

# myExperiment: a repository and social network for the sharing of bioinformatics workflows

Carole A. Goble<sup>1</sup>, Jiten Bhagat<sup>1</sup>, Sergejs Aleksejevs<sup>1</sup>, Don Cruickshank<sup>2</sup>,  
Danus Michaelides<sup>2</sup>, David Newman<sup>2</sup>, Mark Borkum<sup>2</sup>, Sean Bechhofer<sup>1</sup>, Marco Roos<sup>3,4</sup>,  
Peter Li<sup>1,\*</sup> and David De Roure<sup>2</sup>

<sup>1</sup>School of Computer Science, The University of Manchester, Manchester M13 9PL, <sup>2</sup>School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK, <sup>3</sup>BioSemantics group, Human Genetics Department, Leiden University Medical Centre, Albinusdreef 2, 2333 ZA Leiden, and <sup>4</sup>Adaptive Information Disclosure group, Informatics Institute, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

Received February 14, 2010; Revised April 24, 2010; Accepted May 6, 2010

## ABSTRACT

**myExperiment (<http://www.myexperiment.org>) is an online research environment that supports the social sharing of bioinformatics workflows. These workflows are procedures consisting of a series of computational tasks using web services, which may be performed on data from its retrieval, integration and analysis, to the visualization of the results. As a public repository of workflows, myExperiment allows anybody to discover those that are relevant to their research, which can then be reused and repurposed to their specific requirements. Conversely, developers can submit their workflows to myExperiment and enable them to be shared in a secure manner. Since its release in 2007, myExperiment currently has over 3500 registered users and contains more than 1000 workflows. The social aspect to the sharing of these workflows is facilitated by registered users forming virtual communities bound together by a common interest or research project. Contributors of workflows can build their reputation within these communities by receiving feedback and credit from individuals who reuse their work. Further documentation about myExperiment including its REST web service is available from <http://wiki.myexperiment.org>. Feedback and requests for support can be sent to [bugs@myexperiment.org](mailto:bugs@myexperiment.org).**

## INTRODUCTION

The deployment of data and tools as web services has gained increasing popularity over recent years (1). Major data providers such as the European Bioinformatics Institute (2), National Center for Biological Information (<http://eutils.ncbi.nlm.nih.gov>) and the DNA Database of Japan (3), as well as specialist research groups (4–7) have utilized the standardized set of web services protocols to provide much needed programmatic access to their computational resources. This has enabled information to be served directly to applications for performing common bioinformatics analyses such as the derivation of summaries, testing of hypotheses and the search for patterns in data.

Such processes can be represented as workflows that define the sequential flow of data through bioinformatics databases and analytical tools involved in a pipeline. A wide variety of workflow languages have been created to describe data processing pipelines enabling them to be stored for repeated use (8). These processes can be constructed using workflow management software such as Taverna (9,10), Pipeline Pilot (11) and Kepler (12) to combine web services with local tools in a graphical fashion for querying, integrating and analysing data. A plethora of workflows have now been written by developers to form a critical mass of knowledge. For example, bioinformatics workflows have been designed for the analysis of microarray data (10), integration of gene expression levels with systems biology models (13), the extraction and structuring of knowledge from text (14) and

\*To whom correspondence should be addressed. Tel: 0161 306 5139; Fax: 0161 306 4556; Email: [peter.li@manchester.ac.uk](mailto:peter.li@manchester.ac.uk)

the identification of genes associated with diseases (15). This has led to a need for workflows to be shared with colleagues and other interested parties. Sharing supports the reuse and repurposing of workflows for other applications and facilitates the building of workflows. The use of workflows also encourages reproducible research, which is an issue that is becoming important across different scientific domains (16).

Motivated by the needs of scientists and inspired by popular social network websites such as Facebook, the myExperiment (<http://www.myexperiment.org>) project has developed an open source Web 2.0 infrastructure that enables scientific artefacts including workflows to be shared within the life sciences community (17). This infrastructure is comprised of a repository of workflows supported by a social networking environment that facilitates workflow sharing and thus the building of new workflows. This article describes how users can interact with myExperiment to discover workflows for reuse in their research, and how they submit workflows for sharing in a manner controlled by their uploader. We address the relationship between myExperiment, workflow management systems and web service registries, such as BioCatalogue (18). These components perform crucial functions at different stages in the life cycle of workflows that could assist in the reproducibility of data-driven experiments when those experiments are reported in the scientific literature (19).

