An Evidence-Based Review of Academic Web Search Engines, 2014-2016: Implications for Librarians' Practice and Research Agenda

Jody Condit Fagan

ABSTRACT

Academic web search engines have become central to scholarly research. While the fitness of Google Scholar for research purposes has been examined repeatedly, Microsoft Academic and Google Books have not received much attention. Recent studies have much to tell us about Google Scholar's coverage of the sciences and its utility for evaluating researcher impact. But other aspects have been understudied, such as coverage of the arts and humanities, books, and non-Western, non-English publications. User research has also tapered off. A small number of articles hint at the opportunity for librarians to become expert advisors concerning scholarly communication made possible or enhanced by these platforms. This article seeks to summarize research concerning Google Scholar, Google Books, and Microsoft Academic from the past three years with a mind to informing practice and setting a research agenda. Selected literature from earlier time periods is included to illuminate key findings and to help shape the proposed research agenda, especially in understudied areas.

INTRODUCTION

Recent Pew Internet surveys indicate an overwhelming majority of American adults see themselves as lifelong learners who like to "gather as much information as [they] can" when they encounter something unfamiliar (Horrigan 2016). Although significant barriers to access remain, the open access movement and search engine giants have made full text more available than ever.[1] The general public may not begin with an academic search engine, but Google may direct them to Google Scholar or Google Books. Within academia, students and faculty rely heavily on academic web search engines (especially Google Scholar) for research; among academic researchers in high-income areas, academic search engines recently surpassed abstracts & indexes as a starting place for research (Inger and Gardner 2016, 85, Fig. 4). Given these trends, academic librarians have a professional obligation to understand the role of academic web search engines as part of the research process.

Jody Condit Fagan (faganjc@jmu.edu) is Professor and Director of Technology, James Madison University, Harrisonburg, VA.

[1] Khabsa and Giles estimate "almost 1 in 4 of web accessible scholarly documents are freely and publicly available" (2014, 5).

https://doi.org/10.6017/ital.v36i2.9718

Two recent events also point to the need for a review of research. Legal decisions in 2016 confirmed Google's right to make copies of books for its index without paying or even obtaining permission from copyright holders, solidifying the company's opportunity to shape the online experience with respect to books. Meanwhile, Microsoft rebooted their academic web search engine, now called Microsoft Academic. At the same time, information scientists, librarians, and other academics conducted research into the performance and utility of academic web search engines. This article seeks to review the last three years of research concerning academic web search engines, make recommendations related to the practice of librarianship, and propose a research agenda.

METHODOLOGY

A literature review was conducted to find articles, conference presentations, and books about the use or utility of Google Books, Google Scholar, and Microsoft Academic for scholarly use, including comparisons with other search tools. Because of the pace of technological change, the focus was on recent studies (2014 through 2016, inclusive).

A search was conducted on "Google Books" in EBSCO's Library and Information Science and Technology Abstracts (LISTA) on December 19, 2016, limited to 2014-2016. Of the 46 results found, most were related to legal activity. Only four items related to the tool's use for research. These four titles were entered into Google Scholar to look for citing references, but no additional relevant citations were found. In the relevant articles found, the literature reviews testified to the general lack of studies of Google Books as a research tool (Abrizah and Thelwall 2014; Weiss 2016) with a few exceptions concerning early reviews of metadata, scanning, and coverage problems (Weiss 2016). A search on "Google Books" in combination with "evaluation OR review OR comparison" was also submitted to JMU's discovery service,[2] limited to 2014-2016 in combination with the terms. Forty-nine items were found and from these, three relevant citations were added; these were also entered into Google Scholar to look for citing references. However, no additional relevant citations were found. Thus, a total of seven citations from 2014-2016 were found with relevant information concerning Google Books. Earlier citations from the articles' bibliographies were also reviewed when research was based on previous work, and to inform the development of a fuller research agenda.

A search on "Microsoft Academic" in LISTA on February 3, 2017 netted fourteen citations from 2014-2016. Only seven seemed to focus on evaluation of the tool for research purposes. A search on "Microsoft Academic" in combination with terms "evaluation OR review OR comparison" was also submitted to JMU's discovery service, limited to 2014-2016. Eighteen items were found but no additional citations were added, either because they had already been found or were not relevant. The seven titles found in LISTA were searched in Google Scholar for citing references; four additional relevant citations were found, plus a paper relevant to Google Scholar not

[2] JMU's version of EBSCO Discovery Service contained 453,754,281 items at the time of writing and is carefully vetted to contain items of curricular relevance to the JMU community (Fagan and Gaines 2016).

previously discovered (Weideman 2015). Thus, a total of eleven citations were found with relevant information for this review concerning Microsoft Academic. Because of this small number, several articles prior to 2014 were included in this review for historical context.

An initial search was performed on "Google Scholar" in LISTA on November 19, 2016, limited to 2014-2016. This netted 159 results, of which 24 items were relevant. A search on "Google Scholar" in combination with terms "evaluation OR review OR comparison" was also submitted to JMU's discovery tool limited to 2014-2016, and eleven relevant citations were added. Items older than 2014 that were repeatedly cited or that formed the basis of recent research were retrieved for historical context. Finally, relevant articles were submitted to Google Scholar, which netted an additional 41 relevant citations. Altogether, 70 citations were found to articles with relevant information for this review concerning Google Scholar in 2014-2016. Readers interested in literature reviews covering Google Scholar studies prior to 2014 are directed to (Gray et al. 2012; Erb and Sica 2015; Harzing and Alakangas 2016b).

