Haley M. Walton. ORCID Integration Among Publishing and Funding Organizations: An Examination of Process and Rationale. A Master's Paper for the M.S. in L.S. degree. July 2014. 57 pages. Advisor: Kevin L. Smith.

This study examines how a sample of scholarly publishers and granting organizations have integrated the Open Researcher and Contributor ID (ORCID) into their grant application and manuscript submission workflows. The study was conducted to discover what benefits these organizations gain from using the ORCID unique author identifiers and how effective they are at introducing scholars to ORCID as a service.

The data was collected through interviews of representatives from a sample of publishing and funding organizations: the National Institutes of Health, the U.S. Department of Energy's Office of Science and Technical Information, the Wellcome Trust, Autism Speaks, Elsevier, and Oxford University Press. A representative from eJournalPress, a software company that provides manuscript management tools was also interviewed. The result is an analysis of best practices for ORCID integration at these types of organizations and suggestions for improvement. The conclusions drawn are generalizable to other institutions seeking to adopt ORCID themselves.

Headings:

Name authority records (Information retrieval)

Author name disambiguation

Scholarly publishing

Grants management

ORCID INTEGRATION AMONG PUBLISHING AND FUNDING ORGANIZATIONS: AN EXAMINATION OF PROCESS AND RATIONALE

by Haley M. Walton

A Master's paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina

July 2014

Approved by

Kevin L. Smith

Table of Contents

Introduction	2
Related Research	
Methods	13
Funding Organization Integration	
U.S. Department of Energy's Office of Science and Technical Information	18
Wellcome Trust	20
Autism Speaks	22
Analysis	24
Publishing Industry IntegrationElsevier	
Oxford University Press	32
eJournalPress	33
Analysis	35
Conclusion	37
Bibliography	42
Appendix A	
Appendix B	49

Introduction

Author name ambiguity in bibliographic records has been a long-standing problem across the landscape of scholarly publishing. The identification of authors by their surnames and first initial(s) alone, has rendered author name queries nearly meaningless. The records retrieved from a database will not include the body of work of a single "J. White," but of all the scholars sharing that name. Even today's powerful databases continue to use this convention, and as the volume of academic research published annually has skyrocketed—Thompson Reuters, for example, found that its *Web of Science* database ingested over 1 million new records in 2009 alone—the ambiguity problem has only been exacerbated. ¹

Name ambiguity is caused not only by names that are common enough to be shared by several authors, but also by authors with multiple name labels, such as people who have changed their last names or those who have published using a middle initial in some publications and none in others. For example, I could have published under the name "Walton, H." in one journal and "Walton, H.M." in another. These small discrepancies result in two separate records for the same person. Names using non-Latin characters are also problematic. Transliterations, especially of non-Latin characters, can vary across records, again resulting in multiple name records for a single

.

¹ Ellen Rotenberg and Ann Kushmerick, "The Author Challenge: Identification of Self in the Scholarly Literature," *Cataloging and Classification Quarterly* 49, no. 6: 505.

² Hua-Kuang Chen and Chi-Nan Hsieh, "Ambiguity Resolution for Author Names of Bibliographic Data," *Journal of Educational Media and Library Sciences* 49, no. 2 (2011): 216; Andreas Strotmann and Dangzhi Zhao, "Author Name Disambiguation: What Difference Does It Make in Author-Based Citation Analysis?" *Journal of the American Society for Information Science and Technology* 63, no. 9 (2012): 1820.

individual.³ As metadata librarians and catalogers will bemoan, poorly formed records rife with typos and misspellings also create unnecessary individual records.⁴ The complex nature of name ambiguity has made it persistently problematic for information scientists.

Not surprisingly, over the years there have been a number of proposed solutions to facilitate the disambiguation of names. Some of these include methods of automatic name disambiguation. These methods are the product of research into artificial intelligence and machine learning in which algorithms can be "taught" to use context clues in bibliographic records—such as institutional affiliations, co-author names, text from the abstract, addresses, and dates—to identify which publications belong to a particular author. Often researchers will compare two different models of machine learning to discover which is better suited for a certain purpose. One may outperform the other in a search for citations using co-author data, while the other handles keywords in the title of a paper more effectively. Algorithmic methods of disambiguation have proven imperfect, though, in that they necessitate some manner of human involvement to correct errors.

On the opposite end of the spectrum, manual disambiguation relies on humancreated name records (i.e. authority files) for each author. While this method has

-

³ David Stern, "Author as Object: Disambiguation and Enhanced Links," *Online*, November/December 2010, 30.

⁴ Chen and Hsieh, "Ambiguity Resolution for Author Names," 216.

⁵ Martin Enserink, "Are You Ready to Become a Number?," *Science* 323 (2009): 1663.

⁶ Hui Han, Le Giles, Hongyuan Zha, Cheng Li, and Kostas Tsioutsiouliklis, "Two Supervised Learning Approaches for Name Disambiguation in Author Citations," *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries* (2004): 303.

⁷ Chen and Hsieh, "Ambiguity Resolution for Author Names," 231-232; Rotenberg and Kushmerick, "The Author Challenge," 510.

functioned effectively in small to mid-sized library catalogs, it is wholly unfeasible when considering the vast numbers of authors contained in a single online academic database. 8 The manual creation and maintenance of author name records is an immensely labor-intensive and time-consuming exercise that is functionally impossible to implement today.

In recent years, a third approach to disambiguating author names has come to the fore. It requires that authors be assigned a unique identifier, which would, like Digital Object Identifiers (DOIs) for published articles and datasets, be persistent and only ever represent one individual by unifying all of his or her name labels. 9 My identifier would include a full record of all the names I had ever published under, including "Walton, H.," "Walton, H.M.," and any other variations. Searches on this identifier would produce my bibliography alone, and not the work of other Waltons in the database.

A scholar's professional identity is constructed in major part through his or her publications, and "proper attribution and association of one's scholarly output is imperative to professional branding and reputation management." It can be a challenge to collate a comprehensive bibliography for tenure reviews and grant applications if publications are scattered across a number of name labels. Author identifiers not only make it easier for authors to represent themselves, but allow academic institutions to keep better track of their faculty publications, provide "reliable linking to other articles by the same author, promoting the discovery of related scholarly works," and allow publishers

⁸ Denise Beaubien Bennett and Priscilla Williams, "Name Authority Challenges for Indexing and Abstracting Databases," Evidence Based Library and Information Practice 1, no. 1 (2006): 49.

⁹ Martin Fenner, "ORCID: Unique Identifiers for Authors and Contributors," *Information Standards Quarterly* 23, no. 3 (2011): 13. Rotenberg and Kushmerick, "The Author Challenge," 504.

and granting organizations to keep clearer records of the scholars for which they are publishing or providing funding.¹¹

From the early 2000s to today, various bibliographic database providers have been creating unique identifying numbers for scholars—ArXiv, Elsevier (Scopus), CrossRef, and Thompson Reuters among them—each providing their own identification system. Yet, because these identification values are independent of one another, it has been argued that a more centralized, international author identification system should be implemented. This would ensure that that identification services would not be limited to single, proprietary databases or segregated by discipline or by geographic region. 14

In 2012, the International Standards Organization (ISO) created the International Standard Name Identifier (ISNI), which is intended to disambiguate not only the names of authors, but also performers, creators, producers, publishers, researchers, and more. The ISO maintains that its ISNI will serve as a "bridge identifier," one that can be used by various organizations across an industry or in interdisciplinary searches. While this highly ambitious project could eventually evolve into something indispensable, it currently requires that users register for an ISNI through a registration agency, of which there are only three currently listed on their website. While this does lower the chances of accidentally registering twice, it is a complicated process that could not, because of

_

¹¹ Ibid., 11.

¹² Nancy K. Herther, "Who's On First?: Name Disambiguation Theory," *Searcher*, no. 8 (October 2010): 29-30.

¹³ Rotenberg and Kushmerick, "The Author Challenge," 519.

¹⁴ Ibid., 508; Martin Fenner, Consol Garcia Gómez, and Gudmundur A. Thorisson, "Collective Action for the Open Researcher and Contributor ID (ORCID)," *Serials* 24, no. 3 (2011): 277.

¹⁵ Andrew MacEwan, Anila Angjeli, and Janifer Gatenby, "The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control," *Cataloging and Classification Quarterly* 51, no. 1-3 (2012): 55-56.

¹⁶ "Do you have an ISNI?," *ISNI International Agency*, accessed June 10, 2014, http://www.isni.org/do-you-have-an-isni.

time constraints and inconvenience to authors, be feasibly integrated into a manuscript submission or grant application workflow. In addition, it is not specifically targeted at scholarly communication applications and could be more than a researcher requires during his or her career.

Therefore, in 2009, Thompson Reuters, in partnership with the Nature Publishing Group, held a Name Identifier Summit, one of the products of which was a non-profit organization that would serve as an "open and global registry for unique...identifiers" for authors of scholarly work exclusively. ¹⁷ It is intended to be used by all academic disciplines, span national and institutional boundaries, and also interact with existing scholarly author identification systems. ¹⁸ They called the project the Open Researcher and Contributor ID (ORCID).

Based on code licensed from Thompson Reuters' ResearcherID system, ORCID was launched in 2012 as a non-proprietary service and makes its code available under an open source license. ¹⁹ Beyond its comprehensive and cooperative approach to working with other scholarly profile systems (almost anything from Scopus to CrossRef), ORCID has also taken greater steps than its predecessors to establish itself within major scholarly institutions. The staff has acknowledged that in order for the project to succeed, ORCID requires the collective action of a critical number of enabling services and users. ²⁰ To support this process, they placed an emphasis not only on offering the service to individual authors, but turning to publishers, granting organizations, universities, and

¹⁷ Rotenberg and Kushmerick, 509; Fenner, "ORCID," 11.

