

## "I cannot tell what the dickens his name is": Name Disambiguation in Institutional Repositories

Lizzy A. Walker, Michelle Armstrong

---

Walker, LA, Armstrong, M. (2014). "I cannot tell what the dickens his name is": Name Disambiguation in Institutional Repositories. *Journal of Librarianship and Scholarly Communication* 2(2):eP1095. <http://dx.doi.org/10.7710/2162-3309.1095>

© 2014 by the author(s). This open access article is distributed under a Creative Commons Attribution License, which allows unrestricted use, distribution, and reproduction in any medium, providing the original author and source are credited.

JLSC is a quarterly journal sponsored and published by Pacific University Library | ISSN 2162-3309 | <http://jlsc-pub.org>

---

## Rights



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

## Erratum

*Correction [Author-supplied]:* On pp. 2-3, the authors describe ORCID's sources for harvesting works as being limited to Scopus. What is not mentioned is the ability to harvest from CrossRef, available when ORCID first started. Since the initial development of this paper, ORCID has continued to expand its ability to harvest information from other sources such as the Australian National Data Service and DataCite. As noted in the original article, this strategy supports authors in creating lists of their works, but has limitations for scholarship not contained in these sources.

## Authors

Lizzy A. Walker, Michelle Armstrong

# “I cannot tell what the dickens his name is”: Name Disambiguation in Institutional Repositories

Lizzy A. Walker *Assistant Professor, Metadata and Digital Initiatives Librarian, Wichita State University*

Michelle Armstrong *Assistant Professor, Scholarly Communications and Data Management Librarian, Boise State University*

## Abstract

**INTRODUCTION** Authors who publish under more than one form of their name, multiple authors with the same name, and incomplete author information can all create challenges for repository staff when entering metadata. Unless properly addressed, these variations and duplications can result in search and retrieval errors for users. Name disambiguation, the process of identifying, merging, and making names accessible in one standard form, is a vital process repository staff should incorporate into their workflow to address these issues. **DESCRIPTION OF PROGRAM** Staff working with ScholarWorks, Boise State’s institutional repository, are exploring the use of disambiguation tools to solve the issue of name duplication. Systems explored include ORCID, ResearcherID, Scopus, Google Scholar Citations, Names Project, and the Digital Commons’ Author Merge Tool. **NEXT STEPS** Based on this initial assessment, ScholarWorks staff will continue to use the Author Merge Tool on a regular basis and explore ways to document and retain information discovered during the analysis phase. Additionally, they will continue to experiment with emerging name authority tools, such as ORCID. Finally, metadata specialists are encouraged to advocate for international standards that will provide prescribed rules for how metadata is entered into a repository system.

## INTRODUCTION

Name disambiguation is the process of merging variations of an author’s name into one standard form. This provides for consistent identification, improved discovery, and helps publishers present complete author information. When creating author metadata, institutional repositories face unique challenges with name representation. For example, multiple name entries for the same author in browsable lists impedes searching

for a single author. Duplications also create problems when crosswalking metadata between repositories. Often, authors from different institutions collaborate on research and publishing projects. If two repositories have different rules for name entry, this creates problems when searching. For example, Digital Commons, a hosted repository platform, allows for searching across the repositories using their system. If author names are entered differently, there is a possibility that the search will not yield complete information.

Received: 07/08/2013 Accepted: 12/31/2013



© 2014 Walker & Armstrong. This open access article is distributed under a Creative Commons Attribution 4.0 Unported License, which allows unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## LITERATURE REVIEW

### Challenges in name representation

Authority control has been defined as, “the formulation and recording of authorized heading forms in catalog records” (Maxwell, 2002, p.1). Clack (1990) also notes that authority control “involves research, the creation of standardized forms of access points, and linkages to variant forms” (p. 1). These elements of standardization, authorization, and research-based decision-making have helped librarians ensure that users can find exactly what they need when using a traditional library catalog. Additionally, librarians have developed tools to assist with these efforts. For example, catalogers use the Library of Congress Authorities, specifically, the Name Authorities Heading search, to determine the form of name to use for access points in cataloging records. Metadata specialists working with institutional repositories, however, do not have national standards or tools to assist with similar name entry and authority control efforts. As a result, each repository must create its own. This can cause variations in how names are entered and result in name duplication problems.