FINDINGS

Google Books

Google Books (https://books.google.com) contains about 30 million books, approaching the Library of Congress's 37 million, but far shy of Google's estimate of 130 million books in existence (Wu 2015), which Google intends to continue indexing (Jackson 2010). Content in Google Books includes publisher-supplied, self-published, and author-supplied content (Harper 2016) as well as the results of the famous Google Books Library Project. Started in December 2004 as the "Google Print" project,[3] the project involved over 40 libraries digitizing works from their collections, with Google indexing and performing OCR to make them available in Google Books (Weiss 2016; Mays 2015). Scholars have noted many errors with Google Books metadata, including misspellings, inaccurate dates, and inaccurate subject classifications (Harper 2016; Weiss 2016). Google does not release information about the database's coverage, including which books are indexed or which libraries' collections are included (Abrizah and Thelwall 2014). Researchers have suggested the database covers mostly U.S. and English-language books (Abrizah and Thelwall 2014; Weiss 2016). The conveniences of Google Books include limits by the type of book availability (e.g. free e-books vs. Google e-books), document type, and date. The detail view of a book allows magnification, hyperlinked tables of contents, buying and "Find in a Library" options, "My Library," and user history (Whitmer 2015). Google Books also offers textbook rental (Harper 2016) and limited print-on-demand services for out-of-print books (Mays 2015; Boumenot 2015).

In April 2016, the Supreme Court affirmed Google's right to make copies for its index without paying or even obtaining permission from copyright holders (Authors Guild 2016; Los Angeles Times 2016). Scanning of library books and "snippet view" was deemed fair use: "The purpose of the copying is highly transformative, the public display of text is limited, and the revelations do

[3] https://www.google.com/googlebooks/about/history.html

not provide a significant market substitute for the protected aspects of the originals" (U.S. Court of Appeals for the Second Circuit 2015).

Literature concerning high-level implications of Google Books suggests the tool is having a profound effect on research and scholarship. The tool has been credited for serving as "a huge laboratory" for indexing, interpretation, working with document image repositories, and other activities (Jones 2010). At the same time, the academic community has expressed concerns about Google Books's effects on social justice and how its full-text search capability may change the very nature of discovery (Hoffmann 2014; Hoffmann 2016; Szpiech 2014). One study found that books are far more prevalently cited in Wikipedia than are research articles (Kousha and Thelwall 2017). Yet investigations of Google Books' coverage and utility as a research tool seem to be sorely lacking. As Weiss noted, "no critical studies seem to exist on the effect that Google Books might have on the contemporary reference experience" (Weiss 2016, 293). Furthermore, no information was found concerning how many users are taking advantage of Google Books; the tool was noticeably absent from surveys such as (Inger and Gardner's (2016) and from research centers such as the Pew Internet Research Project.

In a largely descriptive review, Harper (2016) bemoaned Google Books' lack of integration with link resolvers and discovery tools, and judged it lacking in relevant material for the health sciences, because so much of the content is older. She also noted the majority of books scanned are in English, which could skew scholarship. The non-English skew of Google Books was also lamented by Weiss, who noted an "underrepresentation of Spanish and overestimation of French and German (or even Japanese for that matter)" especially as compared to the number of Spanish speakers in the United States (Weiss 2016, 286-306).

Whitmer (2015) and Mays (2015) provided practical information about how Google Books can be used as a reference tool. Whitmer presented major Google Books features and challenged librarians to teach Google Books during library instruction. Mays conducted a cursory search on the 1871 Chicago Fire and described the primary documents she retrieved as "pure gold," including records of city council meetings, notes from insurance companies, reports from relief societies, church sermons on the fire, and personal memoirs (Mays 2015, 22). Mays also described Google Books as a godsend to genealogists for finding local records (e.g. police departments, labor unions, public schools). In her experience, the geographic regions surrounding the forty participating Google Books Library Project libraries are "better represented than other areas" (Mays 2015, 25). Mays concludes, "Its poor indexing and search capabilities are overshadowed by the ease of its fulltext search capabilities and the wonderful ephemera that enriches its holdings far beyond mere 'books'" (Mays 2015, 26).

Abrizah and Thelwall (2014) investigated whether Google Books and Google Scholar provided "good impact data for books published in non-Western countries." They used a comprehensive list of arts, humanities, and social sciences books (n=1,357) from the five main university presses in

Malaysia 1961-2013. They found only 23% of the books were cited in Google Books[4] and 37% in Google Scholar (p. 2502). The overlap was small: only 15% were cited in both Google Scholar and Google Books. English-language books were more likely to be cited in Google Books; 40% of English language books were cited versus 16% Malay. Examining the top 20 books cited in Google Books, researchers found them to be mostly written in English (95% in Google Books vs 29% in the sample), and published by University of Malaysia Press (60% in Google Books vs 26% in the sample) (2505). The authors concluded that due to the low overlap between Google Scholar and Google Books, searching both engines was required to find the most citations to academic books.

Kousha and Thelwall (2015; 2011) compared Google Books with Thomson Reuters Book Citation Index (BKCI) to examine its suitability for scholarly impact assessment and found Google Books to have a clear advantage over BKCI in the total number of citations found within the arts and humanities, but not for the social sciences or sciences. They advised combining results from BKCI with Google Books when performing research impact assessment for the arts and humanities and social sciences, but not using Google Books for the sciences, "because of the lower regard for books among scientists and the lower proportion of Google Books citations compared to BKCI citations for science and medicine" (Kousha and Thelwall 2015, 317).

Microsoft Academic

Microsoft Academic (https://academic.microsoft.com) is an entirely new software product as of 2016. Therefore, the studies cited prior to 2016 refer to entirely different search engines than the one currently available. However, a historical account of the tool and reviewers' opinions was deemed helpful for informing a fuller picture of academic web search engines and pointing to a research agenda.

