James Madison University
JMU Scholarly Commons

4-23-2019

The Visibility of Authority Records, Researcher Identifiers, Academic Social Networking Profiles, and Related Faculty Publications in Search Engine Results

Rebecca B. French
James Madison University, frenchrb@jmu.edu

Jody Condit Fagan
James Madison University, faganjc@jmu.edu

Follow this and additional works at: https://commons.lib.jmu.edu/letfspubs

Part of the Library and Information Science Commons

The visibility of authority records, researcher identifiers, academic social networking profiles, and related faculty publications in search engine results

Rebecca B. French and Jody Condit Fagan

James Madison University

Abstract

Information about faculty and their publications can be found in library databases such as the Library of Congress Name Authority File, VIAF, WorldCat, and institutional repositories; in identifier registries like ORCID and ISNI; and on academic social networking sites like Academia, Google Scholar, and ResearchGate, but the way search engines use such identifiers and profiles is unclear. Therefore, researchers at a large comprehensive university conducted several rounds of web searching before and after the creation and modification of faculty authority records. The sample consisted of 24 faculty and the 35 publications associated with their authorities. The researchers searched for the faculty and their publications on the social networking and identity websites directly, then used Google, Bing, and Google Scholar to record which of the faculty members' profiles and publications were found within the top 50 results. Faculty with more profiles were more visible in search engine results, and faculty with authority records ranked more highly in Google. Results related to publication discovery and ranking were more inconclusive, but revealed clear differences between search tools. The implications of this exploratory research can support educational efforts about academic identities and scholarly profiles, begin a research agenda, and inform methodological development surrounding the influence of identity records and academic social networking profiles on web visibility.

Keywords

Academic identities, Authority records, Discoverability, Identity management, Scholarly profiles, Search engine indexing, Search engine optimization, Social media

Introduction

In the current scholarly communication landscape, faculty are encouraged to promote their scholarly work online. Increasing numbers of academic publishers and funders are requiring authors to use identifiers such as ORCID, and librarians are incorporating the topic of identity management into their outreach to faculty. At the same time, interest in making library resources more widely discoverable via the Semantic Web has led libraries to pursue BIBFRAME as a linked data replacement for MARC. Because a linked data environment requires entities to be represented by unique identifiers, librarians are re-envisioning authority control as identity management and considering how other identity systems might complement or interact with traditional library authority files. This work is being led by the PCC Task Group on Identity Management in NACO, which is also investigating how libraries can involve researchers in the management of their own identities (PCC, 2016).

Despite the emphasis on registering for identifiers and publishing linked data, we have little evidence of the influence of these activities on web visibility. To what extent are the records and identifiers that libraries and faculty create being used by search engines? Does the presence of such records raise the visibility of scholarly works? This exploratory study uses a small sample of scholars at one institution to take a first look at the effect of identity records and academic social networking profiles on web discoverability and to test methodology approaches to inform larger studies.

Literature Review

A variety of platforms are available to researchers for managing their online scholarly identities. Authority records and identifiers maintained by other agencies may also be used to disambiguate and identify scholars. In this paper we group together academic social networking

3

(ASN) profiles (Academia, bepress SelectedWorks, Google Scholar, and ResearchGate) and

ORCID records[1] because they are created and maintained by individual scholars. Additionally,

ORCID's researcher pages more closely resemble ASNs than they do authority records. We

grouped authority records (Library of Congress Name Authority File [LCNAF] and Virtual

International Authority File [VIAF]) with International Standard Name Identifier (ISNI) and

WorldCat Identities, because these records are typically not under a researcher's control.

Although WorldCat Identities pages are not authority records or identifiers per se, their

aggregation of bibliographic and authority data for web presentation makes them of interest to

this study. We refer to these two groups as "authority records" and "ASN profiles and ORCID

records" (see Figure 1). Descriptions of each platform and the information recorded in their

records or profiles can be found in the Appendix.

---

[1] ORCID guidelines state that it is not an academic social network, and ask that the phrase "ORCID record"
be used instead of "profile" to refer to an ORCID account containing information about an individual (ORCID,
2018a).
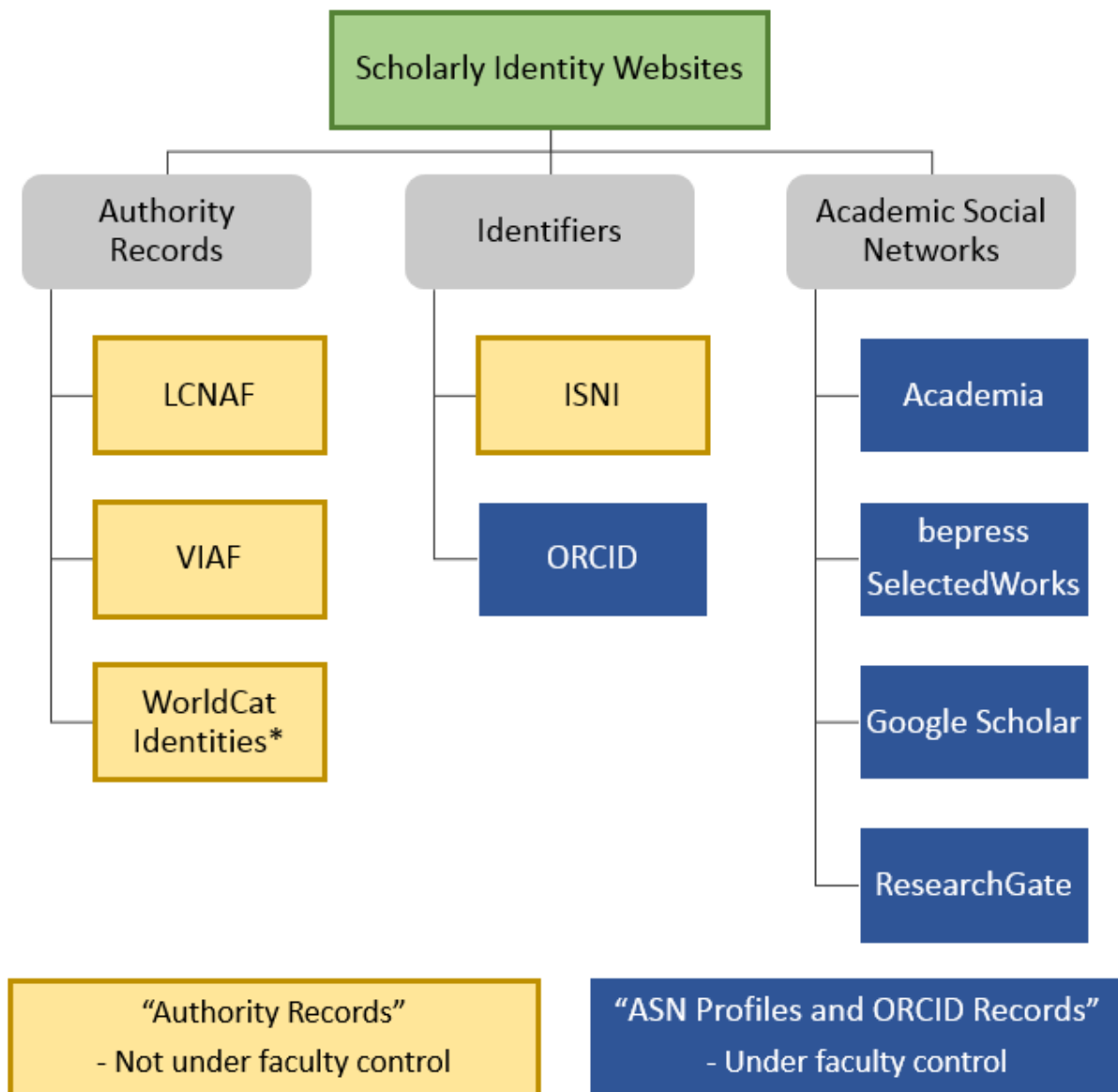
Figure 1. Groupings of scholarly identity websites for the purposes of this study.

Academic Visibility, Researcher Identifiers, and Authority Records

Academics are increasingly aware of the value of becoming visible on the public Web. Tenure-track faculty especially can benefit from publicity about their scholarship as well as themselves as scholars (Bik & Goldstein, 2013; Mauvais-Jarvis, 2016). Common motivations

among scientists for increased visibility include citation frequency and other scholarly performance indicators (Ebrahim et al., 2013; Ward, Bejarano, & Dudás, 2015). Humanities scholars may have different motivations for digital visibility, including public awareness, literature searching, and interactions with peers (Dagienė & Krapavickaitė, 2016; Wu, Stvilia, & Lee, 2017). Several studies have investigated the percentage of scientists and social scientists with social media profiles; results vary widely depending on the sample (Bar-Ilan et al., 2012; Mas-Bleda, Thelwall, Kousha, & Aguillo, 2014).

Academic identifiers are a prosaic but important part of becoming discoverable online. In 2016, three major organizations asked publishers to start requiring ORCID identifiers from their authors via an open letter (ORCID, 2017), and by the end of 2016, 25 associations and commercial companies had signed on, including Wiley and Springer Nature (Meadows, 2017). Other journals have followed suit (Swiontkowski, 2016). Rob Peters, Technical Director at ORCID, said the ORCID database is "definitely indexed by Google and Google Scholar" and that they have been told Google Scholar now uses ORCID identifiers for identifying researchers, noting it is hard to verify this information (personal communication to Fagan, 9/16/17).

Due to librarians' familiarity with issues of name disambiguation, they are well positioned to educate other scholars about managing online academic identities. Libraries promote the use of identifiers to faculty and students through presentations, online guides, flyers, and individual consultations. Outreach activities typically focus on one or more identifiers or profiles that researchers can manage themselves, with ORCID being common (Akers, Sarkozy, Wu, & Slyman, 2016). University promotion of identifiers often involves integrating them into specific information systems like institutional repositories, campus directories, and subject expertise databases (Rosenzweig & Schnitzer, 2015). In one example, Clark and Young (2017)

used ORCID and VIAF identifiers in an implementation of linked data to support their library's staff directory. Librarians at various institutions note that a high degree of personal contact with researchers is key to increasing the quantity of registered identifiers and the amount of data included in profiles (Thomas, Chen, & Clement, 2015).

While there are numerous case studies of activities libraries are undertaking in this area, the profession lacks data on outcomes beyond the number of researchers registered or qualitative feedback from faculty on whether they perceived the information to be helpful. We were unable to find literature on the effectiveness of these activities in boosting researchers' online visibility.

Libraries have also been involved in directly creating identifiers for scholars. In 2013 and 2014, three US institutions created ORCID identifiers for their faculty or student researchers as part of the ORCID Adoption and Integration Program (Brown, Oyler, & Haak, 2015).[2] Libraries have long created authority records for authors, but now that these records are published online and available as linked data, how might they support web visibility of scholars and their publications? Biographical articles in Wikipedia employ an authority control template which displays identifiers from Library of Congress authority records, VIAF, ISNI, and WorldCat Identities, among other sources (Wikipedia, 2017). In a 2014 presentation on increasing web visibility, Schubert and Holloway informed faculty that publishing activities could result in creation of identifiers by agencies like the Library of Congress, VIAF, and ISNI; they appear to assume the utility of such records for faculty visibility. However, we were unable to find literature discussing the influence of authority records on the discoverability of scholars or their publications.

---

[2] Note: ORCID no longer provides this functionality.

Search Engine Indexing and Use of Metadata

Discovering what search engines include can be a challenge, as most search engine companies do not offer specifics. Google Scholar supplements its vague information about coverage with the suggestion to search and find out for yourself (Google, 2018e). WorldCat's FAQ site provides some clues, noting "WorldCat.org has a sitemap that points all search engines to more than 10.5 million records. . . . However, search engines can also freely harvest from the entire database, especially those items on WorldCat.org that are linked to and referenced from other sites. Search engines index everything that is visible on the detailed record pages" (2018a).

Current information about how search engines use metadata is scarce, including the indexing of scholarly resources specifically. Zhang and Dimitroff's landmark study from 2005 about the impact of metadata on visibility in search engine results has been cited 115 times since publication, according to Google Scholar, but seems not to have been reproduced. One article from 2017 investigated the performance of Dublin Core vs. Schema.org with respect to discovery of geospatial resources (Katumba & Coetzee, 2017). The researchers found that pages marked up with Schema.org were more effectively retrieved by Bing than those marked up with Dublin Core, but found no statistically significant difference with Google. Their literature review confirmed "there is currently little work in academic literature that discusses SEO techniques" and noted Zhang and Dimitroff's two-part study to be "among the few scientific publications about search engine optimization" (section 2.2). An informal study by Lotfipanah (2016) showed anecdotal improvement in SEO for a scientific journal, but the methodology was unclear.

Many highly ranked pages concerning SEO of scholarly articles are quite old (e.g., Elsevier Biggerbrains, 2012). More recent publications (Allen & Weber, 2015; Kień, 2014) reference older information (e.g., Beel, Gipp, & Wilde, 2010; Dawson & Hamilton, 2006). The

same is true of tips for SEO practices regarding PDF documents generally. One finds frequent mention of tips in blog posts without mention of their origin (e.g., DeYoung, 2007); it is uncertain if the tips have been verified to still be effective, although perhaps the experience of industry experts lends credibility to their observations. Google's Webmaster Central guide for PDFs is from 2011 (Google, 2011). Google employee John Mueller provided a brief update to the Google Webmaster Central guide by stating in 2016 that Google indexes PDF files as they would any other web page. He noted, "we don't much refresh them as quickly as normal HTML pages because we assume that the PDF files stay stable" (Griffin, 2016). He also confirmed that Google may not index PDFs if links to them are hard to find on the website or are not located in static HTML. Google may also stop crawling content on a website if the engine thinks they have indexed a lot of content from the site already, only including additional content "if we find something really compelling" (Griffin, 2016). In sum, information about SEO is scattershot and often of uncertain provenance.

A small number of studies have specifically examined the indexing of library content in search engines. Onaifo and Rasmussen (2013) presented a study of Canadian public library website search engine rankings and reputation, finding "a moderate positive relationship between the number of indexed pages and daily page views" (p. 102). With respect to institutional repositories, Arlitsch and O'Brien (2012) demonstrated that using Highwire Press tags instead of Dublin Core metadata tags increased their IR's indexing in Google Scholar from 0 to 90%. However, as recently as 2016, Yang (2016) found that search engines are still struggling to discover PDFs and sometimes metadata in IRs. Allen and Weber (2015) offered tips for "coding practices for greater success in search engine retrieval" of open access journals based on corroborating their observations with other studies.

High-quality information about search engine optimization for scholar profiles is also elusive, but blog posts by industry experts suggest SEO for people relies heavily on general social media sites and academic social networking sites, more so than SEO for documents does. Schema.org Person markup (Schema.org, 2019) and personal websites also have the potential to influence discoverability of people in a way that may be different than for documents. SEO consultant Andy Crestodina (2012) outlines fourteen tactics for improving personal SEO on Google, including setting up a G+ profile, using LinkedIn, creating a personal website, and using Google Authorship. Adding even one tweet or post to tell people where to find one's centralized information can boost personal SEO. Academic publishers also provide advice to individuals for promoting themselves on social media, but many of these articles are several years old and may not delve any deeper than industry articles (Thomsen, 2015; Ward & Guest, 2013).

None of the items cited above mentioned scholarly identifiers or authority files as part of SEO strategy. Searches in LISTA and Google Scholar combining the term "search engines" with terms such as "author identifiers," "author identities," and "ORCID" were unsuccessful.

In conclusion, information about how search engines discover scholarly profiles and publications, including the PDF documents, which are a common scholarly format, is somewhat scarce. The influence of scholarly profiles, identifiers, and authority files on search engine rankings has not been studied empirically. Using an exploratory approach, we sought to develop a methodology and gather initial results to inform research concerning the effect academic social networking profiles and authority records might have on improving discoverability.

Methodology

The samples for this pilot study consisted of 24 faculty from James Madison University Libraries and 35 of their publications (between 0 and 3 publications per faculty member). While

these samples are small, our direct knowledge of the participants, access to information about their publications, and ability to create authority records for them meant we would be better able to interpret our results in this exploratory context.

As part of a project to create or update authority records for these scholars, we solicited information from each faculty member, including their preferred form of name, other names they used, previous institutional affiliations, publication citations, and ORCID identifiers. All forms of name provided to us were used in searches on profile sites and in Google and Bing.

In an attempt to avoid having our physical location and institutional affiliation affect our search results (Cazier, 2016), we performed searches using Tor Browser, a version of Firefox which "prevents the sites you visit from learning your physical location" (Tor, 2018). Using Tor Browser also ensured that our full-text results from publication searches would be items available in full text to anyone, not just users with access to our institution's subscriptions. Academia and ResearchGate did not allow us to search for faculty profiles using Tor Browser, so we instead used Firefox for those searches.