According to Smalheiser and Torvik (2009) author name disambiguation “comprises four distinct challenges” (p. 1). First, authors write and publish under more than one form of their name. They may use their full name or, depending on the requirements of publications, their first initial. Some authors also change their names or use variations of their first name. For instance, a professor may publish under *Robert* or *Roberto* depending on the language of the publication.

Another example of published name variations is that of hyphenated names (Scoville, 2003). An author with the name Kurtz-Smith may be represented elsewhere as Kurtz Smith. Secondly, some names are common and there is a chance that there are multiple authors with the same name. Third, metadata entries could be incorrect or incomplete. Smalheiser and Torvik (2009) mention that “some publishers and biographical databases did not record authors’ first names, their geographical locations, or identifying information such as their degrees or positions” (p. 1), all of which assist institutional repositories in metadata entry. Finally, disambiguation is complicated when authors have the same name, but come from different institutions or disciplines. For example, if

there are two authors with the name of Jeffrey Smith from the same institution in different departments, the subject matter of the articles may be the only way to identify the specific author.

A major issue in name representation is addressed by Salo (2009) in her article on name authority and institutional repositories. She notes that if users search for works by an author who has published under various forms of their name, searchers are not able to retrieve a comprehensive list of that author’s works. This happens due to the fact that there is no name authority control in metadata records. Metadata schemas, such as Dublin Core Qualified, are standards, but they make no mention of how to enter names in metadata records. Existing metadata standards, according to Salo, “do not incorporate authority control mechanisms” (p. 254) and most repositories utilize Dublin Core metadata standards (p. 254).

### ORCID: A blossoming possibility

ORCID, or Open Researcher and Contributor ID, launched in October 2012. According to their website, ORCID is “an open, non-profit, community-based effort to create and maintain a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers” (“About ORCID,” 2012). ORCID allows for both author-initiated disambiguation, as well authority-initiated disambiguation. For individual researchers, ORCID provides a registry to secure unique identifiers and “manage a record of activities” (“About ORCID”). Additionally, registered ORCID members, such as universities, can use an application programming interface to create and manage ORCID records on behalf of an author.

Similar to a journal article’s digital object identifier (DOI), ORCID provides a unique persistent alpha-numeric code assigned to a single author. A profile is created providing personal information, such as published names, author website, biography, country, and scholarly works. Publication data can be harvested from Scopus, one of the largest citation databases of peer-reviewed research (“Content Coverage Guide”, 2011, p. 4), into an author’s profile. This aids the author in that s/he does not have to manually enter the information into the ORCID record. While this service may be beneficial to some researchers, it still remains to be seen how the harvesting tool will be utilized by authors who publish in journals that are

not indexed by Scopus. Additionally, there are sections of the author profile that are still being developed, including affiliations, grants, and patents. Stern (2010) notes that ORCID “could be used eventually for many other purposes, such as grant evaluation and tenure considerations” (p. 31). In regard to these cases especially, it is important for an institution to be able to identify the correct author.

Name disambiguation does not just affect academic institutions. Haak, Fenner, Paglione, Pentz, and Ratner (2012) mention that this issue also affects publishers. For example, if they cannot locate the right author history and citation metadata, they can also experience ambiguity problems. Haak et al. (2012) also report ORCID could “serve an important role in supporting efforts in the publishing community, including conflict-of-interest reporting and author role acknowledgment” (p. 259). For instance, when an author creates a record of a work, ORCID allows the author to assign their role such as research, writing, and several other options depending on the type of entry. Authors may also select from Assignee or Co-inventor if the entry is a patent. By having the ability to add this information, ORCID gives the author a valuable tool in helping others identify his or her works twofold. First, the identification number is unique to an author. Secondly, the author facilitates his or her findability by adding his or her own information to their records within the ORCID database. Stern points out that there is the “hope” that authors will actively participate in the project to help create a “more current and accurate database than one developed solely by outside parties” (p. 31). Despite the time that it would take for them to manually enter their data, the authors would have a hand in helping disseminate their scholarship in that end users will find the research more easily.

Smalheiser and Torvik (2009) address some of the challenges with regard to using an ID based system for identifying authors. Specifically, “it fails to take into account the realities of human behavior” (p. 4). For the system to work, Smalheiser and Torvik have a similar view to that of Stern in that authors need to actively participate in the project by manually adding “their own data accurately and periodically” (p. 4). They also point out that all authors would need to participate, even those who are not primary authors, and they must enter all information that would help end users. However, this suggestion may not be appealing to researchers. They

state, “We have not even been able to convince our own colleagues to add their middle initials or suffixes when publishing papers, even though this would take almost no time or effort and would assist in disambiguation” (p. 5). Leaving data entry in the hands of researchers who may have limited time or not want to take part in the project introduces the possibility of incomplete information.