Microsoft Academic was born as Windows Live Academic in 2006 (Carlson 2006), was renamed Live Search Academic after a first year of struggle (Jacsó 2008), and was scrapped two years later after the company recognized it did not have sufficient development support in the United States (Jacsó 2011). Microsoft Asia Research Group launched a beta tool called Libra in 2009, which redirected to the "Microsoft Academic Search" service by 2011. Early reviews of the 2011 edition of Microsoft Academic Search were promising, although the tool clearly lacked the quantity of data searched by Google Scholar (Jacsó 2011; Hands 2012).

There were a few studies involving Microsoft Academic Search in 2014. Ortega and Aguillo (2014) compared Microsoft Academic Search and Google Scholar Citations for research evaluation and concluded "Microsoft Academic Search is better for disciplinary studies than for analyses at institutional and individual levels. On the other hand, Google Scholar Citations is a good tool for individual assessment because it draws on a wider variety of documents and citations" (1155).

[4] Google Books does not support citation searching; the researchers searched for the book title to manually find citations to a book.

As part of a comparative investigation of an automatic method for citation snowballing using Microsoft Academic Search, Choong et al. (2014) manually searched for a sample of 949 citations to journal or conference articles cited from 20 systematic reviews. They found Microsoft Academic Search contained 78% of the cited articles and noted its utility for testing automated methods due to its free API and no blocks to automated access. The researchers also tested their method against Google Scholar, but noted "computer-access restrictions prevented a robust comparison" (n.p.). Also in 2014, Orduna-Malea et al. (2014) attempted a longitudinal study of disciplines, journals, and organizations in Microsoft Academic Search only to find the database had not been updated since 2013. Furthermore they found the indexing to be incomplete and still in process, meaning Microsoft Academic Search's presentation of information about any particular publication, organization, or author was distorted.

Despite this finding, MAS was included in two studies of scholar profiles. Ortega (2015) compared scholar profiles across Google Scholar, Microsoft Academic Search, Research Gate, Academia.edu, and Mendeley, and found little overlap across the sites. They also found social and usage indicators did not consistently correlate with bibliometric indicators, except on the ResearchGate platform. Social and usage indicators were "influenced by their own social sites," while bibliometric indicators seemed more stable across all services (13). Ward et al. (2015) still included Microsoft Academic Search in their discussion of scholarly profiles as part of the social media network, noting Microsoft Academic Search was painfully time-consuming to work with in terms of consolidating data, correcting items, and adding missing items.

In September 2016, Hug et al. demonstrated the utility of the new Microsoft Academic API by conducting a comparative evaluation of normalized data from Microsoft Academic and Scopus (Hug, Ochsner, and Braendle 2016). They noted Microsoft Academic has "grown massively from 83 million publication records in 2015 to 140 million in 2016" (10). The Microsoft Academic API offers rich, structured metadata with the exception of document type. They found all attributes containing text were normalized and that identifiers were available for all entities, including references, supporting bibliometricians' needs for data retrieval, handling, and processing. In addition to the lack of document type, the researchers also found the "fields of study" to be too granular and dynamic, and their hierarchies incoherent. They also desired the ability to use the DOI to build API requests. Nevertheless, the advantages of Microsoft Academic's metadata and API retrieval suggested to Hug et al. that Microsoft Academic was superior to Google Scholar for calculating research impact indicators and bibliometrics in general.

In October 2016, Harzing and Alakangas compared publication and citation coverage of the new Microsoft Academic with Google Scholar, Scopus, and Web of Science using a sample of 145 academics at the University of Melbourne (Harzing and Alakangas 2016a) including observations from 20-40 faculty each in the humanities, social sciences, engineering, sciences, and life sciences. They discovered Microsoft Academic had improved substantially since their previous study (Harzing 2016b), increasing 9.6% for a comparison sample in comparison with 1.4%, 2%, and 1.7% growth in Google Scholar, Scopus, and Web of Science (n.p.). The researchers noted a few

problems with data quality, "although the Microsoft Academic team have indicated they are working on a resolution" (n.p.). On average, the researchers found that Microsoft Academic found 59% as many citations as Google Scholar, 97% as many citations as Scopus, and 108% as many citations as Web of Science. Google Scholar had the top counts for each disciplinary area, followed by Scopus except in the social sciences and humanities, where Microsoft Academic ranked second. The researchers explained that Microsoft Academic "only includes citation records if it can validate both citing and cited papers as credible," as established through a machine-learning-based system, and discussed an emerging metric of "estimated citation count" also provided by Microsoft Academic. The researchers concluded that Microsoft Academic is promising to be "an excellent alternative for citation analysis" and suggested Microsoft should work to improve coverage of books and grey literature.

Google Scholar

Google Scholar was released in beta form in November 2004, and was expanded to include judicial case law in 2009. While Google Scholar has received much attention in academia, it seems to be regarded by Google as a niche product: in 2011 Google removed Scholar from the list of top services and list of "more" services, relegating it to the "even more" list. In 2014, the Scholar team consisted of just nine people (Levy 2014). Describing Google Scholar in an introductory manner is not helped by Google's vague documentation, which simply says it "includes scholarly articles from a wide variety of sources in all fields of research, all languages, all countries, and over all time periods."[5] The "wide variety of sources" includes "journal papers, conference papers, technical reports, or their drafts, dissertations, pre-prints, post-prints, or abstracts," as well as court opinions and patents, but not "news or magazine articles, book reviews, and editorials." Books and dissertations uploaded to Google Book Search are "automatically" included in Scholar. Google says abstracts are key, noting "Sites that show login pages, error pages, or bare bibliographic data without abstracts will not be considered for inclusion and may be removed from Google Scholar."

