

Exploring Researchers' Participation in Online Research Identity Management Systems

Shuheng Wu

Queens College, The City
University of New York
65-30 Kissena Blvd., Queens, NY
Shuheng.Wu@qc.cuny.edu

Besiki Stvilia

Florida State University
142 Collegiate Loop,
Tallahassee, FL
bstvilia@fsu.edu

Dong Joon Lee

Texas A&M University
5000 TAMU, College Station,
TX
djlee@library.tamu.edu

ABSTRACT

Prior studies have identified a need for engaging researchers in providing and curating their identity data. This poster reports preliminary findings of a qualitative study exploring how researchers use and engage in online research identity management (RIM) systems. The findings identify nine activity or task related motivations of using RIM systems. This study also identified three levels of participation in RIM systems: Readers, Personal Record Managers, and Community Members. Most participants of this study fell into the category of Personal Record Managers, who may maintain their own profiles in a RIM system. This suggests that a majority of researchers may be willing to maintain their research identity profiles. Institutional repository managers may consider recruiting researchers as not only research information and data providers, but also curators of their own research identity data.

Keywords

Research identity management systems, motivations, engagement, ResearchGate, Google Scholar.

INTRODUCTION

There are different *research identity management systems*, often referred to as *research information management (RIM) systems*, from publishers, libraries, universities, search engines, and content aggregators (e.g., Google Scholar, ORCID, ResearchGate). These systems employ different approaches to curating research identity information or data: manual curation by information professionals and/or users (including the subject of identity data), automated data mining and curation scripts (aka bots), and some combination of the above. With universities engaging in curating digital scholarship produced by their faculty members, staff, and students through institutional repositories (IRs), some of these universities and their IRs

{This is the space reserved for copyright notices.}

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.

[Author Retains Copyright. Insert personal or institutional copyright notice here.]

try to manage the research identity profiles of their contributors locally (e.g., Expertnet.org, Stanford Profiles). While knowledge curation by professionals usually produces the highest quality results, it is costly and may not be scalable (Salo, 2009). Libraries and IRs may not have sufficient resources to control the quality of large scale uncontrolled metadata often batch harvested and ingested from faculty authored websites and journal databases. They may need help from IR contributors and users to control the quality of research identity data.

The literature on online communities shows that successful peer curation communities that are able to attract and retain enough participants can provide scalable knowledge curation solutions of a quality that is comparable to the quality of professionally curated content (Giles, 2005). Hence, the success of online RIM systems may depend on the number of contributors and users they are able to recruit, motivate, and engage in research identity data curation. There is a significant body of research (e.g., Cosley et al., 2006; Nov, 2007; Stvilia et al., 2008) on what makes peer knowledge creation and curation communities successful. However, most of the previous research has focused on encyclopedia, question answering, and citizen science communities. There has been little investigation on the peer curation of research identity data. This study explores how researchers use and participate in RIM systems. Particularly, it addresses the need to have greater knowledge of how to design scalable and reliable solutions for research identity data curation by examining researchers' perceived value of research identity data and services, and their motivations to engage in online RIM systems. Findings can enhance our knowledge of the design of research identity data/metadata services, and mechanisms for recruiting and retaining researchers for providing and maintaining their research identity data.

RELATED WORK

There have been considerable deliberations on the needs for and uses of research identity data and how to manage that effectively in Library and Information Science (LIS) research and practice communities (e.g., NISO Altmetrics Initiative; Research Data Alliance). An OCLC Task Group aiming to register researchers in authority files identified

five stakeholder groups of research identity data: researchers, funders, university administrators, librarians, and aggregators (OCLC Research, 2014). For the researchers stakeholder group, the Task Group formulated five needs: disseminate research, compile all publications and other scholarly output, find collaborators, ensure network presence is correct, and retrieve others' scholarly output to track a given discipline. This set of needs was compiled based on expert opinions of the Group members, supplemented with a scenario-based analysis. It would be valuable to test this typology empirically and investigate what can be the disincentives for researchers to participate in online research identity data sharing and curation.