As detailed below, we first conducted a series of searches to set a baseline (Round 0 and Round 1). Faculty were searched for on scholarly identity websites and in Google and Bing, and their publications were searched for in Google and Google Scholar. Name authority records (NARs) were then created for faculty who lacked them and updated for faculty who had existing NARs. We then repeated the searches (Round 2).

Searches for Records and Profiles on Scholarly Identity Websites

We thought it important to record whether faculty had scholarly profiles to be able to check for their possible influence on search results as well as to note any additional profiles or records that were created during the course of the study. Data on the existence of LC and VIAF

authority records and ISNI identifiers for each faculty member was provided by the Metadata Strategies department, and the Head of Digital Collections contributed data on bepress SelectedWorks profiles. To set a baseline for the other ASN profiles and ORCID records, we searched for each faculty member in Academia, Google Scholar, ORCID, and ResearchGate, using the faculty member's first and last names as search terms. If the individual had submitted additional names, an additional search on each site was performed using each name.

For each of the names searched, the research team recorded the search date, whether the person was found, the URL of the profile, and the name used on the profile. Similarly to Bar-Ilan et al. (2012), we verified the identities of the scholars according to self-provided details or details known to the authors. Regarding Academia, searches may return names with no further information. We decided Academia results had to have at least one associated paper to count as a profile. To identify whether the individual represented by the Academia profile corresponded to our faculty member, we looked for James Madison University as an institutional affiliation. If no affiliation was present, publication titles were used. With ResearchGate, profiles had to have a URL in the form https://www.researchgate.net/profile/%3cname%3e. We did not count non-profile pages that listed an author's works (https://www.researchgate.net/scientific-contributions/<name>).

After 3 and 6 months had passed,[3] we performed the same procedure to determine whether faculty profiles had been created on these sites since our last visit. Because Google's crawl frequency is algorithmic in nature and depends heavily on the target website's

---

[3] Dates of searches on scholarly identity websites were: Round 0: 7/24/17; Round 1: 10/11/17 – 11/8/17; Round 2: 1/9/18 – 2/20/18.

characteristics (Google, 2018b), we had no hypothesis about how long it might take for the creation of authority records or profiles to influence search results.

Searches for Records and Profiles in Google and Bing

Since our hypothesis was that records and profiles on scholarly identity websites would result in better web visibility for faculty, we investigated how they were ranked in Google and Bing. Using the same names we used in the scholarly identity website searches, we searched each name in Google and Bing, then used Find in Page to locate the websites' hostnames within the first 50 results. The hosts searched for were loc.gov, viaf.org, isni.org, www.worldcat.org/identities/, academia.edu, researchgate.net, orcid.org, scholar.google.com, and bepress.com. For each match we recorded where the site ranked, or "NA" if it was not found within the top 50 results.

After gathering data we noticed our assistant had occasionally listed multiple hits for some sites. Looking more closely, we saw that these could be hits on specific publications (especially with bepress, as some URLs to publications begin with a string similar to profile URLs) or even to another profile for a faculty co-author. We found that the top-ranked hit was most likely to be the profile itself, so we retained the top hit's ranking for further analysis and did not include any additional hits for that profile in our results. Also, to allow for comparison between individuals with different numbers of names searched, we analyzed the top-ranking hit for each individual regardless of which name variant returned that result, except for when we specifically examined whether faculty were found under more than one form of their name.

We repeated these searches two months after the LCNAF records were created and updated, which was three months after our initial round of Google and Bing searches.[4]

Questions we investigated based on these searches included:

- How highly ranked are scholarly identity websites in search engines?

- To what extent do Google and Bing appear to index the various sites?

- Are there differences in the rankings of faculty with and without NARs?

- Do the rankings of scholarly identity websites in search results change after we create or update NARs?

- How do rankings correlate with the existence of scholarly identity websites?

Searches for Publications in Google and Google Scholar

Additional research questions surrounded the extent to which scholarly identity records influence the visibility of faculty publications. We performed several series of searches for publications by our sample of faculty. The publications sample comprised all citations in existing authority records and citations submitted by faculty in response to a survey designed to gather information for creating authority records; therefore it was not a comprehensive list of all their publications. Faculty were instructed to submit up to four publications for the survey.

In the sample of 35 items there were 14 books, 14 journal articles, 3 book chapters, 1 archival collection, 1 music score, 1 thesis, and 1 unpublished paper. Publication dates for these items ranged from 1993 to 2017, with just over half having been published since 2014.

---

[4] Dates of Google and Bing searches were: Round 1: 11/13/17 (before NARs were updated/created); Round 2: 2/1/18 – 2/6/18 (after NARs were updated/created).

To set a baseline for publication ranking, we submitted two searches for each publication in both Google and Google Scholar: a title search (including subtitle) with no quotes, and a keyword+last name search using major words from the title and the author's last name. We examined the first 50 results. Previous studies of known items focusing on retrieval effectiveness have examined as few as the top three results (Ciccone & Vickery, 2015), but since we were interested in how visible sources were, we decided to look more deeply into results sets.

For each search, we recorded the search date, the number of hits found in the first 50 results,[5] and the ranking and URL of the first three hits. A "hit" was defined as either a full-text version of the publication or a structured record. Examples of structured records included library records, Amazon.com book result pages, Google Books result pages, and journal article landing pages. A bibliographic citation (e.g., in a reference list) was not considered a record and was not counted as a hit.

When using Google Scholar, we clicked on "All X versions" to look for full text, because studies have shown full text can be buried under this link (Fagan, 2017; Pitol & De Groote, 2014). We numbered these with decimals, for example if a result ranked second in the main results list and clicking "All X versions" revealed second and third versions of full text, we recorded the rankings as 2, 2.2, and 2.3. Versions were not counted toward the total number of hits in the first 50 results because they do not appear on the main results screen.

After Round 0, we found most of our publications ranked very highly, often appearing as the first or second result. This meant any increase in their ranking would not be detectable. In

---

[5] While we recorded the number of hits found in the first 50 results, we did not end up using this data (see Methodological Development).

Round 1 and Round 2, we decided to perform a search on major title keywords only, without the author's last name included in the search string, in addition to full title and keyword+last name. By repeating our searches before the authority records were created or updated, we were able to compare rankings between Round 0 and Round 1 to see how they might change after a few months, independent of any influence from the authority records. A final round of searches was conducted 1 month after distribution of the new and updated NARs.[6] We hypothesized that if a third round of searches showed a difference after the NARs were created, it seemed more likely that change could be attributed to the authority records. We chose 1 month to try to reduce the effect of other types of changes on ranking.

Questions we investigated based on this series of searches included:

- How much of the sample could be found by Google and Google Scholar?

- How did publications rank in search results, and did the rankings change over time?

- Did the visibility of publications differ based on the type of search (full title, keyword+last name, keyword only)?

- How many of the items found were available in full text to the general public, and did this change over time?

- How did search rankings correlate with the existence of faculty profiles and authority records?

---

[6] Dates of publication searches in Google and Google Scholar were: Round 0: 7/25/17 – 7/27/17; Round 1: 11/13/17 – 12/1/17; Round 2: 1/9/18 – 1/30/18.

Records and Profiles on Scholarly Identity Websites

Prior to this study, 10 of the 24 faculty in our sample had authority records (NARs) in the LCNAF; these had been created between 1996 and 2016. Between Rounds 1 and 2, we updated all 10 existing records and created new authority records for 12 additional faculty members. Authority records were not created for two faculty members due to a lack of bibliographic warrant.

In all rounds of searching, the presence of faculty in VIAF matched that in the LCNAF, as all LCNAF records are contributed to VIAF. While VIAF records can come from sources other than the LCNAF, during this study, no faculty member without an LCNAF record was found in VIAF with a record contributed from another source. Six faculty had ISNIs at the beginning of the study, and no new ISNIs were created during the course of the study (see Table 1).

Table 1. The existence of records and profiles on scholarly identity websites before and after the study.

	Before Study	After Study	Change
Authority Records			
ISNI	6	6	0
LCNAF	10	22	+12
VIAF	10	22	+12
ASN Profiles and ORCID Records			
Academia	5	5	0
bepress SelectedWorks	13	13	0
Google Scholar	7	7	0
ORCID	19	19	0
ResearchGate	6	7	+1

Of the 24 faculty in our sample, by the end of the study 5 had profiles in Academia, 13 had bepress SelectedWorks profiles, 7 had Google Scholar profiles, 19 had publicly visible ORCID records, and 7 had profiles in ResearchGate (see Table 1). Three additional faculty members provided us with their ORCID identifiers but did not have publicly visible ORCID records; they are not included in our results or analysis. Figure 2 shows the number of faculty with zero, one, two, three, four, or five ASN profiles or ORCID records: 4 faculty had none, 5 had one, 6 had two, 4 had three, 3 had four, and 2 had five. One individual gained a ResearchGate profile after our Round 1 searches on the profile sites, but before we conducted Round 1 profile searches in Google and Bing and searches for faculty publications. No other profiles were created or removed for any other faculty members.
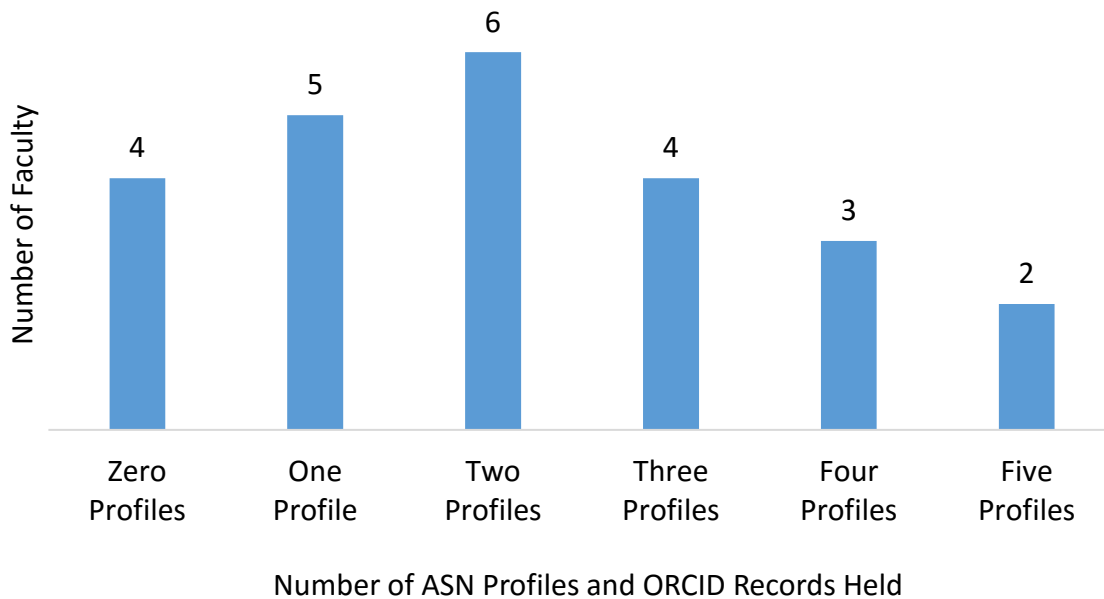

Figure 2. Distribution of the number of ASN profiles or ORCID records held by faculty (n=24).

Over half of the faculty in our sample with an ORCID record used its "Also known as" feature to list additional forms of their names. We were unable to determine whether faculty

included variant names in their Academia, bepress SelectedWorks, and ResearchGate profiles, as those sites do not display this information.

We searched for profiles using variant forms of name provided to us by 19 faculty members. Of the 5 faculty with Academia profiles, 4 were searched for using name variants, and all 4 were found with only one of the names searched. Of the 7 faculty with Google Scholar profiles, 6 were searched for using name variants. Three were found for only one of the names, and 3 individuals were found under more than one form of name. Of the 19 faculty with ORCID records, 15 were searched for using name variants, and all 15 were found for each form of name searched in at least one round. Finally, of the 7 faculty with ResearchGate profiles, 5 were searched for using name variants. Of these, all 5 were found for only one of the names.

Visibility of Records and Profiles in Google and Bing

Most authority records were not very visible in search engine results. Google did not find any records from the Library of Congress in either round of searching. In Bing, one faculty member's LC NAR was found in both rounds and another faculty member's was found only in Round 2. No records from VIAF or ISNI were found in the top 50 results in either search engine.

Google and Bing both found one WorldCat Identities page in Round 1 and two WorldCat Identities pages in Round 2. Because we did not search for these pages for our sample at the beginning of our study and are not able to determine when such pages are updated, we do not know how many faculty had a WorldCat Identities page that could have been found by Google or Bing, and this data cannot be directly compared with the other sites.

ASN profiles and ORCID records were more visible than authority records, but results were highly variable by website and search engine (see Figure 3). Of the 5 faculty with Academia profiles, Google only found one of them in the top 50 results in both rounds. Eleven of

the 13 faculty with bepress SelectedWorks profiles were found by Google in Round 1, and 10

were found in Round 2. Of the 7 Google Scholar profiles, all were found in the top 50 results for

both rounds of Google searching. Seven of the 19 faculty with ORCID records were found using

Google in Round 1, and 8 were found in Round 2. Finally, of the 7 faculty who had

ResearchGate profiles when Google searches were conducted, all 7 were found in the first round

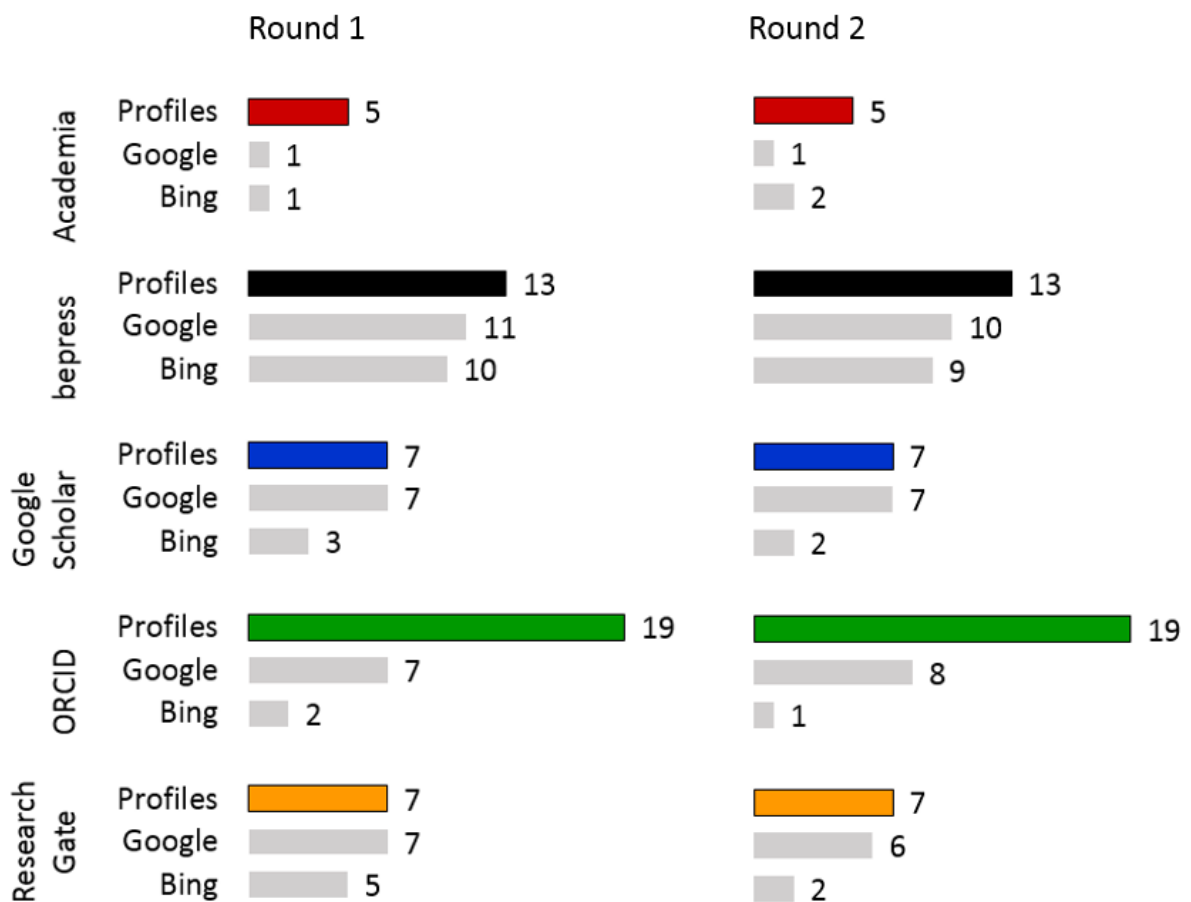of Google searching and 6 were found in the second round.

Figure 3. Visibility of ASN profiles and ORCID records in Google and Bing top 50 results. The

top bar in each group shows the number of profiles that could have been found; the lower two

bars show how many were found by Google and by Bing.