Studies of Google Scholar can be divided in to three major categories of focus: investigating the coverage of Google Scholar; the use and utility of Google Scholar as part of the research process; and Google Scholar's utility for bibliographic measurement, including evaluating the productivity of individual researchers and the impact of journals. There is some overlap across these categories, because studies of Google Scholar seem to involve three questions: 1) What is being searched? 2) How does the search function? and 3) To what extent can the user usefully accomplish her task?

The Coverage of Google Scholar

Scholars want to know what "scholarship" is covered by Google Scholar, but the documentation merely states that it indexes "papers, not journals"[6] and challenges researchers to investigate

[5] https://scholar.google.com/intl/en/scholar/inclusion.html
[6] https://www.google.com/intl/en/scholar/help.html#coverage

Google Scholar's coverage empirically despite Google Scholar's notoriously challenging technical limitations.

While some limitations of Google Scholar have been corrected over the years, longstanding logistical hurdles involved with studying Google Scholar's coverage have been well-documented for over a decade (Shultz 2007; Bonato 2016; Haddaway et al. 2015; Levay et al. 2016), and include:

- Search queries are limited to 256 characters
- Not being able to retrieve more than 1,000 results
- Not being able to display more than 20 results per page
- Not being able to download batches of results (e.g. to load into citation management software)
- Duplicate citations (beyond the multiple article "versions"), requiring manual screening
- Retrieving different results with Advanced and Basic searches
- No designation of the format of items (e.g. conference papers)
- Minimal sort options for results
- Basic Boolean operators only[7]
- Illogical interpretation of Boolean operators: *esophagus OR oesophagus* and *oesophagus OR esophagus* return different numbers of results (Boeker, Vach, and Motschall 2013)
- Non-disclosure of the algorithm by which search results are sorted.

Additionally, one study reported experiencing an automated block to the researcher's IP address after the export of approximately 180 citations or 180 individual searches (Haddaway et al. 2015, 14). Furthermore, the Research Excellence Framework was unable to use Google Scholar to assess the quality of research in UK higher education institutions, because of researchers' inability to agree with Google on a "suitable process for bulk access to their citation information, due to arrangements that Google Scholar have in place with publishers" (Research Excellence Framework 2013, 1562). Such barriers can limit what can be studied and also cost researchers significant time in terms of downloading (Prins et al. 2016) and cleaning citations (Levay et al. 2016).

Despite these hurdles, research activity analyzing the coverage of Google Scholar has continued in the past two years, often building off previous studies. This section will first discuss Google Scholar's size and ranking, followed by its coverage of articles and citations, then its coverage of books, grey literature, and open access and institutional repositories.

Google Scholar Size and Ranking

In a 2014 study, Khabsa and Giles estimated there were at least 114 million English-language scholarly documents on the Web, of which Google Scholar had "nearly 100 million." Another study by Orduna-Malea, Ayllón, Martín-Martín, and López-Cózar (2015) estimated that the total number

[7] E.g., no nesting of logical subexpressions deeper than one level (Boeker, Vach, and Motschall 2013) and no truncation operators.

of documents indexed by Google Scholar, without any language restriction, was between 160 and 165 million. By comparison, in 2016 the author's discovery tool contained about 168 million items in academic journals, conference materials, dissertations, and reviews.[8] Google Scholar's presence in the information marketplace has influenced vendors to increase the discoverability of their content, including pushing for the display of abstracts and/or the first page of articles (Levy 2014). ProQuest and Gale indexes were added to Google Scholar in 2015 (Quint 2016). Martín-Martín et al. (2016b) noted that Google Scholar's agreements with big publishers come at a price: "the impossibility of offering an API," which would support bibliometricians' research (54).

Google Scholar's results ranking "aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature."[9] Martín-Martín and his colleagues (2017, 159) conducted a large, longitudinal study of null query results in Google Scholar and found a strong correlation between result list ranking and times cited. The influence of citations is so strong that when the researchers performed the same search process four months later, 14.7% of documents were missing in the second sample, causing them to conclude even a change of one or two citations could lead to a document being excluded or included from the top 1,000 results (157). Using citation counts as a major part of the ranking algorithm has been hypothesized to produce the "Matthew Effect," where "work that is already influential becomes even more widely known by virtue of being the first hit from a Google Scholar search, whereas possibly meritorious but obscure academic work is buried at the bottom" (Antell et al. 2013, 281).

Google Scholar has been shown to heavily bias its ranking toward English-language publications even when there are highly cited non-English publications in the result set, although selection of interface language may influence the ranking. Martin-Martin and his colleagues noted that Google Scholar seems to use the domain of the document's hosting web site as a proxy for language, meaning that "some documents written in English but with their primary version hosted in non-Anglophone countries' web domains do appear in lower positions in spite of receiving a large number of citations" (Martin-Martin et al. 2017, 161). This effect is shown dramatically in Figure 3 of their paper.

Google Scholar Coverage: Articles and Citations

The coverage of articles, journals, and citations by Google Scholar has been commonly examined by using brute force methods to retrieve a sample of items from Google Scholar and possibly one or more of its competitors. (Studies discussed in this section are listed in Table 1). The goal is usually to determine how well Google Scholar's database compares to traditional research databases, usually in a specific field. Core methodology involves importing citations into software such as Publish or Perish (Harzing 2016a), cleaning the data, then performing statistical tests,

[8] The discovery tool does not contain all available metadata but has been carefully vetted (Fagan and Gaines 2016).
[9] https://www.google.com/intl/en/scholar/about.html

expert review, or both. Haddaway (2015) and Moed et al. (2016) have written articles specifically discussing methodological aspects.