¹⁸ Fenner, "ORCID," 11.

¹⁹ Fenner, et al., "Collective Action," 277; "What is ORCID?," *ORCID*, accessed 10 March 2014, http://orcid.org/content/initiative.

²⁰ Fenner, et al., "Collective Action," 277-278.

independent research institutions, encouraging them to adopt ORCID identifiers, and integrate the registration process into their workflows.

The ORCID team offers a grant, funded by the Alfred P. Sloan Foundation, to universities and professional organizations to defer the costs of integrating the utility into their organization's bibliographic management infrastructure. Boston University, Cornell, Notre Dame, and Purdue are among the institutions that have received the grant. ²¹ In this way, ORCID is gaining a stronger foothold in the unique identifier market than its predecessors and, arguably, its competitor ISNI.

Still, no matter how beneficial an ORCID profile may be for an author, it is often difficult to get researchers to participate in any form of identity management process.²² It takes time away from their research and can be very confusing, as is evident from the many existing identification tools mentioned above. Therefore, it becomes almost more critical to have larger organizations take the lead, requiring scholars to register for an ORCID as part of the process of applying for a grant, creating a CV for tenure review, or during the submission of a manuscript to a journal.

ORCID employs both automatic and manual disambiguation to assist members in building their bibliographies. Authors must sign up for their own identifier and provide some information in their name authority record, such as name variants, email, and current and past institutions where they were employed. These metadata are then applied to algorithmic searches of a number of databases. If an author does not provide reliable, correct metadata about him- or herself for the system, the algorithms will not perform

-

²¹ "Adoption and Integration Program," *ORCID*, accessed 11 March 2014, http://orcid.org/content/adoption-and-integration-program.

²² Rotenberg and Kushmerick, "The Author Problem," 504.

effectively.²³ Even if authors are required to sign up for an ORCID during the application for a grant, if they do not return and manage their profile beyond that single visit, the success of the overall project could be jeopardized. But, because ORCID is designed to integrate with other author identification or profile systems, provide a secure way to manage data through a simple and convenient interface, and increase the findability of their work by peers and other consumers, they will likely be enticed to return to the site and curate an authoritative profile.

The purpose of this study is to examine how a sample of publishers and granting organizations have integrated ORCID into their application and submission workflows in order to discover what benefits these organizations gain from using ORCID identifiers and how effective they are at introducing scholars to ORCID. The result was an analysis of best practices for ORCID integration and identification of where publishers and granting organizations can improve their integrations. The conclusions drawn are generalizable to other institutions seeking to adopt ORCID themselves.

_

²³ Stern, "Author as Object," 31.

Related Research

Information and computer scientists have been conducting research on the problem of name disambiguation for decades. It was, in large part, due to the shortcomings of many of the solutions proposed in that research that ORCID was created. As the introduction suggests, author name disambiguation research falls into three categories: manual, automatic, and a combination of both. The majority of older disambiguation articles examining any of these methods are not current enough to address ORCID specifically, and even the most recent generally only mention it in passing. ²⁴ There are a number of articles centered around ORCID itself, some of which are discussed below. Still, there was no identifiable research into the direct implementation of ORCID among publishers and funding organizations. This study aims to fill that gap. Though this study takes a novel approach, the core concepts of author name disambiguation are built upon the work of previous researchers.

There has been very little recent exploration of manual disambiguation methods because there appears to be consensus that such an approach is unfeasible when addressing today's huge volumes of information. As Bennett and Williams argue: "To meet the needs of the 21st Century...[indexing and abstracting] databases may need to implement options that present a high degree of probability that items have been authored by the same individual, rather than options that provide high precision with the expense

²⁴ Strotmann and Zhao, "Author Name Disambiguation," 1820, 1823; Rotenberg and Kushmerick, "The Author Challenge," 509-510.

of manual maintenance."²⁵ On the other hand, though, Veve claims that current technologies for automatically disambiguating author names and creating authority files based on those names are ineffective and will always require some form of human intervention because "they do not address the issue of how to extract or harvest…names directly from [XML] records and transform them into useful access points."²⁶ Access points are crucial for ensuring the discoverability of research articles, and author names are not usable points of access until they are disambiguated.

While there is a dearth of manual disambiguation research, there are numerous studies on automatic disambiguation, for the purposes of both precise database recall and for bibliometrics. Within the realm of automatic disambiguation there are two major machine learning methods on which processes rely: supervised and unsupervised. Supervised models require a "training set" of labeled data to parse and learn from before they can begin disambiguating general data. Unsupervised methods, on the other hand, base their disambiguation on the data they can observe. There is no training set to provide positive feedback on what makes a good solution (i.e. unique name), so unsupervised algorithms must determine this through clustering of data and other mechanics.²⁷

Supervised learning models generally perform better than unsupervised models when using the most common criteria for disambiguation: co-author names, article title, and journal title. But, as Chen and Hsieh found in their 2011 study, using publication year and number of pages significantly improved the performance of the *k*-means

_

²⁵ Bennett and Williams, "Name Authority Challenges," 49.

²⁶ Marielle Veve, "Supporting Name Authority Control in XML Metadata: A Practical Approach at the University of Tennessee," *Library Resources & Technical Services* 53, no. 1 (2010): 41.

²⁷ Chen and Hsieh, "Ambiguity Resolution for Author Names," 218; Neil R. Smalheiser and Vetle I. Torvik, "Author Name Disambiguation," page 10, retrieved June 25, 2014, http://arrowsmith.psych.uic.edu/arrowsmith_uic/tutorial/ARIST_preprint.pdf.

unsupervised method, making it comparable to the performance of the Naïve Bayes and support vector machines—both supervised methods.²⁸

The majority of automatic disambiguation studies are conducted using supervised models. Han and his colleagues, for example, conducted a seminal study of large-scale author name disambiguation that resulted in 90% accuracy in disambiguating names such as "J Anderson" and "J Smith." They concluded that co-author names "appear to be the most robust attribute for name disambiguation."29 Other studies have reinforced this conclusion by conducting their own experiments on various datasets. Kang et al. found that over 85% of disambiguations in their data were due to a reliance on co-authorship.³⁰ Strotmann and Zhao, too, found that the co-author associations were the strongest attributes for disambiguation, but they caution that with the "increasing globalization of research," even co-author associations are becoming ambiguous. 31 "Given the surge in Chinese and Korean research programs over the last decade," they write, "and the unique cultural and historical family name distributions found in those countries," a co-author with the same surname and first initial could easily be two different people.³²

The third option for disambiguation is, of course, a combination of manual authority records and algorithmic solutions. ORCID, along with a number of other unique

²⁸ Chen and Hsieh, "Ambiguity Resolution for Author Names," 232; In computer science and machine

learning, k-means is a method for vector quantization used for cluster analysis in data mining. Generally, heuristic algorithms are employed for solving complex problems. Naïve Bayes models are part of a family of probabilistic classifiers based on the application of Bayes' theorem for text categorization. Support vector machines are algorithms that, when using a set of training examples, can build a model for assigning data to the categories it "learned" by analyzing the training set data. "k-means clustering," Wikipedia, accessed July 5, 2014, http://en.wikipedia.org/wiki/K-means_clustering; "Naïve Bayes classifier," Wikipedia, accessed July 5, 2014, http://en.wikipedia.org/wiki/Naive Bayes classifier; "Support vector machine," Wikipedia, accessed July 5, 2014, http://en.wikipedia.org/wiki/Support_vector_machine. ²⁹ Han et al., "Two Supervised Learning Approaches," 303.

³⁰ In-Su Kang, Seung-Hoon Na, Seungwoo Lee, Hanmin Jung, Pyung Kim, Won-Kyung Sung, and Jong-Hyeok Lee, "On Co-Authorship for Author Disambiguation," Information Processing and Management 45

³¹ Strotmann and Zhao, "Author Name Disambiguation," 1822.

³² Ibid, 1832.

identifier systems that have been developed—from the arXiv Author ID to the ISNI—are built to require both human interaction with the system (creating an authority file for oneself) and allow for some automatic matching functionality. The majority of literature concerning ORCID is positive, with authors hoping that it will finally unify all the disparate identifiers and offer a real solution to the name ambiguity issue. As Butler wrote in a 2012 essay in *Nature*, "[ORCID] could revolutionize research management, vastly increase the precision and breadth of scientific metrics and help in developing new analyses of, for example, social networks." Thompson Reuters authors Rotenberg and Kushmeric also stated that ORCID could have the ability to "[establish author identification] standards that will benefit all stakeholders in the research community." Fenner and his colleagues also support, in a number of different articles about ORCID and its development, many of the aforementioned services that ORCID will provide. The interaction is an authority in the research of the interaction is an authority in the interaction in the research community.

As evidenced by the literature reviewed here, no study exists that addresses the unique implementations of ORCID among a representative sample of publishers and funders. There are no formal assessments of the benefits these organizations are reaping from integrating ORCID into their workflows or if the integrations are succeeding at increasing the number of ORCID users. This study serves to address this issue and provide recommendations for how publishers and funders beyond those featured here can integrate ORCID effectively and permanently into their own systems of record.

_

³³ Declan Butler, "Scientists: Your Number is Up," *Nature* 485 (May 2012): 564.

³⁴ Rotenberg and Kushmeric, "The Author Challenge," 519.