## USERS AND COMMUNITIES IN MYEXPERIMENT

Users are the heart of myExperiment. They may be developers interested in contributing their workflows into the repository for subsequent sharing with the scientific community. Users may also be scientists wishing to discover workflows to be reused in their own research. myExperiment has built a social community around its repository of workflows which facilitates the sharing of workflows between developers and interested parties. In this respect, myExperiment also acts as a training resource providing exemplar workflows and access to expertise to guide users in the orchestration of web services.

The creation of a social community has necessitated the registration of users in order for individuals to be identifiable in myExperiment. While public content can be freely browsed and downloaded by anonymous users, a richer user experience is available upon registration. This is a simple process in myExperiment, requiring the entry of a username and password or an openID URL (<http://openid.net>). An e-mail address is also requested to confirm the registration of a user. Registration leads to a user profile, which can be edited with further information about contact details and research interests. Furthermore, each profile provides listings of friends, workflows and other digital objects in myExperiment belonging to or valued by the user.

Users in myExperiment can benefit from two mechanisms to form communities. Firstly, a user may request friendship from other registered people. Friendship links lead to the building of a network of trusted individuals.

Users can then opt to restrict the sharing of particular workflows and other associated documents to this trusted network. If data security is of the utmost importance for an organization, then system administrators can deploy their own instance of myExperiment within an Intranet. Secondly, communities in myExperiment can be formed by the creation of groups. A registered user can set up a group for which they become the administrator and they can invite other users to join. Other users can also request to join the group. This type of community is designed to allow people who, for example, want to work on the same project, are at the same institution or have the same research interests to share and manage data in their collaboration.

As well as sharing and maintaining content, users have been central to the design of the site. The philosophy is to facilitate users to work as they wish rather than to impose a particular style of interaction. Hence the interactive interface in particular follows a so-called 'perpetual beta' development model, while programmatic interfaces to the site remain stable. The development road map is published on the myExperiment Wiki (<http://wiki.myexperiment.org>).

## WORKFLOW DISCOVERY

New users of workflows will come to myExperiment wanting to query its repository to look for pre-existing workflows which they can make use of in their own research and data analyses. The workflows home page provides the starting entry point for the discovery of workflows (Figure 1). myExperiment is open to any workflow system, supporting the sharing of workflows written in a range of workflow languages. While, at present, the majority are Taverna workflows, myExperiment contains workflows written in 24 other languages. The workflows web page also categorizes the latest, last-updated, most viewed, most downloaded and most favoured workflows.

The discovery of workflows in myExperiment can be performed in three ways. First, a set of workflows can be selected for browsing based on popular tags that have been used to describe them (Figure 1). Secondly, workflows can be discovered using a keyword search. For example, a search using 'BLAST' currently leads to 102 workflows being found by the keyword. Thirdly, they can be found by association with particular users or groups. Each workflow in myExperiment has a dedicated web page showing descriptive information about its inputs, outputs and the operations it makes on data, as well as a graphical representation of the workflow where possible, and information about credit, attribution and licensing. Feedback can be provided on a workflow. This can be in the form of a rating, review and comment, or simply the marking of a workflow as a favourite by a user. Feedback helps a contributor to build up their reputation. In order for feedback to be given, a closer review of a workflow is required and this depends on it being downloadable for inspection and execution so that a user can understand the operations it makes on data.

The screenshot shows the myExperiment website interface. At the top, there is a navigation bar with 'my experiment' logo and links for 'About', 'Mailing List', 'Publications', 'Logout', 'Give us Feedback', and 'Invite'. Below this is a secondary navigation bar with 'Home', 'Users', 'Groups', 'Workflows', 'Files', and 'Packs'. A search bar is located in the top right of the main content area.

The main content area is titled 'Home » Workflows'. It features a summary of workflow counts by language: Taverna 1 (662), Taverna 2 beta (145), GWorkflowDL (25), Trident (Package) (21), Bioclipse Scripting Language (11), Chemistry Plan (8), BioExtract Server (7), SimileXMLv3 (4), WSVLAM (3), Trident (XOML) (3), scuff (2), Triana (2), Excel 2007 Macro-Enabled Workbook (2), Kepler (2), Nimrod Parameter Sweep (2), XProc (1), Makefile (1), Manual Web Service (1), Pipeline Pilot (1), Stella (1), Diagram (1), Document (1), and SPARQL (1). Below this are buttons for 'Upload New Workflow' and 'View All Workflows'.

