Studying Service and Metadata Models of Research Information Management Systems

Dong Joon Lee[1], Besiki Stvilia[2], Shuheng Wu[3]
[1]Texas A&M University
[2]Florida State University
[3]Queens College, The City University of New York

Abstract

As the uses for research information management (RIM) systems grow, digital curation communities are increasingly interested in building shared, sustainable, community-based models for RIM services and metadata, and connecting those models to researchers' activities and needs. This paper presents RIM service and metadata models and compares the models to researchers' activities and engagement levels in RIM systems. The findings of this study can inform the design of RIM service templates or repertoires for different levels of researcher participation in RIM systems. Future related work will include the collection and analysis of empirical data of the actual uses of RIM services and metadata by researchers.

Contact: djlee@tamu.edu

1 Introduction

Research, as well as the evaluation of scholarly impact for both individual researchers and institutions, has become increasingly data driven. There are growing needs and opportunities to share, reuse, and aggregate data from different contexts. Many different research information management (RIM) systems such as Google Scholar, ResearchGate, VIVO, and ORCID exist, and they all tend to use their own data and metadata models. As the uses for RIM systems grow, digital curation communities are increasingly interested in building shared, sustainable, community-based models for RIM services and metadata, and mapping those models to researchers' activities (Thelwall & Kousha, 2015).

There are many different RIM systems, which support researchers with different functionalities, services, and metadata elements. Likewise, researchers may use and engage with RIM systems differently and have different needs and priorities for RIM services and metadata based on their career status, productivity, and attitudes (Wu, Stvilia, & Lee, 2016). It is essential to explicate and systematically match those needs and priorities to the existing RIM metadata and services to identify potential gaps, and to design effective templates and repertoires of RIM services for different types of researchers. This paper contributes towards that need by examining RIM services and user-editable or user-supplied data in three RIM systems and mapping those to researchers' activities and engagement levels with RIM systems.

2 Research Method

This paper reports on part of a larger study, which examined researchers' use of and engagement with RIM systems. The study used activity theory (AT; Engeström, 1987) and literature analysis to develop an interview protocol. The authors then conducted semi-structured interviews with fifteen researchers from ten institutions regarding their use of and participation in RIM systems. A detailed report of findings from the interviews can be found elsewhere (Wu, Stvilia, & Lee, 2016). This paper analyzed the RIM services and metadata elements of three RIM systems (Google Scholar, ResearchGate, and ORCID) and compared those to researchers' activities and engagement levels identified in the interviews. Since not all three systems specified their RIM metadata and service models, one of the authors manually benchmarked the systems to assemble aggregate sets of RIM services and metadata elements provided by the systems. Next, the authors used conceptual analysis to systematically match those sets to the sets of researchers' activities and engagement levels to address the following questions:

1. What are the services and metadata elements currently used by RIM systems to support researchers' activities?
2. How do RIM services and metadata map with researchers' activities in a RIM system?
3. What is the set of services needed for each type of researcher participation in RIM systems?

3   Findings and Discussion

3.1   Service and metadata models

The analysis of three systems produced a typology of RIM services (see Table 1). The typology comprises three high-level types: research, networks, and systems. Each of the high-level types also included multiple service types and functions. From here on we will refer to the typology as the service model.

	Services	Functions
Research	Profile management	Editing profile; Adding research interests, skills and expertise, teaching experience, education background, employment history, and awards and funding data; Managing contact information; Deleting profile; Exporting profile as a CV; Linking to personal webpage; Adding multiple versions of name; Generating QR code for profile
	Publications	Retrieving data from outside databases; Adding data into outside databases; Adding publications; Self-archiving full text; Providing available PDFs; Linking to full text on the Web; Discovering new publications; Discovering new citations; Recommending readings; Requesting reviews from specialists in relevant fields; Requesting feedback on my paper; Updating the list of publications automatically; Exporting all article metadata from profile; Sorting publication list by published dates, titles, citation counts, and types; Providing co-author networks; Assigning identifiers; Allowing visibility setting
	Metrics	Providing h-index, i10-index, altmetrics score, and citation counts; Generating citation graph by years
	Project management	Creating a project; Editing project details; Adding or removing collaborators; Sharing project updates; Following projects
	My library	Bookmarking articles from the Web in my library
Networks	Connection	Connecting with other researchers; Inviting colleagues; Requesting papers; Endorsing expertise; Sharing to social networking sites; Following publications; Following authors
	Job	Posting a job; Finding the right candidates; Recommending job postings
	Q & A	Adding questions; Answering questions; Editing questions and answers; Following questions
Systems	Account setting	Managing email settings; Setting security and privacy controls; Setting profile as public/private pages; Signing in with a social media account; Signing in using an institutional account; Setting profile photo availability; Adding cover pages to publications
	System integration	Retrieving data from outside databases; Adding data into outside databases
	API	Allowing to collect public data

