

Adding Discovery to Scholarly Search: Enhancing Institutional Repositories with OpenID and Connotea

Ian Mulvany, Manager, Connotea.org (Nature Publishing Ltd.) i.mulvany@nature.com
David Kane, Repository Manager, Waterford Institute of Technology, Ireland dkane@wit.ie

Our Project:

By linking Connotea to institutional repositories through the use of open standards the value of the data on the repositories can be enriched by social data from the service. Conversely, registered users from linked repositories acquire automatic membership of Connotea, boosting the quantity and quality of Connotea's member base with quality membership, I.E. academic authors.

This integration would provide the basis for a new 'discovery and evaluation' layer to be added to institutional repositories globally.

In this scenario, repositories register with Connotea as trusted providers of Identity (Using OpenID).

Aims

- Encourage growth of Connotea's data by making it easy for users of Repositories (like EPrints) to join/use Connotea through an implementation of OpenID.
- Host JavaScripts on Connotea.org that can insert social data directly into record pages in repositories.

Objectives

- Provide for 'discovery' and evaluation of information on top of regular search.
- Increase the amount of quality data in the Connotea database (in support of the previous objective)

Success Indicators

- New members of Connotea begin to join through OpenID
- Web visitor statistics begin to indicate new usage patterns suggesting they are 'discovering' information thanks to the use of social data.

Background:

Apart from archival, the main purpose of research repositories is to increase the findability of knowledge that has been produced by staff in the host institution – speeding up the absorption of that knowledge by the relevant research communities.

The most common use-case scenario for discovery of information presented by repositories is where a searcher types some keywords into Google and links through to a repository metadata record page or to the item itself.

Another goal of repositories is to make knowledge available without restriction – the ideal of the open access movement. Restrictions exist in terms of subscription barriers to journals, but dealing with a vastly increased amount of information is a barrier to its effective use. In order to prevent information overload due to open access (and things do seem to be moving in this direction) it will turn out to be necessary and indeed inevitable that additional tools for search, discovery and evaluation of material should be developed.

One way of overcoming this information overload is to increase the ways in which people may evaluate information. This is done in a number of websites¹ which use 'social data' to

¹ [librarything, amazon, citeulike, connotea etc.]

enhance existing metadata about an item. The value of such social book marking sites scales with the amount of information that they contain. Social data is about the collection of information about users' preferences that, in aggregate, can be used to provide reliable predictions about what an individual might be interested in.

For example, in LibraryThing² when a user visits a page for a 'work', he or she is instantly able to see what users who bookmarked this work also bookmarked. These items are also likely to be of interest to the visitor who is interested in a specific work. This is because these suggested items represent an aggregation of bookmark data generated by users who all have an interest in this particular item (book). The situation is similar for tagging with uncontrolled keywords/phrases. Users who click on 'fishing', for example are presented with a list of books for which 'fishing' is a prominent tag.

The success of LibraryThing is directly attributable to the large amount of data that is present on their servers. They now have over 20 million virtually catalogued books (though not 20 million 'works'). A sufficiently large amount of data enables an application to make better guesses about what might be most relevant apropos a specific item.

If this success is to be repeated across institutional repositories, then what is needed is a large, reliable source of 'social' data. One way of providing this data is to host this data in a central location and share it between a large number of institutional repositories. In this paper we suggest that a relationship can come to exist between an ad-hoc federation of repositories and a central warehouse of social or bookmark data. We demonstrate this, using Connotea as an example.

Connotea is a social bookmarking service for academic publications. It is used by its members to create collections of links to relevant items on the web, which they find helpful to their research. Links to items hosted on institutional repositories are already becoming common in the Connotea bookmark collections (called 'libraries'). This trend can only increase as the use of repositories gains traction in the research community.

It is a good time for repositories to have a function that formalises this relationship more and that implements such measures as would make it easy for authors with accounts on repositories add data to Connotea.

One way that this has been done up to now is through the use of the Connotea TaggingTool, an EPrints plugin that pulls social data about a reference from Connotea. This extension has two limitations. It requires a user to be logged in to Connotea and it requires the reference that the user is currently looking at to be already bookmarked in Connotea. Two aims of this project are

- (1) To see if we can move to a framework of using remote JavaScript to pull social data about any reference based on keyword and tag machine,
- (2) To see if we can encourage repository users to participate in tagging by lowering the barrier to entry to Connotea through implementing a common logon process to an archive and to the connote service via the use of OpenID

Using OpenID with Connotea

This integration would provide the basis for a new 'discovery and evaluation' layer to be added to institutional repositories globally.

The integration should be done using OpenID (<http://openid.org/>). Because of its ease of use, OpenID could pave the way for a great expansion in membership of Connotea.



