

Original citation:

Procter, Robert N., Poschen, Meik, Lin, Yuwei, Goble, Carole and De Roure, David (2009) Issues for the sharing and re-use of scientific workflows. In: 5th International Conference on e-Social Science, Köln, Germany, 24-26 Jun 2009.

Permanent WRAP url:

<http://wrap.warwick.ac.uk/53877>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes the work of researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

A note on versions:

The version presented in WRAP is the published version or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



<http://go.warwick.ac.uk/lib-publications>

Issues for the Sharing and Re-Use of Scientific Workflows

Rob Procter¹, Meik Poschen¹, Yuwei Lin¹, Carole Goble², Dave De Roure³

¹ National Centre for e-Social Science, University of Manchester

² Department of Computer Science, University of Manchester

³ School of Electronics and Computer Science, University of Southampton

Email address of corresponding author: rob.procter@manchester.ac.uk

Abstract. In this paper, we outline preliminary findings from an ongoing study we have been conducting over the past 18 months of researchers' use of myExperiment, a Web 2.0-based repository with a focus on social networking around shared research artefacts such as workflows. We present evidence of myExperiment users' workflow sharing and re-use practices, motivations, concerns and potential barriers. The paper concludes with a discussion of the implications of these our findings for community formation, diffusion of innovations, emerging drivers and incentives for research practice, and IT systems design.

Introduction

The role of the Internet in changing the ways in which scientific knowledge is disseminated has been highlighted in the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities published in 2003.¹ Since then, growth in the use Web 2.0-based tools for scholarly purposes suggests the Open Access movement is entering a new phase as the production of scientific knowledge becomes more interactive and collaborative, and its dissemination more rapid.

Several terms, emphasising different aspects of scientific materials and practices, have emerged in the past year or so for a culture where scientific knowledge is shared and disseminated more openly through new and more 'democratic' knowledge production and publishing processes: "Open Science", "Open Data" or "Science 2.0" (Hull, Pettifer and Kell, 2008; De Roure, 2008; Murray-Rust, 2008a; 2008b). Some see a growing number of "Science 2.0" sites and services as evidence of an increasing willingness among researchers to share knowledge, expertise, techniques, resources and results with one another in new ways, raising the prospect of bringing about fundamental changes in the way research is conducted and disseminated (Waldrop, 2008).

It is, of course, necessary to treat claims of potentially radical changes in scholarly practices with caution, so it is essential to investigate and understand what factors may hamper or even prevent their realisation: having the technology alone is not sufficient (e.g., David, 2004; Schroeder, 2007). The potential of e-Research and of Open Science is only likely to be realised if an alignment of interrelated social, cultural, legal and technical transformations can be achieved. For example, concerns have been raised about whether researchers will expose themselves to risks by sharing expertise and results so openly when building an academic

¹ <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

career rests on traditional peer-reviewed publications. It has also been noted that there are disciplinary differences in the uptake of new research technologies (e.g., Kling and McKim, 2000; Hine, 2005). The significance of such factors has yet to be properly understood or assessed (Waldrop, 2008). It is therefore timely and important to investigate how researchers actually make use of Web 2.0 tools, to make sense of the participatory culture (user-generated content) prevailing in Web 2.0, and how these perceptions and activities may shape the ways in which scientific knowledge is produced and shared.

To understand these and other issues, we are undertaking a series of in depth studies of researchers' use of Web 2.0 tools. One particular study we have been conducting over the past 18 months has focused on researchers' use of a Science 2.0 site, myExperiment. Funded under the UK JISC Virtual Research Environment programme, myExperiment is a Web 2.0-based repository with a focus on social networking around shared research artefacts such as workflows (De Roure et al., 2009). Scientific workflows are fundamental to the automation of complex, computationally-driven research processes, so it seems reasonable to conclude that facilitating workflow sharing will have a major impact both on the e-Research vision of shorter time to discovery (Hey and Trefethen, 2005; NSF, 2009) and the goals of the Open Science movement.