Table 1. Types and functions of the RIM systems

In addition to the aggregate set of RIM services, the authors developed an aggregate set of user-editable metadata elements in the three RIM systems (see Table 2). All three systems did not require their users to have a profile in order to browse the public data on other researchers' profiles. However, ResearchGate required their users to specify a name, an institution, and an email address as a minimum to create a profile, while Google Scholar and ORCID required only a name and an email address for a profile.

Categories	Metadata Elements
Identity	First name; Middle name; Last name; Alternative first name; Alternative middle name; Alternative last name; Published name; Also known as; Institution; Country; Keywords; OtherIDs; Biography; Gender; Profile photo; Email
Employment	Institution/employer; City, State, Country; Department; Role/title; Start date; End date
Research experience	Position; Institutions; Department; Research group; Time period; Location; City; State; Country; Description
Education	Institution; Field of study; Degree; Department; Time period; Location; City; State; Country
Publication	Publication title; Subtitle; Authors; File; Type; Journal referee; Journal title; Volume; Issue; Page; Day; Month; Year; Topics; Abstract; DOI; Citation type; Description; Identifier type; Identifier value; Identifier URL; Relationship; URL; Language used in this form; Work contributors; Source; Publisher; Editor; Edition; ISBN; Chapter; Book title; Repository link; License; Conference; Location; Affiliation; Patent_Ref.No; Poster_ordinal; Grant number; Report number; Supervisor; Degree; Version number; State

Teaching	Institutions; Department; Time period; Location; City; State; Country; Description
Funding	Funding type and subtype; Title; Principal investigator; Co-investigators; Role; Address; Total funding amount; Description; Start date; End date; Funding agency name; Funding agency city, Funding agency region; Funding agency country; Funding external identifier; Funding contributors_orcid; Funding contributors_credit name; Funding contributors_email; Funding contributors_attributes; Funding contributors_role; Secondary institutions; Disambiguated organization identifier; Disambiguation source; Grant number; Grant URL; Alternate URL
Contact	Location; Website; Phone; Mobile; Fax; Twitter; Skype; Instant messenger; Birthday; Language; Scientific societies; Journal referee; Other interest; ORCID; Time zone; Homepage

