A workflow for easier repository deposit using SWORD - EM-Loader project

Fred Howell[1], Ian Stuart[2], Theo Andrew[2] & Mary Robinson[3]

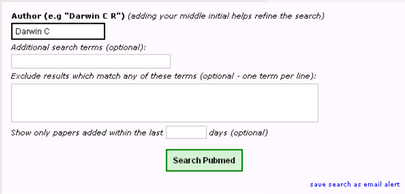
[1] Textensor Limited, [2] EDINA National Data Centre (University of Edinburgh), [3] SHERPA (University of Nottingham)

Introduction

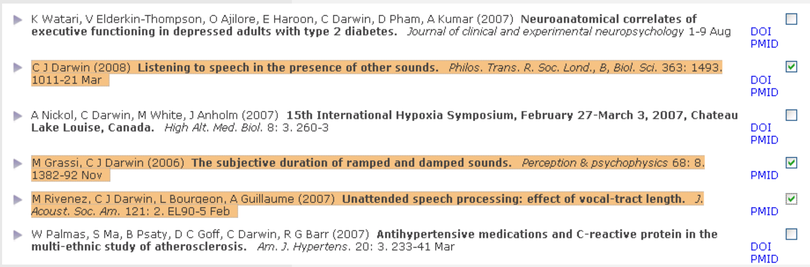
Most academics see the need for maintaining a professional personal web page listing and linking to their publications. However, the time required to add the bibliographic metadata needed for submission of their papers to an open access repository can be prohibitive. The EM-Loader project reduces the effort required for a researcher to submit all their papers to a repository to just a few clicks by linking a system designed for maintaining a personal publications list on a web page (http://publicationslist.org) to a repository, in this case the Depot (http://depot.edina.ac.uk, based on ePrints), using automated interfaces including SWORD.

User workflow

[1] The process starts by registering an account on PublicationsList.org and choosing a user name.

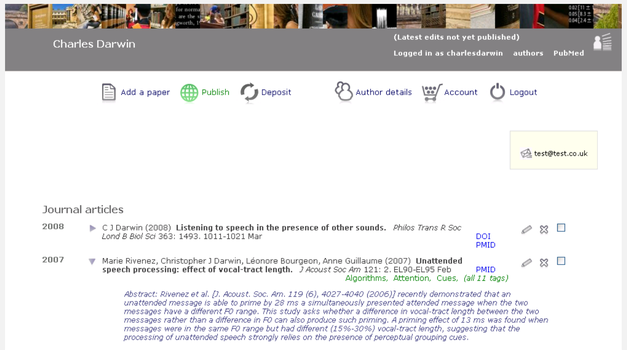


[2] Users can then import their publications metadata in batch mode from bibtex, endnote, web of knowledge, or from the PubMed database by using the integrated author search.



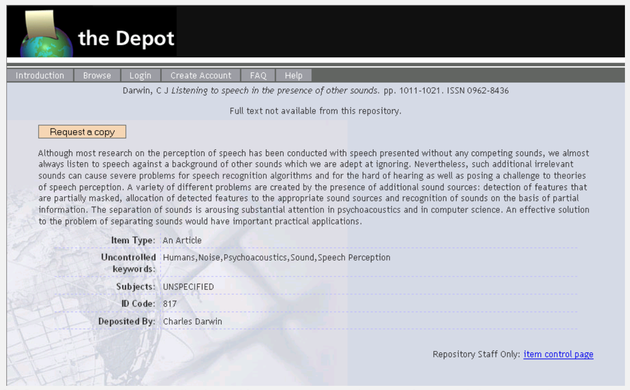
[3] From the metadata returned by the author search users can then select which publications to include in their publications list.

[4] Having imported details, the user gets a publications list which they can publish as static web page (using the 'Publish' button); these items can also be sent to a designated repository using the 'Deposit' button



[5]If the user doesn't have an account with the designated repository, and if authorised to do so by the repository, publicationslist.org can automatically create an account for them. A 'create account' link sends the user's email, full name and publications list user name to the repository, creates a repository account, and stores their repository user name in publicationslist.org.

[6] Once the repository account has been set up, there is a 'Click to send to repository' link. This sends all selected items to the local repository, using the SWORD protocol.



[7] On pressing the button, all selected items are sent to the repository, and the screen shows progress and repository IDs as each item is sent. After sending items, an additional check is performed to fetch the status of each item sent; as some repositories have an approval process. This check updates the publications list entries with a status; 'submitted', 'pending', 'accepted', 'rejected' and a link to the repository entry.

Experiences using SWORD

In theory, using SWORD is an extremely good idea. In practice, we encountered a number of issues during implementation. Hopefully our experience will contribute to improvements in future versions of SWORD:

1. METS is overly complicated for the task - it would be better if a simpler JSON format was supported.
2. SWORD itself does not specify a metadata format that should be used for the transfer. The format we chose is JSON using the field names from bibtex.
3. Using the Atom publishing protocol for posting data to a repository is over complicated. A simpler API just based around HTTP POST of a zip file with JSON metadata would be easier for developers to use, and would have an easier learning curve.
4. SWORD does not support updates and versions of existing items. We implemented a way of doing this with ePrints as it was an essential feature for us, it would be better if this were part of the standard.
5. To fetch the status of deposited items over time, to do author searches of metadata, and to create accounts we needed additional 'fetch' interfaces as well as SWORD.

The future: zero click deposit

The need for the researcher to press the Deposit button each time they add publications could be removed by switching away from a PUSH model where items are sent to the repository (via SWORD) to one where the repository does an automatic periodic PULL from the user's publications list. The system could work as follows:

1. Researcher maintains a publications list web page as at present.
When publicationslist.org makes the HTML version, it also generates two Atom feed XML files; one which contains the metadata in HTML (for news readers), and one which contains the same information in JSON-Bibtex.

2. Researcher tells repository the URL of their publications list
This could be done by pasting a link to the atom feed, or using an API call similar to the 'create repository account' module we wrote for EM-Loader.

3. Repository polls the Atom feed periodically, fetches any new items
This would require a new repository module which runs as a regular cron task - which could poll the registered atom feeds for new items, create new entries

http://publicationslist.org/em-loader/

JISC

EDINA