Using a range of research methods, including interview (face-to-face, telephone, instant messaging and email), we have been undertaking a longitudinal study of myExperiment users' attitudes and their behaviour in relation to the sharing and re-use workflows.

In the paper, we will present evidence from this study of myExperiment users' workflow sharing and re-use practices, motivations, concerns and barriers. In particular, we will explore patterns of sharing and re-use, attitudes towards issues such as trust and provenance and their implications for workflow sharing and re-use both now and in the future. The paper concludes with a discussion of the implications of our findings for community formation, diffusion of innovations, emerging drivers and incentives for research practice, and IT systems design.

myExperiment.org

myExperiment, funded under the second phase of the UK Joint Information Standards Committee (JISC) Virtual Research Environments programme, uses a range of social networking and content sharing tools made popular through sites such as Facebook and MySpace to facilitate the sharing of digital research resources and, in particular, scientific workflows. Its aim is to make it easy for researchers to contribute to a pool of research artefacts such as scientific workflows, build communities and form relationships which will facilitate the sharing, reuse and repurposing of workflows, reduce time-to-experiment, share expertise and avoid reinvention.

Launched in November 2007, myExperiment.org has at the time of writing over 1800 registered users and nearly 700 workflows. So far, its membership has been drawn mainly from the bioinformatics domain. Other domain-specific communities are also emerging, however, including chemistry, social statistics and music information retrieval. It is possible for anyone to access artefacts marked as 'publicly shared', but it is necessary to register to take advantage of the full range of features, including the ability

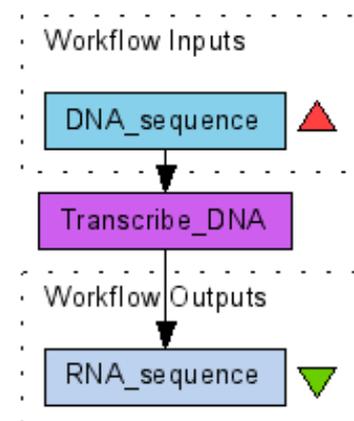


Figure 1: a simple scientific workflow.

to upload and share workflows.

A scientific workflow is a way of expressing in a formal, machine-readable notation the steps involved in the conduct of a particular analytical process (Taylor et al., 2007). A workflow captures what is involved in the performance of each step and its relationship to other steps (i.e., what inputs it requires from preceding steps and what steps consume its outputs) within the overall sequence. Generically, each step in the process may involve invoking some (and possibly) remote service (e.g., a database or analytical tool). A workflow, then, is a computational artefact, i.e., a program, and, once created, it may be executed ('enacted') over and over again. A simple example is shown in Figure 1 where a DNA sequence is retrieved from a database and this is fed into a service which converts it into an RNA sequence.

Research Methods

In our ongoing investigation, we have conducted 35 interviews with 28 myExperiment users through phone, instant messaging, via email and face-to-face. The study has been designed not only to gather data about expectations and motivations as researchers join myExperiment but also to provide a longitudinal perspective which enables us to track myExperiment users' attitudes and behaviour as they become more experienced and to contrast the expectations of new users with those who joined earlier. Interviewees are selected on the basis of myExperiment activity profiles, including workflows uploaded/downloaded; number of friends; group membership; group moderation and discipline.

In total, we expect to conduct between 40-50 interviews with 25-30 respondents in 4-5 waves by September 2009. In each wave, existing respondents have been re-interviewed and new respondents recruited. To date, we have conducted 35 interviews with 28 users. One user has been interviewed three times and five users interviewed twice.

Interviewees are recruited on the basis of myExperiment activity profiles, including workflows uploaded/downloaded; number of friends; group membership; group moderation and discipline – or by snowball sampling (i.e., users suggested by interviewees). In the former, we visited the site from time to time and approached people who had been recently active on myExperiment.

To understand the nature of workflow technologies, how they are used in scientific research and how they are constructed, one of the authors participated in a workflows training workshop for postgraduate students at the University of Manchester.

