

The margins of scholarship

Repositories, Web 2.0 and scholarly practice

A paper for OR08

Richard M. Davis, University of London Computer Centre

Overview

Beginning with a consideration of the significance of the social networking functions associated with Web 2.0 in relation to examples of traditional and historical scholarly practice, and the extent to which such functionality is appropriate for IR web interfaces, this paper will then present the outputs of the SNEEP (Social Networking Extensions for Eprints) project and related discussion and research.

1. "Flickr for Eprints"

In April 2006, Jill Walker (Associate Professor of Humanistic Informatics at the University of Bergen, and an avid blogger since at least 2000) delivered a lively and compelling paper to the DSpace User Group, describing repositories from her perspective as a user - both researcher and depositor.

Here's an example of a document repository that works wonderfully: [Flickr](#). Flickr is an online photo sharing site. People upload their own photos, and the system provides them with so many cool ways of using their photos through this system that there's *a lot of incentive to do so*. You can see your friends' photos, you get comments, you can find other photos of the same things, you can more easily find your own photos. You can have your most recent photos show up on your website (through RSS) or you can write programs to do cool stuff with your own or with other peoples' photos.

Jill also described the attractions of sites like CiteULike and Del.icio.us, bookmark managers that are "like an exoskeleton rather than a repository". As well as storing bookmarks in a centralised database, accessible from anywhere (home, office, library), these systems support tagging, publishing and sharing of bookmarks, in ways that make everyone's bookmarks - and particularly teachers' and researchers' - potentially useful to other users. "Adding social features like these to institutional repositories could make them useful for researchers to actually use, and not simply a smart way of preserving research publications." (Walker, 2006)

Jill created a wishlist for IR functions, the first three of which relate directly to the class of functionality we would now associate with Web 2.0 (a term she doesn't use in this blog post).

1. I want the data I enter to be usable by other web applications - an open API. For instance, Flickr's open API means that I could write a program that uses my [photos].

2. I want RSS feeds that I can put in my website, the way I can put my most recently bookmarked articles from CiteULike there.
3. I want to be able to click a keyword on an article and see other articles not just in my own institution's repository but across repositories.

Two years on, some still question whether this kind of functionality has any relevance in the world of serious academic publishing/archiving, and consider it unnecessary frivolity (detracting from important issues like metadata and preservation). They may consider that blog-like functionality trivialises scholarship by placing the "informal discourse" (Mortensen & Walker) alongside the formal.

But, as Sharon Kopyc (2006) describes, the next generation of students (and researchers) is already upon us, and they have ever more demanding expectations of the media and interfaces they work with: if they find a JSTOR plugin for Facebook useful, and it's trivial to create one, why shouldn't they have it?

2. Scholarly margins

Some aspects of Web 2.0 functionality needn't be considered so strange and new-fangled. In a recent talk, David Pearson, Director of the University of London Library, highlighted the value and interest of annotations in books:

A book which has been annotated can give us a direct insight into the interface between author and reader, can show not only that a text was definitely read and absorbed but also perhaps how it was received.

David gives some examples, such as Thomas Carlyle's marginal comments on Elizabeth Barrett Browning, and Henry VIII's annotations to theological texts on divorce, though he also recognises that this is not a habit librarians tend to encourage.

In a similar vein, from Wikipedia we learn that the historian A. L. Rowse was an habitual and sometimes acerbic annotator of books.

For example, his copy of the January 1924 edition of *The Adelphi* magazine edited by John Middleton Murry bears a pencilled note after Murry's poem *In Memory of Katherine Mansfield*: 'Sentimental gush on the part of JMM. And a bad poem. A.L.R.'

(I have a copy of a Penguin history book - Shakespeare's Plutarch - which was once Rowse's, and it is indeed neatly and pointedly annotated, but alas there is nothing so quotable in its marginalia.)

Recently at ULCC we begun work with The Linnean Society on an online archive of their digitised collections, and have encountered a similar phenomenon in this system. Carl Linnaeus's specimens are abundantly annotated not only by Linnaeus himself, but by his successors, such as Sir Edward Smith. (The history of these annotations is described in Spencer Savage's 1945 catalogue of the Linnean Herbarium.)

As part of the digitisation, manuscript comments on the specimen sheets have been afforded as much attention as the images of the plant specimens themselves: they have not only been photographed, and filed along with the specimen, but also transcribed and the transcription included in the object's metadata.

[We will illustrate this with examples from the Linnean Online collection.]

With this in mind, the Linnean Society's request for a Commenting feature for their online system takes on a new dimension: botanical researchers will now be able to recommence that dialogue-by-annotation that they had lost in the 19th Century when the Society forbade the practice of further annotating Linnaeus's specimens (commentary had to continue on separately filed sheets).