³⁵ Brian Wilson and Martin Fenner, "Open Researcher & Contributor ID (ORCID): Solving the Name Ambiguity Problem," *Educause Review* 47, no. 3 (May/June 2012): 54-55; Fenner, "ORCID," 11-13; Fenner et al., "Collective Action," 277-279.

Methods

The data for this study was collected through remote interviews via telephone or online tools of representatives from a sample of publishers and funding organizations. Initial contact with these representatives was made through an email explaining the nature of the research and asking if they would be willing to participate. For those who responded, 20-60 minute interviews were conducted over a three month period. The participants were asked not to reveal any sensitive or proprietary information during the interviews, in which the audio of the conversation was recorded.

After the data were collected, interviews were reviewed and partially transcribed. Similarities between each of the implementations were aggregated and compared to identify unique practices. From these data, conclusions were drawn about best practices for integrating ORCID into a publishing or grant application workflow, and some recommendations for how to improve these processes are proposed.

Funding Organization Integration

Each year, funding agencies award billions of dollars to researchers and their labs in the form of competitive grants. Consideration for a grant requires an extensive application process that includes a detailed proposal, budget, list of supplies and subjects, and *curricula vitae* of the investigators. For every grant there are almost always numerous applicants, some of whom may be new applicants and others past recipients seeking a new grant from the same organization. Ambiguous and duplicate records for single individuals can make finding a particular author's oeuvre and tracking research progress difficult.

According to Walter Shaffer, Senior Scientific Advisor for Extramural Research at the National Institutes of Health, ambiguous names can have significant financial implications for granting organizations. New researchers could, for example, receive certain benefits that prior grant recipients do not. One could theoretically tweak his or her name by adding or removing a middle initial and a new record would be created, enabling the receipt of new investigator perks. ³⁶ The disambiguation of researchers' names eliminates this and other problems, such as the inability to retrieve name-based search results with both high precision and recall. ORCID is seen by many funding organizations as an effective way to achieve this.

Described below are the integrations of ORCID into the grant management systems in place at four major funding organizations: the National Institutes of Health,

-

³⁶ In conversation with the author, April 1, 2014.

a part of the U.S. Department of Health and Human Services; the Office of Science and Technical Information at the U.S. Department of Energy; the UK-based Wellcome Trust; and Autism Speaks, an organization that funds research into the causes of and treatments for autism. An analysis of their ORCID adoption and use follows, examining common practices between the organizations.

National Institutes of Health

As the United States government's primary biomedical and health-related research agency, the National Institutes of Health (NIH) funds thousands of research projects each year. Eighty percent of the agency's annual budget goes to approximately 50,000 competitive grants for both domestic and foreign labs. ³⁷ In 2013, just over \$20 billion was awarded for research activities. ³⁸ Being one of the oldest federal research agencies, the NIH maintains a database of grant applications and awards that have been processed since its emergence as a funding organization in 1936. Today, the NIH receives approximately 80,000 applications for grants annually, around 10,000 of which receive funding. ³⁹ Disambiguating these millions of records is pivotal to ensuring that public money is being allocated efficiently, and that the NIH's own recordkeeping is accurate. The NIH currently employs a significant number of people whose primary tasks are to de-duplicate database records for single investigators by "collapsing" duplicate

³⁷ Walter Schaffer (Senior Scientific Advisor for Extramural Research, NIH) in conversation with the author, April 1, 2014; "NIH Budget," *National Institutes of Health*, accessed June 17, 2014, http://www.nih.gov/about/budget.htm.

³⁸ "Funding – NIH Awards by Location and Organization," *NIH Research Portfolio Online Reporting Tools (RePORT)*, accessed June 17, 2014, http://report.nih.gov/award/index.cfm. **See Figure 1 in Appendix A** for an image of the query performed on the funding database to retrieve this number. ³⁹ The other (approximately) 40,000 awards are made as continuing funding to labs that have already been given a grant. The majority of NIH grants last for four years, with funds being distributed on an annual basis. Walter Schaffer in conversation with the author, April 1, 2014.

profiles into one unified and disambiguated record. The addition of ORCID is seen as a way to assist in this process.

ORCID is being integrated into a new utility called the Science Experts Network or SciENcy. Built by the NIH's National Center for Biotechnology Information in collaboration with the National Science Foundation (NSF), the U.S. Department of Defense (DOD), the Smithsonian, the U.S. Department of Energy, the U.S. Department of Agriculture (USDA), and the Environmental Protection Agency (EPA), SciENcv helps researchers "to maintain a centralized profile that allows [them] to go in and claim all of their [research] products, including all their papers and data pieces."⁴⁰ Though the project is still in beta, it is intended to store researcher information that can then be formatted to fit the grant applications for the each of the aforementioned agencies. As Walter Schaffer explained, "[researchers will] have one [SciENcv] profile, but they'll be able to produce multiple products. Biosketches...for the NIH and also...for NSF grant applications. We're trying to make this a Fed-wide utility." The association of each profile with ORCID will, when the SciENcv project is complete, make it possible for "all participating federal science agencies" to easily look up a researcher's oeuvre and grant history. 42

Associating an ORCID with SciENcv is a straightforward process. "It's right up front," Mr. Schaffer attested, meaning that the field for entering an ORCID appears early in SciENcv registration process. ⁴³ Researchers can enter an ORCID they already have or

⁴³ In conversation with the author, April 1, 2014.

⁴⁰ "Federal-Wide Researcher Profile Project," *National Institutes of Health*, accessed June 17, 2014, http://rbm.nih.gov/profile_project.htm.

⁴¹ In conversation with the author, April 1, 2014.

⁴² Sally Rockey, "Test Drive SciENcy," *NIH Office for Extramural Research Extramural Nexus*, published November 20, 2013, accessed June 26, 2014, http://nexus.od.nih.gov/all/2013/11/20/test-drive-sciency/.

click a button that will take them to the ORCID website to sign up. Associating an ORCID with a SciENcv profile is voluntary at this point and is unlikely to be made mandatory, but within the beta instance, approximately 20% of users have connected an ORCID to their profiles. He is schaffer projects that this number is not necessarily indicative of future trends: "I expect [ORCID usage] to increase. We'll allow the [scientific] community to make the case for ORCID's use, and the more it's used, the more it will be adopted by the community."

Though Mr. Schaffer is of the opinion that "most people understand what ORCID is and where it's going," it seems that such low usage in the beta version could indicate that more outreach on the part of the NIH could be required. As assing awareness passively though adding an ORCID field to the SciENcv registration and profile management interfaces may not be enough. Fortunately, some official communications from the NIH have been released to external parties who may be looking to apply for a grant, telling them about ORCID, SciENcv, and how they interoperate. ORCID was not launched too long ago, and it is certainly not as well established as it could be in the future. Therefore, it is highly likely that the number of ORCIDs associated with SciENcv will increase as the organization gains traction in the research and academic communities. Still, by implementing ORCID, the highly influential NIH is bringing greater attention to the system. Other funders may be more likely to follow the NIH's lead and integrate ORCID into their own systems.

_

⁴⁴ Walter Schaffer in conversation with the author, April 1, 2014.

⁴⁵ Ibid.

⁴⁶ Ibid.

⁴⁷ Ibid.

U.S. Department of Energy's Office of Science and Technical Information

The Office of Science and Technical Information (OSTI) is the division of the U.S. Department of Energy (DOE) that aggregates, preserves, and distributes the research outputs of the many DOE labs and grant recipients nationwide. While the DOE proper awards the grants, the OSTI does the administration of grant and research information, making it the office that can benefit most directly from ORCID integration. Researchers do not interface directly with the OSTI, however they communicate with the Science and Technical Information Program officers and points of contact within their labs. It is those officers who then submit research outputs to the OSTI for curation and dissemination. This structure allows for the OSTI to simultaneously collect technical documents, conference papers, articles, multimedia, and software, collectively referred to as scientific and technical information from the diverse labs across the national DOE complex. OSTI has a number of ingestion interfaces for the different types of research the DOE funds—all of which are intended to include ORCIDs in the future—but the first to be modified to include an ORCID field was that of the grantees from universities.

According to Jannean Elliottt, Technical Information Specialist at OSTI, even before the organization was a member of ORCID, they "modified [the university grantee interface] so that any grantee who is inputting his technical report info or metadata about a journal article he's written can add in his ORCID number." The ingestion interfaces, too, no longer use text strings for inputting ORCIDs, but instead validate them against the

⁴⁸ "About OSTI," *U.S. Department of Energy Office of Science and Technical Information*, accessed June 26, 2014, http://www.osti.gov/home/about.html.

⁴⁹ Jannean Elliottt (Technical Information Specialist at the U.S. Department of Energy Office of Scientific and Technical Information) in conversation with the author, April 9, 2014. ⁵⁰ "About OSTI."

⁵¹ In conversation with the author, April 9, 2014.

ORCID registry using the OAuth authorization system. OAuth is designed to allow ORCID users to share the publications, datasets, and any other information stored privately on ORCID's servers with third party entities such as the OSTI. The value of validation is immeasurable. To have a plain text field where authors can add in a string of values invites errors that could potentially lead to their work being associated with the wrong ORCID, or if they miss a character, not associated with a well-formed ORCID at all. OAuth eliminates this issue.

Though many of the OSTI-side ingestion interfaces have been updated, each lab has their own information administration system that must be individually updated to include an ORCID field. According to Ms. Elliott, "everybody has different software and platforms. They all work for different contractors and have different budgets. We really have to...wait on them to get changes in their workflow and new metadata from their authors built in to their systems, and that can sometimes take a while." Because of this long transition process, the OSTI has only 30 registered ORCIDs among the thousands of researchers funded by the DOE.

