Purdue University

Purdue e-Pubs

Charleston Library Conference

Supporting Research Information Management in the Research University: Partnerships, Challenges, and Possibilities

Rebecca Bryant
OCLC, bryantr@oclc.org

Beth Sandore Namachchivaya
University of Illinois at Urbana-Champaign, sandore@illinois.edu

Julie Speer
Virginia Polytechnic Institute and State University, jspeer@vt.edu

Follow this and additional works at: https://docs.lib.purdue.edu/charleston

Part of the Scholarly Communication Commons

An indexed, print copy of the Proceedings is also available for purchase at:
http://www.thepress.purdue.edu/series/charleston.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences.

Supporting Research Information Management in the Research University: Partnerships, Challenges, and Possibilities

Rebecca Bryant, Senior Program Officer, OCLC Research

Beth Sandore Namachchivaya, Associate University Librarian for Research and Associate Dean of Libraries, University of Illinois at Urbana-Champaign

Julie Speer, Associate Dean for Research and Informatics, Virginia Tech

Introduction

North American universities and research institutes are increasingly engaging in complex efforts to collect and synthesize information about an institution's research footprint. Broadly defined, research information management (RIM) is the integrated collection of the scholarly outputs of its researchers by a research institution. RIM as defined here is analogous to current research information system (CRIS), a concept widely used in Europe but largely unfamiliar in the United States. RIM is also inclusive of other terms sometimes used within this emerging landscape, such as research networking system (RNS) and research profiling system (RPS).

RIM systems aggregate research metadata from both internal and external data sources. Internally, institutions may pull together information such as job titles and organizational affiliations, courses taught, grants awarded, patents, honors, publications and datasets, and campus committees and service. These data are usually sourced from many different systems that vary broadly by institution but may include the enterprise data warehouse, student information system, campus awards management system, institutional repository, and many others. In addition to system-delivered information, which can be refreshed regularly, manual entry may be needed to provide information that is otherwise unavailable. Many institutions also source data from external services such as Scopus, Web of Science, PubMed, and others, particularly for the collection of publications metadata.

In the United States, research information management implementations can demonstrate at least five specific use cases:

- Public researcher expertise profiles
- Faculty activity reporting (FAR) workflows
- Open access support and integration with campus institutional repositories
- Reporting and benchmarking
- Reuse and interoperability

The adoption of public research expertise profiles such as the Experts system at the University of Minnesota (experts.umn.edu) is one of the most common use cases in the United States. In these systems, the institution aggregates the research outputs of affiliated researchers into a public, searchable portal to increase institutional and researcher visibility and discoverability. In a second, less widely adopted use case, institutions such as the University of Arizona (uavitae.arizona.edu) have implemented campus-level workflows to support faculty review and activity reporting. Within these systems, faculty are incentivized to maintain information about their research outputs within the single integrated system of record. In a third case use, institutions may also use their RIM systems to support both local and federal open access (OA) policies. In these cases, such as https://scholarworks.duke.edu/elements/ at Duke University, the RIM may be used for enhanced identification and tracking of OA-eligible publications. It may also support notifications to researchers to encourage self-archiving of eligible works through integrated workflows that support deposit into the local institutional repository.

As an institution aggregates information about its research footprint, it can also use this information for a fourth use case: Improved reports, dashboards, and benchmarks. Institutions may use RIM

information for customized decision support outputs at the department, college, and campus level, and as institutions increasingly think of RIM as a primary "system of record" or "data warehouse," institutions can save time by entering once and reusing often. RIM information can be reused in many ways, making up a fifth use example. For instance, RIM information can be reused to provide automatic updates to Web pages across the institution. Researchers can extract information to create biosketches and curriculum vitae (CVs). Departments and institutional reporting professionals can extract information to support academic program review and accreditation activities.

Because RIM implementations support numerous campus goals, including open access compliance and enhanced discoverability, faculty reviews, and internal campus reporting, there are many campus stakeholders. Libraries are usually active participants, but not always, and other prominent stakeholders include the vice president for research and research office, provost, chief information officer, as well as data warehouse and institutional reporting professionals. Depending upon campus goals and organization, other stakeholders may include the graduate school, academic colleges and departments, technology transfer office, campus advancement and corporate relations, and campus news bureau.

