

This article was downloaded by: [the Bodleian Libraries of the University of Oxford]

On: 11 April 2014, At: 05:08

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## New Review of Academic Librarianship

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/racl20>

### The Role of an Entity Registry in Scholarly Communication: Exploring Creative Uses of Research Activity Data

Sally Rumsey<sup>a</sup> & Cecilia Loureiro-Koechlin<sup>a</sup>

<sup>a</sup> SERS, Bodleian Libraries, University of Oxford, UK

Published online: 20 Oct 2010.

To cite this article: Sally Rumsey & Cecilia Loureiro-Koechlin (2010) The Role of an Entity Registry in Scholarly Communication: Exploring Creative Uses of Research Activity Data, *New Review of Academic Librarianship*, 16:S1, 17-27, DOI: [10.1080/13614533.2010.505831](https://doi.org/10.1080/13614533.2010.505831)

To link to this article: <http://dx.doi.org/10.1080/13614533.2010.505831>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Versions of published Taylor & Francis and Routledge Open articles and Taylor & Francis and Routledge Open Select articles posted to institutional or subject repositories or any other third-party website are without warranty from Taylor & Francis of any kind, either expressed or implied, including, but not limited to, warranties of merchantability, fitness for a particular purpose, or non-infringement. Any opinions and views expressed in this article are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis.

The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor & Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Taylor & Francis and Routledge Open articles are normally published under a Creative Commons Attribution License <http://creativecommons.org/licenses/by/3.0/>. However, authors may opt to publish under a Creative Commons Attribution-Non-Commercial License <http://creativecommons.org/licenses/by-nc/3.0/> Taylor & Francis and Routledge Open Select articles are currently published under a license to publish, which is based upon the Creative Commons Attribution-Non-Commercial No-Derivatives License, but allows for text and data mining of work. Authors also have the option of publishing an Open Select article under the Creative Commons Attribution License <http://creativecommons.org/licenses/by/3.0/>.

It is essential that you check the license status of any given Open and Open Select article to confirm conditions of access and use.

## THE ROLE OF AN ENTITY REGISTRY IN SCHOLARLY COMMUNICATION: EXPLORING CREATIVE USES OF RESEARCH ACTIVITY DATA

SALLY RUMSEY and CECILIA LOUREIRO-KOECHLIN

SERS, Bodleian Libraries, University of Oxford, UK

*The Oxford University Research Archive (ORA) is the University's open access repository for research outputs. To simplify deposit in ORA, a registry has been created containing data harvested from existing sources to be used as repository metadata. The registry stores publicly available research activity data, that is, data about research including people, projects, and funders. Data held in the registry are available for purposes beyond the repository, particularly as the registry uses semantic web technologies which expedite data sharing. Value is added by aggregating disparate data for discovery, by making creative use of data such as revealing connections between entities and recording data provenance. This article describes the entity registry, its role within ORA, and as a tool to support scholarly communication. The advantages of storing data as entities and gathering, aggregating, and displaying research activity data to assist dissemination of research are explained. Examples of use of the registry and its data are provided that enable easy discovery of researchers and their activities and reveals hidden connections between them, thereby, encouraging communication and collaboration between different subject areas. Development of the registry and examples has been underpinned by extensive stakeholder analysis and user testing.*

**Keywords:** entity registry, semantic web, data aggregation, institutional repository, research activity data

### Introduction

New technologies improve the efficiency of scholarly communication processes and open up possibilities which were not available before. These advances have the potential to transform scholarly communication in a profound manner (Williams & Lawton). Among all these technological innovations, Institutional

---

Address correspondence to Sally Rumsey, ORA Service & Development Manager, SERS, Bodleian Libraries, Osney One Building, Osney Mead, University of Oxford, Oxford, OX2 0EW, UK. E-mail: sally.rumsey@bodleian.ox.ac.uk

Repositories (IRs) have been widely implemented in research intensive higher education institutions (See, e. g., Markey et al. and van Westrienen & Lynch). IRs offered numerous benefits to researchers and their institutions as they capture, preserve, and provide access to the digital work products of a community (Foster and Gibbons). They also help in promoting the institution's research profile and providing global visibility for research outputs.

