

This is a preprint of an article accepted for publication in the Library Hi Tech (Lee, D. J., Stvilia, B., & Wu, S. (in press). Towards a metadata model for research information management systems. *Library Hi Tech*)

Towards a Metadata Model for Research Information Management Systems

Dong Joon Lee

University Libraries, Texas A&M University, 5000 TAMU, College Station, TX 77843-5000. E-mail: djlee@tamu.edu

Besiki Stvilia

School of Information, Florida State University, 142 Collegiate Loop, Tallahassee, FL 32306-2100. Email: bstvilia@fsu.edu

Shuheng Wu

Graduate School of Library and Information Studies, Queens College, 65-30 Kissena Blvd., Queens, NY 11367-1597. E-mail: Shuheng.Wu@qc.cuny.edu

Abstract

Purpose - This paper examined the relationships between researcher characteristics and their use of metadata in their ResearchGate profiles.

Design/methodology/approach - This paper reports on one part of a larger study that examined researchers' use of and engagement with research information management systems. The study's design included qualitative, semi-structured interviews with 15 researchers and a survey completed by 412 researchers. Detailed reports of findings from the interviews and survey can be found elsewhere. This paper reports on the part of the study that analyzed the use of metadata elements in the ResearchGate profiles of 126 survey participants.

Findings - Most researchers shared metadata related to their research rather than their teaching or service. Statistical analyses revealed statistically significant relationships between researchers' metadata use and their participation levels in research information management systems, as well as between metadata use and researchers' seniority.

Originality/value - The study's findings help to identify researchers' priorities for different metadata elements, as well as to construct profile metadata templates for each specific participation level.

Keywords - Research information management systems, RIMS, Metadata, ResearchGate, Identity metadata, Authority control

1. Introduction

Academia is a sprawling metropolis of teachers, students, administrators, and researchers, each with their own interests, yet each dependent on the work of others in order to effectively teach or do their own research. This requires access to the already-existing, interconnected web of knowledge created by past research. Research Information Management System (RIMS) attempt to map that web; they can be defined as systems that aggregate, curate, analyze, and use metadata about research activities (Bryant et al., 2017). Depending on the community of use, different terms are used to refer to RIMS (e.g., Current Research Information System [CRIS], Research Networking System [RNS], or Research Profiling System [RPS]). RIMSs are rapidly gaining in popularity (Martin-Martin et al., 2016), and there have been multiple studies on the needs for and uses of RIMSs from the perspectives of different stakeholder groups (Martin-Martin et al., 2016; Smith-Yoshimura et al., 2014; Thelwall and Kousha, 2017). Many institutions develop and implement their own local RIMSs to support their various stakeholder groups, including researchers, funders, university administrators, librarians, and aggregators (Smith-Yoshimura et al., 2014). Each of the groups has different needs for and uses of RIMSs. Researchers use RIMSs to find relevant literature, document manuscripts, identify other researchers, disseminate research, interact with peers, monitor the literature, evaluate different entities, curate research output, and look for jobs; funders employ RIMSs to track funded research output; university administrators apply RIMSs to generate internal scholarly output reports; librarians use RIMSs to disambiguate names; and aggregators utilize RIMSs to publish metadata for each researcher.

Researchers' needs for research identity information and their priorities for those needs can be identified directly by interviewing and/or surveying them, or indirectly by examining their willingness to share their research identity information through their RIMS profiles. The extent of data or metadata use in the profile can represent its relative importance to the different activities and stakeholders of that data (Greenberg, 2001; Stvilia and Gasser, 2008). Furthermore, metadata use can reflect how disciplinary differences affect research information management practices (David and Thomas, 2015; Faniel and Jacobsen, 2010; Stvilia et al., 2015). Although there is a significant body of literature on metadata use in information systems, there is still a lack of investigation into researchers' use of metadata in RIMSs.

To facilitate the use of research identity data, enhance its quality (e.g., completeness and accuracy), and engage researchers in its curation, it is important to identify researchers' value structures of and priorities for different research identity data elements. The purpose of this study was to examine researchers' practices of using metadata elements in RIMSs. The findings from this study can be used to develop a value model of RIMS metadata elements (Stvilia and Gasser, 2008), which can help institutional repositories (IRs) better align their RIMS metadata models and quality assurance activities with researchers' needs and priorities.

The study specifically examined researchers' metadata use in ResearchGate, one of the most prominent and widely used RIMSs, to address the following research questions:

- Which metadata elements do researchers currently use in their RIMS (i.e., ResearchGate) profiles?
- What are the relationships between the use of metadata elements and levels of researcher participation in RIMS (i.e., ResearchGate)?
- What are the relationships between the use of metadata elements and researcher seniority levels in RIMS (i.e., ResearchGate)?
- What are the relationships between the use of metadata elements and disciplines in RIMS (i.e., ResearchGate)?

2. Literature Review

Previous studies (e.g., Jeng et al., 2017; Van Noorden, 2014) have examined how scholars use online RIMSs such as Google Scholar, ResearchGate, ResearcherID, Mendeley, and ORCID in their professions. These studies showed that scholars use RIMSs to find relevant literature, document manuscripts, identify researchers, disseminate research, interact with peers, monitor the literature, evaluate research, curate identity data, and look for jobs. Along with the use

of RIMs, the literature on impact indicators in RIMs (e.g., citation counts, view counts, number of downloads, h-indices, etc.) showed that RIMs can be used to assess the impact of authors, published items, journals, and publishers (Martin-Martin et al., 2016), as well as to disseminate scholarship to its research communities (Thelwall and Kousha, 2015). Thelwall and Kousha (2017) also found that *view counts* for uploaded articles in ResearchGate profiles are a new audience indicator. In their study, the *view counts* had low-to-moderate positive correlations with the numbers of Scopus citations and Mendeley readers, which reflect the wider audience of ResearchGate.

Although there is a dearth of research on metadata use in RIMs, many studies have investigated metadata use within IRs (David and Thomas, 2015; Zavalina and Kizhakkethil, 2015). For instance, a university library examined their event logs related to metadata changes occurring over time in its IR, which may help them develop in the future an effective and efficient metadata management policy for the IR (Zavalina and Kizhakkethil, 2015). An IR also reviewed metadata use for electronic books to identify and understand the relative importance of metadata elements for its end users (David and Thomas, 2015). They found that *subject*, *title*, and *author* on electronic books are the most important metadata fields for searching and discoverability. The study's result informed the authors on how to develop and implement workflow changes to the metadata quality control process.

Knowledge organization tools, including structured metadata models and controlled vocabularies for RIMs, have been proposed in the community (e.g., Common European Research Information Format [CERIF] or Consortia Advancing Standards in Research Administration Information [CASRAI] Dictionary). CASRAI Dictionary is a comprehensive, controlled vocabulary to support the interoperability of data exchanges among research organizations (CASRAI, n.d.). A large number of research universities in Canada and the United Kingdom are the primary members of CASRAI. CASRAI has 871 terms for the research administration domain and 484 terms for the research data domain. CERIF is a structural metadata model for RIMs created by the European Commission. EuroCRIS (n.d.) maintains this model to represent research activities, outputs, and relationships. Pinto, Simões, and Amaral (2014) reported that RIMs based on the CERIF model focus on researchers, projects and institutions of the research domain, while RIMs without CERIF focus mainly on researchers than other elements and have more flexibility to be localized. The CERIF model was mainly adopted by European countries, i.e., Belgium, Slovenia, Germany, and other countries in the European Network of National Contact Points (Pinto et al., 2014; Quix and Riechert, 2017).

