<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Virtual Research Environments: Balancing Incentives and Closure in an Emergent Information Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors(s)</strong></td>
<td>Shankar, Kalpana</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2012-10-09</td>
</tr>
<tr>
<td><strong>Conference details</strong></td>
<td>Innovation in Information Infrastructures workshop, Edinburgh, 9-11th October, 2012</td>
</tr>
<tr>
<td><strong>Item record/more information</strong></td>
<td><a href="http://hdl.handle.net/10197/4208">http://hdl.handle.net/10197/4208</a></td>
</tr>
</tbody>
</table>
Virtual Research Environments: Balancing Incentives and Closure in an Emergent Information Infrastructure

Kalpana Shankar
School of Information and Library Studies
University College Dublin
Dublin, Ireland
Kalpana.shankar@ucd.ie

Introduction

An emergent model for scientific research is the use of wikis, blogs, and other collaborative tools for research purposes, what some science journalists have called Science 2.0 (Waldrop, 2008), or “Open Science”. One model of Open Science is the Virtual Research Environment (VRE), which comprises online tools, content, and middleware (Voss & Procter 2009) for distributed research.

VREs encompass e-science, cyberinfrastructure, and similar large-scale scientific infrastructures, but this work in progress looks specifically at general use, open science Virtual Research Environments (VREs) as emergent infrastructures for scientific research. It focuses on the mechanisms that VREs use – technical, stated policy, and social/professional norms - to incentivize participation in an open collaborative infrastructure while acknowledging and protecting intellectual property rights. Although there are many unexplored dimensions of the use of VREs, this paper focuses on two “open” VREs for scientific research. The paper reports on preliminary observations of publicly available help pages and community information from two widely-used VREs. To do this, the paper explores the FAQs and community help pages of two VREs that explicitly focus on content standards, information sharing, etiquette, and access.

These are myexperiment.org, a general use VRE for sharing scientific workflows and other digital objects (based in the United Kingdom), and Openwetware.org (OWW), a VRE for biologists and biological engineering (developed in the United States). Structurally, both are Web-based platforms where researchers can communicate, store and share data and other digital research, and manage research projects. Users can create and edit documents, contribute to the community through forums, upload data sets and workflows, and manage projects and teams.

This paper uses these examples to show three interrelated mechanisms - technical infrastructure, local policy, and appeals to professional norms – that must be deployed to balance competing interests for openness/transparency and closure. The paper concludes with future research plans.

Background

As Benkler (2004) writes, “Science has long been the quintessence of nonproprietary production. Academic freedom to choose one's goals and open distribution of the inputs and outputs of the scientific process are its organizational norms.” Like Free/Libre Open Source Software (F/LOSS) and similar movements, Open Science adherents emphasize these organizational norms by advocating for transparency in communication and resource sharing; VREs use Web-enabled public access to middleware, data analysis tools, and data (http://www.openscience.org). However, multiple uneasy tensions persist: personal and professional gain, intellectual property, concerns for misuse/abuse of data”, such that the open sharing of data and information is still surprisingly uncommon (Lakhan et al 2008) in most research disciplines. Thus, while positive incentives for Open Science, such as greater and faster exposure and publication times and reduction of costs for publication and access, have been demonstrated to be useful in numerous research domains, the difficulties of integration into existing work practices potentially complicate the use of the Open Science model.

VREs represent only one form of infrastructure for “doing” Open Science research but one that can be potentially be designed to balance these competing demands. VREs act as
gateways and collections for services, communication, and resource sharing. Entry into a VRE is generally open to anyone who can sign up for an account, but there is no pressure to participate beyond that. The underlying platform is created and managed centrally through institutional arrangements, although users, developers, and administrators collectively create the content, tools, and data within them.

VREs are built to some extent upon the model of the commons for research and resource sharing. Benkler (2003) succinctly defines the knowledge commons as a “particular type of institutional arrangement for governing the use and disposition of resources”, particularly collectively owned and managed information. The commons-based peer production model that Benkler describes refers specifically to the coordination of creative activity without traditional hierarchical organization. Hess and Ostrom (2001) argue that the regulation of the commons is conducted on a spectrum from an “anything goes” (called “open access” commons) approach to highly structured rules (some formal and others socially enforced) for governing resources. The developers of myexperiment.org and OWW suggest that a VRE should act as an open access commons through four mechanisms: the ability to share research objects, professional and reputation-based incentives for making such objects available for others, an open and extensible environment so that researchers can contribute to the maintenance of the portal, and a platform for the creation of customized protocols and tools to do all of these tasks (De Roure et al 2008).