Different units in universities (e.g., office of research) are increasingly interested in collecting and analyzing research output for reporting, accreditation, and organizational reputation management. Those activities and interests overlap with traditional interests of academic libraries, which have to better align their digital services with those broader organizational needs and priorities (Dempsey, 2014; Tenopir et al., 2012). One approach would be to add research identity management services to IRs (Palmer, 2013). There is evidence from the practice that adding research identity management services to an IR might increase researchers' interest in the IR (Dempsey, 2014). However, the increased interest in an IR might not always translate in the increased use of an IR and/or increased engagement in research identity data curation as multiple RIM systems (e.g. ResearchGate, Academia.edu) offer similar services and strive for researchers' attentions and contributions. Relying solely on automated mining, extraction, and aggregation of research identity data might result in poor quality. Libraries need researcher engagement in identity data curation to provide scalable and high quality research identity management services.

The online community literature shows that volunteer knowledge curators in open peer-production systems like Wikipedia are mostly driven by intrinsic motivations such as their interests in specific areas (Nov, 2007; Stvilia et al., 2008). Previous studies also examined user motivations to contribute in other online communities. Ames and Naaman (2007) interviewed 13 'heavy' users of a Flickr application and identified four types of motivations for tagging: self-organization, self-communication, social-organization, and social-communication. A study of Flickr collections by Stvilia and Jørgensen (2009) listed eight motivations members might have when organizing photographs into groups. Nov et al. (2010) found a positive relationship between the motivation of building reputation in the community and the amount of metainformation (i.e., tags) provided. Similarly, in a study examining an online network of legal professionals, Wasko and Faraj (2005) found a significant positive effect of building reputation on the quality and volume of knowledge contribution. The online communities literature provides valuable insights for designing RIM systems and building and maintaining user communities around those systems. However, more

empirical research is needed to understand what motivates researchers to engage in RIM systems.

RESEARCH METHOD

Guided by activity theory (AT; Engeström, 1987), this study employed semi-structured interviews (Blee & Taylor, 2002) to answer the following research questions:

1. How do researchers use online RIM systems?
2. What are the levels of researcher engagement in online RIM systems?

This study defines the research population as employees and students of institutions having an IR and classified as Research Universities in the Carnegie Classification of Institutions of Higher Education. The participants of this study must have at least one peer-reviewed research publication and have used at least one RIM system by the time of interviews. This study used AT and literature analysis to develop an interview questionnaire. The authors conducted semi-structured interviews with eight researchers from four institutions regarding their use of and participation in RIM systems. One participant was a full professor, one was an associate professor, one was an assistant professor, two were postdoctoral researchers, and three were doctoral students. All interviews, ranging from 17 to 68 minutes, were audio recorded, transcribed, and coded with NVivo 10. Two of the authors independently coded all the interviews using an initial coding scheme based on Activity Theory and literature analysis. After comparing, discussing, and resolving any differences in their coding, the two authors formed a new coding scheme with emergent codes and subcategories and recoded all interviews.

FINDINGS

Activities and motivations

From analysis of the interviews we identified the activities in which researchers engaged using RIM systems (see Table 1) and the motives of those activities.

Find relevant literature

One of the most frequent activities in which RIM systems are used is literature search. Outcomes of this activity can be used as input to other scholarly activities such as literature analysis, manuscript writing, or planning a research project. The literature search activity may include four actions: search, determine, select, and obtain. Researchers may use different RIM systems for different types of searches (e.g., known item, subject, navigational searches) based on the strengths or capabilities of RIM systems. One participant explained how he used ResearchGate and Google Scholar for different purposes:

I think they have different functions. Like for ResearchGate I can follow some people, so I can have their most recent papers. But sometimes I also use Google Scholar when I have a specific paper that I want to look for. So if I know the title of the paper, or I know the author, and I want to see their publications, I will use Google Scholar. It's convenient.

Researchers may also use RIM systems to define and manage their own bibliographies by following or 'bookmarking' the core papers of a specific research area. One participant specified how he used ResearchGate to manage and expand his bibliographies:

Some of the big papers were sort of like in everyone's research. These are the cornerstone articles that you base a lot of your research on ... I follow some of those articles [in ResearchGate].