In Bing, of the 5 faculty with Academia profiles, only one was found in Round 1, and 2 were found in Round 2. Ten of the 13 faculty with bepress SelectedWorks profiles were found by Bing in Round 1, and 9 were found in Round 2. Three of the 7 Google Scholar profiles were found in Round 1, and 2 were found in Round 2. Of the 19 faculty with ORCID records, Bing found 2 of them in Round 1 and only 1 in Round 2. Finally, of the 7 faculty with ResearchGate profiles, Bing found five in the first round and two in the second.

Thus, Google was more successful than Bing at finding faculty across all platforms. Profiles from bepress, Google Scholar, and ResearchGate were more visible in Google search results than those from Academia or records from ORCID.

Figures 4 and 5 show how each faculty member's ASN profiles and ORCID records were made visible and ranked by Google. Figures 6 and 7 do the same for Bing.
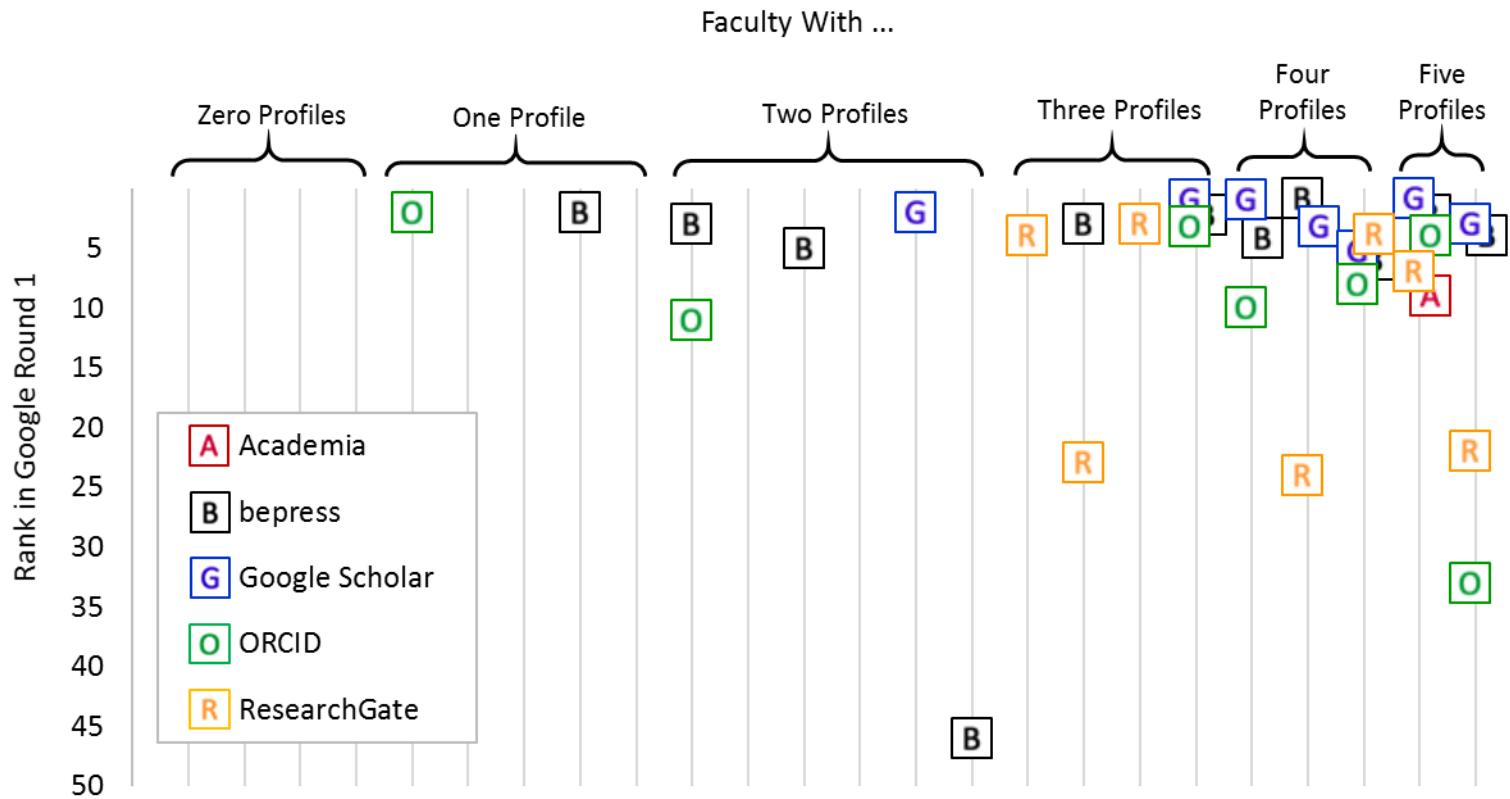
Figure 4. Ranking of top hits for ASN profiles and ORCID records (labeled "Profiles") in Google Round 1, grouped by the number of profiles and records held. There are 33 data points on the chart.

Figure 5. Ranking of top hits for ASN profiles and ORCID records (labeled "Profiles") in Google Round 2, grouped by the number of profiles and records held. There are 32 data points on the chart.
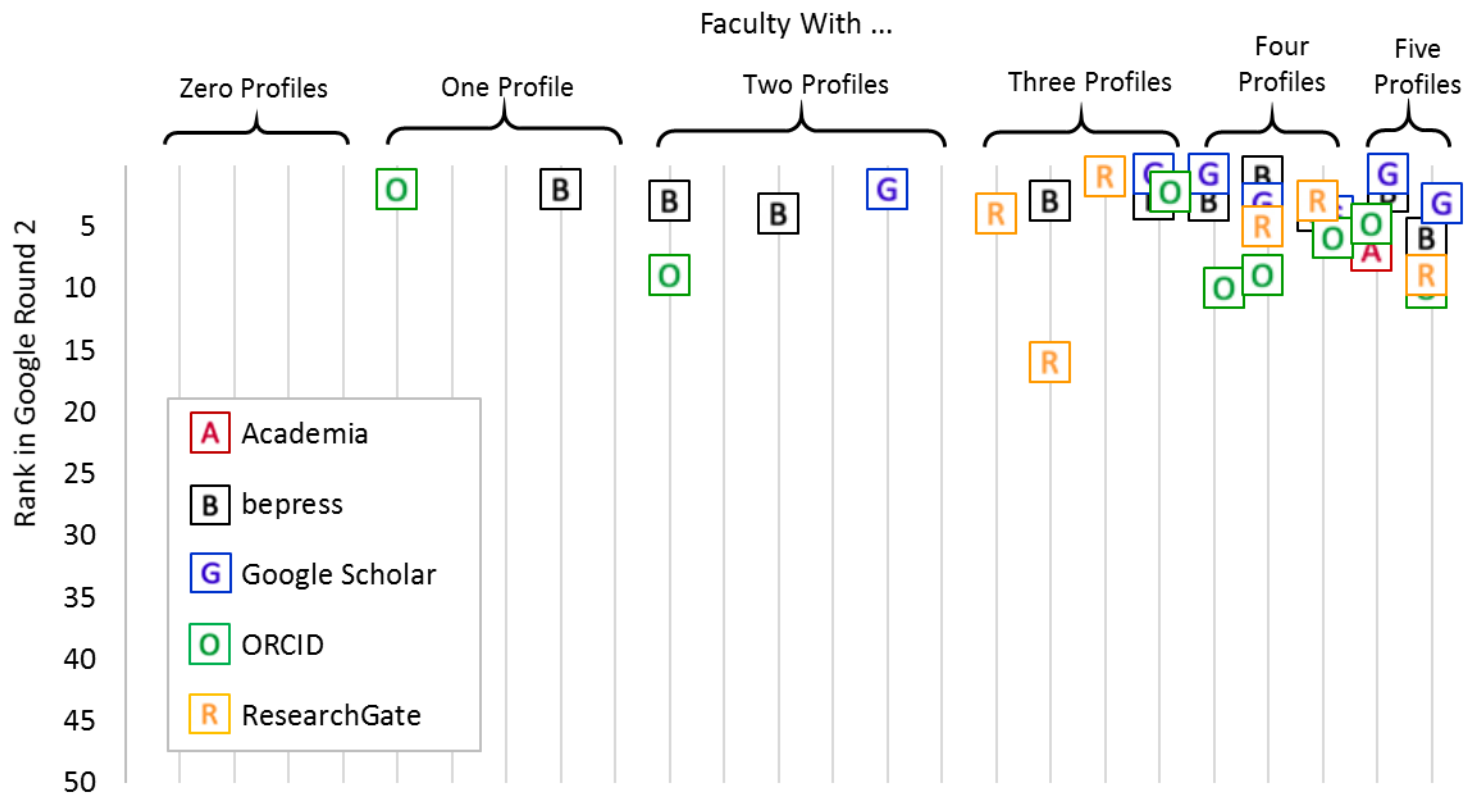
Figure 6. Ranking of top hits for ASN profiles and ORCID records (labeled "Profiles") in Bing Round 1, grouped by the number of profiles and records held. There are 21 data points on the chart.
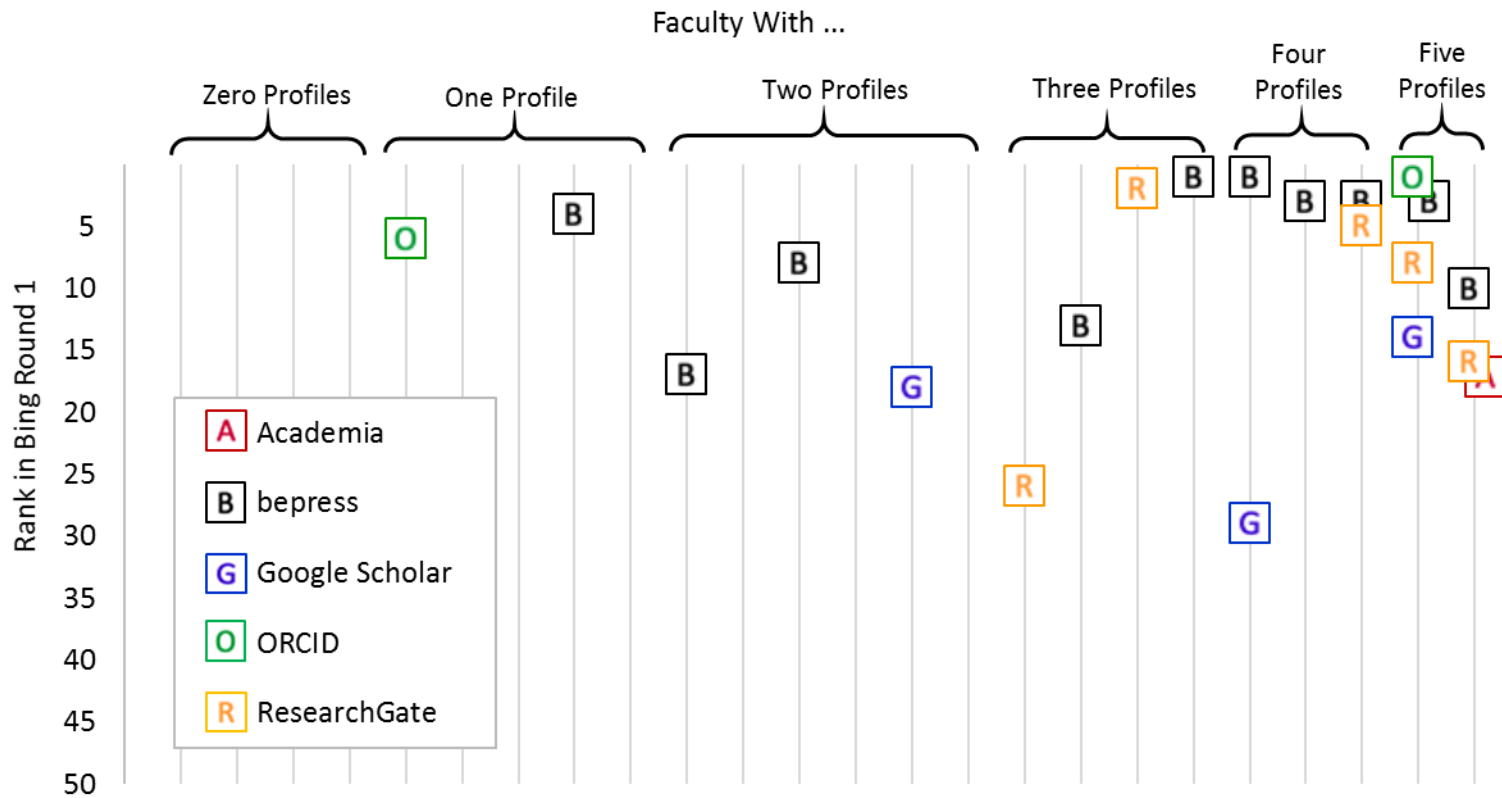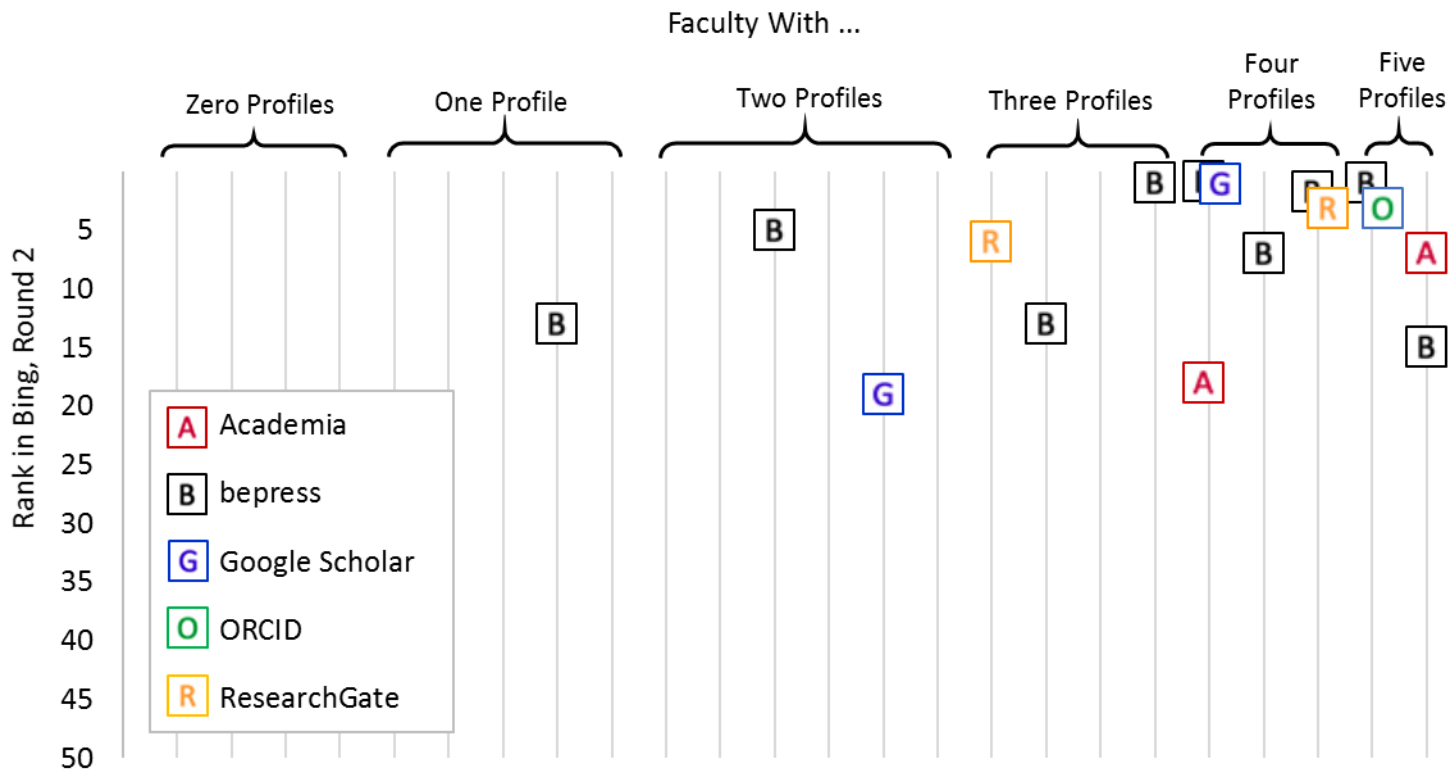
Figure 7. Ranking of top hits for ASN profiles and ORCID records (labeled "Profiles") in Bing Round 2, grouped by the number of profiles and records held. There are 16 data points on the chart.

Google found 33 of 51 ASN profiles and ORCID records in Round 1 (65%) and 32 in

Round 2 (63%). The profiles and records found in Round 1 represented 15 of the 20 individuals

with profiles, whereas the profiles found in Round 2 only represented 14 individuals. In Bing, 21

faculty profiles were found in Round 1 (41%, for 14 individuals), and 16 were found in Round 2

(31%, for 11 individuals) (see Table 2).