This structure is certainly a hindrance for ORCID adoption. It takes much longer to apply an organization-wide policy about ORCID when the labs have so much autonomy. But, as Ms. Elliott described:

Those folks out at the labs [are] working hard to advertise and promote ORCID. I've heard of one or two of them who have had ORCID days, where they've had special events in their technical libraries and encouraged authors to come over to sign up...for the third year in a row we gave presentations on ORCID [at the annual Science and Technical Information Program meeting]: what is happening, what systems have been changed, who's doing what.⁵⁴

-

⁵² "Tokens Through 3-legged OAuth Authorization," *ORCID*, accessed July 5, 2014, http://support.orcid.org/knowledgebase/articles/119676-tokens-through-3-legged-oauth-authorization.

⁵³ In conversation with the author, April 9, 2014.

⁵⁴ Ibid.

Clearly, the OSTI is making a concerted effort to reach out to the Science and Technical Information Program officers, who can speak directly to their research personnel about ORCID.

This year the OSTI received its first public data file from ORCID, the content of which was the information that registered OSTI researchers have included in their ORCID profiles. There was little usable information in the file because the majority of author profiles only contained their names and email addresses, the most basic information for registering for an ORCID. This does not bode well for high ORCID usage, but Ms. Elliottt was optimistic about the future: "Right now we have a lot of labs who are very excited and are...putting...author profile modules in their databases to collect ORCID numbers." ORCID registration is optional for researchers receiving grants through the DOE, but as the SciENcv tool begins to be used across the federal science departments, it is possible that the OSTI could see a significant increase in ORCID usage by its researchers.

Wellcome Trust

The Wellcome Trust was established in 1936 at the behest of Sir Henry Wellcome, whose estate comprises the original endowment. Today the Trust is the United Kingdom's "leading science and research charity," ⁵⁶ awarding approximately £750 million in funding for "the brightest minds in biomedical research and the medical

⁵⁵ In conversation with the author, April 9, 2014.

⁵⁶ Sharmila Devi, "Innovation and Excellence is Developing Across Europe," *Financial Times*, published May 19, 2014, accessed June 26, 2014, http://www.ft.com/intl/cms/s/2/77af146e-c641-11e3-ba0e-00144feabdc0.html#axzz35lrmvRFX.

humanities, with the aim of improving human and animal health."⁵⁷ The Wellcome Trust followed the ORCID project throughout its development and into launch, after which it began integrating ORCIDs into its own grants management and recordkeeping workflows. While the Trust keeps its own unique identifiers for principal investigators, ORCID allows for stronger, clearer linkage between the research and the individuals who produced it by creating a disambiguated, comprehensive record of an author's oeuvre. Wellcome's ORCID integration contributes to streamlining the granting process, and also "help[s]...track researcher's progress and whether [their] grants are producing high quality science."⁵⁸ The latter is done primarily by the evaluation team which, as Jonathon Kram, a team member, explained, "keep[s] track of outputs from [Wellcome's] funding" to ensure the money is being spent wisely.⁵⁹

Wellcome has integrated ORCID into their eGrants management system, through which all of its new grant applications are processed. Registration for an ORCID or verification of a current one is the third step in the eGrants account registration process, making it highly visible to researchers. ⁶⁰ Included is a brief summary of what ORCID is—this links out to the ORCID website for more information—and why the Wellcome Trust endorses it. The clickable buttons to "Verify existing ORCID id" and "Register for ORCID id" are more prominent than the "Skip this for now" link, ⁶¹ signifying that ORCID registration at the Wellcome Trust is voluntary, but highly recommended.

-

61 Ibid. See Figure 2 in Appendix A.

⁵⁷ Jonathon Kram (Research Assistant, Evaluation Team at the Wellcome Trust) in conversation with the author, April 2, 2014; "Funding," *Wellcome Trust*, accessed June 26, 2014, http://www.wellcome.ac.uk/Funding/index.htm.

⁵⁸ Benjamin Thompson, "Distinguishing Researchers with an ORCID," *Wellcome Trust Blog*, published February 15, 2013, accessed June 26, 2014, http://blog.wellcome.ac.uk/2013/02/15/distinguishing-researchers-with-an-orcid/.

⁵⁹ Jonathon Kram in conversation with the author, April 2, 2014.

⁶⁰ "eGrants User Guide – General Information," *Wellcome Trust*, updated May 20, 2014, accessed June 26, 2014, https://grants.wellcome.ac.uk/egrants/General/Help/eGrants%20User%20Guide.pdf.

Wellcome has elected to keep ORCID participation voluntary at this point in time because it is still a relatively new tool and Wellcome, according to Kram, "did not want to make registration to a relatively new entity mandatory until penetration and integration of ORCID [was] more universal." But, the organization does anticipate that with the widespread adoption of ORCID, use of the identifier with become mandatory. As of April 2014, there were 1,556 eGrants users, 1,200 of which included verified ORCIDs in their profiles. 63

The evaluation team at the Wellcome Trust has done at least one survey of its researchers about their reception of ORCID. Most of the feedback was positive, said Kram, though a few participants gave it low marks because they interpreted it to be yet another online author profile that they would have to maintain. ⁶⁴ With the large volume of author identification systems, from Scopus' Author Identifier to Google Scholar, it can be onerous and time-consuming for researchers to complete multiple profiles. This suggests that a few of the survey respondents do not completely grasp the concept of ORCID's ability to interoperate with these systems to reduce the burden on researchers. It could take time and significant effort by both investigators and ORCID member organizations like the Wellcome Trust to convey that message.

Autism Speaks

Founded in 2005, Autism Speaks is the "world's leading autism science and advocacy organization." In 2012, the organization awarded \$24.2 million in grant

⁶² Jonathan Kram, email message to the author, June 30, 2014.

⁶³ Jonathon Kram in conversation with the author, April 2, 2014; Jonathon Kram, email message to the author, June 30, 2014

⁶⁴ Jonathon Kram in conversation with the author, April 2, 2014.

^{65 &}quot;About Us," *Autism Speaks*, accessed June 26, 2014, http://www.autismspeaks.org/about-us.

funding for research into the causes and prevention of autism, treatments for both children and adults affected by the disease, and identifying a cure. ⁶⁶ Unlike the aforementioned funding organizations, the majority of the money Autism Speaks awards comes from donations, especially fundraising walks and private donors. ⁶⁷ This makes them more directly accountable to those donors. While the U.S. government also uses public money to fund its research, there is less active accountability to American taxpayers than there is to a group of dedicated individuals who give their money specifically for research that might directly affect them or someone they know.

Autism Speaks is breaking new ground among funding organization with the integration of ORCID into their Science Grants System. Unlike other funders, though, they are requiring all new grant applicants, most of whom are academic researchers, to have an ORCID. As their policy document reads:

For new applications, Autism Speaks requires all principal investigators, coinvestigators, and mentors to register with ORCID to obtain a unique online identifier and to allow Autism Speaks limited access to their ORCID account... Applications without ORCID accounts for the key personnel listed above will not be reviewed. 68

Those grantees who received their funding before the policy took effect the week of April 7, 2014 are grandfathered into the old policy, but if they receive an official communication from the organization, such as a reminder letter about an upcoming progress report, a line or two is included about the new partnership with ORCID and how highly Autism Speaks recommends creating an ORCID profile. The new policy, though,

^{66 &}quot;Autism Speaks 2012 Annual Report," accessed June 26, 2014,

http://www.autismspeaks.org/sites/default/files/documents/autism_speaks_2012_annual_report.pdf, 22.

⁶⁷ Ed Clayton (Senior Director of Strategic Funding and Grants Administration at Autism Speaks) in conversation with the author, April 14, 2014.

⁶⁸ "Policy on ORCID Integration with Autism Speaks Science Grants System," *Autism Speaks*, email message from Ed Clayton to the author, April 14, 2014.

will not delay the application submission process. If only the principal investigator has registered for an ORCID, they are permitted to submit the application, but when they do, they are alerted that an email will be sent to all of their co-investigators and mentors reminding them that if they do not sign up for their own ORCIDs, their application will not go to review. ⁶⁹ The administration at Autism Speaks are hoping that other funding organizations will follow their lead and start requiring ORCIDs because the project will not be successful without a critical number of users and enabling services. ⁷⁰

Analysis

As evidenced by these four funding organizations, the common driving force behind their adoption of ORCID was their own need to better track research outputs to ensure they are getting the most for their investment. That includes ensuring the discoverability of each investigator's oeuvre by correctly disambiguating names and associating an individual's publications, grant applications, and progress reports with a unique identifier. Having all of their information in a centralized place could make it quicker and easier to transfer information between funders and track an investigator's work from the start of a grant to the final products (data sets, articles, books, etc.).

There is an informal consensus among the funders that ORCID will be most valuable to the research community when it is mandatory for applying for a grant or submitting a manuscript. But, they acknowledge that in these first few years it may not be an achievable goal for large, international funding organizations because ORCID is still within its first two years of service and has not yet been used widely enough to make it

⁶⁹ Ed Clayton in conversation with the author, April 14, 2014. **See Figure 3 in Appendix A.**

⁷⁰ Fenner, Gomez, and Thorisson, "Collective Action," 277.

mandatory. They perceive it to potentially slow researchers down in their already lengthy grant application processes.

Maintaining a voluntary sign up process could foster goodwill among researchers who desire less regulation, but it also hinders progress toward widespread use of ORCID. Many other organizations could see mandatory ORCID use at organizations like the NIH, DOE, and Wellcome Trust as powerful endorsement of the system. Because they have not chosen to do so at this time, though, it opens up the door for smaller funders like Autism Speaks to mandate ORCID use, starting a trend that could encourage larger funders to emulate them. It may begin as a small movement, but it could possibly trigger the communities that smaller organizations fund to spread the word among themselves. If whole scholarly communities start using ORCID *en masse*, it could extend to other, perhaps larger communities as word travels between colleagues.