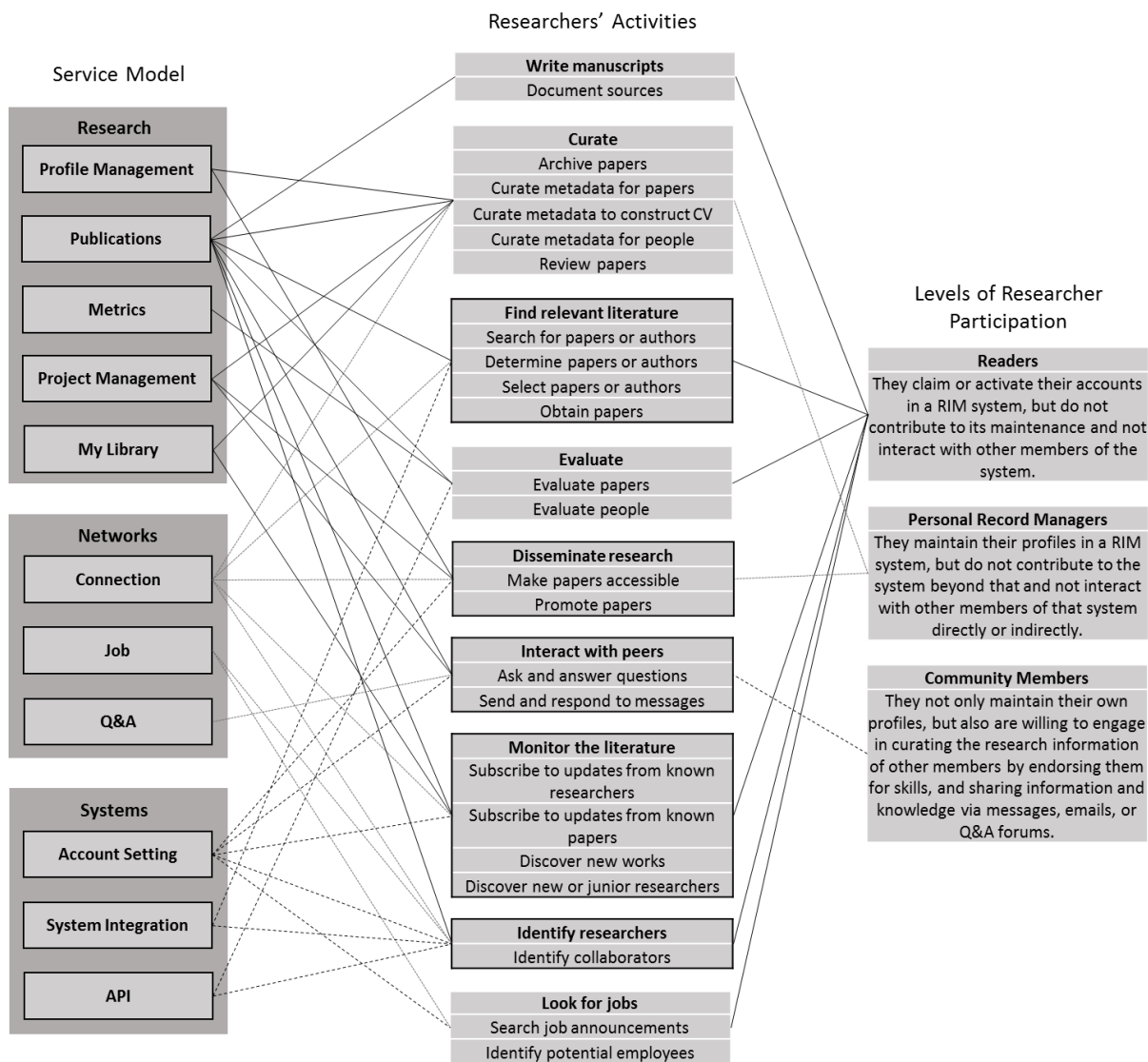Table 2. User-editable metadata elements used by Google Scholar, ResearchGate, or ORCID

Figure 1. RIM system service model mapped to the researchers' activities and levels of engagement in RIM systems

3.2   Mapping

Figure 1 summarizes the mapping between the service model of RIM systems and researchers' activities. Services and metadata grouped under the Publications subcategory can be used to support all of the activities with the exception of the *Look for jobs* activity. The Connection and Account Setting subcategories were the most frequently mapped ones of the Networks and Systems categories. The activities of *Find relevant literature*, *Disseminate research*, *Interact with peers*, *Monitor the literature*, and *Identify collaborators* were mapped to all three categories of RIM system services.

The analysis of interview data (Wu, Stvilia, & Lee, 2016) identified three levels/categories of researcher participation in a RIM system. These categories were determined by clustering/grouping interview participants based on the sets of activities they engaged in when using a RIM system (see Figure 1). Researchers who belong to the "Readers" category may use the smallest set of services. Researchers who are "Personal Record Managers" may use all the subcategories of services that "Readers" use as well as two additional subcategories from the Research category (i.e., *Profile management* and *Project management*). The third group, "Community Members," may use the *Q&A* subcategory of the Networks category along with all the services that "Personal Record Managers" use.

4   Conclusion and Future Research

This paper contributes to digital curation communities' knowledge of designing service and metadata models of RIM systems. It presents the results of conceptual mapping between researchers' activities and the set of existing services and metadata elements provided by RIM systems that can be used to support those activities. Future research will collect and analyze the empirical data of actual uses of RIM metadata and services by researchers.

5   References

Engeström, Y. (1987). *Learning by expanding: An Activity-Theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy.

Thelwall, M., & Kousha, K. (2015). ResearchGate: Disseminating, communicating, and measuring scholarship? *Journal of the Association for Information Science and Technology, 66*(5), 876-889. doi:10.1002/asi.23236

Wu, S., Stvilia, B., & Lee, D. J. (2016). Exploring researchers' participation in online research identity management systems. *Proceedings of the Association for Information Science and Technology (ASIS&T), Copenhagen, Denmark, 52,* Retrieved from http://dl.acm.org/citation.cfm?id=3017447.3017552