ORCID attempts to address this issue in several ways. In addition to providing options for universities and other trusted entities to manage author records, ORCID provides data sharing options for related services. Besides being able to import publication information from SCOPUS, ORCID IDs can also be linked to other author identifier services such as ResearcherID (Notess, p. 64). ORCID is also developing tools to assist publishers in integrating ORCID IDs in the publishing workflow (Haak, 2013). These strategies provide tremendous potential to assist authors in creating a complete list of formally published works, but do not sufficiently address the creation of other types of scholarly works, such as presentations, white papers, and other grey literature. In these cases, the author or an authorized ORCID member would need to manage the ID record. Despite these limitations, ORCID is continuing to grow and has the potential to support name authority control in the institutional repository environment.

### **ResearcherID: What's in a name?**

ResearcherID, created by Thomson Reuters, is a free author initiated service whereby authors register for a unique alpha-numeric identifier and create their own record. The author may enter affiliations, publications, URLs, subjects, and research interest information into their profile. Citations and times cited information, compiled by Web of Science, is automatically updated in ResearcherID providing additional journal article data. Rotenberg and Kushmerick (2011) state that users can “self-classify themselves in a given area of research/expertise through association of user-generated keywords in their profile” (p. 514) providing additional verification of an author's work. Users may also incorporate their unique identifiers into their professional websites using the ResearcherID badge (p. 514-515). They may also use Web of Science to search for specific authors by their unique identifiers. ResearcherID Labs provides a variety of tools that visualize research collaborations and author publications (p. 515). Since ResearcherID is a service that

requires author input, it can benefit researchers by giving them direct control over identifying their work.

One disadvantage of the system is that the author cannot crosslist if s/he has two distinct roles. For example, some academic librarians are also faculty members. However, in ResearcherID, s/he must select one or the other as his or her role. A benefit to using ResearcherID is that after the author is assigned their unique identifier, there is the option to connect with an ORCID account. This allows publication and author data to be shared between the systems, thus eliminating the need to reenter information.

### **Scopus: Author Identifier and Author Profile**

Created by Elsevier, the Scopus Author Identifiers and Profiles “provide a rival to ResearcherID” (Notess, 2013, p. 61). Whereas ResearcherID is author initiated, Scopus is authority initiated. Records have been created in such a way that “profile pages are available for all authors and do not require individual scholars (or their institutions) to enter anything” (p. 62). In addition to assigning unique numeric Author Identifiers, information that is harvested includes alternative names, affiliations, number of documents published, references, h-index information, co-authors, and other information that would help identify an author (p. 63). A list of publications is also provided. The service collocates the authors’ works “based on their similarity in affiliation, publication history, subject, and coauthors” (Moed, Aisati, & Plume, 2013, p. 931). In case the algorithms do not pull all of the name forms an author uses, that author may request a merge of the appropriate profiles (Notess, p. 63). Users without appropriate access rights might not be able to access some articles, but the information is publically visible.

Notess (2013) states that one major advantage of this service is that Scopus is “a much larger database of scholars since it includes many who do not even know that they are included” (p. 62). It may be questioned whether this is a true advantage as it may not be determined how many author profiles exist that require merging or how much information on a profile might be incorrect. Another disadvantage is that the form of author name they use is based on the article from which the profile is created (“Author Identifier,” 2013). This may be problematic in that wrong attribution would be applied in some cases, and in others the Author Profile may be incomplete for a given author (Moed et al., p. 931).

### **Virtual International Authority File: Also known as VIAF**

Utilizing linked data is another viable option for name disambiguation. The Virtual International Authority File (VIAF) is an example of such a linked data repository and is a combined effort between the Library of Congress, OCLC, the Bibliothèque nationale de France, and the Deutsche Nationalbibliothek (Moulaison and Stanley, 2013, p. 45). VIAF is “designed to provide convenient access to the world’s major name authority files” (“VIAF,” 2013). Collected into a central database, national and regional authority files are merged for an author. The combined record is then assigned a unique identifier (Niu, 2013, p. 413). This is done by matching and linking authority files of various libraries and merging these files into a “super” authority record, combining names for an author (“VIAF,” 2013). VIAF provides a “convenient means for a wider community of libraries and other agencies to repurpose bibliographic data produced by libraries serving different language communities” (“VIAF,” 2013).