In terms of the actual integration into funding workflows, it appears that common practice is to put the ORCID section close to the beginning of the registration or submission process. This appears to be a good strategy; at the beginning of the application, the researchers are still sharp and not exhausted from an hour or two of filling in fields. It is more likely that they would be willing take some time and read about ORCID at that point rather than skipping the sign up or just inputting the bare minimum of information to create one. Including a short description of ORCID in the grant application interface and offering a link to the ORCID website appears to be the most common way of educating researchers—while they are working in systems like eGrants and SciENcv—about the identifier and how it can benefit them.

Encouraging new researchers to sign up for ORCID when they apply for a grant has so far proven to be a workable approach, but it leaves out the researchers who have existing profiles. Major communications such as the progress report reminder emails that Autism Speaks sends and the OSTI's library presentations and ORCID days appear to be outstanding ways to reach out to researchers whose grant-funded work is already in progress. While these messages and events might be ignored by those researchers who are not mandated to sign up, they still raise awareness about ORCID and their granting organization's association with it.

Publishing Industry Integration

On the opposite side of the research lifecycle from granting organizations are the publishers; the institutions that take research articles, review them, and make them available in journals, books, and online. Not unlike the process of applying for a grant, the submission of a manuscript to a journal involves multiple steps throughout which maintaining an individual author's identity is vital. Submission is the first stage, in which the research team presents its article to the journal. From there, the manuscript is sent to editors and peer reviewers to be examined and reviewed for correctness. It is more likely at this stage that the manuscript will be returned to authors for revision than it is that it will be immediately accepted. The revision and resubmission process could be repeated several times before the journal is prepared to publish the article.

Just as funding organizations receive thousands of applications each year, publishers must contend with just as many, if not more articles and ambiguous author names. The integration of ORCID into their manuscript processing workflows could significantly decrease the number of duplicate profiles authors maintain and quickly identify the "S Baker" who wrote the papers on physics, not on philosophy. But even beyond that, by adding peer review to the mix, it becomes necessary for publishers to keep track of the people who are currently reviewing one or more articles. This ensures that the reviewers get due credit for their work, that they are not assigned too many articles, and that the reviews they are submitting are of good quality.

Described below are the ORCID integrations at two publishing houses: Elsevier and Oxford University Press. In addition, there is an examination of how eJournalPress, a manuscript processing software vendor, has integrated ORCID into its systems, which must be customizable enough to fit the needs of each one of their many customers.

Elsevier

One of the most recognizable names in scientific and medical publishing, Elsevier produces nearly 2,200 journals, numerous books—including Gray's Anatomy—and offers information solutions used across the globe. 71 Scopus, an indexed bibliographic citation database, and ScienceDirect, a searchable database of full-text journal articles and book sections, are two of Elsevier's best known web-based products, though it provides many more. Elsevier was among the companies present at the 2009 Name Identifier Summit, where the original idea for ORCID was proposed, and it has closely involved with the project since. As Michael Habib, the Senior Product Manager for Scopus, explained: "We had [already been] involved in trying to disambiguate authors...and we had tackled it using matching algorithms, [but] we'd...seen that that will only get us so far and that there does need to be some sort of human element to it."⁷² An internal project that would allow authors to provide manual corrections for their profiles had been in progress at Elsevier long before ORCID came on the scene, but, as Mr. Habib related, "we recognized, coming at it both from [the] Scopus information provider perspective and as a publisher, that the real solution was to have complete endto-end use of an identifier, starting with the funder." As a result, ORCID has been

^{71 &}quot;At a Glance," *Elsevier*, accessed June 27, 2014, http://www.elsevier.com/about/at-a-glance.

⁷² Michael Habib (Senior Product Manager, Scopus) in conversation with the author, April 23, 2014.

⁷³ Michael Habib, in conversation with the author, April, 23, 2014.

incorporated into both Scopus and the manuscript submission processes on the publishing side.

The first integration came in the form of the Scopus-to-ORCID feedback wizard. This tool was designed to allow Scopus access to an author's ORCID profile for the purpose of uploading all of the citations contained in its database associated with a particular author to ORCID. 74 Once the author is signed into ORCID, Scopus will perform a search of its own database on his or her name. The interface allows that author to change his or her name information for the search in the event that Scopus retrieves more than one Author ID profile for different name variations or finds other authors' work. In addition, the author can add institutional affiliations and multiple name variants to help refine the search. 75 From the results, the author can select the correct profile(s) and choose which name variation they want to use for their Scopus ID in the future. 76 Using the profile information, Scopus then aggregates a list of all publications associated with the profile(s). Using this list, authors can select all the publications they know to be theirs and deselect any that were incorrectly associated with them. If there are a significant amount of articles missing, authors can perform another search to locate them.⁷⁷ Once authors have selected their correct publications, the wizard produces a summary of their Scopus profile for them to review before linking it with ORCID. The final step in the process is uploading their citations to the ORCID database.⁷⁸

-

⁷⁴ See Figure 4 in Appendix B.

⁷⁵ See Figures 5 through 6 in Appendix B.

⁷⁶ See Figures 7 through 8 in Appendix B.

⁷⁷ See Figure 9 in Appendix B.

⁷⁸ See Figures 10 through 11 in Appendix B.

As of June 2014, there is a total 68,928 ORCIDs associated with Scopus author profiles. The total number of articles and other works that have been submitted by Scopus to ORCID since the wizard launched on October 14, 2012—the day ORCID went live—is now more than 2.1 million. From this, it is clear that employing a tool that can quickly and (relatively) painlessly upload batches of high-quality citations from a database like Scopus to ORCID could make the process much less onerous for authors. With such large numbers of scholars associating their Scopus Author IDs with ORCID, Elsevier has the potential to have a formidable impact on worldwide ORCID adoption, which benefits not only their organization, but scholars, universities, funders, and other publishing organizations.

The second phase of ORCID integration was in the Elsevier Editorial System (EES), which represents the publication rather than information management side of Elsevier. EES allows authors to create profiles for the journals to which they are submitting their work, of which there are approximately 1,800. EES was originally designed so that each author had to create a different profile for each journal. For instance, if an author published an article in *Animal Behaviour* and then another article in *Zoology*, he or she would have to manage both profiles separately. But, as Getty Bruens, Senior Application Manager for EES, related, Elsevier introduced what are called consolidated accounts. ⁸¹ Similar to the "collapsed" profiles at the NIH, with this one umbrella account authors can access all their other accounts in EES. And Elsevier is only

⁷⁹ See Figure 12 in Appendix B.

^{80 &}quot;Statistics," *Scopus to ORCID*, accessed June 27, 2014, http://orcid.scopusfeedback.com/statistics. N.B. This section of the site is password protected. Please contact Scopus if access is desired.

⁸¹ In conversation with the author, April 30, 2014.

allows ORCIDs to be registered with consolidated accounts. ⁸² This prevents authors from trying to sign up for multiple ORCIDs for each profile. Nearly 40% of current EES account are consolidated and approximately 100,000 of those accounts are associated with an ORCID. ⁸³

From these data, the EES staff has found that the highest number of ORCIDs are added to authors' consolidated accounts when they submit a manuscript rather than when they initially sign up for EES. Only around 10% of new registrants are linking their Elsevier profile to ORCID. 84 Co-authors are also encouraged, but not required to connect an ORCID to their profiles. There is an option in the submission interface that allows the corresponding author to enter the names and email addresses of the co-authors. Filling in this information is not mandatory, and is less likely to be entered as the number of co-authors on a paper increases. If the corresponding author provides the co-authors' email address and names, though, those individuals receive an email invitation add their ORCIDs to the submission. 85

Elsevier is also maintaining, according to Getty Bruens, an highly proactive informational and marketing campaign for ORCID use among its authors. ⁸⁶ Elsevier offers workshops on publishing best practices and ethics to early-career researchers, and ORCID is now included in that training. Outreach to editorial boards and to manuscript reviewers—two groups profoundly tied to the publication process—includes newsletters and in-person events in which ORCID is a feature. These types of communication can be

⁸² In conversation with the author, April 30, 2014.

⁸³ See Figure 12 in Appendix B for an example of how ESS users can link their profiles to ORCID.

⁸⁴ Getty Bruens in conversation with the author, April 30, 2014.

⁸⁵ Ibid. See Figure 13 in Appendix B.

⁸⁶ In conversation with the author, April 23, 2014.

disregarded by busy scholars, but they are still an effective way to insert ORCID into the vocabulary of researchers, reviewers, and editors associated with Elsevier.

Fundamentally, Elsevier got involved with ORCID because, according to Michael Habib, having "clean data [about] who an author is, in a central way, rather than having [a] split identity, is...helpful in all sorts of ways." These can include anything from keeping track of a researcher's grant-funded activities to simplifying the profile and CV maintenance for individual investigators.

Oxford University Press

As the largest university press in the world, Oxford University Press is a global publishing organization that produces books, academic journals, and online resources covering many topics across the vast scholarly milieu. ⁸⁸ One of the primary reasons Oxford looked into partnering with ORCID is because it is, according to Simone Larche, Online Submission Systems Manager at Oxford Journals, the way the industry is trending and Oxford aims to stay abreast of new ideas and trends. ⁸⁹ Before working with ORCID, Oxford was involved mainly with Thompson Reuters' Researcher ID, but as Ms. Larche stated, "ORCID was what everybody was waiting for."