Table 2. Number of ASN profiles and ORCID records and individuals found, with averages and

standard deviations of rankings for all faculty and for groups based on authority record status.

Google	Round 1	Round 2
ASN Profiles and ORCID Records Found (N=51)	n=33 (65%)	n=32 (63%)
Individuals Found (N=20)	n=15 (75%)	n=14 (70%)
Average Rank and SD (all profiles and records)	*M*=7.91, *sd*=10.29	*M*=4.50, *sd*=3.47
Average Rank and SD (no NAR)	n=16, *M*=11.13, *sd*=12.65	
Average Rank and SD (with NAR)	n=17, *M*=4.88, *sd*=5.47	
Average Rank and SD (new NAR)		n=15, *M*=5.40, *sd*=3.91
Average Rank and SD (updated NAR)		n=17, *M*=3.71, *sd*=2.67

Bing	Round 1	Round 2
ASN Profiles and ORCID Records Found (N=51)	n=21 (41%)	n=16 (31%)
Individuals Found (N=20)	n=14 (70%)	n=11 (55%)
Average Rank and SD (all profiles and records)	*M*=9.76, *sd*=8.30	*M*=7.19, *sd*=6.36
Average Rank and SD (no NAR)	n=10, *M*=9.30, *sd*=5.73	
Average Rank and SD (with NAR)	n=11, *M*=10.18, *sd*=9.75	
Average Rank and SD (new NAR)		n=6, *M*=8.83, *sd*=5.11
Average Rank and SD (updated NAR)		n=10, *M*=6.20, *sd*=6.51

Note: "no NAR" means "profiles and records of faculty with no NAR" and "with NAR" means

"profiles and records of faculty who had a NAR." There were 24 profiles and records for faculty

with no NAR in Round 1 and 22 for faculty with a new NAR in Round 2. There were 27 profiles

and records for faculty who had a NAR in Round 1 and 27 for those with an updated NAR in

Round 2 (since all faculty NARs were updated between rounds).

Profiles were ranked more highly in Google than in Bing. In Round 1, the average rank was 7.91 in Google and 9.76 in Bing. In Round 2, the average rank was 4.50 in Google and 7.19 in Bing. Average rank improved from Round 1 to Round 2 in both Google and Bing, moving up about 3.5 ranks in Google and 2.5 ranks in Bing. Both search engines displayed a somewhat similar amount of variance in rankings in Round 1 (standard deviations 10.29 for Google and 8.30 for Bing), but in Round 2, the variance in Google decreased much more dramatically than Bing's (standard deviations 3.47 for Google and 6.36 for Bing).

Nineteen of the 24 faculty in our sample provided us with more than one form of their name. Four individuals with Academia profiles were searched for using name variants, and neither search engine found any individual's Academia profile under more than one name. Nine bepress profiles were searched for under alternate names. Three were found under more than one name in both rounds of Google searching (33%), and four were found in both rounds of Bing searching (44%). Six faculty with Google Scholar profiles were searched for using name variants. In Google, Google Scholar profiles were found under more than one name at least half of the time (3 profiles found in Round 1, 4 in Round 2), but in Bing, Google Scholar profiles were never found under more than one name. Both search engines found 2 out of 5 (40%) ResearchGate profiles under alternate names in both rounds. Google found 3 out of 15 (20%) ORCID records under more than one name; Bing found no ORCID records under multiple names.

Google and Bing failed to find any scholarly identity website within the top 50 results for 9 of the 24 faculty in our sample, despite the fact that all but one had at least one site that could have been found. Four of the nine had neither ASN profiles or ORCID records in either round, but three had authority records in LCNAF and VIAF in Round 2. While in retrospect it is not too

surprising these four were not found, the other five had ASN profiles and ORCID records in addition to four of them having authority records. Of these five, one faculty member had a profile on Academia and an ORCID record in both rounds, one had a bepress SelectedWorks profile and an ORCID record in both rounds, and the other three had ORCID records in both rounds. Having only an ORCID record or authority record may not be sufficient for web discoverability.

Effect of authority records on visibility of scholars.

In analyzing the effect of authority records on visibility of scholars, we focused on NARs in the LCNAF because the presence of our sample in VIAF and ISNI was dependent on whether they were found in the LCNAF. We again compared the rankings of ASN profiles and ORCID records in Google and Bing results across Rounds 1 and 2, this time grouped by authority record status.

Google found a similar percentage of ASN profiles and ORCID records for faculty with and without NARs in Round 1 and a similar percentage for faculty with new and updated NARs in Round 2. In the first round of searches, Google found 17 of the 27 profiles and records for faculty with authority records (63%) and 16 of the 24 profiles and records belonging to faculty without a NAR (67%). In Round 2, 17 of the 27 profiles and records for faculty with existing NARs that had been updated were found (63%) and 15 of the 22 profiles and records for faculty members with new NARs were found (68%). No ASN profiles and ORCID records were found in Round 2 for the two faculty members for whom NARs were not created. The Round 1 Google results were clustered more tightly at the top for those with authority records (average rank 4.88, standard deviation 5.47) than for those without (average rank 11.13, standard deviation 12.65), as shown in Figure 8 and Table 2. In Round 2, there was little difference between those with new

28

NARs (average rank 5.40, standard deviation 3.91) and updated NARs (average rank 3.71,

standard deviation 2.67), as shown in Figure 9.

Figure 8. Ranking of top hits for ASN profiles and ORCID records in Google Round 1, grouped by authority record status. There are 33 data points on the chart.
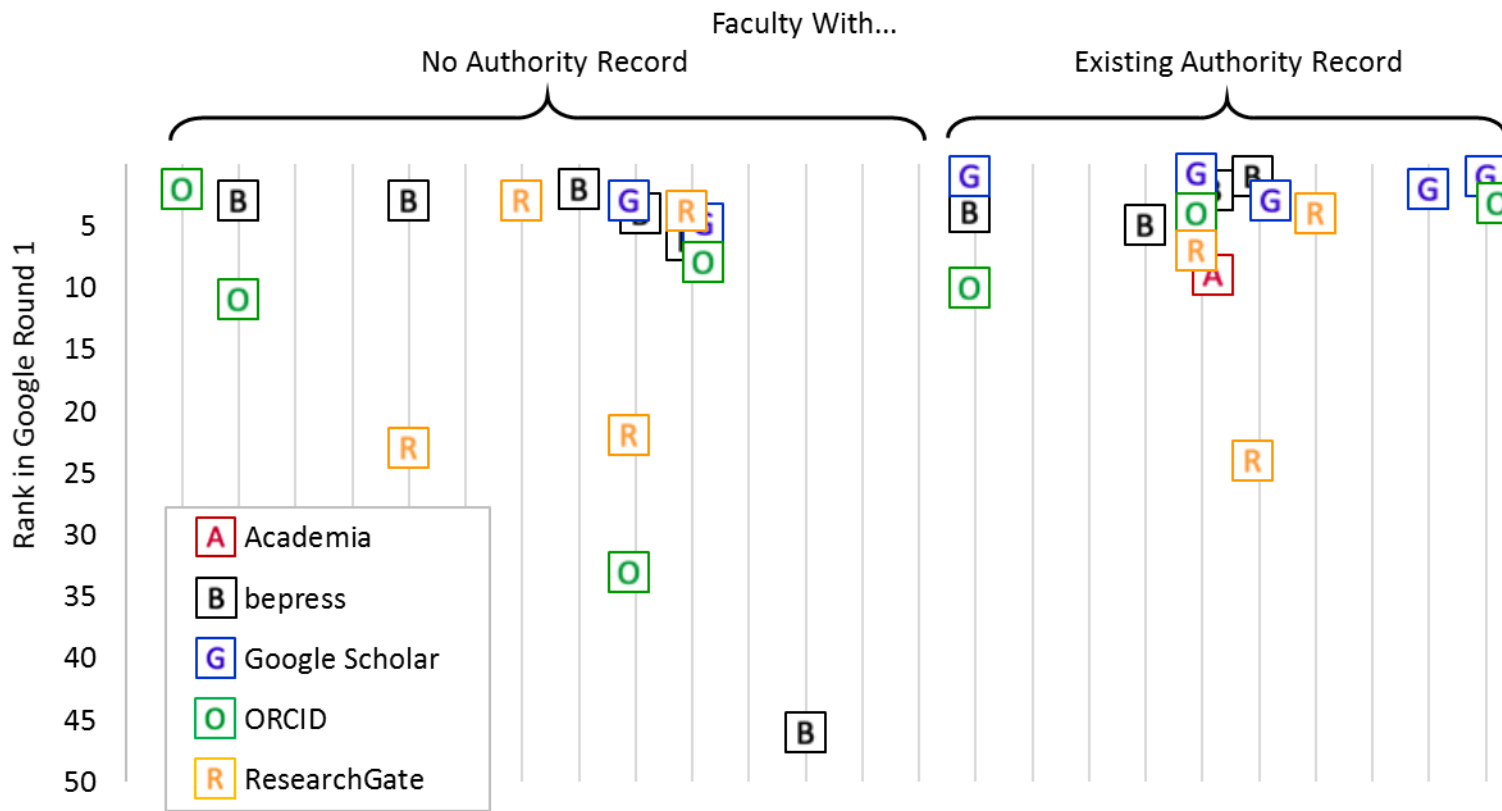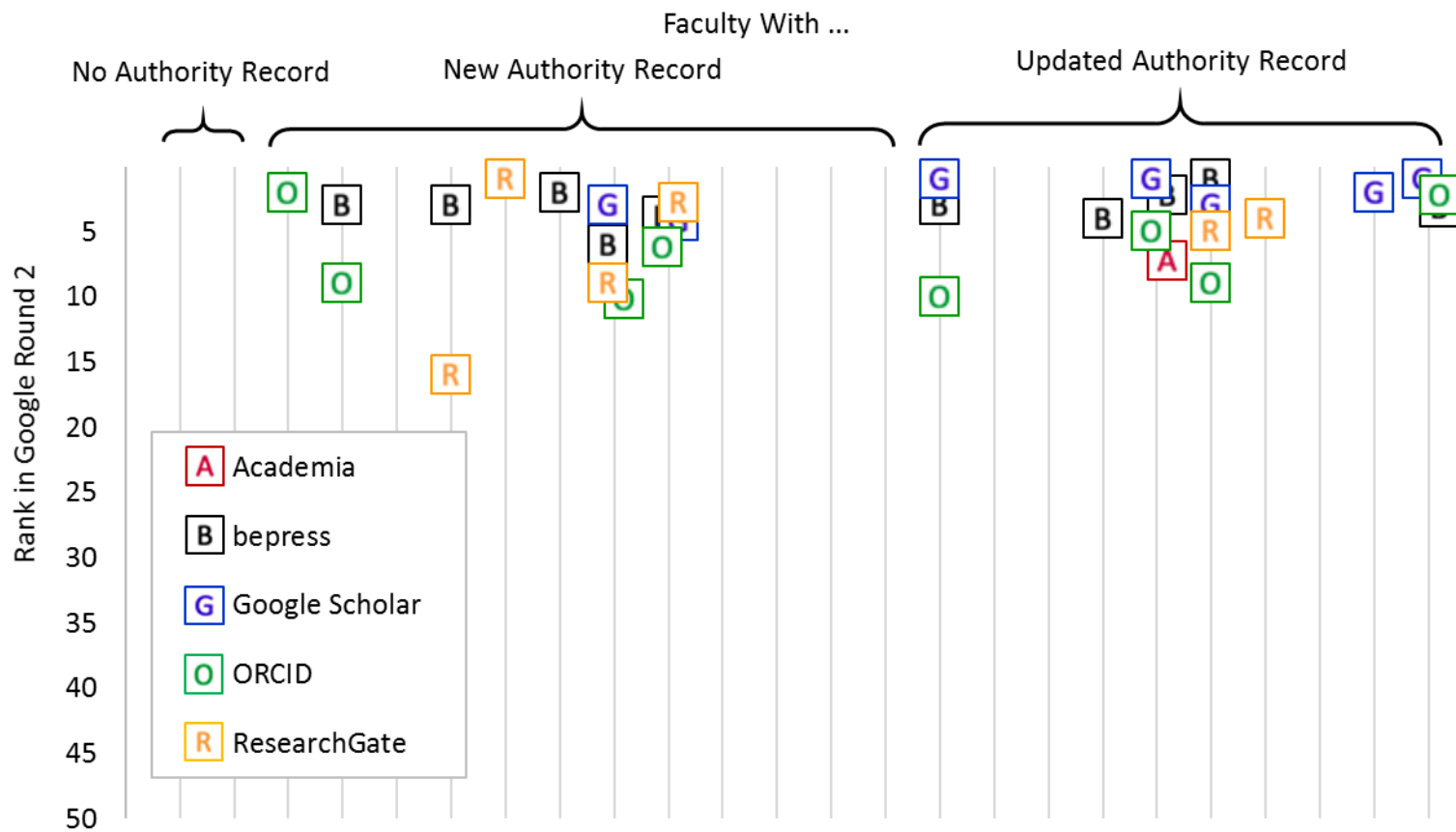
Figure 9. Ranking of top hits for ASN profiles and ORCID records in Google Round 2, grouped by authority record status. There are 32 data points on this chart.

In the first round of Bing searches, faculty with NARs had 11 of their 27 ASN profiles and ORCID records found (41%), whereas those without NARs had 10 of their 24 profiles and records found (42%). Rankings were lower and more variable for those with NARs (average rank 10.18, standard deviation 9.75) than for those without NARs (average rank 9.30, standard deviation 5.73), as shown in Figure 10 and Table 2. In the second round of Bing searches, faculty with new NARs had six of their 22 records and profiles found (27%), while 10 of 27 records were found for faculty with updated NARs (37%) (Figure 11). As with Google, Bing did not find any ASN profiles or ORCID records in Round 2 for the two faculty members without authority records. Although in Round 2 Bing's ranking for faculty with new NARs was higher, the variability was greater (average rank 6.20, standard deviation 6.51) than for those with updated NARs (average rank 8.83, standard deviation 5.11).

Figure 10. Ranking of top hits for ASN profiles and ORCID records in Bing Round 1, grouped by authority record status. There are 21 data points on this chart.