A section titled 'Top 50 tags for Workflows' lists various tags such as 'abstracts', 'african trypanosomiasis', 'AIDA', 'alignment', 'benchmarks', 'bio2rdf', 'BioAID', 'bioinformatics', 'biorange\_nl', 'BLAST', 'cdk-taverna', 'concept', 'd-grid', 'data integration', 'data-driven', 'demo', 'e-science', 'ebi', 'example', 'filter', 'gene', 'genotype', 'graph', 'gves', 'gworkflowdl', 'kegg', 'Kegg Pathways', 'localworker', 'microarray', 'mygrid', 'ondex', 'ovtk', 'oxl', 'pathway', 'pathway-driven', 'pathways', 'phenotype', 'protein', 'pubmed', 'rdf', 'sequence', 'sleeping sickness', 'taverna', 'text', 'text mining', 'text\_mining', 'tutorial', 'uniprot', 'VL-e', and 'workflow pattern'.

The 'Latest' section displays two workflow entries:

- CaliBayes\_Client (v1)**: Created: 03/02/10 @ 22:29:47 | Last updated: 03/02/10 @ 22:44:15. License: Creative Commons Attribution-No Derivative Works 3.0 Unported License. Description: A workflow for consuming CaliBayes web services. CaliBayes: www.callbayes.ncl.ac.uk.
- SBML shorthand to SBML converter (v1)**: Created: 03/02/10 @ 19:19:39 | Last updated: 03/02/10 @ 19:20:34. License: Creative Commons Attribution-No Derivative Works 3.0 Unported License. Description: converts SBML shorthand to SBML document via BASIS web services BASIS: www.basis.ncl.ac.uk:81/Basis SBML shorthand: http://www.staff.ncl.ac.uk/d.j.wilkinson/software/sbml-sh/

On the right side, there is a user profile for 'Peter' with a 'New/Upload' button. Below the profile are links for 'My Profile', 'My Messages', 'My Memberships', 'My History', and 'My News'. Further down, there are sections for 'My Stuff' (15 Friends, 5 Groups, 1 Packs, 1 Files, 5 Workflows), 'Friends' (Andy Tseng, Antoon Goderis, Bayu, Jayawardhana, Carole Goble, David Withers), 'My Favourites' (0 favourites), 'My Tags' (11 tags: beanshell, bioconductor, limma, maxd, microarray, pathway, pathways, pdf, r, sbml, user\_interaction), and 'Popular Tags' (25 tags).

**Figure 1.** A screenshot of the workflows web page showing (A) the number of workflows written in different languages, (B) a cloud of popular tags used to describe workflows and (C) the latest workflows submitted to myExperiment. Panels on the right hand side of web pages in myExperiment are used to (D) upload new workflows and digital documents, create new groups and (E) access content associated with the registered user.

All workflows in myExperiment have a hyperlink so they can be downloaded and opened in their native workflow system for further editing or enactment.

In order to bring myExperiment to potential users, external applications can access its content by making use of its programmatic interface. Designed for ease of reuse and for community development, applications such as wikis and blogs or even mashups can access content in

myExperiment or be augmented with its functionality through a set of read and write RESTful application programming interfaces. This has been used by the myExperiment plug-in for Taverna that allows its workbench to access and download workflows directly from myExperiment (20), integration with BioCatalogue and the development of Facebook applications and Google Gadgets.

## SHARING WORKFLOWS

Authors of workflows share the efforts of their work through the social infrastructure provided by myExperiment around its workflow repository. While workflows have mainly been contributed for philanthropic purposes, they have also been submitted as part of the publication process of papers (10,14,15) or for demonstrating how web services developed by organizations can be used in bioinformatics applications (2). The process of making a workflow shareable begins by selecting 'workflow' on the New/Upload panel (Figure 1). This leads to a workflow submission web page, which requests selection of the workflow file to be uploaded, as well as a title and a description. Workflows in myExperiment are uploaded in their native format. The submission of a workflow also provides the opportunity to tag it with keywords to support discovery, assisted by an autocompletion mechanism and a display of existing tags; additional assistance in structured metadata creation, including controlled vocabularies, is on the development road map.