The digital curation of scholarly content requires more than a collection of individual data elements; it requires contextual understanding of the heterogeneous research activity data aggregated by RIMs for the academic communities (Borgman et al., 2007; Lee and Stvilia, 2017; Stvilia et al., 2015). Jeffery, Houssos, Jörg, and Asserson (2014) proposed a three-layer metadata model for RIMs using CERIF. The first layer of the model, entitled "Discovery Metadata," is comprised of a simple metadata schema (e.g., Dublin Core). The second layer, entitled "Contextual Metadata," includes the CERIF model as a structured contextual metadata schema within the domain of research information. The model captures the semantic relationships between entities of research information (e.g., publications, persons, institutions, and funding). The third layer of the model, entitled "Domain Metadata," discusses subject- or topic-specific metadata models, including Core Scientific Metadata Model (CSMD), Statistical Data and Metadata eXchange (SDMX), Data Documentation Initiative (DDI), etc. According to Jeffery et al. (2014), metadata schema in the second layer, "Contextual Metadata," should support functions from both the first and third layers. CERIF helps users to generate descriptive metadata for their research products and provides pre-defined semantic relationships between its entities.

Jörg, Höllrigl, and Baker (2014) reported a strategic partnership between CASRAI and CERIF by initiating and developing a crosswalk among their models. They presented the ability to create an abridged Curriculum Vitae (CV) using CASRAI concepts within CERIF XML. Although the abridged CVs include a smaller subset of data about researchers' scholarly products than their complete CVs, the abridged CVs demonstrate the capacity of CERIF to describe researchers' formal profiles. In addition, Haak, Baker, and Höllrigl (2014) reported that Open Researcher and Contributor ID (ORCID) has also updated its metadata to be aligned with CASRAI and CERIF. A study has proposed a metadata application profile for

ORCID to fill the gap between practical uses and conceptual understandings of ORCID metadata (Lee et al., 2014). The proposal included a list of functional requirements and a domain model for ORCID.

3. Study Design

This paper reports on one part of a larger study that examined researchers' use of and engagement with RIMs. The study's design included qualitative, semi-structured interviews with 15 researchers and a survey completed by a purposive sample of 412 researchers. The surveyed researchers represent 80 universities in the United States classified as RU/VH (very high research activity) in the Carnegie Classification of Institutions of Higher Education, a leading framework for recognizing and describing institutional diversity in United States higher education. Detailed reports of findings from the interviews and survey can be found elsewhere (Wu et al., 2017; Stvilia et al., 2018). This paper reports on the part of the study that analyzed the use of metadata elements in the ResearchGate profiles of 126 of the survey participants. The study categorized survey participants into three categories: Readers, Personal Record Managers, and Community Members. Table 1 and Table 2 present sampled profiles' frequency on participation, seniority levels, and disciplines, as well as definitions of the three participation categories. The following three activities were used to define three participation levels: maintaining a RIMS profile, answering questions, and endorsing other researchers. The levels were defined in a progressively cumulative manner. Readers had to use a RIMS, but did not perform any of these activities. They may or may have a profile in a RIMS. Researchers who belonged to the Record Manager level had to maintain their profile in a RIMS, but they also did not engage in the other two activities: answering questions or endorsing other researchers. On the other hand, researchers who belonged to the Community Member level had to engage in either of those two activities (i.e., answering questions or endorsing researchers) in addition to maintaining their profile. Whether a particular researcher performed a particular activity or not was determined by her/his response to a related survey question (e.g., "Do you personally maintain or edit your profile in a RIM system(s)?" ; Stvilia et al., 2018). In addition, each of these activities was connected with a specific set of motivations that may spur the activity. The description and analysis of the activity –motivation relationships was reported somewhere else (Stvilia et al., 2018). Fifty survey participants with ResearchGate profiles from both the Personal Record Managers and the Community Members categories were randomly selected. Since ResearchGate did not specify its metadata and service model, one of the authors of the study purposely examined all the metadata fields that ResearchGate provides for individual members' profiles in general and collected the metadata elements in order to assemble an aggregate set of metadata elements provided by ResearchGate.

Table 1. Sampled profiles' frequency on participation and seniority levels

Participation Levels	Freq	Seniority Levels	Freq	%
Readers: They may or may not have a profile in a RIMS, but do not maintain the profile if they have one and do not interact with other members of the system or contribute to the system.	26	Graduate Student	6	23
		Postdoc	6	23
		Assistant Professor	6	23
		Associate Professor	3	12
		Full Professor	5	19
Personal Record Managers: They maintain their profiles in a RIMS, but do not contribute to the system beyond that and do not interact with other members of that system directly or indirectly.	50	Graduate Student	7	14
		Postdoc	8	16
		Assistant Professor	17	34
		Associate Professor	9	18
		Full Professor	9	18
Community Members: They not only maintain their own profiles, but are also willing to engage in curating the research information of other members by endorsing them for skills, or to share their knowledge and information via messages, emails, or Q&A forums.	50	Graduate Student	3	6
		Postdoc	20	40
		Assistant Professor	16	32
		Associate Professor	5	10
		Full Professor	6	12

Table 2. Sampled profiles' frequency on discipline categories

#	Discipline Categories	Readers		Personal Record Managers		Community Members		Total Participants	Total Participants
		Freq	%	Freq	%	Freq	%	Freq	%
1	Humanities	2	7.7	6	12	3	6	11	9
2	Social Sciences	10	38.5	14	28	11	22	35	28
3	Physical Sciences	5	19.2	13	26	13	26	31	24
4	Life Sciences	1	3.8	3	6	15	30	19	15
5	Engineering	8	30.8	14	28	8	16	30	24

4. Findings

4.1. Which metadata elements do researchers currently use in their ResearchGate profiles?

The analysis of ResearchGate metadata elements produced an aggregated set of user-editable metadata elements (Table 3). The individual elements were grouped into nine categories: *person*, *publication*, *research subject*, *research experience*, *teaching experience*, *education*, *award*, *contact*, and *other*. ResearchGate requires its users to specify a name, an institution, and an email address as a minimum to create a profile. However, the RIMS does not require its users to have a profile in order to browse the public data on other researchers' profiles.