To this end, VREs as emergent infrastructures exert various forms of control over their content. In the next sections, the role of three kinds of regulation, technical, policy, and social, is explored vis-à-vis their potential for incentivizing sharing while acknowledging and allowing for long-held community norms to “own” data. It is important to note that as with all kinds of large-scale infrastructures, these mechanisms are interrelated.

Interrelated Technical, Policy, and Social Mechanisms for Access Governance

While they are not universally used or applicable, open VREs offer a potentially useful set of platforms to study scientific infrastructure as commons. To date, most of the research on VREs has been on design, development, data (re)use, and the lifecycle of research (De Roure et al 2003). However, examining incentives and disincentives and the nature of sharing from the perspective of the commons may offer greater insights into the tensions inherent within the use and nonuse of Open. This includes the ability of VREs to enhance or inhibit the nature of collaboration and data sharing for “small science”, still an essential component of scientific research. It also suggests that there are significant insights to be gained from focusing on how infrastructure enforces social norms and policies, an important area of research in the role and function of sociotechnical infrastructures.

VREs must provide allow for user modification and customization, but allow those users to keep closed that which they wish to keep closed. This balance is enforced through multiple mechanisms. To give one example, Myexperiment.org explicitly acknowledges that data creators have the right to govern what data is shared, with whom, and under what circumstances. This is enforced through a “three level mapping scheme” that is part of the structure of the data description (metadata) itself. The system will check the Policy data structure to see if the current user can perform the specified actions (sharing, modifying, etc.) and whether they have the permission to do so for that specific resource.

The wiki-based structure of OWW means that anyone can edit another Wiki page for any reason and add pages, which honors the openness/collaborative nature of the Open Science model, but makes data protection more problematic. Since the Wiki infrastructure cannot enforce content standards, OWW relies on an “Etiquette” page to indicate “preferred” information tags, formatting and commenting suggestions, and links to keep things “tidy”. The Etiquette page makes clear that OWW users should use their real names, science-related content, and “reasoned discourse” to adhere to the norms of scientific communication.
Both platforms encourage users to make all information “open” or public, while acknowledging that that is a lofty goal. In the case of OWW, the loftiness is explicit. The “About Pages” openly state that since biologists generally don’t share information widely or publicly (but there is no further discussion of this topic), it is the intent of the OWW platform developers to provide a platform where scientists can do so if they wish, fostering more openness through a quid-pro-quo model.

Copyright is one policy mechanism by which both VREs balance openness and intellectual property rights; a higher level of regulation in the form of copyright adherence is also maintained by OWW. All materials placed on OWW must be protected with the Creative Commons Attribution Sharealike and Wikipedia’s Gnu Free Documentation License (GFDL) (as these are not compatible with each other, OWW uses both). However, there is no technical mechanism for enforcement; the help wiki suggests that community monitoring is suggested as the only applicable mechanism for policing copyright fraud.

Independent of copyright and technical mechanisms, both VREs acknowledge the tensions inherent in making information “open” (the definition of open left vague) and is best encouraged through social/professional mechanisms. It is assumed that the creators of research objects own them, a reasonable assumption to make in these environments. The developers and “guardians” of both OWW and myexperiment suggest that the ownership of workflows and other research objects is a problematic issue that can be solved by community buy-in and perhaps more interestingly, the development of the technical infrastructure of the workflow and metadata around the workflow. The regulation of resources does depend significantly on norms shared in other aspects of scientific work, such as attribution and citation as well the use of real names, reasoned discourse, and science related content as befits a science portal.

Conclusion

As noted earlier, this project is in its earliest stages. Next steps will include interviews with users, administrators, and if possible, individuals who have used but no longer use VREs. This will also involve more extensive empirical and analytical work on technical and social infrastructure for balancing incentives and disincentives. It will be instructive to learn more about how users and administrators of VREs navigate the “layers” of data management, their own perceptions of openness, and the ways in which such commons-based infrastructure evolve.

Although the emphasis of this paper is on open access VREs, the themes it surfaces are not limited to such. These themes are of equal importance to other forms of data archives, e-science, e-social science, and digital repositories. The sustainability of large-scale infrastructures, why and how they are used or not used, how users navigate multiple tensions inherent within them, are always of pressing concern to those who develop them, invest in them, and depend upon them for their ongoing work. As the push for larger infrastructures and archives continues with a concomitant exhortation to focus on big data, VREs and other forms of knowledge commons potentially offer insights into the nature of collaboration, data sharing, and incentives and pressures for “small science”, still an essential but not to be ignored component of scientific research and the information infrastructure of scientific practice.

References