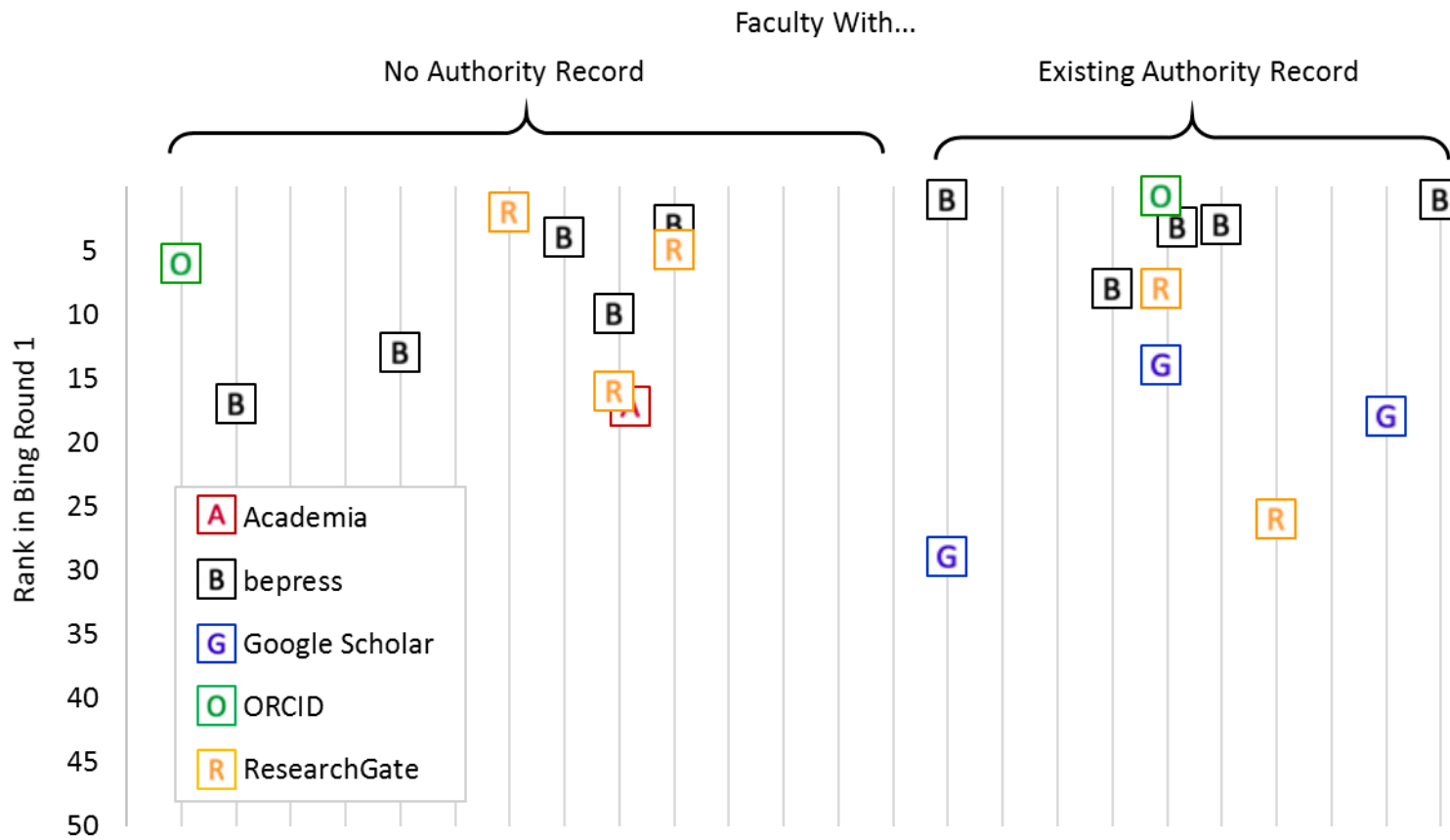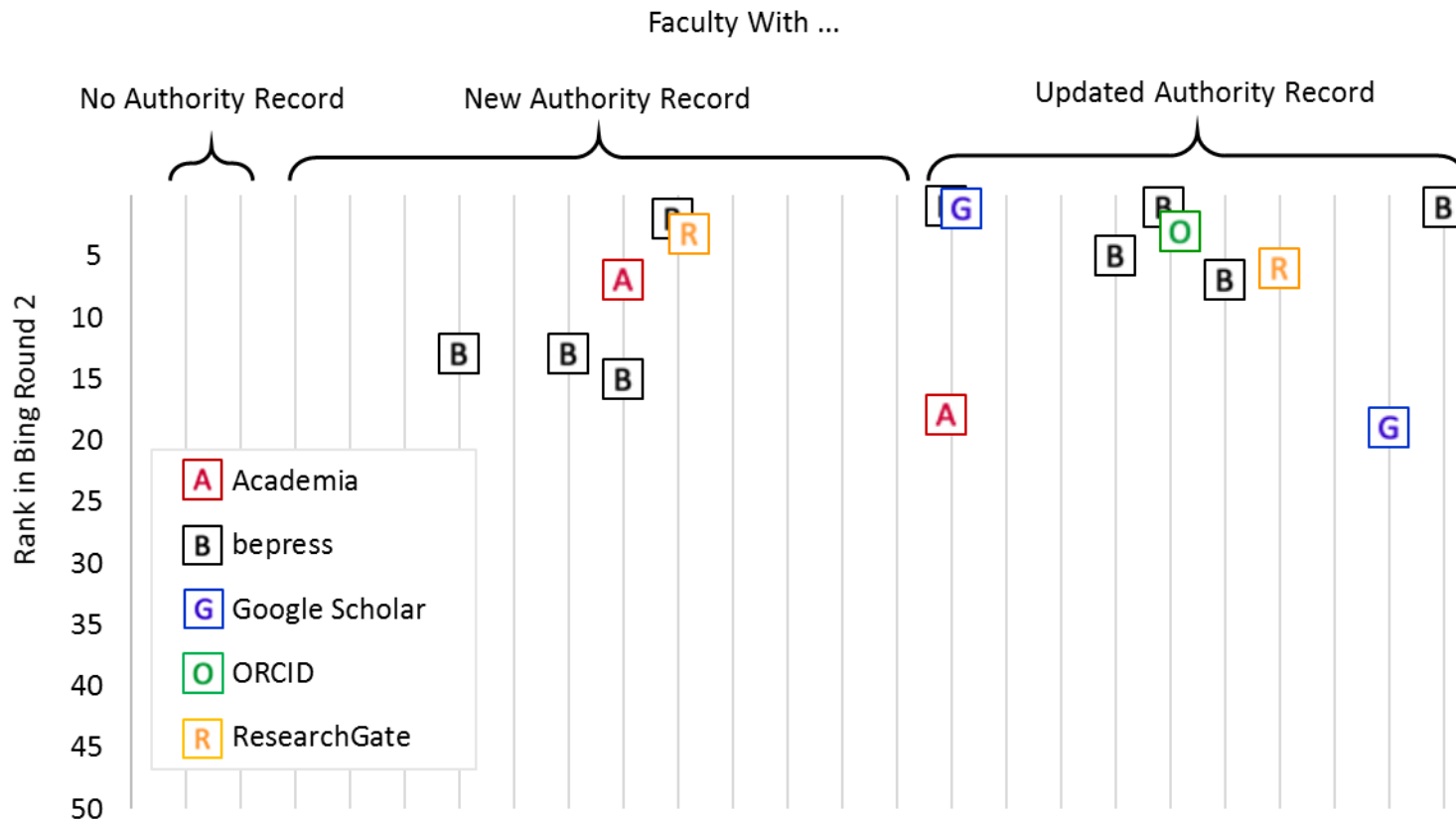Figure 11. Ranking of top hits for ASN profiles and ORCID records in Bing Round 2, grouped by authority record status. There are 16 data points on this chart.

Overall, results in Bing displayed more variability in ranking than did Google results, and Google results appeared closer to the top of the results lists. In all cases, the existence of an authority record alone did not determine whether that scholar's ASN profiles and ORCID records were found by search engines, as there were always some faculty in each category for whom no profiles or records were found. However, in both rounds, profiles and records of faculty with NARs (existing, new, and updated) ranked more highly on average in Google (3.71-5.40) than those without NARs (11.13). In Bing, faculty with existing NARs ranked slightly lower but had greater variability than those with no NAR in Round 1, and in Round 2, faculty with new and updated NARs ranked slightly higher than those in Round 1 with no NARs. ASN profiles and ORCID records for faculty with new NARs in Round 2 ranked more highly in both search engines than they did in Round 1, when they lacked NARs. In Google the average rank of this group increased from 11.13 to 5.40. The increase in Bing was smaller, from 9.30 to 8.83. While the amount of variability decreased significantly in Google, it stayed about the same for Bing (see Table 2).

Effect of academic social networking profiles and ORCID records on visibility of scholars.

The 24 faculty in our sample each maintained between 0 and 5 profiles or records in Academia, bepress SelectedWorks, Google Scholar, ORCID, and ResearchGate. Those with a greater number of these were more visible in Google searches, as shown in Figures 4 and 5. Among the 5 faculty with 4 or 5 ASN profiles or ORCID records, 19 of their 22 profiles or records were found in both Round 1 and Round 2 (86%). The 15 faculty with 1, 2, or 3 profiles or records had 14 of their 29 profiles or records found in Round 1 (48%) and 13 in Round 2 (45%). Rankings were high regardless of the number of profiles and records; they were higher in

Round 2. Faculty with 4 or 5 profiles or records had an average rank of 7.95 in Round 1 and 4.84 in Round 2, compared to faculty with 1, 2, or 3 profiles or records, who had an average rank of 7.86 in Round 1 and 4.00 in Round 2. For both groups, the variance in rankings decreased dramatically from Round 1 to Round 2 (see Table 3).

Table 3. Number of ASN profiles and ORCID records found, with means and standard deviations of rankings for groups based on individuals' number of profiles and records.

Google	Round 1	Round 2
ASN Profiles and ORCID Records Found (1, 2, or 3) (N=29)	n=14 (48%)	n=13 (45%)
ASN Profiles and ORCID Records Found (4 or 5) (N=22)	n=19 (86%)	n=19 (86%)
Average Rank and SD (1, 2, or 3)	*M=7.86, sd=11.94*	*M=4.00, sd=3.98*
Average Rank and SD (4 or 5)	*M=7.95, sd=8.56*	*M=4.84, sd=2.92*

Bing	Round 1	Round 2
ASN Profiles and ORCID Records Found (1, 2, or 3) (N=29)	n=9 (31%)	n=6 (21%)
ASN Profiles and ORCID Records Found (4 or 5) (N=22)	n=12 (55%)	n=10 (45%)
Average Rank and SD (1, 2, or 3)	*M=10.56, sd=8.00*	*M=9.50, sd=6.05*
Average Rank and SD (4 or 5)	*M=9.17, sd=8.12*	*M=5.80, sd=5.79*

Note: "1, 2, or 3" means "sites of faculty who had 1, 2, 3 academic social networking profiles or ORCID records," and "4 or 5" means "sites of faculty who had 4 or 5 academic social networking profiles or ORCID records."

In Bing, scholars with more profiles or records were also more visible than those with fewer (Figures 6 and 7). For the 5 faculty with 4 or 5 profiles or records, 12 of their 22 were found in Round 1 (55%) and 10 in Round 2 (45%). The 15 faculty with 1, 2, or 3 profiles or records had 9 of the 29 found in Round 1 (31%) and 6 in Round 2 (21%). Rankings were higher for faculty with 4 or 5 profiles or records than for those with fewer. Faculty in this group had an average rank of 9.17 in Round 1 and 5.80 in Round 2, compared to faculty with 1, 2, or 3 profiles

or records, who had an average rank of 10.56 in Round 1 and 9.50 in Round 2. Similar to Google, both groups showed a decrease in the variance of rankings between rounds, although Bing's decrease was less dramatic.

Faculty with 4 or 5 profiles had their profiles or records found approximately twice as often by both search engines as did faculty with fewer profiles or records. There were not marked differences in ranking between faculty with 1, 2, or 3 profiles or records and faculty with 4 or 5 profiles or records. On average, Google ranked profiles and records more highly, and Bing's results showed more variability in ranking than Google's.

Visibility of Publications in Google and Google Scholar

Most of the 35 publications in our sample were found in both Google (33 found) and Google Scholar (29 found) in at least one of the search rounds. The music score and archival collection were not found in any round in either search tool. Additionally, three of the books and one book chapter were not found in Google Scholar in any round.

For all search types, Google found a similar number of publications in each round (between 29 and 32). Google Scholar found a similar number of publications for full title and keyword+last name searches in all rounds (between 25 and 28) and found fewer publications for keyword only searches (15 in Round 1 and 14 in Round 2) (see Figure 12).
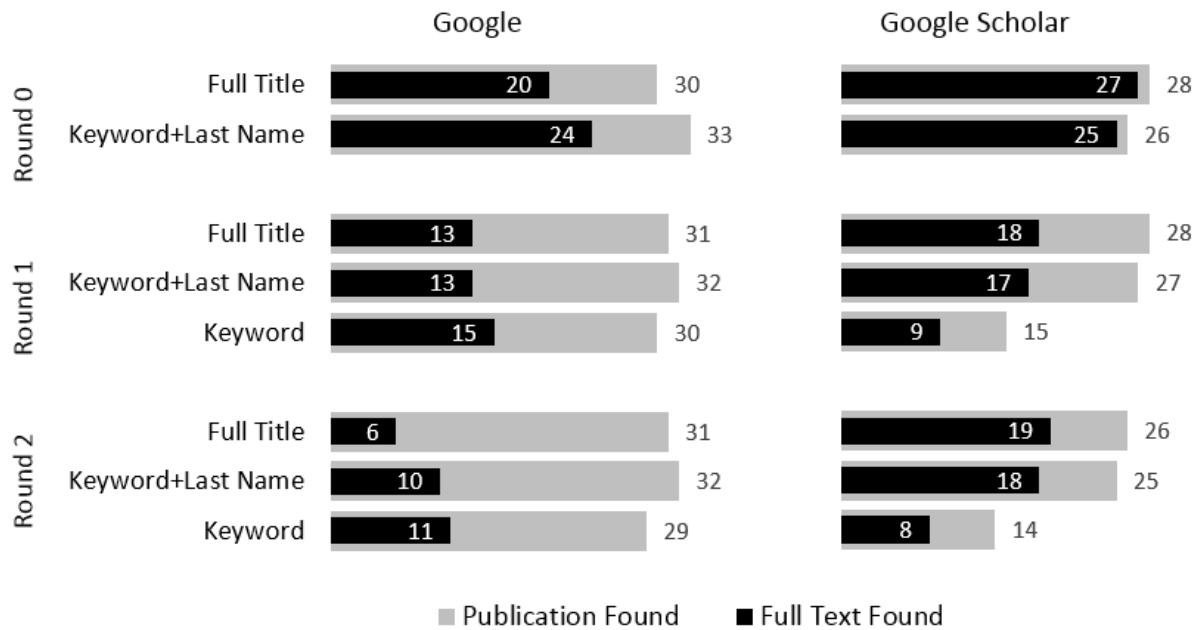
Figure 12. Publications and full text found by Google and Google Scholar. The grey bars show the number of publications found as records or in full text; the black bars show the number of publications found in full text.

In all rounds, when items were found, they typically ranked highly in search results, as shown in Figure 13. At least two-thirds of the first hits for found items ranked in the top five search results. The keyword searches in Google Scholar were the exceptions to this, with only approximately one-third of found items appearing within the top 5 search results. For full title searches, only one item ever ranked outside the top 5 results (for one round in Google Scholar). Keyword+last name searches in Google saw all but one or two items per round ranked in the top 5; in Google Scholar these searches had between 2 and 4 items ranked outside the top 5.
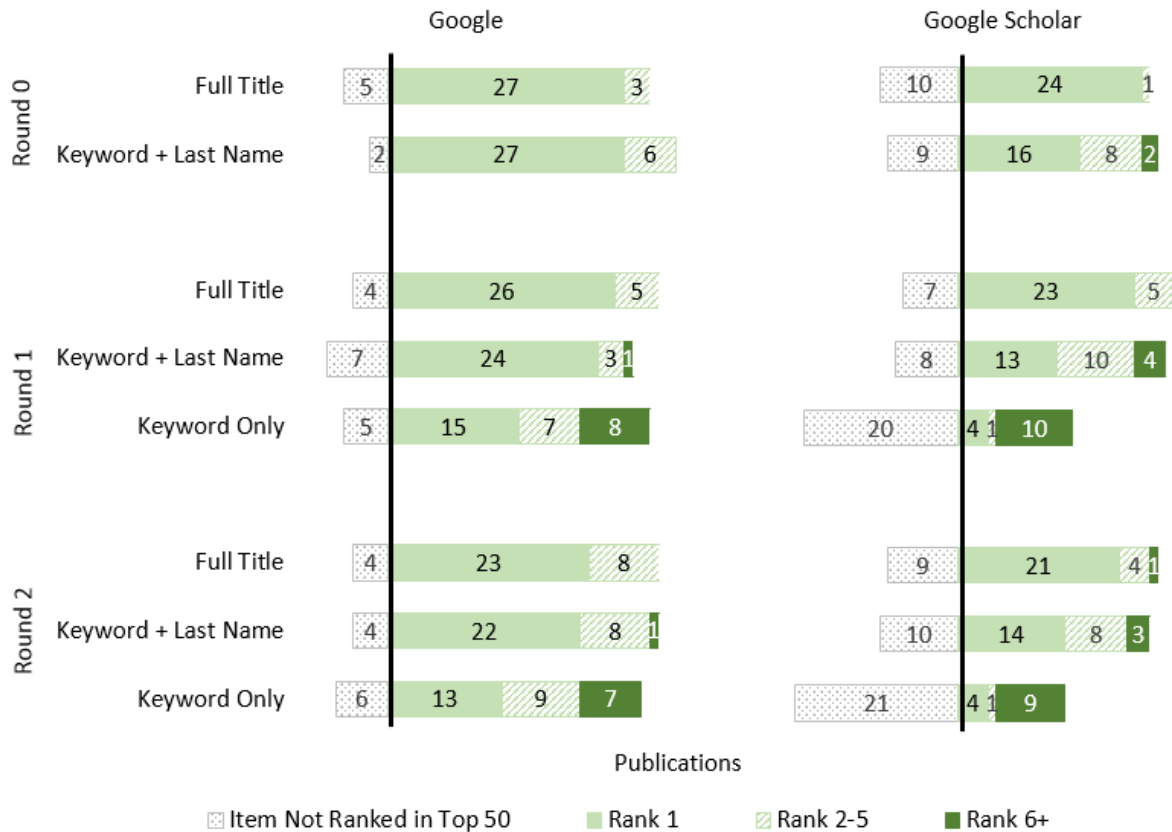
Figure 13. Ranking of top hit for each publication in search results for Google and Google Scholar.

The number of publications found varied only slightly between rounds (increasing or decreasing by 1-2 items or staying the same) (see Figure 12). We also did not find meaningful changes in the ranking of the first hit. Appearances and disappearances of results for individual publications did not seem to form a clear pattern. It is important to remember we only looked at the top 50 results, so "disappearances" could also be thought of as "large decreases" in ranking and "appearances" as "large increases" in ranking.