Recent studies repeatedly find that Google Scholar's coverage meets or exceeds that of other search tools, no matter what is identified by target samples, including journals, articles, and citations (Karlsson 2014; Harzing 2014; Harzing 2016b; Harzing and Alakangas 2016b; Moed, Bar-Ilan, and Halevi 2016; Prins et al. 2016; Wildgaard 2015; Ciccone and Vickery 2015). In only three studies did Google Scholar find fewer items, and the meaningful difference was minimal.[10]

Science disciplines were the most studied in Google Scholar, including agriculture, astronomy, chemistry, computer science, ecology, environmental science, fisheries, geosciences, mathematics, medicine, molecular biology, oceanography, physics, and public health. Social sciences studied include education (Prins et al. 2016), economics (Harzing 2014), geography (Ştirbu et al. 2015, 322-329), information science (Winter, Zadpoor, and Dodou 2014; Harzing 2016b), and psychology (Pitol and De Groote 2014). Studies related to the arts or humanities 2014-2016 included an analysis of open access journals in music (Testa 2016) and a comparison between Google Scholar and Web of Science for research evaluation within education, pedagogical sciences, and anthropology[11] (Prins et al. 2016). Wildgaard (2015) and Bornmann et al. (2016) included samples of humanities scholars as part of bibliometric studies, but did not discuss disciplinary aspects related to coverage. Prior to 2014, the only study found related to the arts and humanities compared Google Scholar with Historical Abstracts (Kirkwood Jr. and Kirkwood 2011).

Google Scholar's coverage has been growing over time (Meier and Conkling 2008; Harzing 2014; Winter, Zadpoor, and Dodou 2014; Bartol and Mackiewicz-Talarczyk 2015, 531; Orduña-Malea and Delgado López-Cózar 2014) with recent increases in older articles (Winter, Zadpoor, and Dodou 2014; Harzing and Alakangas 2016b), leading some to question whether this supports the documented trend of increased citation of older literature (Martín-Martín et al. 2016c; Varshney 2012). Winter et al. noted that in 2005 Web of Science yielded more citations than Google Scholar for about two-thirds of their sample, but for the same sample in 2013, Google Scholar found more citations than Web of Science, with only 6.8% of citations not retrieved by Google Scholar (Winter, Zadpoor, and Dodou 2014, 1560). The unique citations of Web of Science were "typically documents before the digital age and conference proceedings not available online" (Winter, Zadpoor, and Dodou 2014, 1560). Harzing and Alakangas's (2016b) large-scale longitudinal comparison of Google Scholar, Scopus, and Web of Science suggested that Google Scholar's retroactive expansion has stabilized and now all three databases are growing at similar rates.

[10] For example, Bramer, Giustini, and Kramer (2016a) found slightly more of their 4,795 references from systematic reviews in Embase (97.5%) than in Google Scholar (97.2%). In Testa (2016), the music database RILM indexed two more of the 84 OA journals than Google Scholar (which indexed at least one article from 93% of the journals). Finally, in a study using citations to the most-cited article of all time as a sample, Web of Science found more citations than did Google Scholar (Winter, Zadpoor, and Dodou 2014).

[11] Prins et al. classified anthropology as part of the humanities.

Google Scholar also seems to cover both the oldest and the most recent publications. Unlike traditional abstracts and indexes, Google Scholar is not limited by starting year, so as publishers post tables of contents of their earliest journals online, Google Scholar discovers those sources (Antell et al. 2013, 281). Trapp (2016) reported the number of citations to a highly-cited physics paper after the first 11 days of publication to be 67 in Web of Science, 72 in Scopus, and 462 in Google Scholar (Trapp 2016, 4). In a study of 800 citations to Nobelists in multiple fields, Harzing found that "Google Scholar could effectively be 9–12 months ahead of Web of Science in terms of publication and citation coverage" (2013, 1073).

An increasing proportion of journal articles in Google Scholar are freely available in full text. A large-scale, longitudinal study of highly-cited articles 1950-2013 found 40% of article citations in the sample were freely available in full text (Martín-Martín et al. 2014). Another large-sample study found 61% of articles in their sample from 2004–2014 could be freely accessed (Jamali and Nabavi 2015). In both studies, nih.gov and ResearchGate were the top two full-text providers.

Google Scholar's coverage of major publisher content varies; having some coverage of a publisher does not imply all articles or journals from that publisher are covered. In a sample of 222 citations compared across Google Scholar, Scopus, and Web of Science, Google Scholar contained all of the Springer titles, as many Elsevier titles as Scopus, and the most articles by Wolters Kluwer and John Wiley. However, among the three databases, Google Scholar contained the fewest articles by *BMJ* and *Nature* (Rothfus et al. 2016).

Study	Sample	Results
(Bartol and Mackiewicz-Talarczyk 2015)	Documents retrieved in response to searches on crops and fibers in article titles, 1994-2013 (samples varied by crop)	Google Scholar returned more documents retrieved for each crop. For example, "hemp" retrieved 644 results in Google Scholar, 493 in Scopus, and 318 in Web of Science; Google Scholar demonstrated higher yearly growth of records over time.
(Bramer, Giustini, and Kramer 2016b)	References from a pool of systematic reviewer searches in medicine (n=4795)	Google found 97.2%, Embase, 97.5%, MEDLINE 92.3% of all references; When using search strategies, Embase retrieved 81.6%, MEDLINE 72.6%, and Google Scholar 72.8%.
(Ciccone and Vickery 2015)	Based on 183 user searches randomly selected from NCSU Libraries' 2013 Summon search logs (n=137)	No significant difference between the performance of Google Scholar, Summon, and EDS for known-item searches; "Google Scholar outperformed both discovery services for topical searches."
(Harzing 2014)	Publications and citation metrics for 20 Nobelists in chemistry, economics, medicine, physics, 2012-2013 (samples varied)	Google Scholar coverage is now "increasing at a stable rate" and provides "comprehensive coverage across a wide set of disciplines for articles published in the last four decades" (575).
(Harzing 2016b)	Citations from one researcher (n=126)	Microsoft Academic found all books and journal articles covered by Google Scholar; Google Scholar found 35 additional publications including book chapters, white papers, and conference papers.
(Harzing and Alakangas 2016a)	Samples from (Harzing and Alakangas 2016b, 802) (samples varied by faculty)	Google Scholar provided higher "true" citation counts than Microsoft Academic but Microsoft Academic "estimated" citation counts were 12% higher than Google Scholar for life sciences and equivalent for the sciences.