VIAF is not without its flaws. For instance, the system pulls information automatically and matches authority records for the same name. Niu (2013) makes the argument that if multiple authority records belonging to one “bibliographic identity are not matched, each will get a separate ID” (p. 413-414). Despite this, Niu makes the prediction that “identity management systems will be linked to, and aggregated with, library authority databases, and globally unique IDs will be used in place of authorized headings” (p. 418). By adopting these systems, institutional repositories will be able to “improve search precision and recall” (p. 418).

### **Google Scholar Citations: Author-initiated name disambiguation**

In 2011 the Google Scholar Citations service was released, whereby participation is author initiated. All the author needs is an institutional “.edu” account and the willingness to enter his or her data. With the information authors provide, Google harvests scholarly works from the web and populates the profiles. Google’s mission with this service is to “provide a simple way for authors to keep track of citations to their articles” (“Google Scholar Citations,” 2013). Notess (2013) states that “some scholars enter the URL for the Google Scholar Citations



pages as a homepage in ResearcherID” (p. 64). This further assists in name disambiguation in that the author has linked two resources together which ensures that the correct author name has been established.

Unfortunately, name issues still arise using this service. Specifically, a serious disadvantage to this system, as with others that use algorithms to automatically harvest information from the web is the lack of reliability in the results. Google states, “We use a statistical model to try to tell different authors apart but such automatic processes are not always accurate” (“Google Scholar Citations,” 2013). According to Google, the author may opt to review the updates to their profile or enter the citation information manually (“Google Scholar Metrics,” 2013). Despite these alternatives, the dependence on individual author mediation prevents Google Scholars Citation from being a viable tool to assist institutional repositories in name disambiguation.

### **The Names Project: Another strategy**

In 2007, with funding from the JISC Repositories and Preservation Programme, the British Library and Mimas, a national data center in the United Kingdom, began working on the Names Project, an authority initiated program, designed “to investigate requirements for a name authority service for UK repositories,” (Cross, Danskin, Hill, & Needham, 2011, p. 4). It also provides a “prototype name authority service for individuals and institutions in order to demonstrate the feasibility of such a system” (Danskin, Hill, & Needham, 2011, p. 15). This came about, according to Hill (2008), to address the influx of institutional repositories in the United Kingdom and the need for improving name representation in institutional repositories.

To develop such a name authority prototype, the Names Project used data from the British Library’s Electronic Table of Contents (ETOC). Consisting of approximately 38 million records, the data set allowed the project to test and refine its disambiguation algorithm. Further testing utilized a smaller dataset from MERIT and was evaluated with assistance from the British Library (Danskin, Hill, & Needham, p. 16). Not only did the Names Project demonstrate the great potential of such an automated system, it also produced a large set of disambiguated records

which they were able to share with the International Standard Name Identifier (ISNI) initiative (Cross et al., p. 8). Additionally, the Names Project has made a point of engaging stakeholders, including participating in meetings with ORCID (Cross et al., p. 4).

Despite its impressive accomplishments, the Names Project has certain limitations which make it an impractical tool for many institutional repositories. Although the resulting dataset was quite large, almost 47,000 records, the authors included were primarily from the United Kingdom (Cross et al., p. 7). Repositories in other countries would find few of the researchers included in their collections, making the Names Project of limited value. Additionally, even with the improved disambiguation, participants in the Names Project reported that human mediation will continue to be a part of the process. “Future services will require an element of human review to resolve ambiguities and for quality assurance” (Danskin, Hill, and Needham, p. 19).

## **DESCRIPTION OF PROGRAM**

### **ScholarWorks: Boise State’s research showcase**

In 2008, Boise State University launched ScholarWorks, its institutional repository designed to capture and showcase the institution’s scholarship. The following year, ScholarWorks’ staff began uploading content into the system. Using the Digital Commons platform, ScholarWorks’ staff implemented a mediated deposit model where they identified eligible faculty scholarship, reviewed publisher’s copyright policies, solicited author permissions, obtained the correct version of the publication, and uploaded the document and appropriate metadata into ScholarWorks. Although this approach provided a useful service to faculty and assisted in ensuring quality metadata, it still took time for the repository collection to grow. As a result, the problem of name duplication errors within the institutional repository developed slowly over a period of several years.