Looking at the top three hits for each search, Google experienced more changes in rank between Round 0 and Round 1 when using keyword+last name searches (61% of hits changed

rank) as opposed to full title searches (57% of hits changed rank). For these keyword+last name searches, the most common changes were slight increases and slight decreases, whereas for full title searches, the most common change was for a previously not-found item to appear (Figure 14).

Between Rounds 0 and 1, Google Scholar showed more change with keyword+last name searches (31% of hits changed rank) than with full title searches (20% of hits changed rank). For both search types, decreases in rank and disappearances from the top 50 results were more common than appearances, and there were no instances of increases in rank (Figure 14).

Between Round 1 and Round 2, Google Scholar still experienced the least change when using full title searches (25% of hits changed rank). Google experienced slightly more changes with full title searches (62% of hits changed) than with keyword+last name searches (60% of hits changed). Simple keyword searches were performed in both these rounds as well, and Google Scholar experienced much less change (35% of hits changed) compared to Google (78% of hits changed) (Figure 14).
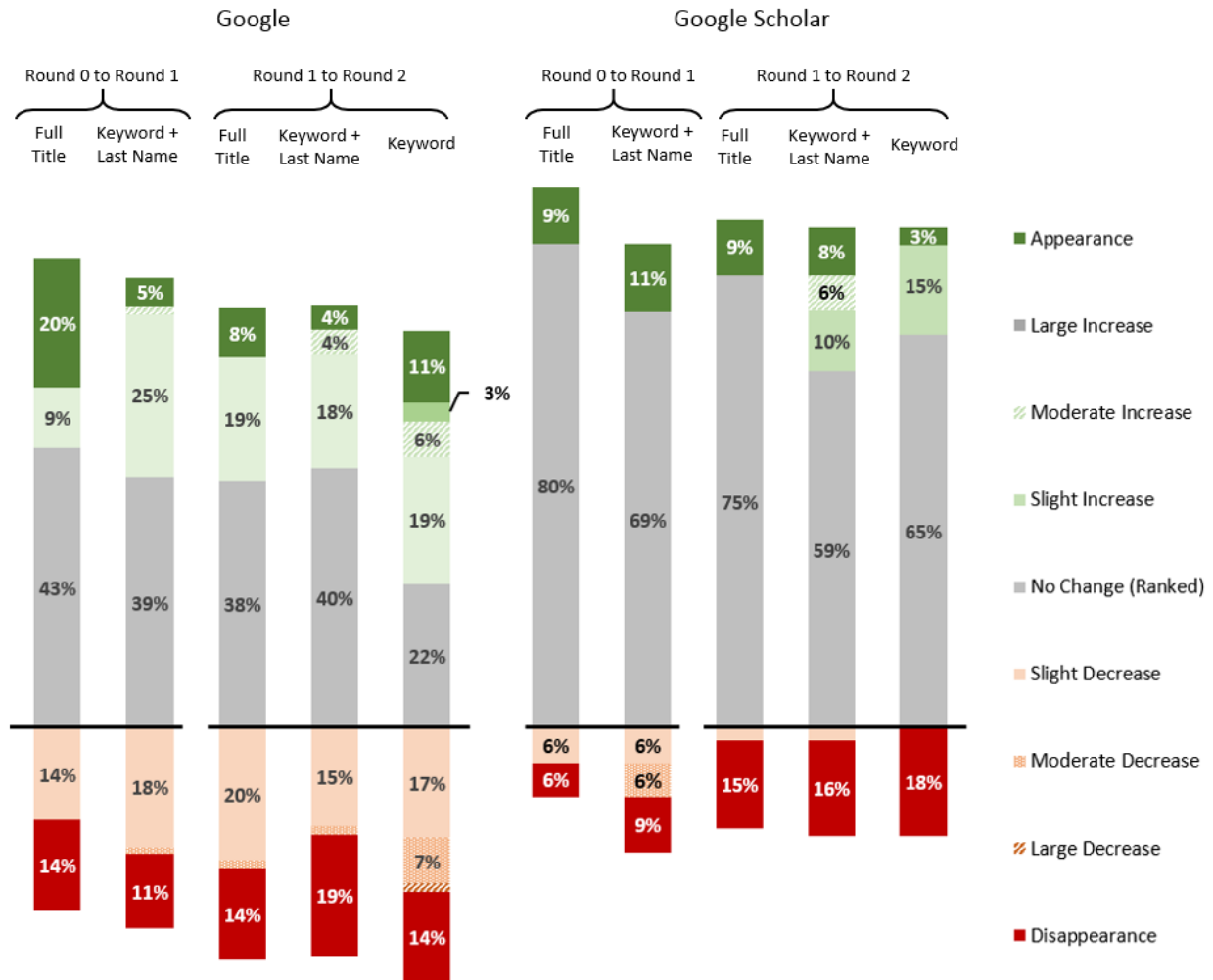
Figure 14. Percent change in hit rank for the top three hits by search type for Google and Google Scholar. Labels of 2% or less not shown. Due to rounding, the sum of each column may not total 100%. "Slight"=1-5 rank change, "Moderate"=6-20 rank change, "Large"=21-50 rank change.

Overall, Google experienced more change (57-61% of hits) between Round 0 and Round 1 than Google Scholar (20-31% of hits), and also between Rounds 1 and 2 (60-78% of hits changed in Google, compared to 25-41% in Google Scholar). Most of the changes seen in Google Scholar were appearances and disappearances rather than changes in rankings.

Finding full text

For each search's first three hits, we investigated whether the result was a full-text version of the publication. Of the 33 items found by Google across all rounds and search types, 30 were found in full text among the first three hits. Of the 29 items found by Google Scholar, 27 were found in full text.

For all search types and in all rounds, Google Scholar found a larger percentage of the found publications in full text than did Google, as shown in Figure 12. Even though Google Scholar only found 27 of the items in the sample in full text (compared to 30 found by Google), Google Scholar found the most full-text hits for searches on publications overall (141 out of 280 searches, or 50%), compared to 112 searches which found full-text hits in Google (40%). In Rounds 1 and 2, keyword searches were more successful than full title and keyword+last name searches at finding full text in Google. In Google Scholar, full title searches were the most successful at finding full text in all rounds, although keyword+lastname searches were only slightly less successful. Google found less full text in each successive round for all search types, whereas Google Scholar's results showed little change between Rounds 1 and 2.

Google and Google Scholar were both successful at providing multiple pathways to full text (see Figure 15). Across all rounds and search types, Google returned three full-text hits for 12 searches (11% of all searches returning full-text hits) and two full-text hits for 33 searches (29%). Google Scholar returned three full-text hits for 26 searches (18% of all searches returning full-text hits) and two full-text hits for 23 searches (16%). In terms of publications, across all rounds and search types, Google returned three full-text hits for 9 of the 35 publications (26%) and at least two full-text hits for 18 of the publications (51%). Google Scholar returned three full-text hits for the same number of publications as Google (9 of 35, or 26%), but only retrieved

15 publications with at least two full-text hits (43%). We did not detect clear patterns between rounds or among search types in the number of pathways to full text for a given item.
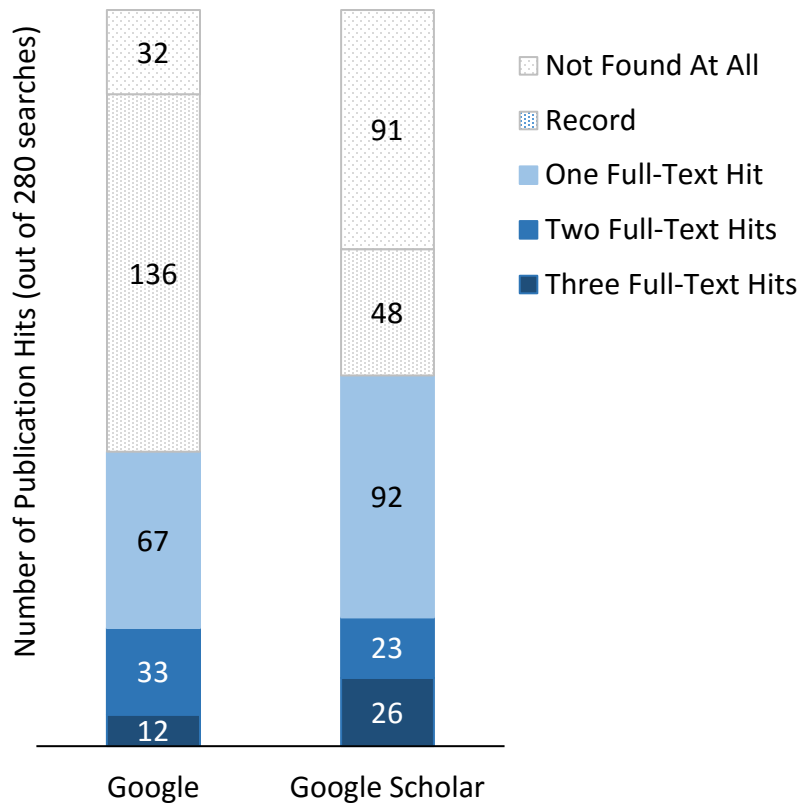
Figure 15. Full-text hits in Google and Google Scholar, all rounds, all searches.

To examine where full-text versions were found, we analyzed the hosts for full-text hits from all rounds. Google Scholar found 216 full-text hits on 25 different hosts (see Table 4), while Google found only 172 full-text hits on 29 hosts (see Table 5). The top full-text hosts in Google Scholar were Google Books (72 hits), JMU Scholarly Commons (our instance of bepress' Digital Commons institutional repository, 30 hits), and Taylor & Francis (26 hits). For Google, the top full-text hosts were Taylor & Francis (34 hits) and Google Books (23 hits). An additional 19 full-text hits were found by Google in bepress SelectedWorks and 19 in JMU

Scholarly Commons. During some rounds of searching, we were unable to access certain hosts

using Tor Browser, so these sites may be underrepresented in the data. Notably, Taylor &

Francis (www.tandfonline.com), ResearchGate (www.researchgate.net), and SAGE Journals

(journals.sagepub.com) all contributed significantly to the full-text hits found across all rounds

despite not having any results in some rounds.

Table 4. Hosts where full text was found in Google Scholar, all rounds, all searches.

Google Scholar				
		Search Type		
Host	Full Title	Keyword + LastName	Keyword	Total
books.google.*	36	26	10	72
commons.lib.jmu.edu	10	14	6	30
www.tandfonline.com	11	13	2	26
bio-diglib.biomedcentral.com	2	4	2	8
pubmedcentralcanada.ca	3	3	2	8
works.bepress.com	4	2	.	6
www.ifla.org	3	3	.	6
crl.acrl.org	3	3	.	6
journals.sagepub.com	2	4	.	6
pdfs.semanticscholar.org	2	2	1	5
www.researchgate.net	3	2	.	5
crln.acrl.org	2	2	1	5
journals.tdl.org	2	2	1	5
www.ncbi.nlm.nih.gov	2	2	.	4
www.journals.ala.org	2	2	.	4
cat.inist.fr	2	2	.	4
(9 hosts with <4 full-text hits each)	*7*	*8*	*1*	*16*
Total	**96**	**94**	**26**	**216**

* any host beginning books.google, for example books.google.fr

Table 5. Hosts where full text was found in Google, all rounds, all searches.

Google				
		Search Type		
Host	Full Title	Keyword + LastName	Keyword	Total
www.tandfonline.com	10	17	7	34
books.google.*	8	7	8	23
works.bepress.com	8	7	4	19
commons.lib.jmu.edu	4	10	5	19
www.ncbi.nlm.nih.gov	7	3	1	11
crln.acrl.org	3	4	2	9
crl.acrl.org	3	3	2	8
journals.tdl.org	2	3	.	5
journals.ala.org	1	3	1	5
www.researchgate.net	2	3	.	5
pdfs.semanticscholar.org	2	2	1	5
journals.sagepub.com	3	1	.	4
digitalcommons.murraystate.edu	2	1	1	4
(16 hosts with <4 full-text hits each)	*5*	*14*	*2*	*21*
Total	**60**	**78**	**34**	**172**

* any host beginning books.google, for example books.google.fr

Discussion

The first three sections of this discussion systematically answer the research questions laid out in the Methodology. The fourth and fifth examine implications for libraries arising from our findings and point to opportunities for future research. The final section offers suggestions for methodological development.

Records and Profiles on Scholarly Identity Websites

Of the 24 faculty in our sample, 20 had at least one ASN profile or ORCID record during this study, and 22 now have authority records. ORCID was the most popular scholarly identity website maintained by JMU librarians, with bepress SelectedWorks the second most common. This may be due to internal training and promotion of these platforms.

ORCID records, and all ASN sites included in this study except for Google Scholar, allow researchers to specify alternate names. This feature was used in the ORCID records of over half our sample, but we were unable to determine whether faculty included variant names on the other sites, as these sites do not display the alternate names entered by users. Academia, bepress, and ResearchGate mention that they use these additional names for matching scholars to their publications, although it is unclear whether they are also used to provide access when searching for profiles. Interestingly, Google Scholar and ORCID were the only sites where any faculty were found under more than one form of name.

Visibility of Records and Profiles in Google and Bing

The relationship between search engines and scholarly identity websites is complex. Using Google and Bing, only 15 of the 24 faculty in our sample were found by name across LCNAF, VIAF, ISNI, WorldCat Identities, Academia, bepress, Google Scholar, ORCID, and ResearchGate, despite the fact that 23 of them had some form of online scholarly identity (by Round 2) and 20 had ASN profiles or ORCID records (in both Rounds 1 and 2). Five of the not-found faculty had ORCID records, suggesting that merely having an ORCID record is not sufficient on its own for web discoverability. We can envision future research investigating how other online faculty behaviors (e.g., including ORCID identifiers in publications, social media use) might influence the visibility of ASN profiles and ORCID records.

Some scholarly identity websites do not seem of much interest to search engines, at least in terms of public search results displays. We do not know if Google and Bing are using them behind the scenes. Both search engines found very few faculty profiles in the top 50 results from the Library of Congress, WorldCat Identities, and Academia. Neither search engine returned results from VIAF or ISNI in the top 50, although when we added "viaf.org" or "isni.org" as a

search term in addition to the name, a result was then found for some searches. Thus, we can conclude these sites are indexed but are not ranked highly. Their indexing may also be inconsistent or incomplete.

Other sites appear more prominently in search engine results. Google and Bing did equally well at finding bepress SelectedWorks profiles. Google was also successful at finding faculty profiles from Google Scholar and ResearchGate; Bing had less success. In both Google and Bing, when ASN profiles or ORCID records were found they tended to rank highly in search results, with Google having higher average rankings than Bing.

On average, rankings of ASN profiles and ORCID records improved, and variance among rankings decreased, between rounds in both Google and Bing. Because we did not record the URLs represented by each hit (as we did with publications), we were unable to further analyze how a specific profile or record changed in rank over time. In both rounds, Bing found fewer ASNs and ORCIDs than Google.

We were unable to determine whether providing alternate forms of name in scholarly identity websites increased findability in Google and Bing. Most profiles were not found by Google or Bing under more than one form of name, except in the case of Google Scholar. It is interesting to note that Google Scholar does not accommodate variant names in its profiles, but Google searches were successful at finding Google Scholar profiles under more than one form of name for at least half of the faculty that we searched for under multiple names. We would need more data on whether faculty included alternate names in their profiles to determine whether that would result in better discovery of other ASN profiles in search engines.

Because this study was conducted in conjunction with a project to create authority records for faculty, Round 1 offered the only opportunity to compare similar groups of faculty

with and without authority records, as most of the sample had NARs by Round 2. In Round 1, although a similar percentage of profiles and records was found for both groups, Google's rankings for faculty with NARs were higher and less variable than for those without NARs. Round 2 Google results for faculty with authority records (both new and updated) were ranked similarly to faculty with NARs in Round 1. Although Bing ranked faculty with NARs about one rank lower than those without NARs in Round 1, in both search engines the average rank for all ASN profiles and ORCID records found increased from Round 1 to Round 2, when more faculty had NARs. Furthermore, those with existing NARs had ranks improve in Round 2, when they had an updated NAR. Those who had NARs created between rounds had ranks improve as well. This indicates that authority records may potentially increase search result rankings in Google and Bing.

ASN profiles and ORCID records for faculty with updated NARs ranked slightly higher than those with new NARs in Google and in Bing. The small magnitude of change leaves uncertainty about whether the existence of an authority record is more important than whether it has recently been updated.

Although Google results in particular ranked highly regardless of the number of ASN profiles or ORCID records held, faculty with a greater number of ASN profiles or ORCID records were almost twice as likely to be found in both Google and Bing than faculty with fewer profiles. If this finding is confirmed by additional research with a larger and more diverse sample, creating profiles could be an effective way for scholars to promote themselves online.