To complete a literature search activity, researchers need to obtain the desired publications. Researchers may be motivated to use RIM systems providing open access to the self-archived versions of publications. RIM systems with social networking features can attract researchers to contact authors and request a copy of publication they cannot access otherwise. One participant indicated what motivated him to use ResearchGate was that it provides open access to his works and allows requesting papers from others:

It's good to have your stuff easily accessible because not everyone has access to databases, but if you're a researcher, it's easy to set up an account on one of these sites and connect with the authors to hopefully get the articles that you want to get.

Document manuscripts

Besides literature search, researchers may use RIM systems in a manuscript writing activity to manage citations. They may use Google scholar to verify bibliographic metadata of the resources cited in their papers, and/or obtain citations in a specific style. One participant revealed his use of Google Scholar when working on the reference list of his paper:

There are times that I need to verify the source just to make sure the title, authors, and year, and just to make sure the information I put in are correct. Google Scholar is doing a good job in accurately reflecting publications, so I use it as a [citation management] resource.

Identify researchers

RIM systems and their citation extraction and analysis functions can be used for identifying potential reviewers, collaborators, letter writers, students, and advisors, who may have similar or specific research interests. One participant explained how she used ResearchGate's citation information to know other researchers and identify potential collaborators:

One of the advantages to using these [RIM] systems is the ability to discover researchers that you may not have known like this ... I'm going to follow this guy from Boston now because apparently he likes my work and I want to be helpful to him, and I want to see what he's doing with the stuff of mine that he's citing, because maybe we could be good collaborators.

A potential future collaboration can be one of the motivations to follow other researchers in RIM systems. One junior researcher stated that she hoped to convert some of the connections she was cultivating with other researchers in ResearchGate into future collaborations.

Activities	Actions	RIM functionalities
Find relevant literature	Search	Search engine; author profile; follow papers; citing & cited papers
	Determine	Author and publication profiles
	Select	Citation count; author impact scores; publication venue impact scores; manuscript status
	Obtain	Download a paper; request a paper from author(s)
Document manuscripts	Document sources	Citation generator
	Verify sources	Author profile
Identify researchers	Identify students, advisors, reviewers, collaborators, & etc.	Citing & cited papers; author profile, follow people; ResearchGate reads statistics
Disseminate research	Make papers accessible	Upload a paper; paper self-archiving status determination
	Promote papers	Recommend papers; recommend people
Interact with peers	Ask and answer questions on forums	Q&A service
	Send and receive private messages	Messaging service
Monitor the literature	Follow known researchers	Receive updates on known researchers
	Follow papers	Receive updates on papers
	Discover new papers	Recommend papers; citing & cited papers
	Discover new researchers	Recommend people; citing & cited papers
Evaluate	Evaluate papers	Citation count; number of reads; manuscript status
	Evaluate people	Author profile; h-index; export a CV; ResearchGate scores
Curate	Archive papers	Upload papers
	Add and modify metadata for papers	Add/update index terms; claim papers; disavow papers; add/update citation information
	Add and modify metadata for people	Create/update profile; merge profiles; add/edit index terms; endorse people for expertise; add/remove suggested co-authors
	Review papers	Open review
Look for jobs	Search	Recommended job postings

Table 1. Activities and RIM system functionalities.

Disseminate research

One of the main motivations of using RIM systems is to disseminate research results. Researchers may use RIM systems to share publications, data, and other research products. Nearly all participants mentioned they used RIM systems to promote their research. The dissemination activity may consist of making research results available and actively promoting those results. To make research results available researchers may upload copies of their

papers, presentations, or data to a RIM system. A service participants found particularly helpful in that action was the one that helped them determine whether a publication could be self-archived in RIM systems based on the publisher's policies. After research results are uploaded to a RIM system, the system then can use push services to promote results to the community.

Researchers may choose a specific RIM system providing more effective mechanisms (e.g., social network) to promote research to the community that they want to reach. One participant emphasized the social network provided by ResearchGate for promoting his research to the peers:

I used ResearchGate besides the Google Scholar because ResearchGate has slightly different methods of constructing the social network and the way they promote research is different – it's more active than Google Scholar. In that sense, it serves my purpose of trying to promote my research in the peers.