Case Study: The University of Illinois' Experts RIM System

The University of Illinois at Urbana-Champaign is one of the original 37 public land-grant institutions established after President Abraham Lincoln signed the Morrill Act in 1862. It is a comprehensive and major public land-grant university (doctoral/research) that is ranked among the best internationally. Illinois' decision to implement a campus-wide RIM system is a natural outgrowth of its strong international reputation as a research-intensive institution, and its interest in making research and educational output accessible to a broad audience of scholars. Institutional context—in this case observing what is important to academic success at an institution and developing services and programs to help drive that success—is a critical element in the development of RIM services. It requires engagement across campus and on many levels—with undergraduates, graduate students and

postdocs, faculty, staff, and administrators. Beginning with the library ROI study that Paula Kaufman and Judy Luther performed at Illinois to determine the value of e-journal access to researchers' grant funding success, the University of Illinois Library began to identify strategic actions that supported the growing research environment at Illinois. Following swiftly on the heels of the ROI study, the libraries at Illinois and the University of Wisconsin collaborated to develop and make available through GitHub an open source "campus research gateway and experts finder" named BibApp.

After a pilot implementation of BibApp in 2011 that was led by the Library, the Office of the Vice Chancellor for Research (OVCR) charged a campus task force that recommended implementing a campus-wide implementation of a RIM system. The Elsevier Pure RIM system was selected in late 2014, led by the OVCR and implemented in late 2015 as a beta service by the library, working in partnership with the campus. Initially named Illinois Research Connections (IRC) and later dubbed "Illinois Experts," the campus articulated several objectives for the Illinois Pure implementation, which is directly focused on making accessible the scholarly accomplishments and expertise of Illinois researchers:

- Showcase Illinois research expertise to external stakeholders.
- Connect researchers with potential collaborators, and encourage interdisciplinary research.
- Automate publication data collection from reliable source(s).
- Enable units and individuals to make timely updates to profiles.

The initial beta launch in late December 2015 made public over 1,800 profiles for tenure-stream faculty primarily in science, technology, engineering, and mathematics (STEM) fields. Approximately 300 profiles for faculty in humanities, arts, and qualitative social sciences disciplines (HASS) were initially not made visible, pending enhancement with citations and references to publications and works that were not indexed in the Scopus database. Illinois aims to add the profiles for several hundred specialized faculty and other professional scholars on campus by mid-2017. The full implementation

will include profiles for up to 2,500 researchers on campus, aggregating scholarly accomplishment information for tenure-stream faculty, specialized faculty, as well as other scholars who are substantially engaged in research. The database encompasses all disciplines, academic colleges, and units, across the campus—over 150 units in all.

Experts Database Use Cases at Illinois

In addition to the several core uses we identified that involved showcasing research within the University and the general public, connecting interdisciplinary collaborators, the Illinois community has started to identify additional productive uses of the Experts database. The database has been used to identify potential reviewers with needed expertise for books and articles, grants, fellowship competitions, and promotion and tenure cases. Students and prospective students at all levels of their studies have begun to use the database to identify faculty advisers, mentors, and dissertation committee members who are doing research in the students' area of interest. The campus' corporate relations unit as well as the campus office of proposal development uses the database to identify scholars whose area of expertise match a funder's research interest. Internal and external media outlets use the database to identify expertise they can interview on current events or research topics. Further, the general public, legislators, alumni, potential corporate partners, the board of trustees— literally anyone can view the breadth, depth, and significance of the institution's scholarship at both the individual and collective levels.

Shaping Content and Policy Through Governance

At the outset of the Experts project, the vice chancellor for research and the university librarian appointed a faculty governance committee to provide feedback on the content and related policies and practice for the database. This initiative was highlighted in the campus strategic plan, with the goal of making scholarship representing all disciplines on campus accessible and searchable from one portal. The governance committee provides invaluable advice and perspective on core issues in the implementation. A topic the committee addressed early in the implementation was how exhaustive the coverage of each profile would be. The Pure system derives citations for scholarly works

from the Scopus database, which provides excellent and exhaustive coverage for STEM scholarship, but it does a less satisfactory job of coverage for the humanities, arts, and social sciences (HASS) disciplines. Consequently, following the initial loading of 1,900 faculty profiles into Pure, we found that approximately 300 profiles for HASS scholars required remediation. We initially set these HASS profiles so that they do not display publicly until we add citation information to each profile that represents the body of work for each scholar. Since the campus focus is on access to Illinois scholarship rather than activity reporting, the project committed to developing a profile for each scholar that includes sufficient citations to accurately represent his or her current research focus. The notion of "representative" vs. "exhaustive" has enabled the project to identify reasonable goals for remediating profiles for HASS scholars that were not representative of their research with the initial loading of data derived from the Scopus database. The governance committee provided additional thoughtful perspective on research metrics that prompted us to initially remove the display of H-index and Altmetric donut information from individual profiles.