Table 3. User-editable metadata categories and elements used in ResearchGate

Categories	Metadata Elements
Person	First name, Middle name, Last name, Alternative first name, Alternative middle name, Alternative last name, Degree, Institution, Department, Position, Time period, Gender, Email address, Profile photo, Time zone, About
Publication	Publication title, Author, File, Type (i.e., book, chapter, code, conference paper, method, patent, poster, proposal, technical report, thesis, working paper), Journal referee, Volume, Issue, Page, Day, Month, Year, Topics, Abstract, DOI, Publisher, Editor, Edition, ISBN, Chapter, Book title, Description, Language(s), Repository link, License, Ref. Number, Ordinal, Grant number, Report number, Supervisor, Degree, Version number, State
Research subject	Topics, Skills & Expertise, Discipline
Research experience	Position, Institution, Department, Research group, Time period, Location, Description
Teaching experience	Position, Institution, Department, Time period, Location, Description
Education	Institution, Field of study, Degree, Time period, Location
Award	Type (i.e., award, grant, scholarship), Title, Start date, End date, Amount, Funding agency, Grant reference, Principal investigator, Research institution, Co-investigator, Secondary institution
Contact	Location, Website, Phone, Mobile, Fax, Twitter, Skype, Instant messenger, Birthday,
Other	Language(s), Scientific societies, Journal referee, Other interest, ORCID

In order to answer the study’s question about how metadata elements are currently being used by researchers, the authors examined user profiles from the sample. Figure 1 presents the use frequencies of metadata categories in the sample. Most profiles used at least one element from the categories of *person*, *publication*, and *research subject*. No profiles contained metadata elements from the *contact* category. In addition, profiles in the Readers group did not include any metadata elements from the categories of *education*, *award*, and *teaching experience*.

Since almost all profiles used metadata elements from the *person*, *publication*, and *research subject* categories, in the next step, the study investigated the use of individual elements within these categories. Figure 2 summarizes the uses of specific metadata elements from the *person* category. More than 90% of the Community Members’ profiles included *first name*, *last name*, *affiliation*, *department*, *photo*, and *position*. More than 88% of the profiles in the Readers group included *first name*, *last name*, *affiliation*, and *department*, while the rest of the metadata elements were used by less than 50% of the Readers’ profiles.

Metadata elements in the *publication* category included *item title*, *author*, *file*, *type*, *journal title*, *volume*, *issue*, *page*, *year*, *abstract*, and *digital object identifier (DOI)*. All of these metadata elements except *file* were highly used (i.e., approx. 60% to 100%, see Figure 3). The element *file* was only used in about 30% of the Readers’ group profiles (see Figure 3).

Figure 4 presents specific uses of metadata elements from the *research subject* category. This category contains only three metadata elements: *skills and expertise*, *topics*, and *discipline*. Among these three elements, *topics* was the most frequently used by all three categories of researchers. All the Community Members’ profiles in the sample contained the metadata element *skills and expertise*, while only 58% of the Readers’ profiles included this element. Similarly, 86% of the Community Members’ profiles had the element *discipline*, and only 65% of the Readers’ profiles contained this element.