Initially, it was not easy to recognize that there was a growing issue with duplicate names. Records for faculty publications are typically uploaded one at a time, and metadata is created using the document as the primary information source. As a result, an author’s name is entered as it appears in the publication. However, as previously mentioned, author names tend to vary and change

over time for a variety of reasons, and ScholarWorks' staff found that this was true for Boise State authors as well. Consequently, as more content was added to the institutional repository, ScholarWorks' staff began to notice problems on the *Browse by Author* page, a complete list of Boise State authors included in the system. These issues ranged from spelling variations, differences due to completeness of the name entered, and multiple listings for authors whose name had changed. Although the author name metadata was correct, the *Browse by Author* list was not.

Overall, the uploading and metadata creation practices used by ScholarWorks' staff were appropriate and resulted in an accurate record for each object ingested into the repository. However, the way that individual metadata was used by the system, and ultimately became discoverable, was problematic. Since the *Browse by Author* page provides a public list of Boise State scholars, display problems were making it difficult to find a comprehensive list of works by a single author and at times presented a confusing interface to the repository content.

**Author Merge Tool: Name disambiguation in practice**

To disambiguate is to make something understandable or clear. In regard to name disambiguation, this means to resolve problems resulting from variations in author

names. One method of accomplishing this goal is to determine the form of the author's name the repository chooses to use and after research and analysis, merge the various names found in the author list. Digital Commons provides tools to assist with such name disambiguation. A growing, hosted platform, the software supports over 250 institutional repositories worldwide, with nearly a million records included in these collections. The name disambiguation tool included in the system is called the *Author Merge Tool* and is an authority initiated process. According to Berkeley Electronic Press (2011), this tool is designed to "unify an author's search results under the author's full professional name" (Introduction, para. 1). The ScholarWorks' staff recently used it and found improvements within the *Browse by Author* list.

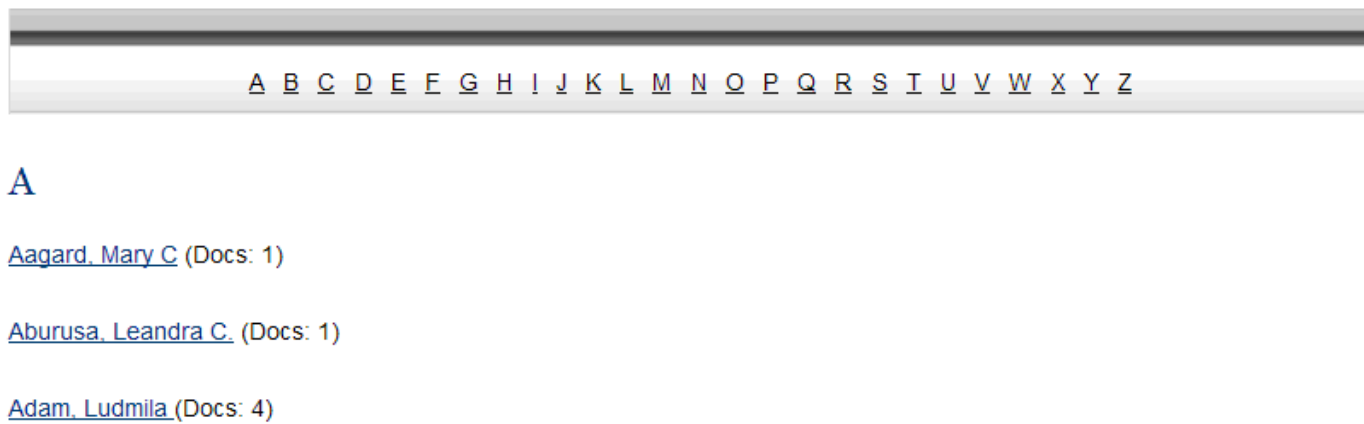
To begin the merging process, the metadata specialist first identifies duplicate names from the *Browse by Author list* (see Figure 1). This is a master list created from the information entered in the *Author* metadata field, which site visitors may use to select the author they would like to browse.

Once the list of names needing to be disambiguated is created, the metadata specialist uses the *Author Merge Tool* available through the *Site Administrator Tools* to conduct another search for authors' last names (see Figure 2, following page).

**Figure 1. Browse by Author list**

**Browse by Author**

Listing of authors who have works in this repository as of March 01, 2013. Click the name of an author to see a listing of that person's work. For more information, see [About the Repository](#).