Oxford's ORCID integration is fairly recent, beginning in earnest in 2013, but is now in use by three journal submissions systems: ScholarOne Manuscripts, Editorial Manager, and BenchPress. ⁹¹ In each of these systems, the author has the ability to

90 Ibid.

⁸⁷ In conversation with the author, April

⁸⁸ "About Oxford University Press USA," *Oxford University Press*, accessed June 27, 2014, http://global.oup.com/academic/aboutus/?cc=us&lang=en.

⁸⁹ In conversation with the author, April 23, 2014.

⁹¹ "ORCID," *Oxford Journals*, accessed June 27, 2014, http://www.oxfordjournals.org/for_authors/orcid.html.

validate his or her ORCID against the official registry using OAuth, reducing human error in data entry. The ORCID field is optional, though, because, as Simone Larche, explained, "The problem with making [ORCID] compulsory is, if you don't have one you won't be able to submit an article...[and] we don't want to stop people submitting articles." From the perspective of publishing houses this is an understandable concern, but ORCID registration is so quick and simple that it would likely not deter many scholars from wanting to submit their articles to such an established and highly regarded publisher. Unfortunately, Oxford still has a number of journals that do not use any form of submission system. Instead the authors simply email a copy of their article to the editor of a particular journal. For these journals it is extremely difficult to even communicate about ORCID registration during the submission process. Emails and flyers are, of course, a secondary option, but having no set field for ORCID iDs severely impedes wider adoption by authors publishing with those Oxford journals. ⁹³

eJournalPress

While not itself a publisher, eJournalPress is a software company that produces customizable, "web-based technology solutions to support manuscript submission, tracking, and peer review" into which ORCID is being integrated. ⁹⁴ Their software, both the peer review management system EJPress and the Journal Production System, a publication tracking workflow engine, is used by a number of journals and publishers, including the Proceedings of the National Academy of Sciences and Nature Publishing

92 Simone Larche in conversation with the author, April 23, 2014.

⁹³ Richard O'Beirne, "OUP and ORCID," *Oxford Journals*, accessed June 27, 2014, http://www.oxfordjournals.com/for_societies/partner_newsletter_orchid.html; Simone Larche in conversation with the author, April 23, 2014.

^{94 &}quot;About Us," eJournalPress, accessed on June 28, 2014, URL. http://www.ejournalpress.com/about.html.

Group, as well as corporate entities. As a service provider to those and many more, eJournalPress has a unique perspective on the needs of publishers and trends in the industry, including ORCID integration.

Soon after ORCID debuted, eJournalPress was, according to Joel Plotkin, CEO of eJournal Press, adding the functionality to its software. ⁹⁵ Configuration settings allow client publishers to decide whether or not ORCIDs will be mandatory and at what point in the publication process to include validation and sign up. As Plotkin explained ORCID can be "required at submission, at manuscript revision, or upon acceptance, [and it can be made] optional at any stage." ⁹⁶ The layout of the ORCID sign up/validation page is very similar to those at Elsevier and Oxford, as well as most of the funders from the previous section. "We...take [authors] to a specific web screen," Plotkin related, "where there's configurable text to explain what ORCID is, the motivation is [for using it], and why they're being asked [by the journal] to fill in an ORCID." ⁹⁷ eJournalPress also validates ORCIDs against the official registry via OAuth rather than relying on the authors to fill in a text field.

Plotkin spoke more than the publishers above about the role of peer reviewers in the publication process. Keeping track of them and all the articles they are reviewing at any given time is necessary to ensure that reviewers are not overworked and that the reviews they are doing are of sufficiently high quality. Having reviewers with ORCIDs will make it much easier to publishers to manage them and guarantee that they get credit for the work they do for their colleagues. ⁹⁸

05

⁹⁵ In conversation with the author, April 16, 2014.

⁹⁶ In conversation with the author, April 16, 2014.

⁹⁷ Ibid.

⁹⁸ Ibid.

As a company whose software also serves the corporate world, eJournalPress must consider an unusual concern about unique identifiers among some research corporations who like to keep their personnel and findings as confidential as possible. Certain for-profit companies—big agricultural conglomerates or pharmaceuticals, for example—were against participating in ORCID, assigning any form of unique identifiers, or merging duplicate accounts because they were afraid that competitors would be so impressed by an article or new discovery that they would attempt to "poach" that employee. According to Plotkin, "Because author names aren't uniquely tracked," it allows [corporations] to be semi-vague about who's doing the work and protect [their] internal employees from poaching." This practice could be highly counterproductive for ORCID adoption because researchers could be required by their employers *not* to apply for an identifier. This is the opposite of most of the publishers and funders examined above, who want greater adoption and openness and discoverability of research.

Still, eJournalPress has closely followed the trends in the publishing industry and tailored its software to fit them. And it is clear from not only the integration of ORCID; but also from the pre-coded settings that allow journals to make registration optional or mandatory that the staff of eJournalPress appear to expect ORCID to become a strong force in scholarly communications.

Analysis

Collectively, the publishers and vendor examined here are optimistic about ORCID and its potential to make the manuscript submission and review process run more smoothly for both them and the authors. But, all seem to agree that ORCID will not be

⁹⁹ In conversation with the author, April 16, 2014.

used to its full potential without making membership mandatory. According to Simone Larche of OUP:

We'd love to make [ORCID] compulsory. Everybody would [because] then it would be like DOIs...but...it's like saying you have to have a Facebook page. What if you don't want one? [If] you have to have [an ORCID] just because you want to submit an article...it is a definite breach of [author] rights. 100

Joel Plotkin of eJournalPress agreed, adding, "We don't want to put up a big brick wall and say, 'Submit to our journal!' and then...[have authors] say, 'Wow, it's so difficult. I'm not going to go submit to you.'" Though this could possibly be true for some of the smaller journals, it is still unlikely to be an impediment to submissions in the long run. If the author must supply in-depth information about his or her work to submit a manuscript, it does not appear that addition of could ORCID be more of a burden. Certainly it takes a few extra clicks and perhaps ten minutes of reading, it does not seem like it could make an already demanding submission form any more difficult.

From the experience of Elsevier, it appears that placing ORCID registration or authentication at the final submission stage could be more beneficial that offering it at registration for an author account through the publisher. It could presumably be jarring and somewhat out of place to ask for it when the review process is complete, but the evidence from Elsevier points to the contrary. Another study would need to be conducted to completely understand where the most ideal placement of ORCID registration is in a funding or manuscript submission workflow, but final manuscript submission appears to be an effective strategy at present.

¹⁰⁰ In conversation with the author, April 23, 2014.

¹⁰¹ In conversation with the author, April 16, 2014.

Conclusion

Overall, the publishers, funders, and vendor studied here present a strong case to their researchers and authors for the use of ORCID by providing adequate information about what it is, why they are affiliated with it, and how to register. Even if researchers choose not to enter an ORCID, the opportunity to do so is still presented, forcing researchers to pause and consider it, even if only for a moment. While a block of descriptive text and a link are informative, the use the Scopus-to-ORCID tool demonstrates that more interactivity with ORCID during the registration and authorization processes increases the use of the identifier. Elsevier has an advantage because the other organizations do not provide extensive databases like Scopus that can upload accurate citations to ORCID, but using the tool encourages authors to familiarize themselves with their ORCID profiles. This could make them more likely to return and enter more data into the authority record. For ORCID to succeed, scholars must take ownership of their identifiers and Scopus-to-ORCID provides a first step toward that goal.

Outreach beyond the submission and application system integration is also a powerful tool for generating interest in ORCID among the scholarly community. While the U.S. Department of Defense's Office of Science and Technical Information cannot integrate ORCID on a system level because of its organizational structure, the staff, both those in the OSTI and the Science and Technology Information Program (STIP) officers, are making a concerted effort to adopt ORCID into their many labs through workshops

and ORCID days, when the library staff or STIP officers help researchers to sign up for an ORCID iD in person. Elsevier's marketing campaign and the inclusion of information about ORCID in official communications from Autism Speaks could also be effective forms of outreach to their communities. Expanding these efforts could boost registration and raise awareness about ORCID use.

All of the organizations except Autism Speaks have chosen to make ORCID voluntary mainly because they do not yet believe that it has enough of a foothold in the scholarly community to support a mandatory implementation. This is understandable to a point. Getting involved with a project that one does not yet feel confident in the success of could be a mistake. As Kram of the Wellcome Trust noted, "We remain committed to ORCID and, once adoption is more widespread, we anticipate that registration and inclusion of an ORCID iD in our grants system will become mandatory." ¹⁰² There is already an expectation that ORCID will be successful enough to merit mandatory use in the future. Because large publishing houses and international funding organizations make up such a major part of the research lifecycle, their influence could shape how individual scholars adopt ORCID. They have the ability to accelerate the process of adoption by mandating it themselves. It is likely that others will follow the lead of the Wellcome Trust, the National Institutes of Health, the U.S. Department of Defense, Elsevier, and Oxford University Press. Making a decisive move to mandate could allow these organizations to create the critical mass of users and services that ORCID requires to be successful.