(Harzing and Alakangas 2016b)	Citations of the works of 145 faculty among 37 scholarly disciplines at the University of Melbourne (samples varied by faculty)	For the top faculty member, Google Scholar had 519 total papers (compared with 309 in both Web of Science and Scopus); Google Scholar had 16,507 citations (compared with 11,287 in Web of Science and 11,740 in Scopus).
(Hilbert et al. 2015)	Documents published by 76 information scientists in German-speaking countries (n=1,017)	Google Scholar covered 63%, Scopus, 31%, BibSonomy, 24%, Mendeley, 19%, Web of Science, 15%, CiteULike, 8%.
(Jamali and Nabavi 2015)	Items published between 2004 and 2014 (n=8,310)	61% of articles were freely available; of these, 81% were publisher versions and 14% were pre-prints; ResearchGate was the top full-text source netting 10.5% of full-text sources, followed by ncbi.nlm.nih.gov (6.5%).
(Karlsson 2014)	Journals from ten different fields (n=30)	Google Scholar retrieved documents from all the selected journals; Summon only retrieved documents from 14 out of 30 journals.
(Lee et al. 2015)	Journal articles housed in Florida State University's institutional repository (n=170)	Metadata found in Google for 46% of items and in Google Scholar for 75% of items; Google Scholar found 78% of available full text. Google Scholar found full text for six items with no full text in the IR.
(Martín-Martín et al. 2014)	Items highly cited by Google Scholar (n=64,000)	40% could be freely accessed using Google Scholar; Nih.gov and ResearchGate were the top two full-text providers.
(Moed, Bar-Ilan, and Halevi 2016)	Citations to 36 highly cited articles in 12 scientific-scholarly English-language journals (n=about 7,000)	47% of sources were in both Google Scholar and Scopus; 47% of sources were in Google Scholar only; 6% of sources were in Scopus only; Of the unique Google Scholar citations, sources were most often from Google Books, Springer, SSRN, ResearchGate, ACM Digital Library, Arxiv, and ACLweb.org.

(Prins et al. 2016)	Article citations in the field of education and pedagogies, and citations to 328 articles in anthropology (n=774)	Google Scholar found 22,887 citations in Education & Pedagogical Science compared to Web of Science's 8,870, and 8,092 in Anthropology compared with Web of Science's 1,097.
(Ştirbu et al. 2015)	Compared # of citations resulting from two geographical topic searches (samples varied)	Google Scholar found 2,732 geographical references whereas Web of Science found only 275, GeoRef 97, and FRANCIS 45. For sedimentation, Google Scholar found 1,855 geographical references compared to Web of Science's 606, GeoRef's 1,265, and FRANCIS's 33; Google Scholar overlapped Web of Science by 67% and 82% for the two searches, and GeoRef by 57% and 62%
(Testa 2016)	Open access journals in music (n=84)	Google Scholar indexed at least one article from 93% of OA journals. RILM indexed two additional journals.
(Wildgaard 2015)	Publications from researchers in astronomy, environmental science, philosophy and public health (n=512)	Publication count from Web of Science was 2-4 times lower for all disciplines than Google Scholar; Citation count was up to 13 times lower in Web of Science than in Google Scholar.
(Winter, Zadpoor, and Dodou 2014)	Growth of citations to 2 classic articles (1995-2013) and 56 science and social science articles in Google Scholar, 2005-2013 (samples varied)	Total citation counts 21% higher in Web of Science than Google Scholar for Lowry (1951) but Google Scholar 17% higher than Web of Science for Garfield (1955) and 102% higher for the 56 research articles; Google Scholar showed a significant retroactive expansion to all articles compared to negligible retroactive growth in Web of Science.

Table 1. Studies investigating Google Scholar's coverage of journal articles and citations, 2014-2016.

Google Scholar Coverage: Books

Many studies mentioned that books, including Google Books, are sometimes included in Google Scholar results. Jamali and Nabavi (2015) found 13% of their sample of 8,310 citations from Google Scholar were books, while Martín-Martín et al. (2014) had found that 18% of their sample of 64,000 citations from Google Scholar were books. Within the field of anthropology, Prins (2016) found books to generate the most citation impact in Google Scholar (41% of books in their sample were cited in Google Scholar) compared to articles (21% of articles were cited in Google Scholar). In education, 31% of articles and 25% of books were cited by Google Scholar (3). Abrizah and Thelwall found only 37% of their sample of 1,357 arts, humanities, and social sciences books from the five main university presses in Malaysia had been cited in Google Scholar (23% of the books had been cited in Google Books) (Abrizah and Thelwall 2014, 2502). The overlap was small: 15% had impact in both Google Scholar and Google Books. The authors concluded that due to the low overlap between Google Scholar and Google Books, searching both engines is required to find the most citations to academic books. English books were significantly more likely to be cited in Google Scholar (48% vs. 32%), as were edited books (53% vs. 36%). They surmised edited books' citation advantage was due to the use of book chapters in social sciences. They found arts and humanities books more likely to be cited in Google Scholar than social sciences books (40% vs. 34%) (Abrizah and Thelwall 2014, 2503).