Interact with peers

Scholarly work may involve interaction. Researchers may interact about any aspect of research such as what design to employ for a particular research problem, what tools to use and how, or how to replicate research results. Researchers may also interact to exchange information about employment opportunities, and to recruit students, collaborators, external reviewers, or letter writers for grant proposals or promotions. Some RIM systems provide researchers with Q&A forums and a direct messaging service to communicate. In some cases, those communications channels become the only means on the Web of reaching a particular researcher. One participant revealed how ResearchGate helped him communicate with a researcher he could not reach otherwise when he was looking for a recommendation letter from the industry:

ResearchGate really gives you a way to connect to the researchers if you somehow cannot find their email address or other contact information from other channels ... I was looking for some recommendation letters for personal use. I wanted one from industry. This company cited my paper ... But for that specific case, the first author's email was not on the paper. And the last author, the corresponding author, actually left the company. So I had nowhere to find them. Then I checked ResearchGate. He was on ResearchGate. So I tried my last resort. I just sent him a message. And surprisingly, he replied.

Monitor the literature

To stay current with the literature, researchers need to monitor the literature for new works and/or contributors. One participant indicated his motivation of using RIM systems was to monitor his network of researchers:

Looking at what people whose work I'm interested in have cited, is useful for me and for following up on and finding out more about information that's useful for me in my own research.

RIM systems can be helpful to a researcher in monitoring literature by sending alerts about new works from the

researcher's network, and recommending new works and authors based on topical or co-citation matches. RIM systems having social networking capabilities enable researchers to connect to and learn about junior researchers' works, which may not be as visible as those of more established researchers. One participant explained how ResearchGate might enable her to know about junior researchers who otherwise could not be reached:

Researchers who are not that famous now, like junior faculty members or doctoral students, who are not big names, I probably cannot find another opportunity to know them all. If they also have a ResearchGate account and have some publications there, I hope this site can give me some automatic suggestions.

Evaluate

Evaluation can be a standalone activity (e.g., benchmarking oneself against other researchers) or part of a research process (e.g., evaluating papers for inclusion in a literature review). The targets of evaluation can be different entities such as a manuscript, a publication venue, an individual researcher, a lab, or an institution. Researchers may play the role of evaluators or be the objects of evaluation by others. If the latter, a researcher still can be an active contributor of a distributed evaluation process by creating and maintaining a profile in a RIM system to support his/her evaluation by others. The context of those evaluation activities may vary. One participant revealed that he created a Google Scholar profile to support his application for an award. Another participant mentioned he used his Google Scholar profile and impact factors as an evidence of his research impact when applying for the U.S. permanent residency.

A researcher's career status may affect the types of evaluation activities she or he may engage in or is asked to perform. Senior researchers may evaluate other researchers for promotion and tenure. Doctoral students, on the other hand, may benchmark themselves against other doctoral students who are at the similar stage in their doctoral programs to assess their competitiveness for the job market. For example, one participant who was a doctoral student illustrated how she used ResearchGate to follow other doctoral students to help prepare herself for the job market:

I followed some students who are at the same stage as myself ... in other schools to see their publication rate, how many publications they will get in one year ... And then I can estimate how much work should be expected for a doctoral student at my stage, so later, when I'm actually in the job market, I will not be too far away.

Curate

Curation of research resources can be defined as a process of managing those resources for discovery and future use (Lord & Macdonald, 2003). The main components of curation activities is quality assurance, which is the process of assuring that the research products, including information resources, meet the needs and requirements of the activities in which they are used (Stvilia et al., 2007). Researchers may use RIM systems to self-archive papers and data and to make them accessible. Researchers may

create and manage metadata for those resources to make them findable and reusable, and also use the metadata to construct a CV for different purposes. RIM systems with social networking capabilities allow researchers to request reviews of the content of their works from their peers. Curation of research information enables all the other activities in which that information is used or reused. Indeed, assuring the quality of their research identity metadata can be a motivation for researchers to establish a profile in a RIM system. For example, one participant created a profile in Google Scholar to correct an error after she found Google Scholar had identified another researcher having the same name as the author of her article.