Larche of Oxford University Press argued another reason for not making ORCID registration compulsory: it would deter authors from submitting manuscripts. It seems

¹⁰² Jonathan Kram, email message to the author, June 30, 2014.

unlikely that one extra step during the already lengthy submissions process would stop anyone from attempting to get their research published in reputable and established journals such as those Oxford produces. Registering for ORCID takes only a few clicks, and with OAuth for validation, there is no data entry beyond the minimum credentials required to sign up for an ORCID iD. It is improbable that this short process could be construed as a roadblock significant enough to prevent scholars from applying for the grant that could fund their research or submitting a paper to a respected journal in their field. If the discussion was about whether or not to mandate the International Standard Name Identifier (ISNI), then there is the potential for a serious roadblock. Using an agency to apply for an ISNI takes time and requires scholars to stop the application or submissions process to seek out help. ORCID is simple and now integrated into those processes. The reasons for not mandating its use cited by the organizations above are controvertible, and more thought should be put into mandating ORCID as Autism Speaks has done.

Despite discussion with representatives of each organization about the issue of where in an application or submission workflow ORCID registration should be placed, it is difficult to draw a conclusion solely from their comments. Elsevier saw an increase in ORCID association with author profiles during the submission stage rather than at initial registration for an Author ID. This suggests that placement during the submission process is more effective than, for example, asking researchers to link their SciENcv to their ORCID. More research would need to be conducted to determine this, though.

The similarities between the publishers and funders examined in this study suggest a number of best practices for ORCID integration. The first is encouraging

researchers to interact with their ORCID profile during registration beyond simply describing it and adding the link to sign up. This could involve a tool such as Scopus-to-ORCID, but could suggest that a researcher permit Google Scholar or another citation aggregator to access their data and begin to fill in their information automatically. This may take more time than funders and publishers are willing to give at the moment, but it has worked spectacularly for Elsevier and could work for them.

The second is designing an outreach campaign beyond the integration itself. The OSTI has a significant limitation to integrating ORCID, but they have looked beyond that to find other avenues of promoting ORCID use at their institution. The Wellcome Trust has included ORCID in its eGrants User Guide. Integration implies making another entity a part of one's own. Making it clear that ORCID is a part of an organization through outreach could be valuable for anyone.

The third is to not overlook those researchers who already have projects in progress or have had a manuscript accepted by a certain publisher before. Making these individuals aware of ORCID through official communications, as Autism Speaks does, or through their lab contacts, as OSTI does, an organization has the potential to increase the number of ORCID iDs they have from current or past researchers. Ideally, as scholars' colleagues begin to register, word will spread, but organizations can help the process along by introducing them to ORCID as well.

ORCID is intended to benefit all the members of the scholarly community, from researchers and universities to publishers and funders. With over 700,000 ORCID iDs issued since its launch in October 2012, the system is growing beyond the initial 300

organizations that supported it. ¹⁰³ The integration of ORCID at the National Institutes of Health, the U.S. Department of Engery's Office of Science and Technical Information, the Wellcome Trust, Autism Speaks, Elsevier, Oxford University Press, and eJournalPress demonstrate the international support for its use and adoption. As more organizations integrate ORCID into their own systems, it could make it possible for it to become the new standard in author identification and disambiguation.

-

 $^{^{103}}$ "ORCID Community," ORCID, accessed July 5, 2014, http://orcid.org/about/community.

Bibliography

- "About OSTI." U.S. Department of Energy Office of Science and Technical Information.

 Accessed June 26, 2014. http://www.osti.gov/home/about.html.
- "About Oxford University Press USA." Oxford University Press. Accessed June 27, 2014. http://global.oup.com/academic/aboutus/?cc=us&lang=en.
- "About Us." *Autism Speaks*. Accessed June 26, 2014. http://www.autismspeaks.org/about-us.
- "Adoption and Integration Program." *ORCID*. Accessed 11 March 2014. http://orcid.org/content/adoption-and-integration-program.
- "At a Glance," *Elsevier*, accessed June 27, 2014, http://www.elsevier.com/about/at-a-glance.
- "Autism Speaks 2012 Annual Report," accessed June 26, 2014,

 http://www.autismspeaks.org/sites/default/files/documents/autism_speaks_2012_a

 nnual_report.pdf, 22.
- Bennett, Denise Beaubien and Priscilla Williams. "Name Authority Challenges for Indexing and Abstracting Databases." *Evidence Based Library and Information Practice* 1, no. 1 (2006): 37-57.
- Butler, Declan. "Scientists: Your Number is Up." Nature 485 (May 2012): 564.
- Chen, Hua-Kuang and Chi-Nan Hsieh. "Ambiguity Resolution for Author Names of Bibliographic Data." *Journal of Educational Media and Library Sciences* 49, no. 2 (2011): 215-240.
- "Do you have an ISNI?" *ISNI International Agency*. Accessed June 10, 2014. http://www.isni.org/do-you-have-an-isni.

- "eGrants User Guide General Information." *Wellcome Trust*. Updated May 20, 2014.

 Accessed June 26, 2014. https://grants.wellcome.ac.uk/egrants/General/Help/
 eGrants%20User%20Guide.pdf.
- Elliott, Jannean. Interview by Haley Walton. Telephone Interview. April, 9, 2014.
- Enserink, Martin. "Are You Ready to Become a Number?," *Science Magazine* 323, (March 2009): 1662-1664.
- "Federal-Wide Researcher Profile Project." *National Institutes of Health*. Accessed June 17, 2014, http://rbm.nih.gov/profile_project.htm.
- Fenner, Martin. "ORCID: Unique Identifiers for Authors and Contributors." *Information Standards Quarterly* 23, no. 3 (2011): 11-13.
- Fenner, Martin, Consol Garcia Gómez, and Gudmundur A. Thorisson. "Collective Action for the Open Researcher and Contributor ID (ORCID)." *Serials* 24, no. 3 (2011): 277-279.
- "Funding." *Wellcome Trust*. Accessed June 26, 2014. http://www.wellcome.ac.uk/Funding/index.htm.
- "Funding NIH Awards by Location and Organization." *NIH Research Portfolio Online Reporting Tools (RePORT)*. Accessed June 17, 2014.

 http://report.nih.gov/award/index.cfm.
- Habib, Michael. Interview by Haley Walton. Telephone interview. April 23, 2014.
- Han, Hui, Le Giles, Hongyuan Zha, Cheng Li, and Kostas Tsioutsiouliklis. "Two
 Supervised Learning Approaches for Name Disambiguation in Author Citations."
 Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries
 (2004): 296-305.

- Herther, Nancy K. "Who's On First?: Name Disambiguation Theory." *Searcher*, no. 8 (October 2010): 24-33.
- Kang, In-Su, Seung-Hoon Na, Seungwoo Lee, Hanmin Jung, Pyung Kim, Won-Kyung Sung, and Jong-Hyeok Lee. "On Co-Authorship for Author Disambiguation." *Information Processing and Management* 45 (2009): 84-97.
- Kram, Jonathon. Interview by Haley Walton. Google Hangouts interview. April 2, 2014. Larche, Simone. Interview by Haley Walton. Telephone interview. April 23, 2014.
- MacEwan, Andrew, Anila Angjeli, and Janifer Gatenby. "The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control."

 Cataloging and Classification Quarterly 51, no. 1-3 (2012): 55-71.
- "NIH Budget." *National Institutes of Health*. Accessed June 17, 2014. http://www.nih.gov/about/budget.htm.

http://www.oxfordjournals.org/for_authors/orcid.html.

- O'Beirne, Richard. "OUP and ORCID." Oxford Journals. Accessed June 27, 2014.

 http://www.oxfordjournals.com/for_societies/partner_newsletter_orchid.html.

 "ORCID Community," ORCID, accessed July 5, 2014, http://orcid.org/about/community.

 "ORCID." Oxford Journals. Accessed June 27, 2014.
- Rockey, Sally. "Test Drive SciENcv." *NIH Office for Extramural Research Extramural Nexus*. Published November 20, 2013. Accessed June 26, 2014. http://nexus.od.nih.gov/all/2013/11/20/test-drive-sciency/.
- Rotenberg, Ellen and Ann Kushmerick. "The Author Challenge: Identification of Self in the Scholarly Literature." *Cataloging and Classification Quarterly* 49, no. 6: 503-520.

- Schaffer, Walter. Interview by Haley Walton. Telephone interview. April 1, 2014.
- "Send Scopus Author details and publication list to ORCID," *Scopus to ORCID*, accessed June 27, 2014, http://orcid.scopusfeedback.com/.
- Sharmila Devi. "Innovation and Excellence is Developing Across Europe." *Financial Times*. Published May 19, 2014. Accessed June 26, 2014. http://www.ft.com/intl/cms/s/2/77af146e-c641-11e3-ba0e-00144feabdc0.html#axzz35lrmvRFX.
- Smalheiser Neil R. and Vetle I. Torvik. "Author Name Disambiguation." Retrieved June 25, 2014. http://arrowsmith.psych.uic.edu/arrowsmith_uic/tutorial/ARIST_preprint.pdf.
- "Statistics." *Scopus to ORCID*. Accessed June 27, 2014. http://orcid.scopusfeedback.com/statistics.
- Stern, David. "Author as Object: Disambiguation and Enhanced Links." *Online*, November/December 2010.
- Strotmann, Andreas and Dangzhi Zhao. "Author Name Disambiguation: What Difference

 Does It Make in Author-Based Citation Analysis?" *Journal of the American*Society for Information Science and Technology 63, no. 9 (2012): 1820-1833.
- Thompson, Benjamin. "Distinguishing Researchers with an ORCID." *Wellcome Trust Blog*. Published February 15, 2013. Accessed June 26, 2014. http://blog.wellcome.ac.uk/2013/02/15/distinguishing-researchers-with-an-orcid/.
- "Tokens Through 3-legged OAuth Authorization." *ORCID*. Accessed July 5, 2014. http://support.orcid.org/knowledgebase/articles/119676-tokens-through-3-legged-oauth-authorization.