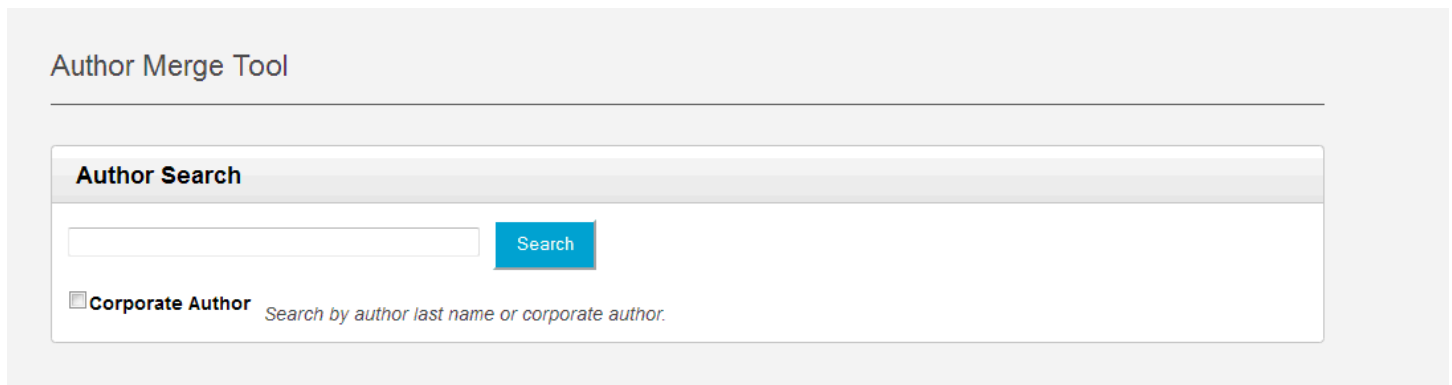
The metadata specialist reviews the resulting list and selects the name that indicates the *Primary Author* (see Figure 3).

Ways of verifying author names include consulting the author's curriculum vitae, researching departmental websites or university documents, consulting with Human Resources staff or systems, utilizing library liaisons' relationships with faculty, or contacting the researcher directly. Institutional repository staff are in a unique position as they have the advantage of knowing many of the authors and their research interests. This is especially true for institutional repositories utilizing a mediated deposit model where IR staff carry out all ingest processes. In particular, the mediated deposit model often requires staff to be knowledgeable of local research initiatives and have established communication mechanisms in place to

facilitate the procurement of permissions and files. These existing processes can facilitate a direct connection with the authors reducing the time it takes to research and verify author name forms. For example, in the case of a repository platform such as Digital Commons, which allows harvesting of metadata from one repository to another, authors with identical names may be accidentally ingested into the wrong repository. When discovered, IR staff can utilize their existing ingest process and connections with local authors to verify the eligibility of the publication and manage any needed withdrawals.

Once any needed research is completed, and the name is selected, it appears in the *Browse by Author* list rather than having multiple names for one author. This disambiguates, or merges, the names to be included in the *Browse by Author* list under the *Primary Author*.

**Figure 2. Author Merge Tool**



Author Merge Tool

**Author Search**

**Corporate Author** *Search by author last name or corporate author.*

**Figure 3. List of names for disambiguation purposes from the last name of Park.**

Merge		Reset		Authors					
Primary Author	Merge	First Name	Middle Name	Last Name	Institution	Corporate Author	Email	Submission	
<input type="radio"/>	<input type="checkbox"/>	Hyungsung		Park	Kyonggi University			<a href="#">A New Approach Toward Digital Storytelling: An Activity Focused on Writing Self-Efficacy in a Virtual Learning Environment</a>	
<input type="radio"/>	<input type="checkbox"/>	Jordan		Park	Department of Philosophy, Boise State University			<a href="#">The Unreliability of Philosophical Intuitions</a>	
<input type="radio"/>	<input type="checkbox"/>	Sae-Hoon		Park	Chonbuk National University			<a href="#">Two Libraries on Two Continents: How Do They Compare?</a>	

Ideally, use of the *Author Merge Tool* solves the problem of duplication in ScholarWorks. For example, the names “Mullner” and “Müllner” were both originally displayed in the authors list, which was indeed a duplication. As the faculty member’s actual name included the umlaut, the decision was made to disambiguate under that name. After updating the ScholarWorks system, the merge was found to be successful. Now when an author search was conducted using either “Müllner” or “Mullner” forms of the name, all of Müllner’s articles were displayed, regardless of how his last name was spelled.