According to Roosendaal and Geurts, there are four functions in scientific communication: registration, certification, awareness, and archiving. Among these functions, the "awareness" function, "the real engine in the communication process," is the most difficult to accomplish. Awareness allows actors to remain informed of new claims and findings (Van de Sompel et al.) IRs play an important role here as they facilitate the dissemination of research outputs. However, there is scope for utilizing related technologies to fulfill the awareness function. Systems complementary to IRS could be used to improve awareness, and thus scholarly communication, by creating alternative channels of communication and by using contextual information, which is easier to distribute but relevant to the understanding of how research outputs are produced.

Van de Sompel et al. and Rieger state that any system designed to support scholarship should be aligned with scholarly endeavors, reflecting the kinds of data scholars create as well as the ways in which they use them. This will contribute to making IRs more useful. Ismail et al. suggested IRs should connect to other kinds of information to assist users, particularly novice researchers, get insights into research topics, as well as experts. This information should also reflect the different aspects and stages of scholarship activities, one of which is access to and dissemination of relevant literature. To this end, contextual information held in IRs can contribute to this supporting environment and be used for other related purposes.

This paper presents an initiative to support research processes and scholarly communication activities which uses and goes beyond institutional repositories. The Building the Research Information Infrastructure<sup>1</sup> (BRII) project at the University of Oxford aims to collect and preserve data that describe research taking place at the University. These data are stored in an entity registry and made available for sharing with a variety of services.

Information about research is made available not only for researchers but to support the management of research across the University, for administrative and academic departments and groups, and senior members of the University, making it easier to discover and re-use these data. This, we expect, will complement current efforts at facilitating and disseminating research outcomes by providing overviews of research activities in an institution. The technology underpinning the service described here introduces innovative solutions for data gathering and processing currently not common in IR systems.

The next sections introduce the Oxford University Research Archive (ORA) and the entity registry containing research information. The paper then discusses ways in which the registry and its data can be used to aid university administrators, academics, and strategists carry out research related activities.

### **The Oxford University Research Archive**

The University of Oxford has implemented an institutional repository, ORA for the preservation, publicity, access, and delivery of research outputs. ORA is the first of Oxford's federated repositories built within the Oxford DAMS (Digital Asset Management System). The DAMS comprises a storage layer below the digital object management layer with services added as required. The software packages used in the DAMS are selected to serve specific purposes. This results in a modular structure that is flexible and easy to maintain and update.

In an effort to simplify and encourage more deposits in ORA, it was deemed advantageous to add existing data automatically to metadata fields without effort on the part of the depositor. These data could be used to pre-fill fields or within an auto-complete function. The data being considered were those already held within the University in sources such as websites which promote people and activities in departments, research groups, and projects. Other publicly available sources of information include funders' websites, and academic journals and databases. The information held in those sources not only resembles the kind of information needed to deposit items in IRs but has the potential to offer much broader benefits. However, those sources are diverse, dispersed, and each is constructed to address local needs. Use of a

registry to gather these data had been partially outlined for ORA to facilitate deposit and enhance data exchange and research visibility between the institution and the wider world. This vision was advanced and transformed into a service with much broader innovative purpose.

The BRII project at the University of Oxford was undertaken to develop an entity registry to gather dispersed data and to create additional services that take advantage of the architecture and technologies of the DAMS repository system. The aim of the project is the efficient sharing of Research Activity Data (RAD) using semantic web technologies (Rumsey 175). RAD comprised standard information that described the researcher and their research: people, projects, funding agencies, and so on. The objective is to collect and process RAD from all these disparate sources and make them available for re-use. By doing this, value is added to the data which can then be used to support research dissemination and publicity by being shared.

### **The Entity Registry**

The registry is a tool to enable research activity data to be harvested, processed, stored, and re-used. Data are harvested from existing sources and mirrored in the registry. Working this way respects the Oxford devolved model of governance and allows data providers to retain ownership of their data. Entities are extracted from the harvested data, dividing data into their constituent parts, which are then deposited into an entity store. Entities are basic “types” of data, for example, person, funder, or project. New entity types can be added as required. Handling the data in this way provides a flexible system where data can be manipulated and processed. The registry provides the ability to aggregate and re-constitute it in different combinations to fulfill different needs. Data harvested from departments and stored in the registry can be pulled back by those same departments to be re-combined in new ways.