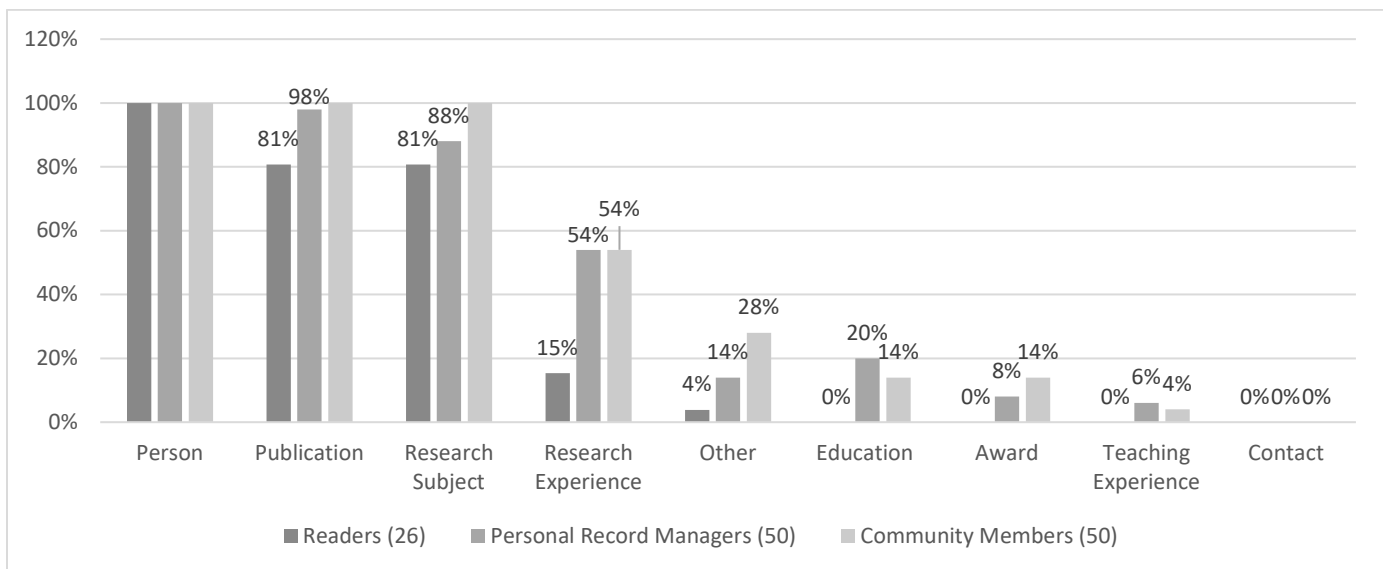


Figure 1. Uses of metadata categories



Figure 2. Uses of metadata elements from the *Person* category

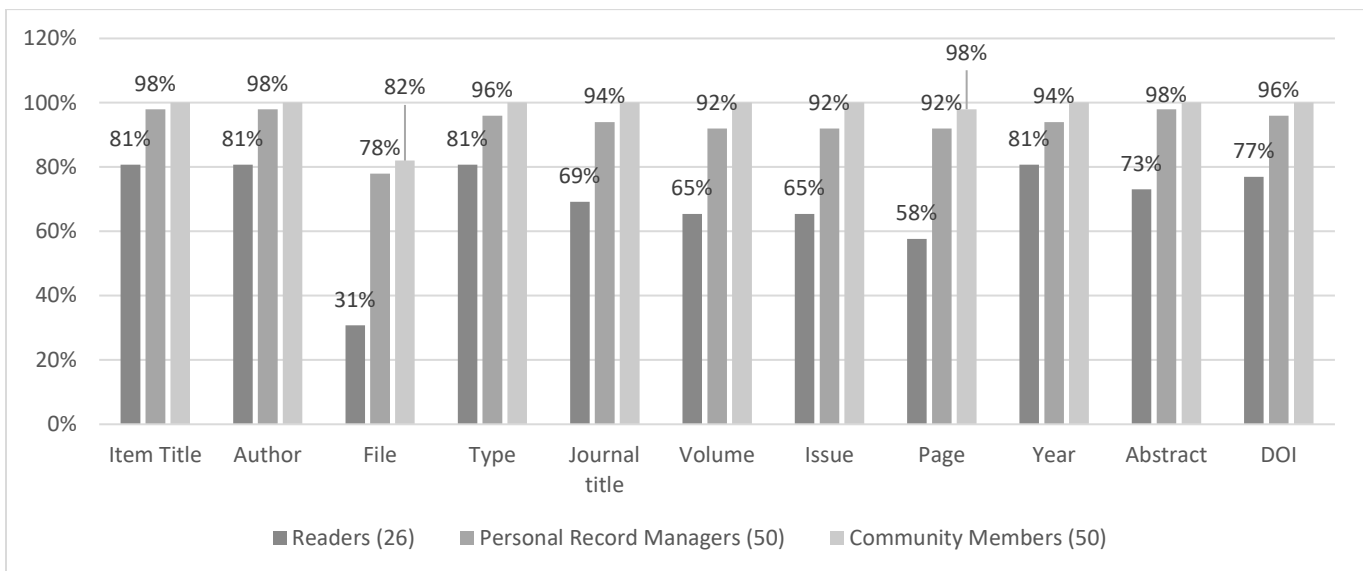


Figure 3. Uses of metadata elements from the *Publication* category



Figure 4. Uses of metadata elements from the *Research Subject* category

4.2. What are the relationships between metadata element use and the levels of researcher participation in ResearchGate?

In order to compare the dependence of metadata elements on researcher participation levels, a statistical analysis was completed using the Kruskal-Wallis test, which indicated several significant relationships (see Table 4). The analysis revealed significant differences between researcher participation levels in terms of *profile views, followers, skills and expertise, following, followed items, topics, photo, position, department, project, research experience, scientific societies, and advisors* (see Table 4).

Table 4. Metadata elements' relationships to levels of researcher participation in ResearchGate, levels of researcher seniority, and categories of discipline

Metadata elements	Levels of Researcher Participation			Levels of Researcher Seniority			Discipline Categories		
	Chi-Square	df	Asymp. Sig.	Chi-Square	df	Asymp. Sig.	Chi-Square	df	Asymp. Sig.
RG Score	2.212	2	0.331	31.874	4	0.001	5.890	4	0.208
Reads	5.081	2	0.079	41.682	4	0.001	9.526	4	0.049
Citations	1.983	2	0.371	50.125	4	0.001	7.143	4	0.129
Profile views	8.187	2	0.017	14.969	4	0.005	4.645	4	0.326
Followers	10.391	2	0.006	33.448	4	0.001	7.964	4	0.093
Research items	1.665	2	0.435	47.585	4	0.001	7.370	4	0.118
Skills and expertise	34.626	2	0.001	6.638	4	0.156	8.422	4	0.077
Following	26.560	2	0.001	4.058	4	0.398	5.624	4	0.229
Followed items	10.822	2	0.004	6.361	4	0.174	2.256	4	0.689
Topics	10.371	2	0.006	4.721	4	0.317	2.733	4	0.603
Awards and achievements	4.315	2	0.116	4.467	4	0.346	0.477	4	0.976
Photo	24.330	2	0.001	8.149	4	0.086	8.017	4	0.091
First name	0.001	2	1.000	0.001	4	1.000	0.001	4	1.000

Last name	0.001	2	1.000	0.001	4	1.000	0.001	4	1.000
Position	30.689	2	0.001	5.037	4	0.284	4.886	4	0.299
Department	7.719	2	0.021	7.293	4	0.121	3.600	4	0.463
Affiliation	0.001	2	1.000	0.001	4	1.000	0.001	4	1.000
Project	6.659	2	0.036	4.397	4	0.355	3.898	4	0.420
About	4.229	2	0.121	7.117	4	0.130	2.113	4	0.715
Research experience	12.288	2	0.002	3.470	4	0.483	4.334	4	0.363
Teaching experience	0.279	2	0.870	2.562	4	0.634	6.394	4	0.172
Education	5.834	2	0.054	1.331	4	0.856	2.879	4	0.578
Language	1.900	2	0.387	6.479	4	0.166	2.286	4	0.683
Scientific societies	6.757	2	0.034	4.648	4	0.325	2.524	4	0.640
Advisor	6.221	2	0.045	4.233	4	0.375	3.515	4	0.476
Journal referee	2.101	2	0.350	3.087	4	0.543	1.635	4	0.802
Question	2.377	2	0.305	2.139	4	0.710	2.863	4	0.581

Note: Significant relationships are in boldface italics.

In addition, to identify the pairwise relative importance of metadata elements among the researcher participation levels, the study analyzed the metadata elements that exhibited significant differences in the researcher participation levels using either the Bonferroni tests of post-hoc pairwise comparisons, or a binary logistic regression analysis. The different tests were used according to the types of data collected. When the outcome variable was count (e.g., number of items), the study used the Bonferroni multiple comparison test; and when the outcome variable was binary (e.g., 0 or 1), the study used binary logistic regression. The analyses revealed statistically significant differences among the participation levels for the following metadata elements: *photo*, *position*, *research experience*, *project*, *profile views*, *skills and expertise*, *following*, *followers*, *followed items*, and *topics* (see Tables 5 and 6).

Table 5. Binary variables' regression analyses among researcher participation levels^a

Metadata elements	Model fit	Coefficient (p-value), when	
	χ^2 / p-value	baseline 0	baseline 1
		Indicator: 1 / 2	
		Indicator: 2	
Photo	29.01 / 0.0001	0.85 (0.090) / 3.89 (0.001)	3.04 (0.004)
Position	31.50 / 0.0001	1.66 (0.002) / 3.25 (0.001)	1.60 (0.008)
Research experience	13.57 / 0.0011	1.87 (0.002) / 1.87 (0.002)	4.65 (1.000)
Project	6.69 / 0.0352	0.47 (0.692) / 1.83 (0.090)	1.37 (0.049)
Department	2.94 / 0.0864	1.85 (0.117) / 0	0
Scientific societies	1.99 / 0.1580	-0.81 (0.169) / 0	0.81 (0.169)
Advisor	2.64 / 0.1042	1.09 (0.123) / 0	-1.09 (0.123)

^a Indicators 0, 1, and 2 specify Readers, Personal Record Managers, and Community Members respectively.

Note: Significant relationships are in boldface italics.

Table 6. Count variables' Bonferroni multiple comparison tests among researcher participation levels^a

Metadata elements	Mean Ranks	Sample 1 - Sample 2	Std. Error	Adj. Sig.
	0/1/2			
Profile views	45.96/	0 – 1	9.025	0.812

	55.92/ 69.81	0 – 2	8.970	0.024
		1 – 2	6.971	0.139
Followers	44.38/	0 – 1	8.828	0.076
	64.13/ 72.81	0 – 2	8.828	0.004
		1 – 2	7.303	0.704
Skills and expertise	39.69/	0 – 1	8.803	0.378
	53.16/ 86.22	0 – 2	8.803	0.001
		1 – 2	7.281	0.001
Topics	44.62/	0 – 1	8.782	0.083
	63.94/ 72.88	0 – 2	8.782	0.004
		1 – 2	7.264	0.655
Following	40.94/	0 – 1	8.824	0.284
	55.69/ 83.04	0 – 2	8.824	0.001
		1 – 2	7.299	0.001
Followed items	54.40/	0 – 1	8.578	1.000
	55.42/ 76.31	0 – 2	8.578	0.032
		1 – 2	7.095	0.010

^a Indicators 0, 1, and 2 specify Readers, Personal Record Managers, and Community Members respectively.