Further research should be done to investigate whether these findings are true more broadly, but these findings point to the potential value of librarians creating authority records for faculty and educating scholars about the value of creating ASN profiles themselves.

Visibility of Publications in Google and Google Scholar

Google and Google Scholar were both effective at finding most of the publications in the sample, and items typically ranked highly in both search tools. Rankings in Google were more variable over time than in Google Scholar, which could be due to the larger scope and number of items included in this search engine, and possibly differences in algorithms and frequency of indexing.

Full title and keyword+last name searches performed similarly on the number of publications found. Keyword searches in Google found slightly fewer publications than other search types, and Google Scholar's keyword searches were noticeably less successful at finding items. As expected, full title searches returned more publications as the top search result in both search tools than did either type of keyword search. The less specific nature of keyword search strings and Google Scholar's use of citation frequency in its relevance algorithm are two possible explanations for these findings.

Google Scholar found more full-text publications than Google. Both search tools were similarly effective at providing multiple pathways to full text, although more often only one full-text option was found in the top three hits. Keyword searches returned more full text in Google, while full title searches were more effective at finding full text in Google Scholar. The fact that we found full text for many items despite using Tor Browser to simulate public access supports

previous research (Jamali & Nabavi, 2015; Martín-Martín et al., 2014) that recent full text is often publicly available through search engines.[7]

Perhaps unsurprising given our sample population of librarian faculty, our bepress institutional repository and SelectedWorks were common full-text hosts. Although our faculty are making their publications available online through these venues, it remains to be seen how the breakdown of popular full-text sources would change for scholars in other disciplines with different publishing opportunities and expectations.

One of our research questions concerned the relationship between ranking of scholars' publications and the existence of their scholarly profiles and authority records. Because our publication sample was so highly ranked, the restricted range of values prevented analysis of how the existence of authors' profiles and records might influence ranking of their publications. However, the high visibility of publications and availability of full-text versions across our sample might indicate that simply posting works online is an effective strategy for increasing web visibility of scholarly publications.

Academic Visibility, Researcher Identifiers, and Authority Records

Although our sample was small and contained only library faculty, this study offers concrete evidence that authority records may increase the visibility of scholars in search engine results. We advocate for academic libraries to continue NAR projects at their institutions, and to use those projects as research opportunities to further illuminate this finding. Demonstrating a

---

[7] Martín-Martín et al.'s (2014) large-scale, longitudinal study of highly-cited articles 1950-2013 found 40% of them to be public full text, while Jamali and Nabavi (2015) found 61% of articles in their 2004–2014 sample to be public full text.

visible increase in search engine rank to scholars would support their participation in institutional identifier and authority record initiatives, and prompt institutions to devote resources toward such efforts.

The work of organizations such as ISNI, LC, ORCID, VIAF, and WorldCat to support clarification of name variants through authority control was also shown to be valuable: the for-profit scholarly identity sites do not seem to support the use of name variants, as shown by the fact that we could not find individuals using name variants on any ASN. ORCID's position as an identifier system with robust metadata is enhanced by the ability for scholars to edit their own data, giving it some visibility to public search engines (unlike the authority record databases) while encouraging emerging standards. Libraries should therefore continue to focus on ORCID in education and outreach efforts to scholars.

Institutional repositories and the scholarly profiles in associated software (such as bepress SelectedWorks) have a similar position to ORCID by providing the opportunity to combine a structured bibliographic system with the ability for researchers to upload publications and curate scholarly profiles. Previous research has found some IRs are not set up well to support digital visibility (Fagan & Willey, 2018; Orduña-Malea & López-Cózar, 2015; Yang, 2016). The results of this study show that IRs, associated scholarly profiles, and full-text publications can be ranked very highly in public search engines, competitive with for-profit websites like Google Scholar, ResearchGate, and Taylor & Francis. If disseminating research and showcasing talented faculty are priorities for institutions, SEO needs to receive more research attention. An interesting line of research could explore the extent to which the addition to IRs of identifiers and links to authority records influences the findability of scholars and their publications. When they are visible to

search engines, ORCID identifiers and IRs have the potential to increase the usefulness of underlying authority record and identifier systems to the broader public web.

Search Engine Indexing and Use of Metadata

Our exploratory research with a small sample offers several implications related to search engine indexing and use of metadata. First, for finding scholarly identity websites, we found Bing to be a less reliable search tool than Google, with worse rankings and greater variability across rounds. While we would not recommend Bing for students or scholars to use, its inclusion in bibliometric research studies offers the important ability to contrast with Google results. We did not use Bing for publication searches, but could see future studies including it as well as Microsoft Academic.

If scholars want their profiles to be findable by Google, then bepress, Google Scholar, and ResearchGate seem to be important places to create scholarly profiles. Academia and the authority record and identity websites (ISNI, LCNAF, VIAF, WorldCat Identities) did not show up much in Google results, and ORCID showed up less than half the time. This finding illuminates what sites to discuss with scholars in outreach and education efforts. While Google Scholar and ResearchGate improve digital visibility, several scholars have described problems with these for-profit sites, including predatory practices, skewed metrics, fake accounts, and lack of features to control plagiarism (Bond, 2017; Hall, 2015). However, it is important to remember that not all scholars have an affiliation with an academic library (Nicolaides, 2018), and in the developing world, even the affiliated may not have access to an IR. This study found that having more ASN profiles correlates to the likelihood of finding those profiles, suggesting that scholars eager to be visible should take the time to create multiple profiles. Future research could

investigate how faculty can boost the visibility of their ORCID records and test some of the assertions of industry experts (e.g., Crestodina, 2012) about how to increase their visibility.

This study's publication search results confirm previous research findings that academic libraries should continue to increase attention to teaching students and faculty how to use tools like Google and Google Scholar effectively (Fagan, 2017). While any type of search in Google seems to be effective, Google Scholar performs best at finding full text when the full title is used. If results aren't found, adding the last name to keywords seems important for finding known items.

Although much is still unknown about how search engines use authority records, identifiers, and ASN profiles, our findings point to several actionable recommendations and future research directions to further refine those recommendations. Due to the potential of NARs to increase visibility of scholars and the lack of authority control for name variants provided by other ASN sites, institutions should devote resources to creating NARs for their faculty and incorporating faculty identifiers into their institutional repositories. Librarians can also educate faculty on the role of ASN profiles and identifiers in increasing their online visibility and can focus their outreach by highlighting the differences in search engine visibility among ASN sites. Although most publications were found (often in full text) and ranked highly in search results, the differences observed between search tools demonstrate a continuing need to teach library users how to use these tools effectively.

Methodological Development

We identified several implications for future research and opportunities for methodological development. This study's samples were drawn from one academic department (Libraries). This approach had the advantage of including almost all faculty in the department,

suggesting that findings would apply similarly to any JMU librarian; however, the extent to which findings would be generalizable to other academic disciplines or institutions is unknown. Future research could repeat this study with similar dedication to a specific (but different) discipline, with a specific discipline at multiple institutions, or with multiple academic departments at the same institution. The sample of publications should also be expanded to include more formats and publication venues and a wider range of publication dates to better reflect the diversity of scholarship in the disciplines studied.

Due to difficulties accessing some of the profile and search engine sites with Tor Browser, we would use a different method to simulate public access and prevent our location from influencing search results. Performing searches off-campus with location services disabled or altered would provide the public user experience (Cazier, 2016). By eliminating the problem of some publication sites not being accessible via Tor Browser, we would be able to track how these various sources changed rank over time and determine whether certain hosts are more consistently visible than others. Another option would be conducting web scraping using a headless browser (e.g., Headless Chrome, Firefox's headless mode, or PhantomJS) as a more efficient way to implement comprehensive data collection for a larger sample while reducing the potential for human error and facilitating reproducibility. Such automation would make it possible to take an inductive approach in which the sites to look for would not have to be predetermined, and it would also aid later reanalysis as additional research questions arise. However, some pilot tests would need to be performed to ensure the chosen headless browser is emulating what the user would experience in a full browser.

In this study, we tried three search variants when looking for publications: full title, keyword+last name, and keyword only. Performing both full title and keyword searches is

important when conducting this type of research, not only because users may search either way (cutting and pasting full titles or typing in keywords) but also because search engines vary in which approach is more beneficial. Google found more full text for keyword searches, while Google Scholar found more full text for full title searches. However, there was not much difference between full title and keyword+last name search results, and non-librarian users are less likely to conduct keyword+last name searches, so we would recommend not including keyword+last name searches as an extra step in research protocols.

For publication searches, while we had recorded the number of hits found in the first 50 results, we did not end up using this data despite a few attempts at analysis. Because we only recorded URLs for the top three hits, if an item had more than 3 hits, we had no further information about why that might be, and therefore that challenged interpretation. In the future we would either record more information about the subsequent hits or simply not gather this data without a clearer hypothesis.

The reasons why search engines might produce different results at different times for the same searches are unclear, and further research is needed to explore this finding. Changes to study design that could improve future investigations, beyond those mentioned above, include:

- Including other academic scholarly networking sites and researcher identifiers, such as Mendeley and ResearcherID;
- Increasing the number of rounds of searching and/or the amount of time between rounds;
- In conjunction with NAR creation projects, randomly assigning faculty to groups so that some faculty have no NARs until the study is complete, to permit a more direct comparison;

- Recording URLs for each hit in search engine results for name searches, to enable analysis of how ranks changed over time; and

- Surveying faculty about activities they engage in that may affect their web visibility, such as adding publications to their institutional repository or writing blog posts.

We remain uncertain of the appropriate time period to wait between search rounds for an intervention to take effect. Waiting too long increases the influence of factors unaccounted for, but enough time must pass for search engine re-indexing. One month is our current best guess at an appropriate amount of time, based on Google's advice to webmasters (Google, 2018a).

While this study involved mostly commercial and third-party tools, in the chance that researchers have webmaster-level access to the sites where scholarly profiles or publications are hosted (e.g., institutional repositories), both Bing and Google offer sophisticated reporting tools that would provide analytics and recommendations about the indexing and SEO of specific websites and their pages (Bing, 2019; Google, 2019a; Google, 2019b). Limited SEO diagnostic information is also available to the public using Google's Lighthouse tool (2018d). The additional information gathered from these tools would complement and further illuminate search result data.

Finally, given the collation and disambiguation functions of authority records and researcher identifiers, it would be instructive to look in more depth at the forms of name scholars use online and in their publications. We do not know to what extent search engines are using identifiers and linked data to collocate and disambiguate researchers, or to match researchers with their publications. Possible questions to investigate include:

- Are scholars consistent in the forms of name used in their profiles and publications? Does the most commonly used name in a faculty member's publications and profiles match

what the faculty member told us was their preferred form of name? Do faculty understand the concept of a "preferred name"?

- Do records representing a publication use the same form of name as found on the publication itself? If not, are differences due to the website's style guide, algorithmic matching, or other factors?

- Does the form of name used in publications have an effect on search engine ranking of publications?

Conclusion

Many colleges and universities are urging faculty to create researcher identifiers and academic scholarly networking profiles and to use identifiers in publications. Some academic libraries are even creating authority records for their institution's faculty. In their role as discipline-inclusive information literacy specialists, librarians can play a strategic role on campus by creating and updating instructional materials and workshops about how to increase the visibility of scholarly profiles and publications. This study's findings suggest such activities should continue in order to increase web discoverability. Our results suggest that faculty with more ASN profiles were more visible in search engines and faculty with authority records ranked more highly in Google, although findings would need to be reproduced in other samples to support firmer conclusions. As trends in identifier use continue, more research is needed to inform librarian and faculty practice regarding authority records, researcher identifiers, and academic social networking profiles.

References

Academia. (2018a). About Academia.edu. Retrieved December 16, 2018 from

https://www.academia.edu/about

Academia. (2018b). Hiding my profile. Retrieved December 16, 2018 from

http://support.academia.edu/customer/en/portal/articles/2339531-hiding-my-profile

Akers, K. G., Sarkozy, A., Wu, W., & Slyman, A. (2016). ORCID author identifiers: A primer

for librarians. *Medical Reference Services Quarterly, 35*(2), 135-144.

doi:10.1080/02763869.2016.1152139

Allen, E. J., & Weber, R. K. (2015). An exploration of indexed and non-indexed open access

journals: Identifying metadata coding variations. *Journal of Web Librarianship, 9*(2-3), 65-

84. doi:10.1080/19322909.2015.1020185

Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low

indexing ratios of IRs in Google Scholar. *Library Hi Tech, 30*(1), 60-81.

doi:10.1108/07378831211213210

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond

citations: Scholars' visibility on the social web. 17th International Conference on Science

and Technology Indicators, Montreal, Canada, September 5-8. Retrieved December 10,

2018 from https://arxiv.org/abs/1205.5611

Beel, J., Gipp, B., & Wilde, E. (2010). Academic search engine optimization (ASEO) optimizing

scholarly literature for Google Scholar & co. *Journal of Scholarly Publishing, 41*(2), 176-

190. https://doi.org/10.3138/jsp.41.2.176

Bik, H. M., & Goldstein, M. C. (2013). An introduction to social media for scientists. *PLoS Biology, 11*(4). Retrieved December 10, 2018 from

https://doi.org/10.1371/journal.pbio.1001535

Bing. (2019). Webmaster help & how-to. Retrieved February 24, 2019 from

https://www.bing.com/webmaster/help/getting-started-checklist-66a806de

Bond, S. (2017, January 23). Dear scholars, delete your account at Academia.edu. *Forbes,* Retrieved December 10, 2018 from

https://www.forbes.com/sites/drsarahbond/2017/01/23/dear-scholars-delete-your-account-at-academia-edu/#6ea426cb2d62

Brown, J., Oyler, C., & Haak, L. (2015). ORCID adoption and integration program report. Retrieved December 10, 2018 from https://doi.org/10.6084/m9.figshare.1290632.v1

Cazier, C. (2016). How to localize Google search results. Retrieved December 10, 2018 from

http://searchengineland.com/localize-google-search-results-239768

Ciccone, K., & Vickery, J. (2015). Summon, EBSCO Discovery Service, and Google Scholar: A comparison of search performance using user queries. *Evidence Based Library & Information Practice, 10*(1), 34-49.

Clark, J. A., & Young, S. W. (2017). Linked data is people: building a knowledge graph to reshape the library staff directory. *Code4Lib Journal*, 36. Retrieved December 10, 2018 from http://journal.code4lib.org/articles/12320

Crestodina, A. (2012). Personal SEO: 14-point checklist to dominate your personal brand on Google. Retrieved February 24, 2019 from https://neilpatel.com/blog/personal-branding-seo/

Dagienė, E., & Krapavickaitė, D. (2016). How researchers manage their academic activities. *Learned Publishing, 29*(3): 155-163. https://doi.org/10.1002/leap.1030

Dawson, A., & Hamilton, V. (2006). Optimising metadata to make high-value content more accessible to Google users. *Journal of Documentation*, 62(3), 307-327.

DeYoung, G. (2007). Eleven tips for optimizing PDFs for search engines. Retrieved December 10, 2018 from http://searchengineland.com/eleven-tips-for-optimizing-pdfs-for-search-engines-12156

Ebrahim, N. A., Salehi, H., Embi, M. A., Habibi, F., Gholizadeh, H., Motahar, S. M., & Ordi, A. (2013). Effective strategies for increasing citation frequency. *International Education Studies, 6*(11), 93-99. Retrieved December 10, 2018 from https://ssrn.com/abstract=2344585

Elsevier Biggerbrains. (2012). Get found — Optimize your research articles for search engines. Retrieved December 18, 2018 from https://www.elsevier.com/connect/get-found-optimize-your-research-articles-for-search-engines

Fagan, J. C. (2017). An evidence-based review of academic web search engines, 2014-2016: Implications for librarians' practice and research agenda. *Information Technology and Libraries, 36*(2). Retrieved December 10, 2018 from https://doi.org/10.6017/ital.v36i2.9718

Fagan, J. C., & Willey, M. (2018). The discoverability of award-winning undergraduate research in history: Implications for academic libraries. *College & Undergraduate Libraries, 25*(2), 164-186. https://doi.org/10.1080/10691316.2018.1456994

Fitzpatrick, K. (2015, October 26). Academia, not edu [Blog post]. Retrieved December 16, 2018 from https://kfitz.info/academia-not-edu/

Google. (2011). PDFs in Google search results. Retrieved December 18, 2018 from

 https://webmasters.googleblog.com/2011/09/pdfs-in-google-search-results.html

Google. (2019a). Google search console. Retrieved February 24, 2019 from

 https://search.google.com/search-console/welcome

Google. (2019b). Search engine optimization (SEO) starter guide. Retrieved February 24, 2019

 from https://support.google.com/webmasters/answer/7451184?hl=en

Google. (2018a). Ask Google to recrawl your URLs. Retrieved December 15, 2018 from

 https://support.google.com/webmasters/answer/6065812?hl=en

Google. (2018b). How Google search works. Retrieved December 18, 2018 from

 https://support.google.com/webmasters/answer/34439?hl=en

Google. (2018c). Inclusion guidelines for webmasters. Retrieved December 15, 2018 from

 https://scholar.google.com/intl/en/scholar/inclusion.html#content

Google. (2018d). Lighthouse. Retrieved February 24, 2019 from

 https://developers.google.com/web/tools/lighthouse/

Google. (2018e). Search tips: Content Coverage. Retrieved December 14, 2018 from

 https://scholar.google.com/intl/en/scholar/help.html#coverage

Griffin, J. (2016). Google: We index PDF's just like any other webpage. Retrieved December 10,

 2018 from https://www.thewebmaster.com/seo/2016/feb/24/google-we-index-pdfs-just-like-

 any-other-web-page/

Hall, G. (2015, October 18). Does Academia.edu mean open access is becoming irrelevant?