Methods for co-referencing are being incorporated to identify individuals accurately. As data are pulled into the registry, entities are grouped together based on pre-set rules, heuristics, or user-driven choices. This work is still in progress and is being

informed by developments at the University of Southampton (Glaser, Jaffri, and Millard).

Semantic web technologies are employed in the registry. Entities and the relationships between them can be described in ways that both humans and machines can understand. Ontologies and vocabularies such as SKOS (Simple Knowledge Organization System) and AIISO (Academic Institution Internal Structure Ontology) are used for this purpose. The ARPFO<sup>2</sup> (Academic Research Project Funding Ontology) ontology has been created as part of the project. These vocabularies, together with the fact that data are provided as linked data,<sup>3</sup> support data re-use, and manipulation. They also provide a machine readable structure for the data and explicit meaning for terms and relationships that support scholarly communication. The registry is designed so that data can be easily extracted using common formats including RSS/Atom feeds.

Easier deposit in ORA will be achieved by the repository “recognizing” a person when they log-on and then pre-populating records. It will enable the repository to offer an auto-complete function to assist with completion of fields.

### **Uses of the Entity Registry and RAD**

Currently, websites are used as key vehicles across the University to promote research. Departments and researchers may have more than one website containing new and/or duplicated information about their work. This is particularly common when researchers participate in interdisciplinary research. Because of this, a user might have to access a number of different websites to find out about a person and their research. This could be a time consuming task and there is no guarantee that the user will find all the relevant information needed.

To demonstrate the benefits of the registry and use of the RAD held in it, the BRII project has developed two examples: the Oxford Blue Pages and a themed website.<sup>4</sup> These two examples show how RAD can be used on their own beyond the scope of ORA. Feedback from a stakeholder analysis (Loureiro-Koechlin “Stakeholder Analysis”) was used in the design and development of these applications. The Oxford Blue Pages is a directory of research expertise that accesses the entity registry and offers several

ways to view and search through its information. The Blue Pages displays the registry's entities as objects representing different aspects of research: people (i.e., Oxford academics), research projects (or research activities), funders (or sponsors), and academic units (e.g., departments, institutes).

It is expected that the Blue Pages will play a role in the dissemination of Oxford's research outputs, as well as the discovery and sharing of research knowledge and expertise. Being a single point of access for all RAD in the registry, the Blue Pages could become a gateway to information which is originally stored in departmental and other websites across the University. Researchers, administrators, and strategists currently access RAD from their own local original sources. However, as this information is dispersed and disparate it becomes difficult to discover and access when outside the local community. The Blue Pages addresses this problem by aggregating RAD and building connections between research entities. For example, by connecting and displaying in one place biography, research interests, publications, and collaborations which belong to the same researcher but which were collected from different sources. Several uses were identified for these functionalities during the user testing of the Blue Pages (Loureiro-Koechlin, "Uncovering User Perceptions"). These uses are:

- (1) Identification of particular expertise and resources within the University.
- (2) Identification of research opportunities at individual academic, departmental, and University levels by connecting information from different areas and subject fields (e.g., potential collaborations and interdisciplinary research).
- (3) Automatic data aggregation which save administrators and strategists' time and make some administrative procedures (e.g., writing reports) more efficient.
- (4) Improvement of academics, departments, and University's research exposure nationally and internationally among research communities and sponsors.
- (5) As it provides an institutional and standardized view of all research taking place within the University, the Blue Pages will facilitate access to RAD to non experts such as students.



The themed website developed as part of BRII uses information about research opportunities in the Medical Sciences division to target potential graduate students. Registry data is easily accessed and displayed via an API (Application Programming Interface). The API accesses subsets of data to create entire websites, webpages or sections within a webpage to target different audiences such as Oxford researchers, funders, and students. The impact of this and other APIs created by BRII is significant, particularly to assist the creation and maintenance of departmental or project websites. Most importantly, data from the registry can be used to create themed websites which combine data gathered from multiple sources. By searching through the registry, data can be gathered about a particular research topic and selected to be published. This includes links to publications and their full texts if available. The API increases efficiencies as data can be acquired from multiple sources with a few clicks. The BRII project has developed examples to show this, including an exemplar “widget” to enable the easy creation and updating of websites. A short screen-cast demonstration of this in action can be seen in the Medical Sciences website (Medical Sciences Division 2008).

