

Automated IR deposit via the SWORD protocol: an MIT/BioMed Central experiment

The Massachusetts Institute of Technology (MIT) Libraries report on an experiment with the SWORD protocol in which articles published by BioMed Central (BMC) are automatically deposited into MIT's open access institutional repository, DSpace@MIT. Included is a description of the background, workflow, challenges, current status and expected outcomes of this experiment.



ELLEN FINNIE DURANCEAU

Program Manager, Scholarly Publishing and Licensing
MIT Libraries



RICHARD RODGERS

Head, Software Development Group
MIT Libraries
USA

Introduction

The Massachusetts Institute of Technology (MIT) Libraries and BioMed Central (BMC) have been working together to explore the potential of the SWORD protocol to automatically deposit articles in MIT's open access institutional repository (IR), DSpace@MIT. This experiment is still not complete, but our work to date has offered proof of concept and leaves us optimistic about the prospect of efficiently and automatically populating a repository with research articles through publisher/library partnerships using the Simple Web-service Offering Repository Deposit (SWORD) protocol.

Background

The MIT faculty's Open Access Policy (<http://libraries.mit.edu/oapolicy>), established in March of 2009, expresses the faculty's commitment to 'disseminating the fruits of its research and scholarship as widely as possible' by making their scholarly articles openly available on the web. Though an open access policy is of course not a prerequisite for taking advantage of automated deposit, the creation of MIT's faculty policy

provided a new impetus for finding efficient methods for acquiring MIT-authored scholarly articles and making them available through MIT's IR, DSpace@MIT.

Since the creation of the policy, the Libraries have been administering it under the guidance of a standing faculty committee, devising workflows to make the implementation 'as convenient for the faculty as possible', as called for in the policy. The faculty policy calls for sharing the author's final manuscript, post-peer-review. However, the faculty committee overseeing implementation has indicated that if the final published version can be legally posted, that is the faculty's preference. With that directive, and with convenience and efficiency as our guiding principles, we concluded that working with publishers and automating deposit would be an essential element of our implementation. We therefore began conversations with publishers whose policies were consistent with posting the final published version in an institutional repository. Early on, we contacted BioMed Central, the largest open access publisher, in the hope that we could work together to achieve efficiencies. The availability of the SWORD protocol held promise for automated deposit with BMC.

The SWORD protocol

Essentially, the SWORD protocol is a feed that allows MIT-authored articles from any of BioMed Central's journals to be deposited automatically in the 'Open Access Articles' collection in DSpace@MIT (<http://dspace.mit.edu/handle/1721.1/49432>) as soon as they are published. This takes the burden off the author to arrange for deposit under the faculty policy. It also takes the burden off the Libraries to acquire, process, catalog and deposit each article individually.

SWORD formalizes a very simple dialog between a depositor and a repository: the depositor identifies itself, and the repository replies with a document (known as the 'service' document) detailing the specific permissions and restrictions binding that depositor. Since almost all repository content is organized by collections, the permissions are typically expressed as a set of collections the depositor may deposit to. The depositor then transmits the content and metadata – packaged in a standard format – to one of the approved collections. When received and validated, the repository acknowledges successful deposit.

While each aspect of this process is based on open and well-documented standards to ensure interoperability, SWORD allows considerable flexibility of practice. For example, the content package may conform to any mutually acceptable format, but the SWORD specification encourages the use of pre-existing, proven standards such as the IMS content package or the Library of Congress METS.

SWORD is also unapologetically built on and for the World Wide Web: in this it differs from many information exchange protocols arising out of the library/repository domain (e.g. Z39.50). In fact, SWORD at a technical level really is just a 'profile' of an existing web-publishing standard known as AtomPub. This latter arose from efforts to unify the burgeoning but fragmented world of blogs and syndication formats. As a result, it is easy to leverage existing web infrastructure and tools to support SWORD clients and servers (depositors and repositories): there is even a SWORD client to allow deposit from Facebook.

SWORD is a transmission protocol, and as such does not directly concern itself with metadata, but much of the initial and ongoing adoption (including the experiment described here) has focused on 'e-prints' – scholarly, peer-reviewed

articles or other publications. Thus another JISC-funded effort known as SWAP (Scholarly Works Application Profile) provided the basis for metadata standardization in SWORD deposits. SWAP is a Dublin Core profile that attempts to embody an FRBR data model for scholarly works. In the context of our open access repository, the ability to distinguish and express relationships among various versions (published, preprint, etc.) of a work led us to adopt elements of the SWAP vocabulary to enhance the basic Dublin Core profile used in the off-the-shelf DSpace SWORD service.

