] Sanderson, R., and Sompel, H.V.d.: Making web annotations persistent over time. Proc. Proceedings of the 10th annual joint conference on Digital libraries, Gold Coast, Queensland, Australia2010 pp. 1-10

[22] Bizer, C., Heath, T., and Berners-Lee, T.: Linked Data - The Story So Far, International Journal on Semantic Web and Information Systems (IJSWIS), 2009

[23] Newman, D.R., Bechhofer, S., and De Roure, D.: myExperiment: An ontology for e-Research. Proc. Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), Washington DC, USA, 26 October 2009

[24] De Roure, D., Goble, C., Aleksejevs, S., Bechhofer, S., Bhagat, J., Cruickshank, D., Fisher, P., Hull, D., Michaelides, D., Newman, D., Procter, R., Lin, Y., and Poschen, M.: Towards open science: the myExperiment approach, Concurrency and Computation: Practice and Experience, DOI 10.1002/cpe.1601

[25] De Roure, D., and Goble, C.: Anchors in Shifting Sand: the Primacy of Method in the Web of Data. Proc. WebSci10: Extending the Frontiers of Society On-Line, Raleigh, NC, US, April 26-27 2010

[26] Shotton, D., Portwin, K., Klyne, G., and Miles, A.: Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article, PLoS Comput Biol, 2009, 5, (4), pp. e1000361

[27] Bechhofer, S., De Roure, D., Gamble, M., Goble, C., and Buchan, I.: Research Objects: Towards Exchange and Reuse of Digital Knowledge. Proc. The Future of the Web for Collaborative Science (FWCS 2010), Raleigh, NC, USA, April 2010

[28] De Roure, D., and Goble, C.: Software Design for Empowering Scientists, IEEE Software, 2009, 26, (1), pp. 88-95

[29] O'Reilly, What is Web 2.0? http://oreilly.com/web2/archive/what-is-web-20.html