The idea for the Medical Sciences graduate opportunities website arose from a strategic priority expressed by divisional staff in charge of recruitment and who had identified needs of potential students. Due to the large numbers of websites and sources of information in the division, students and new researchers may find it difficult to search for and find information including relevant publications. In particular, students would like to learn about specific topics of research and doctoral research opportunities, as well as supervisors. Before applying for a course, students need to understand the areas of research being undertaken by members of the division, grasp the basics of specialist language, and learn about the experts and projects within the division. The graduate opportunities website offers non-researchers an easy way to find information from across the Medical Sciences Division. This is expected to encourage applications and help admission processes.

The themed website reorganizes and presents sets of aggregated medical sciences data to audiences different than the initially intended by the data owners. Although that the original, individual intentions may have not been the ones of recruitment (but promotion or dissemination among academic circles, for

example); these sets of data have been put together and re-used for a different common purpose.

In addition to the aforementioned, the registry could, in future, provide contextual metadata for research output datasets, such as author (creator), affiliation, title of dataset, and funding agency. By saving the data creator effort when creating metadata, the registry offers direct support for the emerging landscape of data storage and access as more researchers retain and provide access to datasets.

### **Dissemination and Publicity of Research**

The aforementioned section has presented some examples of how RAD can be used to promote and provide overviews of research to wider audiences. These can be useful at academic, administrative, and strategic levels. At academic levels, information about experts and their activities can be obtained to help researchers keep up with their fields of research. This information can be used to build up interdisciplinary collaborations, as it helps researchers access information that is beyond the boundaries of their fields. On the other hand, researchers can use RAD and the registry to promote their own activities. By regularly updating their information in their departmental and/or project websites, their activities and publications can be disseminated through the registry within and outside the University and across research fields.

At administrative levels, the registry can facilitate and make the collection of information about research more efficient. This information can be useful to help researchers in their activities by discovering research opportunities across departments and for administrative processes such as writing reports. Having their data in the registry makes departments and their staff more accessible and visible. This means that departments can promote their strengths, their researchers, and resources, and publicize their activities with little effort. Additionally, connections between their staff and activities can be made with others in other departments. These connections are of interest to administrators in University departments who support and facilitate inter-disciplinary collaborations of their research staff. At a departmental and University level information can be collected from the registry to write

reports, research profiles, brochures, and publicity leaflets for recruitment and promotional activities.

At strategic levels, RAD can be used to assess the performance of researchers, levels of success with funders, and identify research strengths and weaknesses. This information is useful to plan future avenues for development in research groups and academic departments.

## Conclusions

While the entity registry shares technology with ORA, conceptually it is a different system which has complementary purposes to ORA's. Its construction means it supports research processes and scholarly communication activities both of which use and go beyond the institutional repository. The registry is aligned with scholarly endeavors as it harvests existing data created by scholars and others which can be exploited to enhance existing modes of dissemination and to offer new ways of sharing research information. This contributes to scholarly communications by providing access to outputs and by providing descriptions of the topics and experts involved in research. This is a valuable and efficient method to extend the ability to inform and engage with colleagues and potential collaborators. Information about research is made available not only for researchers but in other research related purposes, including for administrative, management, and strategic ends.

The use of advanced semantic web technologies, including linked data, to develop the registry and to manage research activity data means that data can be harvested from a number of disparate and dispersed sources. These technologies enable provenance information to be captured and stored with the object. One benefit of this is that numerous sources of data which were previously seen as disconnected and disparate are organized and aggregated, representing one cohesive picture of research activities in the University. This picture keeps its connections with the original sources, thus, allowing data owners to retain ownership and the system to update information on a regular basis. Also, the use of APIs to access the registry allows interested parties to obtain and display the data in ways that are more convenient to them. This saves time and effort for data users and

allows them freedom in how they use the data. In addition, the use of entities to organize data facilitates data management and building connections between digital objects (e.g., researcher X “is part of” project A; researcher Y “has collaborators” B, C, D.) These connections help to reveal research trends and previously unknown common interests as well as past, current, and potential collaborations and interdisciplinary research. However, most importantly, entities help users to make sense of the vast amount of data offered. This is because users can see the entities as the different aspects of real life research. In this way, entities can bring some order to the complex academic information environment.