Note: Significant relationships are in boldface italics.

For binary variables, the study ran two regression analyses with different baselines to identify relative relationships or importance (see Table 5). The analyses showed that the presence of *photo* in a researcher's profile increased the odds that the researcher is a Community Member rather than either a Reader or a Personal Record Manager. The presence of *position* in a researcher's profile increased the odds that the researcher is a Community Member or a Personal Record Manager rather than a Reader. The presence of *position* also increased the odds that a researcher is a Community Member rather than a Personal Record Manager. With regard to *research experience*, comparisons were only significant between the Readers and the Personal Record Managers, as well as between the Readers and the Community Members. The presence of *research experience* in a researcher's profile increased the odds that the researcher is either a Personal Record Manager or a Community Member rather than a Reader. For *project*, a significant relationship only existed between the Personal Record Managers and the Community Members.

For count variables, the Bonferroni tests of post-hoc pairwise comparisons were used (see Table 6). *Profile views* and *followers* were automatically generated and updated by the RIMS. Based on the pairwise comparisons, the profiles of Community Members had significantly higher mean ranks for *profile views* and *followers* than did the profiles of Readers. *Skills and expertise* and *topics* were user-supplied metadata elements. The author who benchmarked researchers' profiles manually counted the numbers of *skills and expertise* and *topics* metadata that the owners of the profiles supplied to their profiles. The comparison tests revealed that the profiles of Community Members had significantly higher mean ranks for the numbers of keywords that they added to either *skills and expertise* or *topics* than did the profiles of Readers. The numbers of keywords in *skills and expertise* also showed significantly higher mean ranks in the profiles of Community Members than in the profiles of Personal Record Managers. Likewise, the profiles of Community Members had significantly higher mean ranks for *following* and *followed items* than did the profiles of Readers and Personal Record Managers. Figure 5 summarizes the relationships between the use of different metadata elements and the levels of researcher participation.

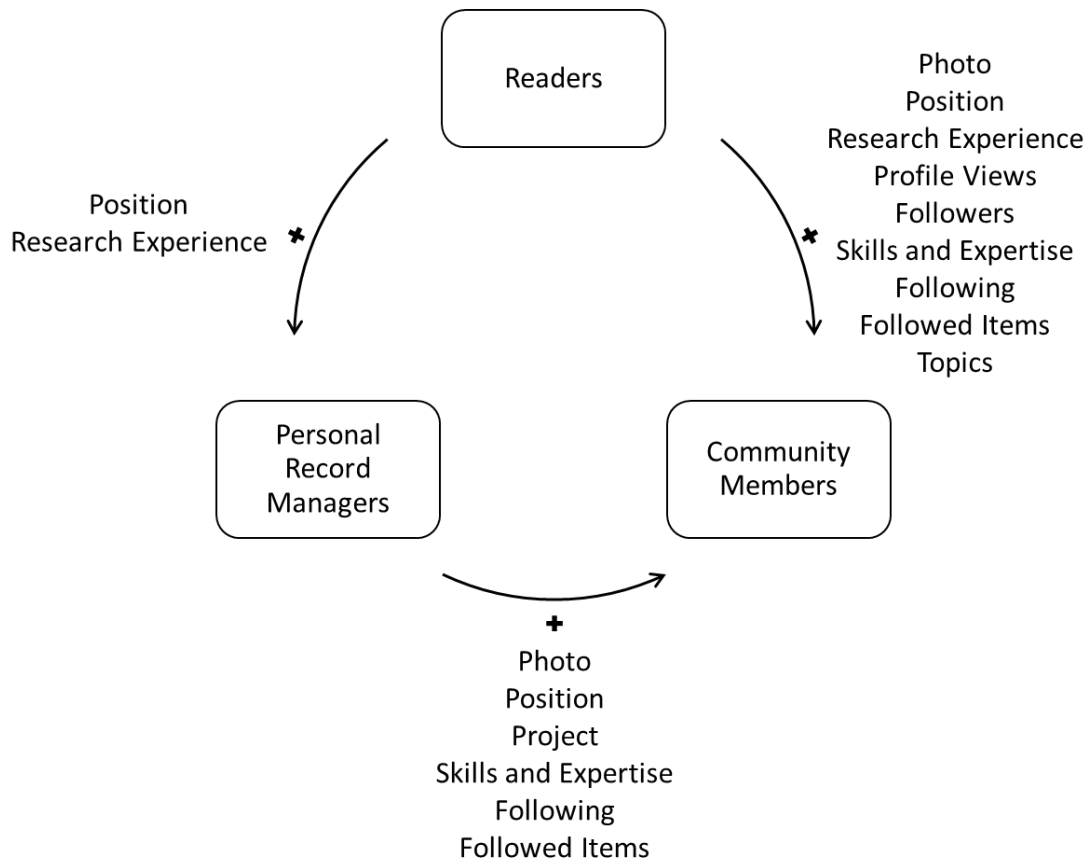


Figure 5. Relationships between the use of metadata elements and the levels of participation. Only statistically significant variables are included. The arrows point toward greater usage of metadata elements

4.3. What are the relationships between metadata element use and researcher seniority levels in ResearchGate?

In addition to the relationships between metadata element use and researcher participation levels, the authors also investigated the relationships between metadata element use and researcher seniority levels. The Kruskal-Wallis tests showed significant differences for *RG Score*, *reads*, *citations*, *profile views*, *followers*, and *research items* based on researcher seniority levels (see Table 4).

The Bonferroni multiple comparison tests indicated multiple significant differences in the mean ranks of metadata element use based on researcher seniority levels (see Table 7). Full professors had significantly higher mean ranks for the *RG Score*, *reads*, *citations*, *followers*, and *research items* than graduate students, postdocs, and assistant professors, while associate professors had significantly higher mean ranks for the same metadata elements than graduate students. For *profile views*, graduate students had significantly lower mean ranks than full professors and postdocs. The *Reads* element showed a significant mean rank difference between full professors and associate professors. Lastly, assistant professors had significantly higher mean ranks for the *followers* and *research items* than graduate students.