Findings

Becoming a myExperiment User

All our interviewees report using myExperiment for publishing and disseminating workflows:

“The ability to be able to publish the workflows [...] It's much more convenient and more organised than just publishing the XML file on my research group's site [...] People who have the same interests or use the same components [...] they'll more likely to find it on myExperiment. So, basically, for things like dissemination it's fantastic, makes it easy to point people out which I think it's useful.”

[Interviewer]: “So [publishing on myExperiment] is an alternative way of getting your work known? [Interviewee]: “Yes [...] in the future, I'll be writing several workflows. I'll cite them and myExperiment URLs in the paper. It's just a personal policy of mine. It's a bit like gene or protein sequences: I submit my workflows.”

Some interviewees express a more collaborative ethos:

“When my solutions work, it would be great to share with others.”

“[...] I have the feeling that there’s a better opportunity for sharing scholarly work in that way too.”

Other interviewees are aware of possible ‘network effects’ and their potential significance for dissemination:

“[...] we started to realise very quickly that it’s going to be the next kind of big area for different kinds of research [...] many people say to us ‘hey, we’re using myExperiment. Can we have access to various different pieces?’”

One interview explained the potential of myExperiment for collaborative project work:

“Used myExperiment also for the (...) project [...] myExperiment will be great for staying in contact with other people for new collaboration [...]”

myExperiment Communities

From a combination of usage statistics and interview data, we are able, tentatively, to identify five (overlapping) communities of users:

- Those who use myExperiment daily or a few times a week, as this fits into their practice of constantly using and researching workflows.
- Those who use myExperiment when they are actually building workflows, so they have a look at myExperiment from time to time to check on new workflows or on the contributions from well known myExperiment users or ‘champions’.
- Those for whom the focus is on using myExperiment for publishing their workflows and other research objects (e.g., collections of related files) might not use it actively over a long period, but reference their workflows in publications or on web pages, blogs, etc.
- Those with a strong developer focus who are looking for possibilities to mash-up myExperiment content through the myExperiment API.
- Those whose methodologies and practices may not yet be ready to exploit myExperiment, but who anticipate benefits materialising in the future.

Interestingly, membership of these communities does not necessarily align along disciplinary boundaries. Where disciplinary differences are in evidence (e.g., bioinformatics is the largest disciplinary group), this is attributable to the availability of datasets and services within those disciplines.

Issues for Workflow Discovery

The availability of both workflow provenance (e.g., who built it) and behaviour (what it does and how) is acknowledged by our interviewees as a key requisite for facilitating discovery (and re-use). From our interviewees’ comments, myExperiment users are still learning what constitutes good ‘curation’ practice (Goble et al., 2008):

“It’s a compromise how much time a contributor would like to invest to make sharing workflow feasible. If I have to be honest, my own workflows are not documented properly either. My workflow is used by a small group of people. Once in a while when I present my work, that’s how they know how my workflow works.”

“[...] it depends on how well people label various components. We’ve seen quite a few models with no descriptions of what input and output are, or what kind of format and what assumptions they made.”

“The functionality is there, but [...] people are being lazy or people don’t understand different ways they can put information in to make that available to other users [...] people make too many assumptions over what anyone else knows [...] It happens everywhere.”

“[...] I found browsing for things to be one of the functionality areas I think need help because I find it difficult to find something [...] some features that make it easier to find workflows would be useful.”

“[...] you can start asking questions to collaborators [...] there’s flexibility there but it’s something that can be improved [...] it’s users rather than myExperiment’s fault.”

Annotation and tagging (i.e., the adding of descriptive labels or metadata) is one example of how myExperiment users are grappling with establishing good curation practice:

“One of the issues we found is everyone kinds of puts things down as a text mining tag which is not very specific in order to get the different bits and pieces that we need. So users need to put in or people who actually put things there need to start considering how to label things in order to have them being found.”

“I always think of tags as supplements for navigation because you can’t really rely on people conforming to tagging standard ways that other people doing it. So if it’s a low frequency tag you are not necessarily going to get everything even related to it.”