We expect that the registry will benefit not only Oxford’s internal audience (academics, administrators, etc.) but also individuals and institutions outside who are not familiar with the organization and its structure (funders, external scholars). This service constitutes a user-friendly face of the University, benefiting particularly the novice (i.e., graduate students) as expert information will be easily searchable and accessible through the Blue Pages. Also, in the future, data could be exported for sharing with government and other bodies for improved scholarly communications.

### Notes

1. More information about the BRII Project can be found at the project’s website <http://brii.bodleian.ox.ac.uk/> and weblog <http://brii-oxford.blogspot.com/>
2. The ARPFO ontology can be found at <http://vocab.ox.ac.uk/projectfunding/schema>
3. <http://linkeddata.org/>
4. At the time of writing the Blue Pages and the themed website are in a pilot stage.

### References

- Foster, Nancy F., and Susan Gibbons. “Understanding Faculty to Improve Content Recruitment for Institutional Repositories.” *D-Lib Magazine* 11.1 (2005). Print.
- Glaser, Hugh, Affraz Jaffri, and Ian Millard. “Managing Co-reference on the Semantic.” WWW2009 Workshop: Linked Data on the Web. (LDOW2009) 20 (2009) Madrid, Spain. Web. 10 March 2010.

- Ismail, M. A., M. Yaakob, and S. A. Kareem. "Semantic Support Environment for Research Activity." *Journal of US-CHINA Education Review* 5 (2008): 36–51. Print.
- Loureiro-Koechlin, Cecilia. "Stakeholder Analysis: Exploratory Study into the Requirements and Uses for Research Activity Data at the University of Oxford," 2009. Web. 10 March 2010. <<http://ora/objects/uuid%3A1df69991-cd37-445b-a4c7-3573ce80c36e>>
- . "Uncovering User Perceptions of Research Activity Data." *Ariadne* 62 (2010). Web. 10 March 2010. <<http://www.ariadne.ac.uk/issue62/loureiroKoechlin/>>
- Markey, Karen, Soo Y. Rieh, Beth St. Jean, Jihyun Kim, and Elizabeth Yakel. *Census of Institutional Repositories in the United States: MIRACLE Project Research Findings*. Washington, DC: Council on Library and Information Resources, 2007. Print.
- Medical Sciences Division, University of Oxford. *Integration with the Oxford Research Archive Registry*. (2008) Web. 10 March 2010. <<http://webteam.medsci.ox.ac.uk/about-us/activities/directory/oraintegration/>>
- Rieger, Oya. "Opening Up Institutional Repositories: Social Constitution of Innovation in Scholarly Communication." *Journal of Electronic Publishing* 11 (2008). Web. 10 March 2010.
- Roosendaal, Hans E., and Peter A. Th. M. Geurts. "Forces and Functions in Scientific Communication: An Analysis of Their Interplay." *Proceedings of the Conference on Co-operative Research in Information Systems in Physics*. September 3, 1997, University of Oldenburg, Germany. Print.
- Rumsey, Sally. "A Case Analysis of Registering Research Activity for Institutional Benefit." *International Journal of Information Management* 30 (2010): 174–179. Print.
- Van de Sompel, H., S. Payette, J. Erickson, C. Lagoze, and S. Warner. "Rethinking scholarly communication: Building the System that Scholars Deserve." *D-Lib Magazine* 10.9 (2004). Web. 10 March 2010.
- van Westrienen, Gerard, and Clifford A. Lynch. "Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005." *D-Lib Magazine* 11.9 (2005). Print.
- Williams, Susan P., and Fides D. Lawton. "eScholarship as Socio-Technical Change: Theory, Practice and Praxis." 3rd International Evidence Based Librarianship Conference, 2005. Brisbane, Queensland, Australia. Print.