Table 7. Count variables' Bonferroni multiple comparison tests among researcher seniority levels^a

Metadata elements	χ^2 /Mean Ranks	Sample 1 - Sample 2	Std. Error	Adj. Sig.	Metadata elements	χ^2 /Mean Ranks	Sample 1 - Sample 2	Std. Error	Adj. Sig.
	0 1 2 3 4					0 1 2 3 4			

RG Score	31.874 / 29.00 49.57 53.61 72.88 88.63	0 – 1	11.167	0.655	Profile views	14.969 / 30.27 64.44 59.61 62.97 73.78	0 – 1	10.850	0.016
		0 – 2	10.843	0.232			0 – 2	10.559	0.055
		0 – 3	12.557	0.005			0 – 3	12.398	0.083
		0 – 4	11.981	0.001			0 – 4	11.783	0.002
		1 – 2	8.262	1.000			1 – 2	8.399	1.000
		1 – 3	10.411	0.252			1 – 3	10.619	1.000
		1 – 4	9.708	0.001			1 – 4	9.894	1.000
		2 – 3	10.062	0.555			2 – 3	10.322	1.000
		2 – 4	9.334	0.002			2 – 4	9.574	1.000
		3 – 4	11.280	1.000			3 – 4	11.570	1.000
Reads	41.682 / 30.67 51.97 56.43 66.15 102.05	0 – 1	11.031	0.535	Followers	33.448 \\ 31.63 55.22 63.17 68.79 99.23	0 – 1	11.070	0.330
		0 – 2	10.695	0.160			0 – 2	10.840	0.036
		0 – 3	12.425	0.043			0 – 3	12.718	0.035
		0 – 4	11.980	0.001			0 – 4	12.247	0.001
		1 – 2	8.488	1.000			1 – 2	8.567	1.000
		1 – 3	10.585	1.000			1 – 3	10.846	1.000
		1 – 4	10.059	0.001			1 – 4	10.289	0.001
		2 – 3	10.234	1.000			2 – 3	10.612	1.000
		2 – 4	9.689	0.001			2 – 4	10.042	0.003
		3 – 4	11.570	0.019			3 – 4	12.045	0.115
Citations	50.125 / 28.57 46.13 58.13 73.09 103.55	0 – 1	11.030	1.000	Research items	47.585 \\ 27.30 47.13 59.53 71.03 102.05	0 – 1	11.027	0.722
		0 – 2	10.693	0.057			0 – 2	10.691	0.026
		0 – 3	12.423	0.003			0 – 3	12.420	0.004
		0 – 4	11.978	0.001			0 – 4	11.976	0.001
		1 – 2	8.487	1.000			1 – 2	8.485	1.000
		1 – 3	10.583	0.109			1 – 3	10.581	0.239
		1 – 4	10.058	0.001			1 – 4	10.056	0.001
		2 – 3	10.232	1.000			2 – 3	10.230	1.000
		2 – 4	9.688	0.001			2 – 4	9.686	0.001
		3 – 4	11.568	0.085			3 – 4	11.566	0.073

^a Indicators 0, 1, 2, 3, and 4 specify graduate students, postdocs, assistant professors, associate professors, and full professors respectively.

Note: Significant relationships are in boldface italics.

Figure 6 illustrates the relationships between metadata element use and researcher seniority levels by showing both the mean rank and average value of these metadata elements for each seniority level. The results show that there could be some variations on the values of the metadata elements according to the profile owner's seniority.

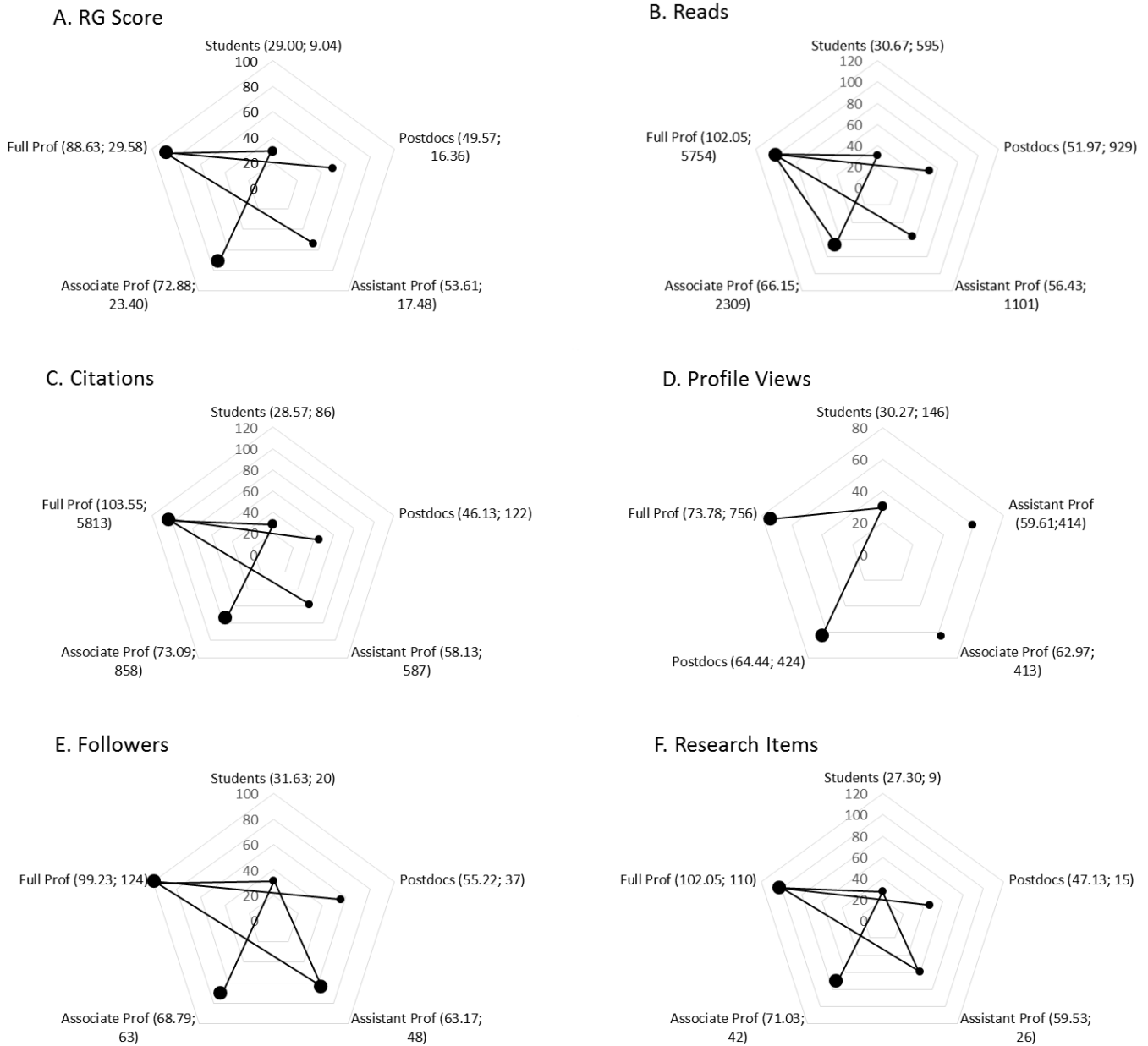


Figure 6. Relationships between metadata element use and the levels of researcher seniority. Numbers in parentheses indicate the mean ranks and the average value of the metadata elements for seniority groups. An edge between a pair of nodes on the graph indicates a statistically significant difference between the seniority groups for the metadata elements ($p < 0.05$).

4.4. What are the relationships between metadata element use and the categories of discipline in ResearchGate?

From the results of the Kruskal-Wallis test (see Table 4), the authors identified that the only statistically significant relationship between discipline and metadata elements in ResearchGate was in the *Reads* element ($\chi^2 = 9.526$, $df = 4$, $p = 0.049$). The authors used the Bonferroni multiple comparison tests as a follow-up to examine the pairwise relationships

between individual disciplines (see Table 8). The follow-up tests revealed that there were no significant differences between individual disciplines for the metadata element *Reads*.