Although these efforts corrected many name duplication issues, there were several challenges that came up when merging names. When using the search function within the *Author Merge Tool* itself, the last name alone is the only way to conduct a search with this tool. If the first full name or first initial of the author’s name is also entered into the search, it yielded no results. This created a problem if the author had a common last name. In these situations, search results yielded each record the repository held that contained that last name. The search for a common last name, combined with an author who has published a large number of articles, may create an extensive list of entries to review. At times, this was even true of less common names. For example, when conducting a search for the last name “Lamb,” twelve results came up, only three of which were the names needing to be merged.

Similarly, since the search feature in the *Author Merge Tool* only provided results when searching an author’s last name, there was no solution when trying to merge records for an author who published using different last names. For instance, in the case of an author who changes their last name due to marriage, the *Author Merge Tool* cannot be used to search for both names simultaneously. As a result, if the metadata specialist finds that the names cannot be merged, they will need to contact the Digital Commons technical support staff to request that the names be disambiguated.

Finally, ScholarWorks staff noted two other limitations of the *Author Merge Tool*. First, there is no way to document the research and merging within the software. This makes it difficult to correct future disambiguation errors. Second, these improvements are limited only to the institutional repositories within the Digital Commons system. Essentially, this means that the solution for ScholarWorks cannot be crosswalked from the Digital

Commons platform to other institutional repository software tools.

## NEXT STEPS

Without international metadata entry standards, repository managers and staff are left to develop their own guidelines and procedures to ensure the creation of accurate bibliographic information and discovery of all the works by an author. Because of this, ScholarWorks staff will continue to use the *Author Merge Tool* on a regular basis. Employing the previously described steps, they will identify possible duplications, and after the required research, merge the selected author names. ScholarWorks staff will also need to find ways to document and retain information discovered during the analysis phase. This will improve the overall workflow as it reduces research time in the future as author names continue to be merged.

Additionally, it will also be important for further experimentation or implementation with emerging tools, such as ORCID, and currently existing services, such as ResearcherID. Although the *Author Merge Tool* provides an immediate benefit, making it easier for site visitors to retrieve available works by a specific author, these benefits do not extend beyond the Digital Commons’ repository itself. Since Digital Commons is a proprietary system, other platforms are not able to utilize the *Author Merge Tool* and consequently the efforts taken by repository staff. As ORCID and other researcher identification services develop, ScholarWorks staff will explore the different features and possible advantages, allowing them to incorporate these benefits into the repository’s policies and procedures.

Given the growing number of repositories and unique types of content, the issue of name duplication is only going to continue. Since there are multiple platforms used to host repositories, it would benefit libraries to request, create, or improve current name disambiguation resources. As institutional repositories are starting to include original, unpublished material, they must catch up to traditional bibliographic description and discovery practices. For example, catalogers use various tools including the Authority File and entry standards like RDA. It would benefit metadata specialists to have similar resources. Although some standards exist such as Dublin Core and OAI, metadata specialists should advocate for international standards that will provide

prescribed rules of how metadata is entered into a repository system. Additionally, tools should continue to evolve so that they can assist with author identification and name disambiguation. Finally, name duplication should be prioritized by institutional repository managers to improve discovery and access.

## CONCLUSION

With the growing number of institutional repositories, it is clear that they are becoming an important part of the scholarly communication system. Millions of records have been created and will continue to be produced as new works are added. Consequently, authority control for author names will be critical to the discovery of these works. ScholarWorks staff's investigation into this issue and experimentation with the different systems discussed, revealed that although progress is being made, no single approach was wholly satisfactory. Name disambiguation, such as Digital Commons' *Author Merge Tool*, is one option for institutional repositories to use as platforms begin to include these features in their software. However, institutional repositories are encouraged to continue to research and collaborate, in order to develop methods for resolving name duplication. These efforts will provide a more efficient and successful discovery experience for the end user.