For several formats, myExperiment has automated parsers that extract metadata from the uploaded workflow file and autogenerate a workflow diagram displayed in the interface. For example, when a Taverna workflow is uploaded, a workflow description can be extracted along with a list of web services (or other types of resources) used by the workflow. If there is not an automated parser available, or if more information is required, users can manually add more metadata as well as upload a workflow diagram generated by the native system. Full details on the registration process can be found on the myExperiment wiki. If the workflow is based on previous work present in myExperiment, then this can also be acknowledged by crediting the relevant person or group in myExperiment. It is also possible to credit other users if the uploaded workflow was a collaborative effort. Furthermore, other workflows or digital documents in myExperiment can be attributed if they were reused in the creation of the uploaded workflow.

The privacy of data in the scientific community can be a serious concern (21). myExperiment provides a flexible authorization model and allows any uploaded content to be made available with varying levels of sharing permissions. The people who can view, download and update a given workflow can be configured based on their relationship with the uploader. For example, maximum security can be placed on a workflow by setting it as private so that it is only accessible by the uploader, whereas the most open option is to allow a workflow to be viewed and downloaded by anyone. The rights by which a user can use a downloaded workflow are governed by the licensing assigned to it in myExperiment, of which there are several choices such as Creative Commons and a GNU General Public Licence.

Since workflows may be the subject of a scientific paper (10) or may have been used in the analysis of published data (15), it is possible to associate workflows with citations. A complementary approach to tying workflows with publications and other types of digital documents is

to use a myExperiment pack. Packs are collections of items such as example enactment input data, Powerpoint slides and PDF files of scientific papers that have been uploaded into myExperiment, as well as URL links to data on the web which can be bound with a workflow. Since packs can be the subject of sharing, tagging and discovery, they extend the application of myExperiment beyond workflows to any type of digital object associated with a scientific experiment involving computation.

## THE WORKFLOW LIFE CYCLE AND REPRODUCIBLE SCIENCE

The life of a workflow extends beyond its initial construction and execution followed by its deposition in a repository. Its reuse also involves the discovery of existing and relevant designs, editing the workflow to repurpose it by the addition or removal of services, trying out the workflow and then re-registration of the workflow as a new version (22). A workflow repository, such as myExperiment and a workflow construction environment such as Taverna, represent two components in the workflow life cycle. Existing workflows can be discovered through myExperiment and then downloaded and edited in their native workflow system. If the repurposing of downloaded workflows requires the addition of other web services, then their discovery can be aided by using service directories such as BioCatalogue (18) and the EMBRACE registry (23). It is also possible to discover which services are used in each workflow to enable searching of workflow content. Current work to provide closer integration of myExperiment and BioCatalogue will increase this functionality, allowing users to find all workflows containing particular services or services with a particular function.

Once a workflow is updated, it can be deposited in myExperiment with a link back to the original, allowing the evolution of workflows to be traced. This does not have to be performed by the original uploader of the original workflow since other members on myExperiment can contribute new versions depending on the access permissions of the initial workflow they have reused.

A workflow repository and construction tool provides two components targeted towards improving the reproducibility of data-driven research involving a combination of software packages that is now conducted in contemporary science (19). Such analyses are often repeated several times with modification of the parameters until the final results are produced. While these results are reported in scientific papers, the actual process of computation is often neglected and makes replication of the computational analysis by an independent scientist difficult if not impossible. Mesirov (19) proposes the use of a reproducible research system (RRS) to enable reproducible science. This RRS is comprised of a reproducible research environment (RRE) to perform the computational analysis and a reproducible research publisher (RRP) that is responsible for the preparation of a document describing the results of the computation.

The infrastructure provided by myExperiment and Taverna, together with the BioCatalogue registry of web services, can offer some of the functionality required for an RRS to replicate analyses of data. The analysis of data is described in a step-by-step manner as a workflow that can be constructed and enacted using Taverna, and Taverna is also responsible for recording the execution provenance in a separate repository (24). The published workflow can be deposited in myExperiment and the web services it uses are described in BioCatalogue. While a document preparation system to complete the proposal by Mesirov (19) is not yet offered directly, this type of component could be provided in the future, perhaps by making use of myExperiment packs for packaging workflows with provenance, input data and final results for redistribution with published papers.

myExperiment is a general repository for workflows and related research objects regardless of their format or native platform. The focus is to enable sharing and reuse of digital experimental protocols and support reproducible science. Some platform providers support workflow libraries, restricted to their own systems. Some are public, such as the Pipeline Pilot script and component libraries (<http://accelrys.org/pipelinepilot/index.html>). Others are restricted to projects, enterprises or platform licence holders, such as InforSense's Community Hub (<http://chub.inforsense.com/>). The newly formed GenomeSpace project plans a repository in the style of myExperiment sometime in the future. Other popular workflow platforms such as Kepler (<https://kepler-project.org/>), Knime (<http://www.knime.org>) and the LONI pipeline (<http://pipeline.loni.ucla.edu/>) have community forums but no community repositories.