The experiment

The first steps in implementing the SWORD protocol with BMC involved establishing search protocols to most accurately identify MIT-authored articles, and deciding on the timetable for feeds. We determined that waiting for SWORD to support a 'push' approach, where nightly feeds would be automatically delivered to DSpace@MIT, would serve all parties best. Once this option was available, the main thrust of effort was developing an appropriate crosswalk to transfer from BMC's data schema (METS) to ours (Qualified Dublin Core). This work included building in local metadata fields (e.g. peer-reviewed status; paper version; content type) in addition to those supplied by BMC, affording additional labor savings.

Key steps in working through the data packages from BMC included reviewing the packages and our transformation process, for accuracy and completeness; reviewing the accuracy of BMC's affiliation searching for MIT; and reviewing the content types received to ensure we were, in so far as possible, receiving only papers appropriate to deposit in our Open Access Articles Collection under the MIT policy. For example, we wanted to avoid receipt of papers whose format did not meet our deposit guidelines, such as poster session reports. During the test phase, which we have just completed as of September 2010, papers were deposited to a test server. In the future, we expect that papers will be delivered directly into the production server.

Challenges

One of the goals of our experiment has been to find the best balance possible between receiving some

papers that are not appropriate for deposit, and missing papers that would have been. The latest test review indicated that some 36% of papers were not ultimately depositable, primarily for reasons outside of BMC's control. For example, some papers associated with a lab that has since separated from MIT were unavoidably included in the set, and we had to manually identify and reject those papers. Some parameters, however, were adjustable by BMC, such as excluding some content types and narrowing the parameters for what papers are considered 'MIT'. Once BMC deploys the changes we have suggested for search parameters, current data suggests we can expect the false hits to drop to something like 20-25%. Given the overall efficiencies we will gain with automated deposit, this seems to us at this stage to be a reasonable match rate. Future work may suggest ways to improve the match rate further.

Outcomes

Although we are still not quite in full production, we are seeing the potential for significant savings from using SWORD for deposit. Automated delivery saves us from having to identify, copy and deposit each paper individually. It also reduces the amount of time we have to spend adding local metadata. So we expect the project to pay off in saved time and a faster-growing repository collection. We have benefited from BioMed Central's willingness to experiment and partner with us, and appreciate their effort to support libraries by finding solutions that speed deposit.

From the BioMed Central perspective, the hope was to pioneer a new methodology through the SWORD integration that would work for other publishers and libraries. Publisher Matthew Cockerill expressed their goals this way: *"We know that campus open access policies have widespread support from faculty members, but the burden of obtaining and uploading the author's version of published articles can limit compliance rates in practice. By automating the deposit of articles published in BioMed Central's open access journals, we aim to help to address*

this problem and to save time for both researchers and administrators. We hope that this collaboration between MIT and BioMed Central will serve as a model for cooperation between research institutions and publishers to deliver a coordinated approach to open access."

Our aim at MIT Libraries has been very much the same: to automate deposit to make it easier for our faculty to make their work openly available through MIT's repository, and to leverage the experiment with one publisher – BioMed Central — so that we could use the SWORD protocol to work efficiently with many publishers. Towards that goal, our next step will be to begin discussions with other open access publishers, and other publishers friendly to sharing the final published version in an institutional repository.

Automation is a key to success in achieving the goal of university and funder deposit requirements: making scholarship and research more widely available. To achieve scalable workflows in support of this goal, we will need automated deposit to operate widely between publishers and repositories. Our experiment with the SWORD protocol has left us optimistic that we are taking a substantial and significant step in that direction.

Acknowledgement

The authors would like to extend a special thanks to UKSG for making this article openly available from the time of publication.

Article © Ellen Finnie Duranceau and Richard Rodgers

■ **Ellen Finnie Duranceau**
Program Manager, Scholarly Publishing and Licensing
MIT Libraries
Tel: 617 253 8483
E-mail: efinnie@mit.edu
<http://libraries.mit.edu/scholarly>

■ **Richard Rodgers**
Head, Software Development Group
MIT Libraries
Tel: 617 253 1663
E-mail: rrogers@mit.edu

To view the original copy of this article, published in *Serials*, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=23&issue=3&spage=212>

The DOI for this article is 10.1629/23126. Click here to access via DOI:

<http://dx.doi.org/10.1629/23126>