Profile Remediation

With approximately 300 HASS profiles requiring remediation, we performed some initial analysis of our options for efficient and bulk upload of citation data. The Pure system supports three methods of data import: Manual entry, import using a bibliography in RIS or BibTeX format, or automatic import using Pure's import module or the bulk import feature. Manual creation of citations using "cut-and-paste" of citations from an existing curriculum vitae is time consuming but accurate in most cases. The project team would need to obtain curriculum vitae from each of the 300 scholars. In the second case, importing an existing bibliography is a viable and time-saving option if a scholar maintains a Google Scholar or Mendeley profile. However, each scholar would have to provide this file to us, and we have found that few arts and humanities scholars appear to use either Google Scholar or Mendeley. The Pure import module allows an individual researcher to search external data sources such as Online Computer Library Center (OCLC) Worldcat, CrossRef, and several other sources in a rudimentary way and add citations one

by one. Researchers can also turn on automatic searches from Web of Science and Mendeley, but so far, this has not shown many benefits. Preliminary analysis suggested that we focus on importing citations to books, monographs, and book chapters for HASS scholars. For this reason, we targeted the OCLC Worldcat database as a rich source of data for these types of publications. We developed a set of scripts that first allowed us to verify an author's name form using the virtual international authority file (VIAF) prior to searching the OCLC Worldcat database using the Worldcat Discovery API for publications for which there was an exact match with the VIAF-verified author name form.

The combination of scripts we used to search the VIAF and the OCLC Worldcat database retrieved 3,216 publication records for 497 individual researchers in the Illinois Experts database. Of those 497 individuals, we were specifically focused on retrieving publications for the humanities, arts, and qualitative social scientists, and those researchers comprised about 340 of the 497 people for whom we imported records. In that HASS subgroup, we were able to import 681 records for 340 people—that is roughly two records per person. While these results were useful, producing accurate citations to works that were not previously imported in the Scopus data, they did not generate the expected volume of citations for HASS scholars that would have resulted in significant increases in profile content. An analysis of the scripts we used to search the VIAF registry suggests that the scripts were too restrictive, focusing on exact matches in the VIAF. We are currently revising the scripts to incorporate "near match" heuristics that can identify results for human review and validation prior to loading this information into the Pure database.

The task of populating each scholar's profile with accurate and complete information is one that ought to be addressable today using automated methods, re-purposing existing publication metadata as far as possible. We intend to continue focusing effort on developing automated methods to identify and import citations and advocating with the Pure development team to enhance the utility of the Pure import module. A reasonable goal for libraries and publishers should be to work collectively (not at the individual institution level) to address the challenges inherent in using existing vetted metadata to build and sustain accurate and representative profiles of scholarly works.

Work in Progress

Illinois continues to augment the content of the Experts database, with plans to add patents and grants in the near future. In late fall 2016, Pure anticipates adding news and media information. Further, the library has initiated marketing as well as training workshops for library subject liaisons, communications staff in academic units that support research, and individual faculty to enhance the content of researcher profiles. A campus implementation of ORCID in 2017–2018 will improve our ability to disambiguate author name information. The library has identified several early adopter academic units on campus that are eager to ingest researcher information from the Pure database into their local data sources and Web pages, thereby eliminating the need for individual units to duplicate this information locally.

Translational Uses

Even at this early stage of implementation, we can identify translational uses of the database as a discovery entry point, as well as use of the database itself to replace previous services and systems. One research institute on campus recently decided to use the Pure database to aggregate its researchers' scholarly output on an annual basis, replacing an arduous manual publication tracking process supported by the institute librarians. Another translational use is bearing out in the area of discovery. Libraries and users are accustomed to discovery that keys primarily on publication data; however, RIM systems enable research discovery that keys on thematic areas, individual researchers, research centers, and academic units. In other words, RIM systems support discovery that is *not* driven primarily by publication. The flexibility of RIM systems also supports the integration of data *about* the scholarly activities of researchers as analytics within the system, enabling creation of on-the-fly visualizations of collaboration networks. These and other enhancements likely to be introduced point to the rise of content discovery layers that flexibly re-purpose underlying bibliographic and other data for viewing through the lens of a particular group of users.