Furthermore, the quality of information determines the outcome of an activity using that information. Concerns on the quality of an activity's outcome using research information and its possible effects on a researcher can be a strong motivator for the researcher to engage in curating his/her research identity profile. One participant noted:

If you don't maintain it [research identity profile], then it gives people an inaccurate view of your productivity, so you run the risk of potentially sending a signal about your productivity that's not accurate.

All participants indicated that they maintained their own personal profiles at least in one RIM system. Their maintenance included adding bibliographic metadata and subject index terms to their publications, uploading full-text articles, and endorsing colleagues for their skills.

Look for jobs

RIM systems may serve as a social network for researchers to look for job information or find job candidates. One participant mentioned she used ResearchGate's Jobs service to look for relevant job postings. Another participant described how he used ResearchGate's messaging services to help a researcher in another country to find a job:

For the messages I received, the only one that's not requesting a paper is the one from an Italian researcher. She told me she's going to graduate, and she's applying for a postdoctoral position. She's personally asked me if I knew any positions in the United States. So I replied her message, gave her some suggestions.

Engagement

Of the eight participants, four had public profiles in Google Scholar, seven used ResearchGate, and three had profiles in Academia.edu. Only one participant mentioned that he had an ORCID account. When asked why they participated in a particular RIM system, some participants recalled incidents that led them to create a profile in that system. Some of them did not purposefully create profiles in a RIM system to meet their research identity management needs, but the profiles were automatically generated and pushed on them. Others mentioned they acted on a recommendation from friends, colleagues, or advisors when creating profiles.

Researchers can be also introduced to a RIM system by another information system such as a search engine. They then perceived the value of membership after observing specific benefits provided by the system. For example, one participant revealed:

I first came to ResearchGate, because a paper I was looking for at that time only had full-text version on ResearchGate ... Then I noticed that's a benefit. I should create an account there.

Levels of engagement

The data analysis identified three levels or categories of researcher participation in a RIM system. Researchers belonging to the first category have claimed or activated an account in a RIM system but do not maintain it or not interact with other members of the system. This category was called Readers as they use RIM systems mostly to access the literature.

Researchers in the second category may maintain their profiles in a RIM system, but do not contribute to the system beyond that and not interact with other members of that system directly or indirectly. That is, they don't ask or answer questions in Q&A forums, endorse other members for their skills, send emails, or respond to other members' emails or requests. This category was labeled as Personal Record Managers. A majority of the participants (four out of eight) were grouped under this category.

Researchers in the third category not only maintain their own profiles, but also are willing to curate research information of other members by endorsing them for skills, and sharing information via messages, emails, or Q&A services. This category was labeled as Community Members, who may be motivated by the feeling of reciprocity and being 'a good member' of the community.

DISCUSSION AND CONCLUSIONS

As mentioned above, an OCLC Task Group formulated researchers' five needs for research identity data (OCLC Research, 2014), which can be nicely mapped to five of the motivations of using RIM systems identified in the current study: find relevant literature, disseminate research, curate, identify researchers, and monitor the literature. However, the empirical data collected from the current study identified four more motivations of using RIM systems or research identity data: document manuscripts, interact with peers, evaluate, and look for jobs. Indeed, most of the IRs do not support those four activities (Lee, 2015). Several participants of the current study mentioned they used ResearchGate as it provided a social network allowing them to follow and communicate with other researchers and look for jobs. IRs may consider incorporating these functionalities to allow their users to communicate with each other, and generate profiles to support different evaluation activities (e.g., self-evaluation, annual review).