“If they can make it easier to discover workflows, like better using the metadata, promoting users to use metadata, like a paper repository and then being able to search other things, that would be very helpful to me.”

Our interviewees also reveal how myExperiment affords a social solution to the challenges of curating workflows:

“I included the workflow [...] in the paper and [...] on the myExperiment webpage there is also a link to the paper. That might be the reason for the spike of interest [...] I had a few emails from people about how they can get this workflow [...] they got in touch to know how the workflow worked. I put the extra instruction on my webpage [...] they just need more help.”

Issues for Workflow Re-Use

We identified a spectrum of activities which can be considered as re-use on myExperiment. The first kind of re-use is what we might call ‘direct’ in that it involves re-use of a complete workflow. Among our sample, it was rare for a specialised (and typically complex) workflow to enacted more or less ‘off the shelf, unless users were already collaborating on a project:

“Used myExperiment also for the (...) project. This is private data, we will create a new group and the members can then download the workflows (...)

The challenges for direct re-use seem to be particularly acute in bioinformatics:

“The problem in re-using workflows is a bioinformatics problem, in having so many identifiers – every database uses its own identifiers – usually biological databases are rather dirty. That means that connecting services [...] that you might think would be able to work with that data, but doesn’t. And there is another problem that is really prominent, which is web service annotation. Many services have parameters as their input, but this

doesn't tell you anything which you need for input. And that is really hampering re-use, something might be useful, but then you can't really tell what you need."

Another interviewee commented on the lack of adequate workflow curation:

"In general, the re-usability is still very hard [...] We see very limited number of re-usability because those workflows are built by researchers. They don't pay attention to annotation and comments. So it's hard to re-use these workflows because they can't find any annotations how to re-use it. This is why I think proper documentation annotation of the workflows. For example, the workflow I've uploaded, a few users have given me attention to documentation, making annotation to the workflows then we can re-use it. In general it's hard to re-use each other's workflows."

The direct re-use of less specialised workflows was more common. One interviewee gave an example of re-using a "well-described and "pretty straightforward" utility workflow as part of a larger workflow the interviewee had written. A variant on direct re-use reported by interviewees occurred where a part of a workflow was re-used:

"Something that I notice is that quite a few people are borrowing some bits and pieces from different workflows."

In other words, users are able to disaggregate large (and possibly specialised) workflows and identify re-usable (and more generic) components. It was noted by several interviewees that workflows contributed as part of an initial 'seeding' effort had started to have an impact by reducing the time needed to build a workflow from scratch.

Based on the above, we can discern two distinct myExperiment communities when it comes to direct re-use. The community which we will refer to as workflow 'consumers' prefer larger workflows ready to be down loaded and enacted; the community of workflow 'builders' prefer smaller, modularised workflows which they can customised and use as the building blocks for their own workflows:

"I upload the complete thing and I also upload its parts. The idea I have is that workflow creators will more look for components [...]. As the end user you probably look for the larger workflows."

"A larger workflow might be too specific for my needs, it might be more worthwhile to look for the smaller parts, to adjust them to my needs. Things in bioinformatics are often so specific, it can be difficult to find the right thing, smaller things are easier to evaluate."

We might conclude that workflow consumers see myExperiment as a workflow 'supermarket', whereas workflow builders see it as a 'toolbox'.

myExperiment users are beginning to realise that tackling barriers to direct re-use will require adopting new and more standardised ways of building workflows:

"Up until now it has been normal to produce something very specific, to create a whole tool and components really for that one tool, not bothering too much about re-use. We are coming from a period where nobody had to care too much about re-use and without re-use nobody had to talk to the creator. I think this is relatively new. So hopefully now, with the changing practice and [...] myExperiment people will be more interested in this (in re-use). This will also change the way people make things (workflows) [...], hopefully. But it's not going to come easy."

One user distinguished her strategies of building modularised workflows from the usual "putting components and combining them":

"The kind of modularisation I was talking about was [...] rather than separating individual items themselves and processing and get outputs [...] it allows easier ways to create

interactions between different workflows [...] So, it allows you to create different files conversion or input translation rather than completely redesign the workflow. You just have to mark something and you can mark out something completely different.”