² <http://librarything.com/>

OpenID is a single sign-on solution to enable a user to authenticate once and gain access to the resources of multiple software systems. The user logs into a website by authenticating against a 3rd party server. Because OpenID is an open standard, it can be implemented by anyone.

Consider the following scenario:

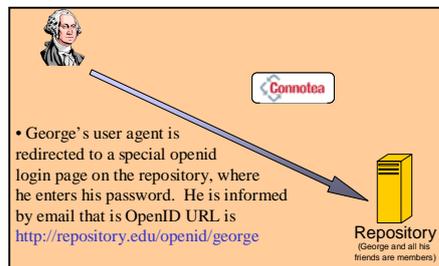
George is an academic in an institution that has an institutional repository. He is the author of many papers; most of them are actually on the repository.

His institution has just installed an OpenID modification for his repository, which means he can now has a passport that makes it easy for him to use a number of services on the web, including Connotea.

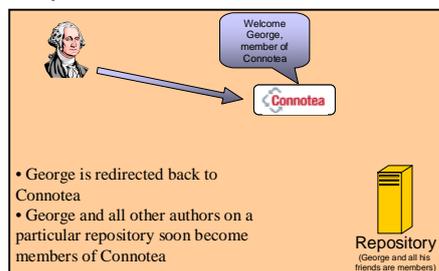
Knowing nothing about Connotea, he is intrigued by the recommendations that are appearing under some of his own and other colleagues' articles in the repository. These recommendations come either from the 'TaggingTool' plug-in, or from another channel, which appears at the bottom all record pages in the site. Some of the recommendations he recognizes as being excellent articles that he already knows of. There are one or two other articles that he has not read, but is keen to do so because of the overall quality of the recommendations that Connotea has provided him with. Keen to know more, he signs up to Connotea. OpenID makes the process very easy and he quickly creates his own library on Connotea. He is able to flag some papers as his own. And Connotea benefits by having many quality authors like him join. He connects with many like-minded academics that he did not know before. In them he can see future research partners and collaborators.

This is how the integration between OpenID and Connotea could happen in practice. The steps happen very quickly.

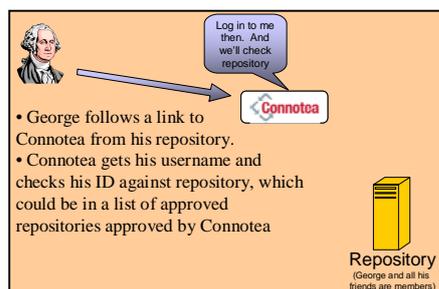
Step 1



Step 2



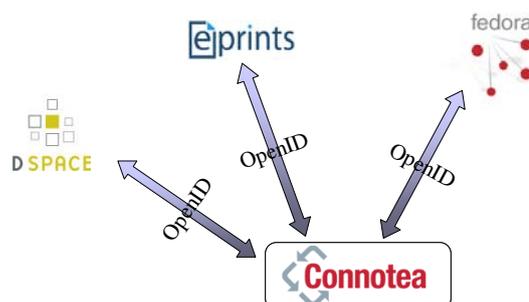
Step 3



This is how the OpenID integration could be done:

1. Develop extensions for each of the main open access repositories that make them function as OpenID servers. These could include Southampton Eprints, MIT Dspace and Fedora.
2. Connotea develops functionality that allows authentication of new users against OpenID servers (the institutional repositories).

By developing OpenID extensions for all the main repositories, it becomes possible for Connotea to augment its user base by orders of magnitude. See below.



Individual repository entries could be enhanced with:

- Recommendations for similar articles
- Individuals who were interested in those articles

Connotea would benefit by:

- Increased membership due to barriers to entry being lowered by
 - Promotion on all participating repositories
 - Lack of need to create/remember new password
- Resulting increase in data

Conclusion

So, in addition the traditional metadata associated with each record, like author, abstract, date etc. There is a new level of data that includes information about who else has the article and what other articles are similar as well as perhaps direct links to the full-text of those articles where institutional subscriptions permit.

Social data immensely quickens the rate at which relevant pieces of information are sifted from bad information relative to a particular context and constitutes a kind of 'grass roots' peer review.

A scholarly communications system that adopts the features that distinguish many of these social web applications will be far more compelling and useful than a passive search portal. Adding a 'social' dimension to scholarly communication will have a synergistic effect – increasing the ease with which relevant materials can be sought and discovered by researchers. By the same token, potential for collaboration may also be more easily identified.

A key foundation for this would be integration of Connotea with the most commonly used repository software, EPrints, using OpenID.