Google Scholar Coverage: Grey Literature

Grey literature refers to documents not published commercially, including theses, reports, conference papers, government information, and poster sessions. Haddaway et al. (2015) was the only empirical study found focused on grey literature. They discovered that between 8% and 39% of full-text search results from Google Scholar were grey literature, with the greatest concentration of citations from grey literature on page 80 of results for full-text searches and page 35 for title searches. They concluded "the high proportion of grey literature that is missed by Google Scholar means it is not a viable alternative to hand searching for grey literature as a stand-alone tool" (2015, 14). For one of the systematic reviews in their sample, none of the 84 grey literature articles cited were found within the exported Google Scholar search results. The only other investigation of grey literature found was Bonato (2016), who after conducting a very limited number of searches on one specific topic and a search for a known item, concluded Google Scholar to be "deficient." In conclusion, despite much offhand praise for Google Scholar's grey literature coverage (Erb and Sica 2015; Antell et al. 2013), the topic has been little studied and when it has, grey literature results have not been prominent.

Google Scholar Coverage: Open Access and Institutional Repository Content

Erb and Sica touted Google Scholar's access to "free content that might not be available through a library's subscription services," including open access journals and institutional repository coverage (2015, 48). Recent research has dug deeper into both these content areas.

In general, OA articles have been shown to net more citations than non-OA articles, as Koler-Povh, Južnic, and Turk (2014) showed within the field of civil engineering. Across their sample of 2,026 scholarly articles in 14 journals, all indexed in Web of Science, Scopus, and Google Scholar, OA articles received an average of 43 citations while non-OA articles were cited 29 times (1039). Google Scholar did a better job discovering those citations; in Google Scholar the median of citations of OA articles was always higher than that for non-OA articles, wheras this was true in Web of Science for only 10 of the 14 journals and in Scopus for 11 of the 14 journals (1040). Similarly, Chen (2014) found Google Scholar to index far more OA journals than Scopus and Web of Science, especially "gold OA."[12] Google Scholar's advantage should not be assumed across all disciplines, however; Testa (2016) found both Google Scholar and RILM to provide good coverage of OA journals in music, with Google Scholar indexing at least one article from 93% of the 84 OA journals in the sample. But the bibliographic database RILM indexed two more OA journals than Google Scholar.

Google Scholar indexing of repositories may be critical for success, but results vary by IR platform and whether the IR metadata has been structured according to Google's guidelines. In a random sample from Shodhganga, India's central ETD database, Weideman (2015) found not one article had been indexed in full text by Google Scholar, although in many cases the metadata was indexed, leading the author to identify needed changes to the way Shodhganga stores ETDs.[13] Likewise, Chen (2014) found that neither Google Scholar nor Google appears to index Baidu Wenku, a major full-text archive and social networking site in China similar to ResearchGate, and Orduña-Malea and López-Cózar (2015) found that Latin American repositories are not very visible in Google or Google Scholar due to limitations of the description schemas chosen as well as search engine reliability. In Yang's (2016) study of Texas Tech's DSpace IR, Google was the only search engine that indexed, discovered, or linked to PDF files supplemented with metadata; Google Scholar did not discover or provide links to the IR's PDF files, and was less successful at discovering metadata.

When Google Scholar is able to index IR content, it may be responsible for significant traffic. In a study of four major U.S. universities' institutional repositories (three DSpace, one CONTENTdm) involving a dataset of 57,087 unique URLs and 413,786 records, researchers found that 48%–66% of referrals came from Google Scholar (Obrien et al. 2016, 870). The importance of Google Scholar in contrast to Google was noted by Lee et al. (2015), who conducted title searches on 170 journal articles housed in Florida State University's institutional repository (using bePress's Digital Commons platform), 100 of which existed in full text in the IR. Links to the IR were found in Google results for 45.9% of the 170 items, and in Google Scholar for 74.7% of the 170 items. Furthermore, Google Scholar linked to the full text for 78% of the 100 cases where full text was available, and even provided links to freely available full text for six items that did not have full

[12] OA articles on publisher web sites, whether the journal itself is OA or not (Chen 2014)
[13] Most notably, the need to store thesis documents as one PDF file instead of divided into multiple, separate files, to create HTML landing pages as per Google's recommendations, and to submit the addresses of these pages to Google Scholar.

text in the IR. However, the researchers also noted "relying on either Google or Google Scholar individually cannot ensure full access to scholarly works housed in OA IRs." In their study, among the 104 fully open access items there was an overlap in results of only 57.5%; Google provided links to 20 items not found with Google Scholar, and Google Scholar provided links to 25 items not found with Google (Lee et al. 2015, 15).

Google Scholar results note the number of "versions" available for each item. In a study of 982 science article citations (including both OA and non-OA) in IRs, Pitol and DeGroote found 56% of citations had between four and nine Google Scholar versions (2014, 603) Almost 90% of the citations shown were the publisher version, but of these, only 14.3% were freely available in full text on the publisher web site. Meanwhile, 70% percent of the items had at least one free full-text version available through a "hidden" Google Scholar version. The author's experience in retrieving full text for this review indicates this issue still exists, but research would be needed to formulate reliable recommendations for users.

Use and utility of Google Scholar as part of the research process

Studies were found concerning Google Scholar's popularity with users and their reasons for preferring it (or not) over other tools. Another group of studies examined issues related to the utility of Google Scholar for research processes, including issues related to messy metadata. Finally, a cluster of articles focused specifically on using Google Scholar for systematic reviews.