Case Study: Virginia Tech's Faculty Activity Reporting-Based RIM System

Faculty activity reporting (FAR) is the primary use case for Virginia Tech's (VT's) research information management implementation. However, this discussion will describe how VT's FAR-based RIM implementation demonstrates all five RIM use cases: Researcher profiles, faculty activity reporting, open access (OA) support and repository integration, reporting, and reuse and interoperability.

Virginia Tech, a comprehensive public land-grant university with over $500 million in research expenditures, offers approximately 250 degree programs to over 33,000 undergraduate and graduate students. Virginia Tech Libraries support the teaching, research, and outreach missions of the university through service and innovation in learning spaces, teaching, learning, and literacies, collections access, and research and curation. Research and curation initiatives are largely offered through the Libraries' Research and Informatics division, which is comprised of units responsible for the management and delivery of new digital research and scholarship services (including repositories and OA support) as well as with effecting strategies for mapping and integrating the libraries' resources, services, and expertise to the university's research enterprise.

In 2012, with faculty input on scenarios for reducing barriers to open access (OA) and based on informal surveying of the research information landscape, the libraries initiated a partnership with the Provost's Office and the Office of the Vice President for Research and Innovation (OVPRI) to explore, in concert with other university data management initiatives, strategies for embedding the libraries' VTechWorks open repository service in existing faculty workflows. The partnership led to the libraries' involvement in selection of a new platform for the electronic faculty activity reporting system (EFARS), a process that involved stakeholders from across the university: The Provost's Office (as leaders of the EFARS initiative), Office of the Vice President for Research and Innovation (OVPRI), Information Technology, and University Libraries. Symplectic Elements was selected as the new EFARS platform in 2013, and implementation began in early 2014. Launch of the new Elements-based EFARS began in the fall of 2016.

Faculty are encouraged to curate their own data within EFARS, which can be used to generate several kinds of faculty activity reports, as well as to populate other external university websites and public facing profile systems. Much like other RIM implementations, a large portion of the EFARS faculty activity data are harvested into the system automatically via external publisher data sources to avoid unnecessary manual data entry; data sources include, but are not limited to Web of Science, SSRN, DBLP, PubMed Central, arXiv, ORCID, and Altmetrics. Campus data sources also offer job title, grants, teaching, and some legacy publications and professional activity data. Custom reports developed by the Office of Institutional Research and Effectiveness are based on university guidelines and offer faculty the ability to use EFARS to generate annual activity reports and promotion and tenure dossiers, in addition to the platform's default curriculum vitae and funding agency biosketch reports.

Repository integration, a key feature of the new EFARS platform, provides faculty users responsible for curating their profile data with a way to quickly and easily deposit works to VTechWorks without introducing separate external submission workflows. The platform displays publisher self-archiving copyright policy information provided by SHERPA/RoMEO, as well as VTechWorks service information provided by the libraries' repository team, to assist faculty in understanding which version of a work can be shared publicly in an institutional repository. Files and metadata are then deposited to a single collection in the repository, and uniform resource identifiers (URIs) are sent back to EFARS to be included in faculty members' publication records. Repository team members then map publications to appropriate college, institute, or department level collections in VTechWorks' DSpace-based repository system.

One of many positive outcomes of the repository integration is the ability to automatically include links to VTechWorks OA publications in any reports generated using EFARS and in any public-facing researcher profile systems that consume EFARS data. CollabVT, the library-managed VIVO implementation, is one example of such a system. CollabVT is a public-facing researcher profile application based on EFARS curated data. Links to

VTechWorks records will be accessible in CollabVT profiles. While still a work in progress, CollabVT will offer well-structured and faculty-curated grants and publication data in an openly accessible platform that can be used to identify collaborators and showcase research expertise.

Implementation of EFARS, VT's primary RIM use case, is supported by the libraries in a variety of ways. The libraries manage local data feeds, provide general application administration, help troubleshoot external publication data feed issues, manage repository and VIVO integrations, provide user support for publication and ORCID issues, create educational content in the form of instructional videos and LibGuides, and offer workshops on related scholarly communication topics such as researcher profile systems, open repositories, altmetrics, and open access. Engaging with the university's RIM initiative has not only provided the libraries with an opportunity to strengthen the university's overall research infrastructure, but to deepen our engagement with faculty in scholarly communication issues and accomplish our goal to effectively map and integrate the libraries' digital research and scholarship services, expertise, and resources to new university research environments.