A different kind of workflow repository focuses on protocol design. The Workflow Patterns repository (<http://www.workflowpatterns.com/>) records abstract, generic workflow patterns. ProtocolDB [<http://bioinformatics.eas.asu.edu/site/ProtocolDB/project/ProtocolDB.htm>; (25)] supports ontology-guided workflow designs that are subsequently mapped onto real services, such as BioMOBY (26). Semantic descriptions offer greater scope for workflow comparison, but at the price of a much higher overhead for metadata capture. myExperiment plans to incorporate richer semantics through controlled vocabulary tagging and integration with BioCatalogue. However, the emphasis will remain a mix of ontology-based, free tagging and community-based reviews, comments and ratings that do not discourage contribution or participation.

## FUTURE WORK

The next phase of features in myExperiment address content discovery as the workflow collection increases in breadth and volume. This includes controlled vocabularies in tagging, sophisticated search capabilities, and mechanisms to encourage and facilitate more extensive workflow descriptions and metadata. Many of these ideas have been trialled in other projects derived from myExperiment.

Packs in myExperiment can, to some extent, encapsulate digital objects such as input data, final results and provenance that are associated with workflows for distribution with published papers. However, these collections of data do not contain enough information about each object nor their relationships with one another to adequately describe an *in silico* experiment on data and make it reproducible. Future work in myExperiment will evolve workflow packs into linked research objects whose properties are self-describing (27). To this end, a prototype service is already deployed to deliver myExperiment content in RDF format based on a modularized ontology drawing on concepts from the Dublin Core, FOAF, and OAI Object Reuse and Exchange vocabularies. RDF content in myExperiment is queryable from a SPARQL end point available at <http://rdf.myexperiment.org/sparql>.

## CONCLUSIONS

Since its introduction in November 2007, myExperiment has over 3500 registered users and contains over 1000 workflows. In this period, around 28 000 visits have been made by returning visitors coming from 168 countries. By showing how potential users can interact with myExperiment, we hope this will provoke further interest from bioinformaticians in myExperiment, and that they will share their knowledge with the wider community or be able to support them in the reuse of workflows required for their research.

## SUPPLEMENTARY DATA

The myExperiment software can be downloaded as open source at <http://rubyforge.org/projects/myexperiment>. Extensive documentation and help pages are available at <http://wiki.myexperiment.org>, and requests for support can be sent to [bugs@myexperiment.org](mailto:bugs@myexperiment.org). Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors would like to thank Paul Fisher and Andrea Wiggins for their help in the development of myExperiment.

## FUNDING

UK Engineering and Physical Sciences Research Council; UK Joint Information Systems Committee; Microsoft Technical Computing Initiative. Funding for open access charge: JISC myExperiment Repository Enhancement project.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Stockinger, H., Attwood, T., Chohan, S., Côté, R., Cudré-Mauroux, P., Falquet, L., Fernandes, P., Finn, R., Hupponen, T.,

