

BibApp 1.0 and Beyond: Developing a Piece of the Scholarly Communication Toolkit

Sarah Shreeves, Bill Ingram, Eric Larson, Dorothea Salo, and Tim Donohue

Research institutions, particularly universities and colleges, often face challenges in understanding the range of research, publications, and collaborations occurring within their boundaries. They often struggle to keep up with the research and collaborations happening. Faculty and researchers are sometimes at a loss to find fruitful collaborations on campus. Libraries often lack the data to truly understand the publication patterns and trends among faculty. Repository managers spend much time trying to identify publications that can go into a repository. Faculty and departmental web pages are inconsistently complete and sometimes out of date; annual reporting processes are still sometimes paper-based or are not integrated into an institution-wide workflow. Grants and contracts offices generally can only provide a view of the departments that are heavily reliant on grants. There is seldom one place where an administrator, a faculty member, a funder, a potential graduate student, a subject librarian can go to better understand the research occurring on campus.

Within this environment a number of tools are in development to help fill gaps in managing, displaying, searching, and mining the publication and citation data that are byproducts of the scholarly communication process. Cornell's VIVO (<http://vivo.cornell.edu/>), Harvard's Catalyst (<http://catalyst.harvard.edu/>), MIT's Citeline (<http://citeline.mit.edu/>), and the BibApp, the subject of this paper, from the University of Illinois at Urbana-Champaign and the University of Wisconsin at Madison (<http://bibapp.org> with a pilot installation at <http://connections.ideals.illinois.edu/>) are all examples of such tools. In addition, publishers and aggregators are developing products that mine the publication and citation data of an institution (for example, Elsevier's SciVal product at <http://www.scival.com/>) to provide information on areas of strength, weakness, as well as comparative data.

At its base, the BibApp helps collect, refine, organize, and present information about scholarly output. Through BibApp, institutions can create a campus expert gateway through a mash-up of publication histories and directory information. College and university libraries can use BibApp to better understand publication patterns of faculty and easily identify material for an institutional repository. Developed with feedback from focus groups at UW-Madison, BibApp is really a mash-up of different data sources and open APIs; it utilizes structured citations (from whatever source), directory information pulled from LDAP, publisher policies on self-archiving from the SHERPA/RoMEO API, OpenURLs and the Google Book API in order to facilitate access to the articles and books that are not in the repository, and SWORD, an open protocol that facilitates repository deposit.

More specifically, BibApp accepts citation data in the RIS and BibTeX formats often used by abstracting-and-indexing databases and personal citation managers, as well as through manual input. Citations may be added by faculty themselves, by librarians, by department administrative staff, or by anyone else a faculty member cares to deputize. This helps ameliorate a long-standing problem with citation-gathering applications: most faculty do not care to put in the work to keep their data up-to-date. The BibApp takes a librarian's jaundiced view of the adequacy and accuracy of provided metadata. It flags duplicate

citations for review and deletion. It is also well aware that article citations often do not contain the preferred form of an author's, publication's, or publisher's name, maintaining internal mappings between "pen names" and preferred representations. As the institutional-repository manager's friend, the BibApp checks journal titles and publishers from citations against the SHERPA/RoMEO database of publisher open-access policies (<http://www.sherpa.ac.uk/romeo/>). Given a manuscript or publisher's typeset PDF that is legal for repository deposit, the BibApp can perform repository deposit on behalf of the author automatically via the Simple Web Service Offering Repository Deposit (SWORD) protocol, complying effortlessly with publisher requirements for acknowledgement of the published version as listed on SHERPA/RoMEO.

BibApp is currently in its second release candidate and is expected to be in a 1.0 release in mid-March 2010. We will discuss here the technical specifications of BibApp and will outline its development path past the 1.0 release. Given that BibApp was first presented at Open Repositories 2007 we will touch on some of the struggles of a collaboration that was often resource poor in terms of development. We will give a demonstration of how BibApp can be used by repository manager and discuss the practicalities of populating such a system in terms of data quality issues and copyright. Finally we will discuss some of the challenges (including issues of control over bibliographic and citation data) and opportunities (the range of similar and complementary tools in development) of the environment in which BibApp sits.