Table 8. Bonferroni multiple comparison test for Reads among disciplines^a

Metadata element	Sample 1 – Sample 2	Std. Error	Adj. Sig.
Reads	0-1	13.189	0.485
	0-2	13.330	1.000
	0-3	14.193	0.637
	0-4	13.330	0.378
	1-2	8.848	0.406
	1-3	10.101	1.000
	1-4	8.848	1.000
	2-3	10.283	0.734
	2-4	9.056	0.289
	3-4	10.283	1.000

^a Indicators 0, 1, 2, 3, and 4 specify Humanities, Social Sciences, Physical Sciences, Life Sciences, and Engineering respectively.

5. Discussion

5.1. Which metadata elements do researchers currently use in their ResearchGate profiles?

This study investigated which metadata elements were used by researchers in a selected RIMS, as well as the relationships between researchers' characteristics and their use of metadata elements in their profiles. Researchers mainly used the metadata elements in the categories of *person*, *publication*, and *research subject*. The metadata elements related to *teaching experience* and *contact* were used the least (see Figure 1). RIMSs aggregate different types of metadata about scholarly activities, which generally include *research*, *grant*, *award*, *teaching*, and *service* (Bryant et al., 2017). However, this study found that the actual use of metadata elements shows an imbalance in favor of providing support for research activities (see Figure 1). Many research institutions prioritize research over teaching and service in their evaluation models (Fairweather and Beach, 2002). The priority of collecting and managing research content in RIMSs also seems to affect how these systems are usually defined: (1) systems for research, (2) systems for researchers, or (3) systems for scholarly activities, e.g., research, teaching, and service (OCLC Research, 2017). These definitions may also suggest to users that RIMSs are intended to be used for managing research content only, and so discourage them from sharing metadata and content related to their teaching and service activities (e.g., a list of the courses they teach, course syllabi and slides, or information about their professional services).

This study found that the selected RIMS, ResearchGate, curated metadata elements about scholarly activities, including research, teaching, and professional activities and service. However, the metadata supporting research publication was well represented, and the metadata used for other scholarly activities was sparse: these metadata elements had only been used between zero and 28 percent of the time (see Table 3 and Figure 1). Research and teaching are intertwined in academia. However, academics often teach in areas that are not related to their active research projects. That is, their teaching identity may not be the same as their research identity. Hence, it is important for academics to share information about their teaching activities so that RIMSs can provide them with content recommendations relevant to their teaching interests or activities. Furthermore, building and maintaining comprehensive scholarly activity profiles could benefit academics as well as their institutions. For example, many universities implemented their local and public RIMSs (e.g., Griffith Experts[1], Scholars@Duke[2], Scholars@TAMU[3]) as institution-level enterprise systems that include information about other identity data, such as teaching and service. These systems can support the needs of their diverse user communities, including students, faculty, administrators, and librarians.

5.2. What are the relationships between metadata element use and researchers' participation levels and demographic information in ResearchGate?

Counting the metadata elements used by researchers revealed the following: (1) Community Members are more willing to share their personal information (i.e., photo and position) than Readers and Personal Record Managers (see Figure 2); and (2) compared to Readers, Community Members and Personal Record Managers are more willing to provide the full texts of their works (as downloadable files) on ResearchGate (see Figure 3). Along with these findings based on descriptive statistics, the findings of the multiple statistical analyses (e.g., Kruskal-Wallis test, Bonferroni multiple comparison test, and logistic regression; see Figure 5) enabled the authors to construct a metadata profile for each researcher participation group (see Figure 7). These profiles present sets of metadata elements that have significantly higher odds of being used by a particular RIMS participation group than by the other groups. For example, the profile of the Readers group has the least number of metadata elements. The Community Members group profile has the highest number of metadata elements, including *photo*, *first name*, *last name*, *affiliation*, *position*, *research experience*, *project*, and *skills and expertise*. These profiles can be used to identify the overall value researchers place on RIMS metadata, as well as the metadata needs of researchers at different levels of RIMS participation. In addition, the significant relationships between metadata use and the levels of researcher seniority (see Figure 6) can also be supportive materials to help identify the value structure. The findings can be used to construct researcher identity metadata templates and infrastructure services in RIMS that are aligned with researchers' metadata needs and priorities (Foster et al., 2004; Stvilia and Gasser, 2008). Alternatively, these findings can help identify the features that can be used in a machine learning algorithm(s) to classify researchers automatically by the RIMS participation levels. Predicting a researcher's participation level can help RIMS librarians or curators craft tailored messages that suggest performing the tasks associated with the participation level, and connect to and enhance the researcher's motivations for those tasks (Stvilia et al., 2018). As a result they can become more successful in engaging the researcher in the RIMS. The literature shows that tailoring system design and communication messages to users' motivations has proven to be successful in assisting users' information and behavior changes (e.g., Kreuter et al., 1999).


	Readers	Personal Record Managers	Community Members
Most common metadata elements supplied by users	First Name, Last Name Department	First Name, Last Name Department Position Research Experience	 First Name, Last Name Department Position Research Experience Project Skills and Expertise

Figure 7. Relationships between researchers' participation levels and their use of metadata in researchers' profiles

6. Conclusion

This paper examined the relationships between researcher characteristics and their use of metadata in their ResearchGate profiles. To the best of our knowledge this is the first study that provided an empirical evidence of differences in the metadata use of researchers at different levels of their careers. In addition, the study contributed to the literature by conceptualizing the levels of researcher participation in RIMSs and constructing a metadata profile for each of those levels. Most researchers shared metadata related to their research rather than their teaching or service. Statistical analyses revealed statistically significant relationships between researchers' metadata use and their participation levels in RIMSs, as well as between metadata use and researchers' seniority. The study's findings help to

identify researchers' priorities for different metadata elements, as well as to construct profile metadata templates for each specific participation level (see Figure 7).

The study has limitations. Since ResearchGate does not publish its metadata and service schemas, the authors had to construct its metadata model for research identity information by examining the ResearchGate profiles. Hence, this set might not be a complete set of elements used by ResearchGate. Furthermore, the study used a sample of 126 researchers representing a single RIMS which limits the generalizability of its findings.

The future research possibilities created by this study include the investigation of the priorities of other RIMS stakeholder groups for research identity metadata (e.g., administrators). Future research should also examine how different tailored communication strategies can increase researchers' sharing of their research identity metadata in RIMs.

Acknowledgments

This research is supported by an OCLC/ALISE Library and Information Research Grant for 2016 and a National Leadership Grant from the Institute of Museum and Library Services (IMLS) of the U.S. Government (grant # LG-73-16-0006-16). This article reflects the findings and conclusions of the authors, and does not necessarily reflect the views of IMLS, OCLC, and ALISE.

