Self Introduction: MacKenzie Smith

MacKenzie Smith is the Associate Director for Technology at the MIT Libraries, where she oversees the Libraries' technology strategy and its digital library research and development program. Her research agenda there focuses on Semantic Web applications for scholarly communication, distributed digital library architectures, and digital data curation including long-term data preservation. She was the Project Director at MIT for the DSpace open source digital archiving platform, and has considerable expertise developing successful open source software communities. Prior to joining MIT, MacKenzie was the Digital Library Program Manager for the Harvard University Library, and held several other technology positions at the Harvard and University of Chicago Libraries. Her academic background is in Library and Information Science, and her research interests are in information management, and particularly digital libraries and archives.

http://www.mit.edu/~kenzie/

Publications


Data or Software You Serve

http://dspace.org/
http://simile.mit.edu
http://facade.mit.edu
http://dspace.mit.edu/

General Questions

1) What is (are) your main interest(s) in attending the workshop?

MIT’s 2009 faculty policy mandating Open Access to research publications has created a new requirement that the library accurately know what the faculty have published in traditional journals. While this publication data can be gathered from publishers and aggregators, disambiguating MIT authors is a complex and labor-intensive process that has proven difficult to automate. While MIT faculty have unique identifiers within the context of the institution, the lack of global standard identifiers prohibit automatically matching new incoming publications to known authors. New initiatives like ORCID from the scholarly
publishing industry and ISNI from the library and rights management communities show promise for improving this situation, but would benefit from a better understanding of the use cases and requirements of the organizations and systems that will consume these identifiers in support of new services.

2) What expertise do you have in identifiers for people, scholarly works and/or their association?

My background is in libraries and authority control for published authors; many years of work on the Simile Project to build out the Semantic Web for metadata, including citation data; experience developing workflows for capturing publication data for MIT faculty-authored papers from the major publishers and aggregator databases.

3) What is the most significant effort in identifiers and disambiguation you know?

There are many small initiatives to create identifiers for the research community, usually directed at a specific research domain, but what is really needed is a global, cross-sector, cross-institutional system that research institutions and all types of publishers can share. Two initiatives working on globally unique identifiers for researchers are ORCID and ISNI. Widespread adoption of such common identifier frameworks could provide an automated mechanism for author disambiguation in the future.

Also, scholarly publishers have developed sophisticated internal mechanisms to disambiguate authors who submit articles for publication, but these mechanisms are proprietary and cannot be leveraged by research institutions today. Adopting a common identifier framework such as ORCID could motivate publishers and institutions to share tools and techniques for author disambiguation of the historical literature.

4) What would you like to learn / achieve at the workshop?

Identifying the use cases and requirements for researcher identifiers and disambiguation to feed into initiatives like ORCID and ISNI would be very valuable.

Sharing information about current tools and techniques for disambiguating authors found in citation data from major publishers and aggregators (e.g. PubMed) to see if there would be value in creating a virtual community to share this information more broadly, e.g. by a DuraSpace Solution Community.