Veve, Marielle. "Supporting Name Authority Control in XML Metadata: A Practical Approach at the University of Tennessee." *Library Resources & Technical Services* 53, no. 1 (2010): 41-52.

"What is ORCID?" ORCID. Accessed 10 March 2014. http://orcid.org/content/initiative.

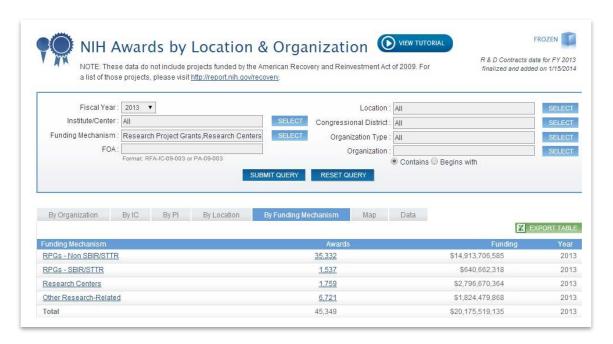
Wilson, Brian and Martin Fenner. "Open Researcher & Contributor ID (ORCID):

Solving the Name Ambiguity Problem." *Educause Review* 47, no. 3 (May/June 2012): 54-55.

Appendix A

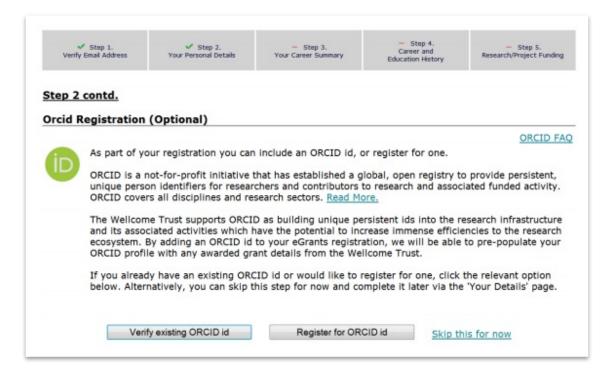
Figures from the Funding Organization Integration section.

Figure 1



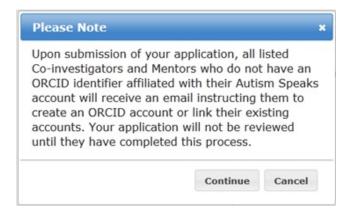
This result was achieved by querying the NIH Research Portfolio Online Reporting Tools (RePORT) awards database using the "By Funding Mechanism" option and using the search terms "2013" and "Research Project Grants, Research Centers."

Figure 2



The Wellcome Trust's ORCID registration interface with the eGrants system.

Figure 3

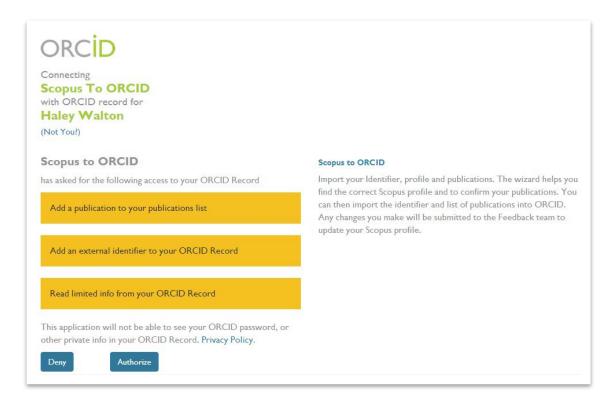


Warning message in Austism Speaks' submission interface. Source: Email message to the author from Ed Clayton, April 14, 2014.

Appendix B

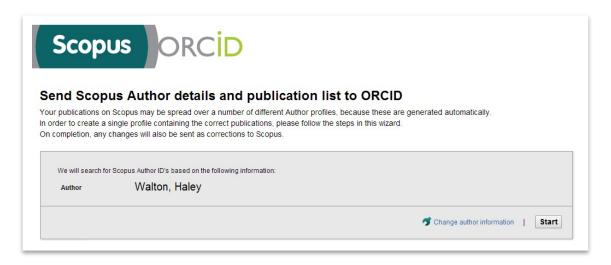
Figures from the Publisher Integration section.

Figure 4



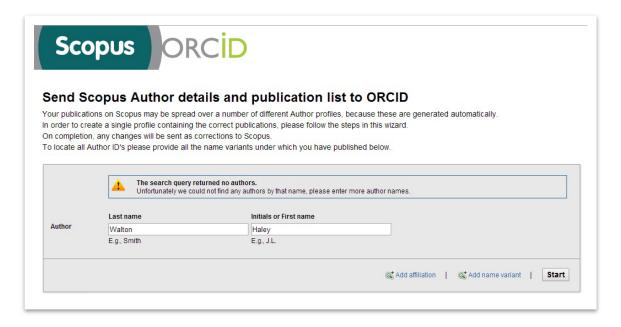
The Scopus to ORCID wizard. After signing into my ORCID account, this page appears, detailing the information in my account that Scopus would like to have access to.

Figure 5



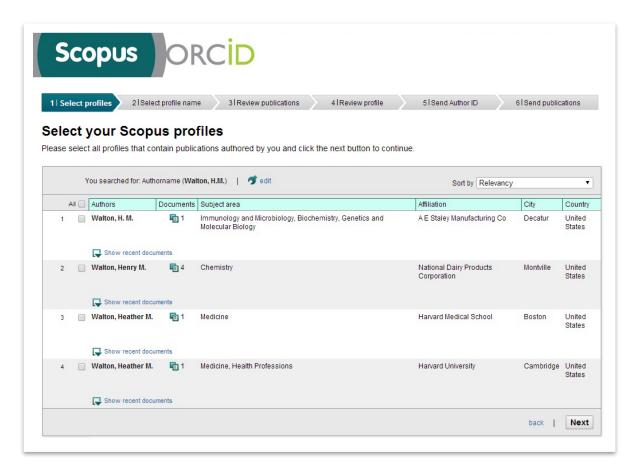
Allowing Scopus the requested access to my profile allows me to advance to the next step of the wizard. Here it states what information Scopus will use to search its database. Since that information is correct, I advanced to the next step by pressing the "Start" button. Source: http://orcid.scopusfeedback.com.

Figure 6



If the query does not return any results or returns too many, the author can add more name variants and institutional affiliations to refine the search. Because I have not published anything that would be in Scopus at this time, the query did not return any results for me. For the sake of this example, I chose to use my initials, H. M., instead of my first name and ran the search again. Source: http://orcid.scopusfeedback.com.

Figure 7



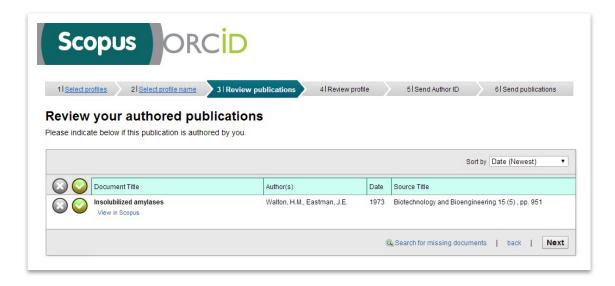
As picture above, the search on H. M. Walton returned four Scopus ID profiles. Of course, none of these are actually mine and should not be associated with my ORCID profile, but for the example I chose the first profile and advanced to the next step by pressing "Next." Source: http://orcid.scopusfeedback.com.

Figure 8



Here the author selects the name variant he or she prefers for his or her Scopus profile. As I have selected a different H. M. Walton's paper for this example, I will select his or her name variant. Source: http://orcid.scopusfeedback.com.

Figure 9



In this step the author reviews the publications Scopus found and marks them as his or her own. If they do not belong to the author, but someone with a similar name, the author can mark the "X" to reject them. The author can perform another search to locate any articles that might be missing from the list. I claimed this document as mine temporarily for the example. Source: http://orcid.scopusfeedback.com.

Figure 10



In this fourth step of the process, the author can review their profile information again to ensure that it is accurate before continuing on to the last two steps. Source: http://orcid.scopusfeedback.com.

Figure 11



The fifth step requires the author to enter his or her professional or institutional email address so that the Scopus ID profile can be linked with ORCID. This is as far as I can go in the process because I do not want to falsely claim this profile as mine. But, after the author submits the email, he or she can then move to the last step of submitting his or her Scopus citations automatically to ORCID. Source:

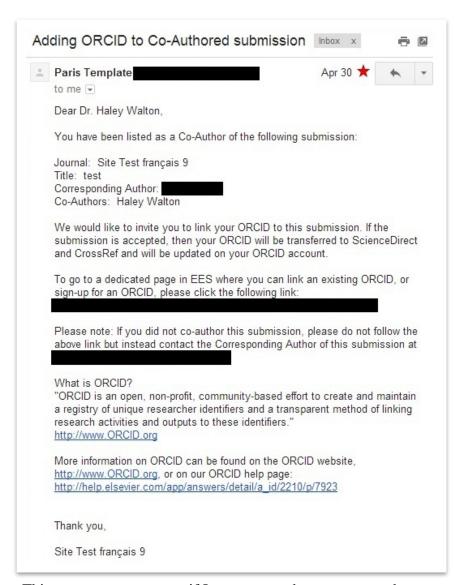
http://orcid.scopusfeedback.com.

Figure 12



N.B. This site is only accessible through Elsevier and was part of a WebEx conversation with Getty Bruens on April 30, 2014. The personal username and URL have been redacted.

Figure 13



This message was sent as if I were a coauthor on a paper that was submitted to EES. It comes from the test version of the site and all personal or private information has been redacted.