“Workflows are often not abstract enough [...], so they can only be re-used in one system [...] software components often only work properly in a specific context. They are often not generic [...] the coupling between software components is often too tight to be re-used in another context.”

We found evidence that myExperiment community members are responding to the need to improve practice in workflow building. One of our interviewees has uploaded a relatively large number of workflows (20+) “to distribute and share best-practice workflow examples”.

We found rather more instances of what we might call ‘indirect’ re-use, that is where workflows published on myExperiment are exploited by users as a source of ideas, in some cases serving as templates or patterns for problem solving while, in others, having a more general instructional value. One interviewee, for example, reported that the workflows on myExperiment sometimes ‘inspired’ him to find solutions to his problems:

“Actually, while I was working on one of these workflows, I had a problem – I couldn’t get the iteration strategy right. So I also checked myExperiment for examples of workflows that did similar things and there I found the solutions. I also used it to look for examples.”

Another interviewee provided evidence of the value of myExperiment as a more general learning resource:

“I like looking at other people’s workflows and leaving comments. I would like to improve my skills in designing workflows, to learn.”

Discussion

Our findings suggest that, to-date, a major motivation for researchers contributing workflows to myExperiment is to acquire ‘social capital’ among their peers, i.e., they use it as a reputation-building tool. At the same time, we also find examples of workflow re-use, though, on the evidence of our interviewee sample, the scale of what we term ‘direct’ re-use is modest. However, this alone does not capture the full extent of re-use, as we emphasise below.

One factor that inhibits re-use of workflows is inadequate curation, a problem for which there is no ‘quick fix’. Researchers will take time to adapt practices for documenting scientific knowledge such that these are compatible with the sharing and re-use of complex artefacts such as workflows. The experience of open source software and open content communities suggests that the quality of workflow curation will improve over time through self-regulation and peer-review (e.g., DiBona et al., 1999; Feller et al., 2005; Stvilia et al., 2008). At the same time, these community-led processes will need to be complemented by progress in establishing what constitutes good practice for the curation of research artefacts and processes in general (see, e.g., Goble et al., 2008).

One perspective on curation is that its purpose is to facilitate judgements about the trustworthiness of artefacts: “trust indicates a positive belief about the perceived reliability of, dependability of, and confidence in a person, object or process” (Fogg and Tseng, 1999). The question would then be: what documentation would satisfy a potential user that a particular workflow a) matches what they are looking for and b) that it works reliably? As Luhmann (1990) reminds us, however, to rely on documentation alone is to fail to pay attention to “the social mechanisms which generate trust” (1990:95). Our study confirms that myExperiment

users view being able to move from workflow description to workflow author to be an important mechanism for consolidating trust. It may be the extent to which sites such as myExperiment are able to provide the right support mechanisms for ‘social curation’ (De Roure et al., 2009) which ultimately proves to be the decisive factor in their success.

A second factor that our study suggests will determine the re-usability of workflows is how they are built: our interviewees emphasised the value for re-use of making workflows more generic and more modular. To understand the significance of these properties, it is instructive to reflect on contrasting examples of the re-usability of computational artefacts. The market for software packages exists because their producers strive to make them generic, modular and ‘mobile’ (Pollock and Williams, 2008). After the first spreadsheet tools appeared in 1979, expectations grew that spreadsheet programs would also be shared and re-used (see, e.g., Nardi, 1993). Yet while spreadsheet programs remain a powerful example of the value of ‘end user’ programming for personal productivity enhancement, the vision of spreadsheet sharing and re-use has not materialised.