## REFERENCES

- About ORCID. (2012). Retrieved April 1, 2013, from <http://about.orcid.org/about>
- Author identifier. (2013). Retrieved September 13th, 2013, from <http://www.info.sciverse.com/scopus/scopus-in-detail/tools/authoridentifier>
- Berkeley Electronic Press. (2011). Author merge tool: How to combine author records in Digital Commons. Retrieved from <http://digitalcommons.bepress.com/reference/37/>
- Clack, D. (1990). Authority control: Principles, applications, and instructions. Chicago: American Library Association.
- Content Coverage Guide. (2011). Retrieved September 14, 2013, from [http://www.info.sciverse.com/UserFiles/sciverse\\_scopus\\_content\\_coverage\\_0.pdf](http://www.info.sciverse.com/UserFiles/sciverse_scopus_content_coverage_0.pdf)
- Cross, P., Danskin, A., Hill, A. & Needham, D. (2011). Names Project, Phase Two (Final report). Retrieved from [http://names.mimas.ac.uk/files/Final\\_Report\\_Names\\_Phase\\_Two\\_September\\_2011.pdf](http://names.mimas.ac.uk/files/Final_Report_Names_Phase_Two_September_2011.pdf)
- Danskin, A., Hill, A., & Needham, D. (2011). The Names Project: A new approach to name authority. *Information Standards Quarterly*, 23(3), 14-19. <http://dx.doi.org/10.3789/isqv23n3.2011.04>
- Google Scholar Citations. (2013). Retrieved September 14, 2013, from <http://www.google.com/intl/en/scholar/citations.html>
- Google Scholar Metrics. (2013). Retrieved September 14, 2013, from <http://www.google.com/intl/en/scholar/metrics.html>
- Haak, L. (2013, August 8). Suggested practices for collection and display of ORCID iDs in publishing workflows. Retrieved from <http://orcid.org/blog/2013/08/08/suggested-practices-collection-and-display-orcid-ids-publishing-workflows>
- Haak, L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing*, 25(4), 259-264. <http://dx.doi.org/10.1087/20120404>
- Hill, A. (2008). "What's in a name?": Prototyping a name authority service for UK repositories. Proceedings of the 10th International Conference of the International Society of Knowledge Organization. Retrieved from <http://names.mimas.ac.uk/files/>
- Maxwell, R. (2002). Maxwell's guide to authority work. Chicago: American Library Association.
- Moed, H., Aisati, M., & Plume, A. (2013). Studying scientific migration in Scopus. *Scientometrics*, 94(3), 929-942. <http://dx.doi.org/0.1007/s11192-012-0783-9>
- Moulaison, H. & Stanley, S. (2013), Beyond failure: Potentially mitigating failed author searches in the online library catalog through the use of linked data. *Journal of Web Librarianship*, 7(1), 37-57. <http://dx.doi.org/0.1080/19322909.2013.738562>
- NISO (2004). *Understanding metadata*. Bethesda, MD: NISO Press. Retrieved from <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- Niu, J. (2013). Evolving landscape in name authority control. *Cataloging & Classification Quarterly*, 51(4), 404-419. <http://dx.doi.org/10.1080/01639374.2012.756843>
- Notess, G. (2013). Searching for scholars. *Online Searcher*, 37(1), 61-64. Retrieved from <http://www.infotoday.com/OnlineSearcher/Articles/On-the-Net/Searching-for-Scholars-87407.shtml>
- Rotenberg, E. & Kushmerick, A. (2011). The author challenge: Identification of self in the scholarly literature. *Cataloging & Classification Quarterly*, 49(6), 503-520. <http://dx.doi.org/10.1080/01639374.2011.606405>
- Salo, D. (2009). Name authority control in institutional repositories. *Cataloging & Classification Quarterly*, 47(3-4), 249-261. <http://dx.doi.org/10.1080/01639370902737232>

Scovile, C., Johnson, E., & McConnell, A. (2003). When A. Rose is not A. Rose. *Medical Reference Services Quarterly*, 22(4), 1-11. [http://dx.doi.org/10.1300/J115v22n04\\_01](http://dx.doi.org/10.1300/J115v22n04_01)

Smalheiser, N. & Torvik, V. (2009). Author name disambiguation. *Annual Review of Information Science and Technology*, 43(1), 1-43. <http://dx.doi.org/10.1002/aris.2009.1440430113>

Stern, D. (2010). Author as object: Disambiguation and enhanced links. *Online*, 34(6), 29-33. Retrieved from <http://pqasb.pqarchiver.com/infotoday/>

VIAF (The Virtual International Authority File). (2013). Retrieved September 13th, 2013, from <http://www.oclc.org/research/activities/viaf.html>

## CORRESPONDING AUTHOR

Lizzy A. Walker  
*Assistant Professor*  
*Metadata and Digital Initiatives Librarian*

University Libraries  
Wichita State University  
1845 Fairmount  
Wichita, KS 67260-0068

[lizzy.walker@wichita.edu](mailto:lizzy.walker@wichita.edu)