Perhaps in other fields too, this approach can give back to readers the simple pleasure of adding their own observations in the empty white fields that surround texts - something gradually eroded by the rise of word-processors and electronic documents.

3. Web 2.0 progress

If, therefore, we conclude, in Sellar & Yeatman fashion, that this kind of feature probably is "a good thing", how far have we got with this in Eprints in particular, and across Open Repositories in general? "Web 2.0" as a fully-fledged concept and buzzword is still barely two years old. The new paradigms it embodies for working with the Web are already quite mature in other web applications with similar functionality, notably Flickr, which is at the end of the day just another kind of repository. (I realise the comparison is a little unfair, since Flickr very much defined expectations of Web 2.0.)

In their 2007 JISC report, *Web 2.0 for content for Learning and Teaching in Higher Education*, Franklin and Van Harmelen called for projects be undertaken to investigate how institutional repositories can be made more accessible for learning and teaching through the use of Web 2.0 technologies. This was the motivation behind the SNEEP - Social Networking Extensions for Eprints. SNEEP is a small project with a modest brief - to create plugins for Eprints to implement Comments, Bookmarks and Tagging (CBT).

The first two plugins (Comments and Bookmarks) derive directly from work already done for the Linnean Society. The third, Tagging, also has a potential use for the Linnean Society, who may use it as part of a collaborative project to collect vernacular names for the specimens. (At the moment, these are primarily referenced by their Linnean, Latin names, of course.)

Whether, and how such features may be used or useful in IRs at large, we can't know. Given that many of our users already have Connotea, CiteULike, Del.icio.us, etc, (not to mention Rich Tags!), demand may not be as overwhelming as in Flickr's case. Nevertheless, it's my view that, particularly in larger repositories, including meta-repositories, there is a good case for providing these functions within the IR, rather than depending on users to go to other services. Much as with the VLE's, it seems responsible and desirable to provide a base-line functionality for all users, while yet enabling power-users who want to use their preferred tools and methods for managing their information.

At the very least, basic CBT functionality is not hugely complex to provide, so there is no harm in trying: if it becomes core to Eprints and its peers, then it can be left to the community to decide how to take it further. The goal is to make the IR environment as *susceptible* as possible to social applications and uses.

[SNEEP is due to be completed by April 2008, and the remainder of this paper will be an opportunity to review the findings of its consultation with existing and potential Eprints users, and the outcomes of the project itself - with reference to Jill Walker's 2006 wishlist.]

Principal references

T. Franklin & M. Van Harmelen (2007). Web 2.0 for content for Learning and Teaching in Higher Education. Bristol: JISC.

http://www.jisc.ac.uk/media/documents/programmes/digital_repositories/web2-content-learning-and-teaching.pdf

Sharon Kopyc (2006). Enhancing Teaching with Technology: Are We There Yet? *Innovate*, 3(2).

<http://www.innovateonline.info/index.php?view=article&id=74>

T. Mortensen & J. Walker (2004). Blogging thoughts: personal publication as an online research tool. (in *Researching ICTs in Context: a research report from InterMedia University of Oslo, Chapter 11*), Chapter 11). <http://imweb.uio.no/konferanser/skikt-02/docs/Researching ICTs in context-Ch11-Mortensen-Walker.pdf>

David Pearson (2007). Libraries as history: the importance of libraries beyond their texts. In *Proceedings Senate House Library Friends: Charles Holden Lecture*, Institute of Advanced Legal Studies, University of London (UK).

<http://eprints.rclis.org/archive/00012026/>

Spencer Savage (1945). *A catalogue of the Linnaean Herbarium*. London, Linnean Society of London, 1945.

Jill Walker (2006). repository (sic). In *jill/txt* (entry for 19/4/2006)

<http://jilltxt.net/?p=1658>

About the Author

Richard Davis is an IT specialist at the University of London Computer Centre. He has worked closely with archives and library professionals for ten years, variously as data and systems analyst, project manager and web developer. Projects include the UK National Digital Archive of Datasets (NDAD), Practice-as-Research In Performance Online (PRIMO), Social Networking Extensions for E-Prints (SNEEP) and Linnean Online, an image repository of the collections of the Linnean Society of London. He has participated in many conferences and workshops on digital preservation and metadata standards, including those of ERPANET, DLM-Forum, DCC and DRH. Presentations include "Digital Preservation In Theory And Practice" for AIIM RMLG conference, "A Bluffer's Guide To Institutional Repositories" for the 2007 WESLINE colloquium, and a repository case-study for the 2007 DSpace User Group. Richard has degrees in English Literature and Computing Science, and is currently studying for a further Master's degree in E-Learning at Edinburgh University.