A question we might legitimately pose is: when it comes to re-usability, are workflows more like software packages or spreadsheet programs? On the one hand, workflows are produced in conditions that are arguably much closer to the world of end user programming than commercial software production. On the other hand, the success of Open Source software demonstrates that commercial software production doesn’t have a monopoly on quality, re-usable software. We might also expect that, as products of research (and hence subject to peer review), workflows are more likely to be trustworthy than spreadsheet programs, whose unreliability is well-known (Panko and Sprague, 1998). Then, again, we can see potential parallels between spreadsheet programs and workflows in the comments of our interviewees about the importance of following a modular approach to workflow construction: the lack of spreadsheet program re-use has been attributed, in part, to the failure of their authors to do this (Nardi and Miller, 1991).

Perhaps, however, the most important reason for treating the spreadsheet analogy with caution is the evidence that, for myExperiment users, re-use has a richer range of meanings, including, acting as a source of ideas, serving as templates or patterns for problem solving, and a more general instructional value. Our findings suggest that, in providing its users with opportunities to learn from others how to build workflows and do so more skilfully, myExperiment is already serving as a useful tool for community capacity building.

Finally, given that researchers will take time to adapt and possibly to redefine currently accepted ways of producing and sharing scientific knowledge, it is critical that tools such as myExperiment be designed with sensitivity for researchers’ concerns, the research practices and cultures within which they work (David, 2004; Hine, 2005; Kling and McKim, 2000). Most importantly, myExperiment must continue to engage closely with its users so that these practices and the tools that support them are able to co-evolve. Findings from our parallel study of the myExperiment development process reveal a very strong commitment to close and continuous user engagement, both in principle (De Roure and Goble, 2009) and in practice (Lin et al., 2008).

Conclusions

In this paper we have summarised preliminary findings from an ongoing study of myExperiment, a Web 2.0-based repository designed to facilitate the sharing and re-use of research artefacts such as scientific workflows. These findings are consistent with a community in formation, where members’ attitudes, intentions and expectations are diverse and evolving. As with any innovation, significant changes in attitudes, intentions and expectations are likely to occur as the activities of individuals filter through the networks of

relationships, reinforcing successful innovations and establishing good practice. Such ‘social learning’ (Williams et al., 2004) is often crucial for the diffusion of innovations.

Our findings also show that the benefits of workflow sharing extend beyond direct re-use. myExperiment users find value in others’ workflows in that these provide templates or simply resources for learning how to build their own workflows. This community capacity building seems to us to be particularly important but is clear that establishing good practice in workflow construction and curation will be important if its impact is to be maximised.

Whether the challenge is to encourage good practice in workflow construction or curation, success is more likely when individual and community interests align. Some evidence for what might drive this alignment is beginning to emerge: a recent study of citations in cancer microarray clinical trial papers showed that those that were linked to publicly available microarray data received more citations (Piwowar, Day and Fridsma, 2007). It may be that such incentives will be important for a growing range of research artefacts as increasing emphasis is placed on publishing the provenance of research findings (Wong et al., 2005). In the meantime, the success of myExperiment in growing a community of researchers and in bootstrapping network effects represent important steps towards exploring the meanings and practices of social curation.

Bibliography

- David, P. (2004). Towards a cyberinfrastructure for enhanced scientific collaboration: providing its ‘soft’ foundations may be the hardest part. Oxford Internet Institute Research Report No. 4.
- De Roure, D. (2008). Web 2.0 and the Democratisation of e-Research. International Summer School on Grid Computing, Hungary, July. Available at <http://www.semanticgrid.org/presentations/issgcWeb3.ppt>
- De Roure, D. and Goble, C. (2009). Software Design for Empowering Scientists. *IEEE Software*, 26(1): 88-95.
- 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-W and Poschen, M. (2009). Towards Open Science: The myExperiment Approach. Available at <http://eprints.ecs.soton.ac.uk/17270/>
- DiBona, C., Ockman, S. and Stone, M. (Eds.) (1999). *Open Sources: Voices from the Open Source Revolution*. Sebastopol, CA : O’Reilly Media.
- Feller, J, Fitzgerald, B, Hissam, S, and Lakhani, K. (Eds) (2005). *Perspectives on Free and Open Source Software*. Cambridge, MA: The MIT Press.
- Fogg, B.J. and Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, May, p. 80-87. doi:10.1145/302979.303001
- Goble, C., Stevens, R., Hull, D., Wolstencroft, K. and Lopez, R. (2008). Data curation + process curation=data integration + science. *Briefings in Bioinformatics* 2008 9(6):506-517.
- Hey, T. and Trefethen, A. (2005). Cyberinfrastructure for e-Science. *Science* Vol. 308. no. 5723, pp. 817 - 821.
- Hine, C. (2005). Material Culture and the Shaping of e-Science. In *Proceedings of the 1st International e-Social Science Conference*, 22-24 June, Manchester, UK.
- Hull D., Pettifer S.R. and Kell D.B. (2008). Defrosting the Digital Library: Bibliographic Tools for the Next Generation Web. *PLoS Computational Biology* 4(10): e1000204 doi:10.1371/journal.pcbi.1000204