 Media Gifts, Retrieved December 10, 2018 from

 http://www.garyhall.info/journal/2015/10/18/does-academiaedu-mean-open-access-is-

 becoming-irrelevant.html

Hickey, T. B., & Toves, J. A. (2014). Managing ambiguity in VIAF. *D-Lib Magazine, 20*(7/8). doi:10.1045/july2014-hickey

ISNI. (2018a). International standard name identifier. Retrieved December 16, 2018 from http://www.isni.org/

ISNI. (2018b). Contributors providing data via VIAF. Retrieved December 16, 2018 from http://www.isni.org/sites/default/files/data-contributors-via-viaf.pdf

ISNI. (2018c). How ISNI works. Retrieved December 16, 2018 from http://www.isni.org/how-isni-works

ISNI. (2018d). ISNI and ORCID. Retrieved December 16, 2018 from https://web.archive.org/web/20130304210352/http://www.isni.org/isni_and_orcid

Jamali, H. R., & Nabavi, M. (2015). Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics, 105*(3): 1635-1651. https://doi.org/10.1007/s11192-015-1642-2

Katumba, S., & Coetzee, S. (2017). Employing search engine optimization (SEO) techniques for improving the discovery of geospatial resources on the web. *ISPRS International Journal of Geo-Information, 6*(9), 284. https://doi.org/10.3390/ijgi6090284

Kieńć, W. (2014). Why and how should you optimize academic articles for search engines? Retrieved December 10, 2018 from http://openscience.com/optimize-academic-articles-search-engines/

Kramer, B., & Bosman, J. (2016). Innovations in scholarly communication. Retrieved December 10, 2018 from https://101innovations.wordpress.com/

Library of Congress. (2018). Library of Congress names. Retrieved December 16, 2018 from http://id.loc.gov/authorities/names.html

Lotfipanah, M. (2016). Does search engine optimisation (SEO) increase scientific journals visibility, prestige, and impact factor as a new method? *International Journal of Knowledge Management and Practices, 4*(1). Retrieved December 10, 2018 from https://ssrn.com/abstract=3042779

Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics, 101*(1), 337-356. Retrieved December 10, 2018 from https://link.springer.com/article/10.1007/s11192-014-1345-0

Matthews, D. (2016, April 7). Do academic social networks share academics' interests? Retrieved December 10, 2018 from https://www.timeshighereducation.com/features/do-academic-social-networks-share-academics-interests

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? Granada: EC3 Working Papers, 18, March 25. Retrieved December 14, 2018 from https://arxiv.org/pdf/1410.8464

Mauvais-Jarvis, F. (2016). Developing academic visibility in the medical sciences. *The Ochsner Journal, 16*(3), 208-209.

Meadows, A. (2017, April 20). It takes a village: One year of journals requiring ORCID iDs. Retrieved December 10, 2018 from https://scholarlykitchen.sspnet.org/2017/04/20/takes-village-one-year-journals-requiring-orcid-ids/

Nicolaides, B. (2018, August 6). Locked out: Research access as a challenge for the discipline. *Perspectives on History.* Retrieved December 10, 2018 from https://www.historians.org/publications-and-directories/perspectives-on-history/september-2018/locked-out-research-access-as-a-challenge-for-the-discipline

OCLC. (2018a). Worldcat.org frequently asked questions. Retrieved December 14, 2018 from

https://www.oclc.org/support/services/worldcat-org/faq.en.html

OCLC. (2018b). WorldCat Identities. Retrieved December 15, 2018 from

https://www.oclc.org/research/themes/data-science/identities.html.

Onaifo, D., & Rasmussen, D. (2013). Increasing libraries' content findability on the web with

search engine optimization. *Library Hi Tech, 31*(1), 87-108.

doi:10.1108/07378831311303958

ORCID. (2017). Requiring ORCID in publication workflows: Open letter. Retrieved December

12, 2018 from https://orcid.org/content/requiring-orcid-publication-workflows-open-letter

ORCID. (2018a). Trademark and iD display guidelines. Retrieved December 17 from

https://orcid.org/trademark-and-id-display-guidelines#02-c

ORCID. (2018b). What is ORCID. Retrieved December 18, 2018 from https://orcid.org/about

ORCID. (2018c). Visibility settings. Retrieved December 18, 2018 from

https://support.orcid.org/hc/en-us/articles/360006897614-Visibility-settings

Orduña-Malea, E., & López-Cózar, E. D. (2015). The dark side of open access in Google and

Google Scholar: The case of Latin-American repositories. *Scientometrics, 102*(1), 829-846.

Pitol, S. P., & De Groote, S. L. (2014). Google Scholar versions: Do more versions of an article

mean greater impact? *Library Hi Tech, 32*(4), 594-611. doi:10.1108/LHT-05-2014-0039

Program for Cooperative Cataloging (PCC). (2016). *Charge for PCC task group on identity

management in NACO.* Retrieved December 12, 2018 from

https://www.loc.gov/aba/pcc/documents/Identity-management-NACO-PCC-TG.pdf

ResearchGate. (2018a). About. Retrieved December 12, 2018 from

https://www.researchgate.net/about

ResearchGate. (2018b). Profile information and visibility. Retrieved December 16, 2018 from

https://explore.researchgate.net/display/support/Profile+information+and+visibility

Rosenzweig, M., & Schnitzer, A. E. (2015). An initiative to address name ambiguity:

Implementing ORCID at a large academic institution. *College & Research Libraries News,*

76(5), 260-264. https://doi.org/10.5860/crln.76.5.9312

Schema.org (2019). Person. Retrieved February 24, 2019 from https://schema.org/Person

Schubert, C., & Holloway, S. W. (2014, May 14). Give yourself an ORCID: Boosting your

visibility through researcher profiling networks. Presentation at James Madison University

Center for Faculty Innovation May Symposium, Harrisonburg, VA. Retrieved from

https://works.bepress.com/steven_holloway/1/

Swiontkowski, M. (2016). ORCID registration required in 2017. *Journal of Bone & Joint*

Surgery, 98(24), 2035. doi:10.2106/JBJS.16.01165

Thomas, W. J., Chen, B., & Clement, G. (2015). ORCID identifiers: Planned and potential uses

by associations, publishers, and librarians. The Serials Librarian, 68(1-4), 332-341.

https://doi.org/10.1080/0361526X.2015.1017713

Thomsen, C. (2015). 5 tips for promoting your research through Facebook. Retrieved February

24, 2019 from https://www.wiley.com/network/researchers/promoting-your-article/5-tips-

for-promoting-your-research-through-facebook

Tor. (2018). Tor FAQ: What is Tor. Retrieved December 15, 2018 from

https://www.torproject.org/docs/faq.html.en#WhatIsTor

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature,*

512(7513), 126-129. doi:10.1038/512126a

Ward, J., Bejarano, W., & Dudás, A. (2015). Scholarly social media profiles and libraries: A review. *Liber Quarterly, 24*(4), 174-204.

Ward, A., & Guest, R. (2013, March 1). Making the most of social media. Retrieved February 24, 2019 from https://www.elsevier.com/authors-update/story/social-media/making-the-most-of-social-media

Wikipedia. (2017). Wikipedia: Authority control. Retrieved December 12, 2018 from https://en.wikipedia.org/wiki/Wikipedia:Authority_control

Wu, S., Stvilia, B., & Lee, D. J. (2017). Readers, personal record managers, and community members: An exploratory study of researchers' participation in online research information management systems. *Journal of Library Metadata, 17*(2):1-34. https://doi.org/10.1080/19386389.2017.1348783

Yang, L. (2016). Metadata effectiveness in internet discovery: An analysis of digital collection metadata elements and internet search engine keywords. *College & Research Libraries, 77*(1), 7-19. doi:10.5860/crl.77.1.7

Zhang, J., & Dimitroff, A. (2005). The impact of metadata implementation on webpage visibility in search engine results (part II). *Information Processing & Management, 41*(3), 691-715. doi:10.1016/j.ipm.2003.12.002

Appendix: Authority Files, Researcher Identifiers, and Academic Social Networking Platforms

Library of Congress Name Authority File

The Library of Congress Name Authority File (LCNAF) exists to establish authorized forms of names for "persons, organizations, events, places, and titles" for "uniform access to bibliographic resources" (Library of Congress, 2018). Authority records are created by the Library of Congress and libraries that are members of the Program for Cooperative Cataloging's Name Authority Cooperative Program. Records in the LCNAF can be viewed through Library of Congress Authorities (https://authorities.loc.gov/) and are also published as linked data through the LC Linked Data Service (http://id.loc.gov/).

Name authority records establish the authorized, or preferred, form of a name. A record also contains references to variant forms of the name and citations recording where the various names were found. Authority records may also contain a variety of additional elements that further assist in disambiguating the person from others with the same or similar names; such elements may include dates associated with the person, occupation, field of activity, associated places or groups, language, and other identifiers.

Virtual International Authority File

The Virtual International Authority File (VIAF) is a shared international authority file maintained by OCLC with contributions from over 50 national libraries, research organizations, and cultural institutions. The Library of Congress and ISNI are two contributors to VIAF.

Entities from the contributed authority files are matched and assigned a VIAF ID. The file is updated with new data monthly, and records are re-clustered at that time (Hickey & Toves, 2014). A VIAF record includes the authorized heading used in each authority file, along with

variant forms of name and additional data contributed from the various sources. The web display of a VIAF record includes links to the record in each of the source authority files, aggregated data on selected co-authors, and publication statistics such as a timeline of publication history, countries and regions of publication, and selected publishers. The page also provides links to other profiles, including ORCID and WorldCat Identities.

WorldCat Identities

WorldCat Identities uses information from the WorldCat bibliographic database to create web pages for people, organizations, and subjects. Names are algorithmically matched to authority records from the LCNAF, and Identities pages are also created for non-authoritative names in WorldCat. Although WorldCat Identities pages are not authority records or identifiers, their aggregation of bibliographic and authority data for web presentation makes them of interest to this study. WorldCat Identities is updated quarterly (OCLC, 2018b).

A WorldCat Identities page for a person includes the primary and alternative forms of their name as taken from bibliographic and authority records. Pages are focused on aggregating information about the person's publications. They include the number of works and publications, a timeline of publications by and about the person, a list of the most widely held works, and the languages, genres, and roles in which the person has published. A word cloud of associated subjects and a list of related Identities (e.g., co-authors and other collaborators) are also included. A "Useful Links" section provides links to authority records in the LCNAF and VIAF. Icons indicating whether the profile is based on a controlled vocabulary or is derived solely from bibliographic data are visible in the list of search results but absent from individual profile pages.

International Standard Name Identifier

The International Standard Name Identifier (ISNI) is an ISO standard number used to identify "contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, and more" (ISNI, 2018a). ISNI data is aggregated from 47 sources, including national libraries, professional societies, and rights management and publishing organizations. VIAF is one of these data sources, and Library of Congress data is contributed through VIAF (ISNI, 2018b). ISNIs can only be created through ISNI Registration Agencies, although anyone may contribute "enrichments, clarifications and corrections" to the database (ISNI, 2018c). Data from multiple sources is matched algorithmically but the process is not completely automated. Identifiers "are assigned when there is a high level of confidence in matching new names to existing names in the database" (ISNI, 2018c), so the presence of an individual in an ISNI data source like VIAF does not mean that they are automatically assigned an ISNI.

An ISNI record contains the identifier (a 16-digit number), multiple forms of name without preference for any one form, and additional information on the person as provided from various data sources. This additional information may include associated dates, types of materials created, roles, publication titles, related names, and free-text notes.

ORCID

ORCID is a nonprofit organization dedicated to ensuring that "all who participate in research, scholarship and innovation are uniquely identified and connected to their contributions and affiliations" (ORCID, 2018b). It is primarily an identifier system that also provides a web platform where researchers can link their publications to their ORCID and enter profile information. ORCID identifiers comply with the ISNI standard (ISNI, 2018d). In addition to the

assigned identifier, ORCID records contain user-editable fields for alternate names ("Also known as"), Country, Keywords, Websites, Emails, Biography, and other IDs. Users can list educational affiliations, dates, and degrees; employment history; and "Works," which can be added via bulk import, manually, or linked via searches of numerous other vendors and services (e.g., DataCite, Scopus). The ORCID identifier is publicly visible even if the record is not, and users can control the visibility of all other profile elements independently (ORCID, 2018c).

Although ORCID defines itself as an identifier system, we grouped it with academic social networks because the information is controlled by the individual; the type of metadata the individual can include and the way it is displayed is similar to other ASNs.

Google Scholar

Google Scholar is the most familiar and most used academic social networking site, according to a 2014 Nature survey (Van Noorden, 2014). The platform presents itself primarily as a search engine for scholarly articles with an added profile feature. Google Scholar provides links to publications on a variety of commercial, open access, and personal websites, provided the content is scholarly in nature and includes an abstract or full text (Google, 2018c). Thus, unlike other ASN sites, users do not include publications directly. Some opportunities, biases, and limitations of Google Scholar were described in a literature review by Fagan (2017).

The Google Scholar profile is by far the shortest of these sites, including free-text fields for one name and affiliation, areas of interest, email address, a homepage URL, and a checkbox to make the profile public or private. Scholars can add publications and identify co-authors using a combination of manual and automatic features.

bepress SelectedWorks

bepress is a private company that produces the institutional repository software Digital Commons as well as SelectedWorks, where scholars can create profiles with biographical and bibliographic information. Although Digital Commons and SelectedWorks are separate sites, individuals can import any of their works from their institution's Digital Commons repository to their SelectedWorks profile, in addition to uploading publications directly to SelectedWorks.

SelectedWorks profiles include the user's name and as many alternate names as they wish, using prefix, first name, middle initial, last name, and suffix fields. Users can set their preferred name and upload a CV. Additional fields include positions held, academic disciplines, research interests, grants, activities and affiliations, honors and awards, courses, educational history, contact information, links to other sites, and ORCID identifier. Users can make their entire profile totally public or private and can add others to their account as "delegates" to help them manage their information.

Academia

Academia describes itself as "a platform for academics to share research papers" and its mission is "to make every scholarly and scientific paper available for free on the internet and to enhance academic discussion and collaboration" (Academia, 2018a). Users can create a profile, upload their work, and use social networking features to connect with other researchers. Despite its .edu domain, which was registered before regulations required educational affiliation, Academia is a for-profit company (Fitzpatrick, 2015).

Academia profiles contain fields for first, middle, and last names, a profile picture, affiliations, biography, email address, phone number, links to social media profiles and other websites, research interests, and a CV. Papers can be uploaded en masse using BibTeX and

organized into sections. Users can add name variants, which Academia says it uses to identify publications and other "Mentions" of the author on the site. Users choose their institutional and department affiliations by selecting from a list of existing names or adding a new name. Users who register an affiliation are given a URL such as https://jmu.academia.edu/<name>; otherwise the URL takes the form https://independent.academia.edu/<name>. Academia profiles cannot be made private or hidden from search engines (Academia, 2018b).

ResearchGate

ResearchGate's mission is "to connect the world of science and make research open to all", and the for-profit company describes itself as "the professional network for scientists and researchers" (ResearchGate, 2018a). The Innovations in Scholarly Communication global survey found that ResearchGate was used by twice as many published scholars as Academia (Kramer & Bosman, 2016, as cited in Matthews, 2016).

Upon creating an account, ResearchGate shows the user sets of publications with possible matches to their names and options to identify as that researcher or not. The user then selects academic disciplines, skills and expertise, and a profile picture. After the initial account creation, the user can add an ORCID identifier, affiliations, experience, education, achievements, and social networking accounts. The ORCID identifier does not seem to display or to be connected with publications within ResearchGate. Currently it is not possible to upload a CV document, as it is with Academia and bepress SelectedWorks (ResearchGate, 2018b). The user can disable their profile's search engine visibility and can customize how much information is displayed publicly. Users can enter multiple alternate names, which ResearchGate says it uses to suggest possible author matches.