Note

1. Griffith Experts at Griffith University, Brisbane, Australia (<https://experts.griffith.edu.au/>)
2. Scholars@Duke at Duke University, North Carolina, USA (<https://scholars.duke.edu/>)
3. Scholars@TAMU at Texas A&M University, Texas, USA (<http://scholars.tamu.edu/>)

References

- Borgman, C.L., Wallis, J.C., Enyedy, N., 2007. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *Int. J. Digit. Libr.* 7, 17–30. <https://doi.org/10.1007/s00799-007-0022-9>
- Bryant, R., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., Mercer, H., Missingham, R., Oxnam, M., Rauh, A., Wright, J., 2017. Research information management: Defining RIM and the library's role (No. 1005988188), OCLC Research Position Paper. OCLC Research, Dublin, OH.
- CASRAI, n.d. CASRAI dictionary [WWW Document]. URL http://dictionary.casrai.org/Main_Page (accessed 1.4.18).
- David, R.H., Thomas, D., 2015. Assessing Metadata and Controlling Quality in Scholarly Ebooks. *Cat. Classif. Q.* 53, 801–824. <https://doi.org/10.1080/01639374.2015.1018397>
- Fairweather, J.S., Beach, A.L., 2002. Variations in Faculty Work at Research Universities: Implications for State and Institutional Policy. *Rev. High. Educ.* 26, 97–115. <https://doi.org/10.1353/rhe.2002.0024>
- Faniel, I.M., Jacobsen, T.E., 2010. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Comput. Support. Coop. Work CSCW* 19, 355–375. <https://doi.org/10.1007/s10606-010-9117-8>
- Foster, I., Jennings, N.R., Kesselman, C., 2004. Brain meets brawn: Why grid and agents need each other, in: *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004. Presented at the Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004*, pp. 8–15.
- Greenberg, J., 2001. A quantitative categorical analysis of metadata elements in image-applicable metadata schemas. *J. Am. Soc. Inf. Sci. Technol.* 52, 917–924. <https://doi.org/10.1002/asi.1170>
- Haak, L., Baker, D., Hoellrigl, T., 2014. CASRAI and ORCID: Putting the Pieces together to Collaboratively Support the Research Community. *Procedia Comput. Sci.*, 12th International Conference on Current Research Information Systems, CRIS 2014 33, 284–288. <https://doi.org/10.1016/j.procs.2014.06.045>
- Jeffery, K., Houssos, N., Jörg, B., Asserson, A., 2014. Research information management: the CERIF approach. *Int. J. Metadata Semant. Ontol.* 9. <https://doi.org/10.1504/IJMSO.2014.059142>

- Jeng, W., DesAutels, S., He, D., Li, L., 2017. Information exchange on an academic social networking site: A multidiscipline comparison on researchgate Q&A. *J. Assoc. Inf. Sci. Technol.* 68, 638–652. <https://doi.org/10.1002/asi.23692>
- Jörg, B., Höllrigl, T., Baker, D., 2014. Harmonising and Formalising Research Administration Profiles CASRAI / CERIF. *Procedia Comput. Sci.*, 12th International Conference on Current Research Information Systems, CRIS 2014 33, 95–102. <https://doi.org/10.1016/j.procs.2014.06.016>
- Kreuter, M.W., Farrell, D.W., Olevitch, L.R., Brennan, L.K., 1999. *Tailoring Health Messages: Customizing Communication With Computer Technology*, 1 edition. ed. Routledge, Mahwah, N.J.
- Lee, D.J., Stvilia, B., 2017. Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLOS ONE* 12, e0173987. <https://doi.org/10.1371/journal.pone.0173987>
- Lee, D.J., Witt, M., Urban, R., Plale, B., 2014. A Metadata Application Profile for ORCID. https://docs.lib.purdue.edu/lib_fspress/79/
- Main features of CERIF | euroCRIS [WWW Document], n.d. URL <https://www.eurocris.org/cerif/main-features-cerif> (accessed 1.4.18).
- Martin-Martin, A., Orduna-Malea, E., Ayllon, J.M., Lopez-Cozar, E.D., 2016. The counting house: measuring those who count. Presence of Bibliometrics, Scientometrics, Informetrics, Webometrics and Altmetrics in the Google Scholar Citations, ResearcherID, ResearchGate, Mendeley & Twitter. *ArXiv160202412 Cs*.
- Pinto, C.S., Simões, C., Amaral, L., 2014. CERIF – Is the Standard Helping to Improve CRIS? *Procedia Comput. Sci.*, 12th International Conference on Current Research Information Systems, CRIS 2014 33, 80–85. <https://doi.org/10.1016/j.procs.2014.06.013>
- Quix, C., Riechert, M., 2017. Modelling National Research Information Contexts Based on CERIF. *Procedia Comput. Sci.*, 13th International Conference on Current Research Information Systems, CRIS2016, Communicating and Measuring Research Responsibly: Profiling, Metrics, Impact, Interoperability 106, 253–259. <https://doi.org/10.1016/j.procs.2017.03.023>
- Smith-Yoshimura, K., Altman, M., Conlon, M., Cristán, A.L., Dawson, L., Dunham, J., Hickey, T., Hook, D., Horstmann, W., MacEwan, A., Schreur, P., Smart, L., Wacker, M., Woutersen, S., 2014. Registering Researchers in Authority Files (No. 893635903). OCLC Research, Dublin, OH.
- Stvilia, B., Gasser, L., 2008. Value-based metadata quality assessment. *Libr. Inf. Sci. Res.* 30, 67–74. <https://doi.org/10.1016/j.lisr.2007.06.006>
- Stvilia, B., Hinnant, C.C., Wu, S., Worrall, A., Lee, D.J., Burnett, K., Burnett, G., Kazmer, M.M., Marty, P.F., 2015. Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *J. Assoc. Inf. Sci. Technol.* 66, 246–263. <https://doi.org/10.1002/asi.23177>
- Stvilia, B., Wu, S., Lee, D.J., 2018. Researchers’ participation in and motivations for engaging with research information management systems. *PLoS One*. Vol. 13 No. 2, e0193459. <https://doi.org/10.1371/journal.pone.0193459>
- Thelwall, M., Kousha, K., 2017. ResearchGate articles: Age, discipline, audience size, and impact. *J. Assoc. Inf. Sci. Technol.* 68, 468–479. <https://doi.org/10.1002/asi.23675>
- Thelwall, M., Kousha, K., 2015. ResearchGate: Disseminating, communicating, and measuring Scholarship? *J. Assoc. Inf. Sci. Technol.* 66, 876–889. <https://doi.org/10.1002/asi.23236>
- Van Noorden, R., 2014. Online collaboration: Scientists and the social network. *Nat. News* 512, 126. <https://doi.org/10.1038/512126a>
- Wu, S., Stvilia, B., Lee, D.J., 2017. Readers, personal record managers, and community members: An exploratory study of researchers’ participation in online research information management systems. *Journal of Library Metadata*. Vol. 17 No. 2, pp. 57-90. <https://doi.org/10.1080/19386389.2017.1348783>
- Zavalina, O., Kizhakkethil, P., 2015. Exploration of Metadata Change in a Digital Repository [WWW Document]. IConference 2015 Newport Beach Calif. U. S. URL <https://digital.library.unt.edu/ark:/67531/metadc503265/> (accessed 1.4.18).