- Korpelainen, E. *et al.* (2008) Experience using web services for biological sequence analysis. *Brief. Bioinform.*, **9**, 493–505.
2. McWilliam, H., Valentin, F., Goujon, M., Li, W., Narayanasamy, M., Martin, J., Miyar, T. and Lopez, R. (2009) Web services at the European Bioinformatics Institute-2009. *Nucleic Acids Res.*, **37**, W6–W10.
  3. Kwon, Y., Shigemoto, Y., Kuwana, Y. and Sugawara, H. (2009) Web API for biology with a workflow navigation system. *Nucleic Acids Res.*, **37**, W11–W16.
  4. Wang, J. and Mu, Q. (2003) Soap-HT-BLAST: high throughput BLAST based on web services. *Bioinformatics*, **19**, 1863–1864.
  5. Jacobsen, A., Krogh, A., Kauppinen, S. and Lindow, M. (2010) miRMAid: a unified programming interface for microRNA data resources. *BMC Bioinformatics*, **11**, 29.
  6. Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J. and Rojas, I. (2006) SABIO-RK: Integration and Curation of Reaction Kinetics Data. *Lecture Notes in Bioinformatics*, **4075**, 94–103.
  7. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
  8. van der Aalst, W. (2003) Don't go with the flow: web services composition standards exposed. *IEEE Intell. Syst.*, 72–76.
  9. Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P. and Oinn, T. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.*, **34**, W729–W732.
  10. Li, P., Castrillo, J., Velarde, G., Wassink, I., Reyes, S., Owen, S., Withers, D., Oinn, T., Pocock, M., Goble, C. *et al.* (2008) Performing statistical analyses on quantitative data in Taverna workflows: an example using R and maxdBrowse to identify differentially-expressed genes from microarray data. *BMC Bioinformatics*, **9**, 334.
  11. Kappler, M. (2008) Software for rapid prototyping in the pharmaceutical and biochemistry industries. *Curr. Opin. Drug Discov. Devel.*, **11**, 389–392.
  12. Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B. and Mock, S. (2004) Kepler: an extensible system for design and execution of scientific workflows. *Proceedings of the 16th International Conference on Scientific and Statistical Database Management*. IEEE Computer Society, Washington, DC, pp. 423–424.
  13. Li, P., Oinn, T., Soiland, S. and Kell, D. (2008) Automated manipulation of systems biology models using libSBML within Taverna workflows. *Bioinformatics*, **24**, 287–289.
  14. Roos, M., Marshall, M., Gibson, A., Schuemie, M., Meij, E., Katrenko, S., van Hage, W., Krommydas, K. and Adriaans, P. (2009) Structuring and extracting knowledge for the support of hypothesis generation in molecular biology. *BMC Bioinformatics*, **10**(Suppl. 10), S9.
  15. Fisher, P., Hedeler, C., Wolstencroft, C., Hulme, H., Noyes, H., Kemp, S., Stevens, R. and Brass, A. (2007) A systematic strategy for large-scale analysis of genotype–phenotype correlations: identification of candidate genes involved in African trypanosomiasis. *Nucleic Acids Res.*, **35**, 5625–5633.
  16. Editorial. (2010) Supporting data. *Nat. Med.*, **16**, 131.
  17. De Roure, D., Goble, C. and Stevens, R. (2009) The design and realisation of the myExperiment virtual research environment for social sharing of workflows. *Future Gen. Comput. Syst.*, **25**, 561–567.
  18. Goble, C., Belhajjame, K., Tanoh, F., Bhagat, J., Wolstencroft, K., Stevens, R., Nzuobontane, E., McWilliam, H., Laurent, T. and Lopez, R. (2008) Biocatalogue: A Curated Web Service Registry for the Life Science Community. *Microsoft eScience Workshop 2008*. Indianapolis, IN, USA.
  19. Mesirov, J. (2010) Accessible reproducible research. *Science*, **327**, 415–416.
  20. De Roure, D., Goble, C., Aleksejevs, S., Bechhofer, S., Bhagat, J., Cruickshank, D., Fisher, P., Hull, D., Michaelides, D., Newman, D. *et al.* (2010) Towards open science: the myExperiment approach. *Concurr. Comput.*
  21. Waldrop, M. (2008) Big data: wikiomics. *Nature*, **455**, 22–25.
  22. Wroe, C., Goble, C., Goderis, A., Lord, P., Miles, S., Papay, J., Alper, P. and Moreau, L. (2007) Recycling workflows and services through discovery and reuse. *Concurr. Comput.*, **19**, 181–194.
  23. Pettifer, S., Thorne, D., McDermott, P., Attwood, T., Baran, J., Bryne, J., Hupponen, T., Mowbray, D. and Vriend, G. (2009) An active registry for bioinformatics web services. *Bioinformatics*, **25**, 2090–2091.
  24. Missier, P., Paton, N. and Belhajjame, K. (2010) Fine-grained and efficient lineage querying of collection-based workflow provenance. *Proceedings of the 13th International Conference on Extending Database Technology*, Lausanne, Switzerland.
  25. Aziz, M. and Lacroix, L. (2008) ProtocolDB: classifying resources with a domain ontology to support discovery. *Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services*. ACM, Linz, Austria, pp. 462–469.
  26. Wilkinson, M. D. and Links, M. (2002) BioMOBY: an open source biological web services proposal. *Brief. Bioinform.*, **3**, 331–341.
  27. De Roure, D. and Goble, C. (2009) Lessons from myExperiment: Research Objects for Data Intensive Research. *Microsoft eScience Workshop 2009*. Pittsburgh, PA, USA.