Preece and Shneiderman (2009) presented a framework to describe user engagement in online social communities consisting of four levels: Reader, Contributor, Collaborator, and Leader. The current study identified three levels of

participation in RIM systems: Readers, Personal Record Managers, and Community Members. These three categories can be mapped to the first three levels of engagement of Preece and Shneiderman's framework. Most participants of the current study fell into the category of Personal Record Managers, who may maintain their profiles in a RIM system, but do not contribute to the system beyond that nor interact with other members of that system. A study of data curation practices in IRs found that IR staff's curation activities focused on ensuring the quality of publication metadata for the long-term preservation of publications to increase their reusability (Lee, 2015). Findings of the current study suggest that a majority of researchers may be willing to maintain their research identity profiles. IR managers may consider recruiting researchers as not only research information/data providers, but also curators of their own research identity data.

This study provides rich qualitative data regarding how researchers use and participate in online RIM systems. Still, this poster is limited as it reports preliminary findings based on interviews with eight participants from four institutions. More interviews will be conducted with researchers from other institutions and disciplines to gain different perspectives. Based on findings of interviews, we will develop and implement a survey to reach more researchers, and develop a quantitative model of researcher participation in RIM systems.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to the researchers who participated in the study. This research is supported by an OCLC/ALISE Library and Information Research Grant for 2016 and the National Leadership Grants from the Institute of Museum and Library Services (IMLS). The article reflects the findings and conclusions of the authors, and do not necessarily reflect the views of OCLC, ALISE, and IMLS.

REFERENCES

- Ames, M., & Naaman, M. (2007). Why we tag: Motivations for annotation in mobile and online media. In B. Begole & S. Payne (Eds.), *Proceedings of the SIGCHI* (pp. 971-980). New York, NY: ACM.
- Bauin, S., & Rothman, H. (1992). "Impact" of journals as proxies for citation counts. In *Representations of science and technology* (pp. 225-239). Leiden: DSWO Press.
- Blee, K. M., & Taylor, V. (2002). Semi-structured interviewing in social movement research. In B. Klandermans & S. Staggenborg (Eds.), *Methods of social movement research* (pp. 92-117). Minneapolis, MN: University of Minnesota Press.
- Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2006). Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1037-1046). New York, NY: ACM.
- Dempsey, L. (2014). *Research information management systems - a new service category?* Retrieved from <http://orweblog.oclc.org/archives/002218.html>
- Engeström, Y. (1987). *Learning by expanding: An Activity-Theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy.
- Giles, J. (2005). Internet encyclopedias go head to head. *Nature*, 438(7070), 900-901.
- Lee, D. J. (2015). *Research data curation practices in institutional repositories and data identifiers* (Doctoral dissertation). Retrieved from http://purl.flvc.org/fsu/fd/FSU_migr_etd-9638
- Lord, P., & Macdonald, A. (2003). *E-Science curation report: Data curation for e-Science in the UK: An audit to establish requirements for future curation and provision*. Bristol, UK: JISC.
- Nov, O. (2007). What motivates Wikipedians. *Communications ACM*, 50(11), 60-64.
- Nov, O., Naaman, M., & Ye, C. (2010). Analysis of participation in an online photo-sharing community: A multidimensional perspective. *Journal of the American Society for Information Science & Technology*, 61(3), 555-566.
- OCLC Research. (2014). *Registering researchers in authority files*. Retrieved from <http://www.oclc.org/research/themes/research-collections/registering-researchers.html>
- Palmer, D. (2013). *The HKU Scholars Hub: Reputation, identity & impact management*. How librarians are raising researchers' reputations. Retrieved from <http://hub.hku.hk/bitstream/10722/192927/1/Reputation.pdf>.
- Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1), 13-32.
- Salo, D. (2009). Name authority control in institutional repositories. *Cataloging & Classification Quarterly*, 47(3-4), 249-261.
- Stvilia, B., Twidale, M., Smith, L. C., Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science & Technology*, 59(6), 983-1001.
- Stvilia, B., Gasser, L., Twidale, M., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science & Technology*, 58, 1720-1733.
- Stvilia, B., & Jørgensen, C. (2009). User-generated collection-level metadata in an online photo-sharing system. *Library & Information Science Research*. 31(1), 54-65.
- Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future*. Retrieved from

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.3251&rep=rep1&type=pdf>

Wasko, M. M., & Faraj, S. (2005). Why should I share?
Examining social capital and knowledge contribution in

electronic networks of practice. *MIS quarterly*, 29(1), 35-57.