- Kling, R. and McKim, G. (2000). Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science* 51(14): 1306-1320.
- Lin, Y., Poschen, M., Procter, R., Voss, A., Goble, C., Bhagat, J., De Roure, D., Cruickshank, D. and Rouncefield, M. (2008). Agile Management: Strategies for Developing a Social Networking Site for Scientists. In Proceedings of the 4th International e-Social Science conference, 18-20 June, Manchester, UK.
- Luhmann, N. (1990). Familiarity, Confidence, Trust: Problems and Alternatives. In Gambetta, D. (Ed.): *Trust: Making and Breaking Cooperative Relations*, Oxford. Basil Blackwell.
- Murray-Rust, P. (2008a). Chemistry for everyone. *Nature*, 451: 648-651.
- Murray-Rust, P. (2008b). Open data in science. *Serials Review* 34(1): 52-64.
- Nardi, B. and Miller, J. (1991). Twinkling lights and nested loops: distributed problem solving and spreadsheet development. *International Journal of Man-Machine Studies*, v.34 n.2, p.161-184.
- Nardi, B. (1993). *A Small Matter of Programming: Perspectives on End User Computing*. MIT Press.
- National Science Foundation report Cyberinfrastructure Council (2009). NSF-Supported Research Infrastructure: Enabling Discovery, Innovation and Learning. Available at: http://www.nsf.gov/news/nsf09013/nsf_09013.pdf
- O'Reilly, T. (2005). What Is Web 2.0? Available at <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>
- Panko, R.R. and Sprague, Jr., R.H. (1998). Hitting the wall: errors in developing and code inspecting a 'simple' spreadsheet model. *Decision Support Systems*, April; 22(4): 337-353.
- Piwowar, H., Day, R. and Fridsma, D. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *Nature Precedings*: doi:10.1038/npre.2007.361.1
- Pollock, N. and Williams, R. (2008). Global Software and its Provenance: Generification Work in the Production of Organisational Software Packages. In Voss, A., Hartswood, M., Procter, R., Slack, R., Rouncefield, M. and Büscher, M. (Eds.) *Configuring user-designer relations: Interdisciplinary perspectives*, Springer.
- Schroeder, R. (2007). e-Research Infrastructure and Open Science: Towards a New System of Knowledge Production. *Prometheus*, 25(1): 1-17.
- Stvilia, B., Michael B., Twidale, L. and Gasser, L. (2008). Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology* 59(6):983-1001.
- Taylor, I.J., Deelman, E., Gannon, D.B. and Shields, M. (Eds.) (2007). *Workflows for e-Science: Scientific Workflows for Grids*, Springer.
- Waldrop, M. (2008). Science 2.0: Great New Tool, or Great Risk? *Scientific American*. Available at <http://www.sciam.com/article.cfm?id=science-2-point-0-great-new-tool-or-great-risk>
- Williams, R., Stewart, J, Slack, R. (2004). *Social Learning in Technological Innovation*, Cheltenham, Edward Elgar.
- Wong, S. C., Miles, S., Fang, W., Groth, P. and Moreau, L. (2005). Provenance-based Validation of E-Science Experiments. In: 4th International Semantic Web Conference (ISWC), 6-10 November, Galway, Ireland.