Popularity and User Preferences

Several studies have shown Google Scholar to be well-known to scholarly communities. A survey of 3,500 scholars from 95 countries found that over 60% of 3,500 scientists and engineers and over 70% of respondents in the social sciences, arts, and humanities were aware of Google Scholar and used it regularly (Van Noorden 2014). In a large-scale journal-reader survey, Inger and Gardner (2016) found that among academic researchers in high-income areas, academic search engines surpassed abstracts and indexes as a starting place for research (2016, 85, Figure 4). In low-income areas, Google use exceeded Google Scholar use for academic research. Major library link resolver software offers reports of full-text requests broken down by referrer. Inger and Gardner (2016) showed a large variance across subjects for whether people prefer Google or Google Scholar: "People in the social sciences, education, law, and business use Google Scholar more to find journal articles. However, people working in the humanities and religion and theology prefer to use Google" (88). Humanities scholar use of Google over Google Scholar was also found by Kemman et al. (2013); Google, Google Images, Google Scholar, and YouTube were used more than JSTOR or other library databases, even though humanities scholars' trust in Google and Google Scholar was lower.

User research since 2014 concerning Google Scholar has focused on graduate students. Results suggest Scholar is used regularly but the tool is only partially sufficient. In their study of 20 engineering masters' students' use of abstracts and indexes, Johnson and Simonsen (2015) found

that half their sample (n=20) had used Google Scholar the last time they located an article using specific search terms or criteria. Google was the second most-used source at 20%, followed by abstracting and indexing services (15%).

Graduate students describe Google Scholar with nuance and refer to it as a specific part of their process. In Bøyum and Aabø's (2015) interviews with eight PhD business students and Wu and Chen's (2014, 381) interviews with 32 graduate students drawn from multiple academic disciplines, the majority described using library databases and Google Scholar for different purposes depending on the context. Graduate students in both studies were well aware of Google Scholar's use for citation searching. Bøyum and Aabø's (2015) subjects described library resources as more "academically robust" than Google or Google Scholar. Wu and Chen's (2014) interviewees praised Google Scholar for its wider coverage and convenience, but lamented the uncertain quality, sometimes inaccessible full text, too many results, lack of sorting function (document type or date), finding documents from different disciplines, and duplicate citations. Google Scholar was seen by their subjects as useful during early stages of information seeking. In contrast to general assumptions, more than half the students (Wu and Chen 2014, 381) interviewed reported browsing more than 3 pages' worth of Google Scholar results. About half of interviewees reported looking at cited documents to find more, however students had mixed opinions about whether the citing documents turned out to be relevant.

Google Scholar's "My Library" feature, introduced in 2013, now competes with other bibliographic citation management software. In a survey of 344 (mostly graduate) students, Conrad, Leonard, and Somerville found Google Scholar was the most-used (47%) followed by EndNote (37%), and Zotero (19%) (2015, 572). Follow-up interviews with 13 of the students revealed that a few students used multiple tools, for example one participant noted he/she used "EndNote for sharing data with lab partners and others "across the community"; Mendeley for her own personal thesis work, where she needs to "build a whole body of literature"; and Google Scholar Citations for "quick reference lists that I may not need for a second or third time."

Messy Metadata

Many studies have suggested Google Scholar's metadata is "messy." Although none in the period of study examined this phenomenon in conjunction with relative user performance, the issues found could affect scholarship. A 2016 study itemized the most common mistakes in Google Scholar resulting from its extraction process: 1) incorrect title identification; 2) missing or incorrectly assigned authors; 3) book reviews indexed as books; 4) failing to group versions of the same document, which inflates citation counts; 5) grouping different editions of books, which deflates citation counts; 6) attributing citations to documents that did not cite them, or missing citations that did; and 7) duplicate author profiles (Martín-Martín et al. 2016b). The authors concluded that "in an academic big data environment, these errors (which we deem affect less than 10% of the records in the database) are of no great consequence, and do not affect the core system

performance significantly" (54). Two of these issues have been studied specifically: duplicate citations and missing publication dates.

The rate of duplicate citations in Google Scholar has ranged upwards of 2.93% (Haddaway et al. 2015) and 5% (Winter, Zadpoor, and Dodou 2014, 1562), which can be compared to a .05% duplicate citation rate in Web of Science (Haddaway et al. 2015, 13). Haddaway found the main reasons for duplication include "typographical errors, including punctuation and formatting differences; capitalization differences (Google Scholar only), incomplete titles, and the fact that Google Scholar scans citations within reference lists and may include those as well as the citing article" (2015, 13).

The issue of missing publication dates varies greatly across samples. Dates were found to be missing 9% of the time in Winter et al.'s study, although it varied by publication type: 4% of journals, 15% of theses, and 41% of the unknown document types" (Winter, Zadpoor, and Dodou 2014, 1562). However Martin-Martin et al. studied a sample of 32,680 highly-cited documents and found that Web of Science and Google Scholar agreed on publication dates 96.7% of the time, with an idiosyncratically large proportion of those mismatches in 2012 and 2013 (2017, 159).

Utility for Research Processes

Prior to 2014, studies such as Asher, Duke, and Wilson's 2012 evaluated Google Scholar's utility as a general research tool, often in comparison with discovery tools. Since 2014, the only such study found was Namei and Young's comparison of Summon, Google Scholar, and Google using 299 known-item queries. They found Google Scholar and Summon returned relevant results 74% of the time; Google returned relevant results 91% of the time. For "scholarly formats," they found Summon returned relevant results 76% of the time; Google 79%; and Google 91% (2015, 526-527).

The remainder of studies in this category focused specifically on systematic reviews, perhaps because such reviews are so time-consuming. Authors develop search strategies carefully, execute them in multiple databases, and document their search methods and results carefully. Some prestigious journals are beginning to require similar rigor for any original research article, not just systematic reviews (Cals and Kotz 2016). Information provided by professional organizations about the use of Google Scholar for systematic reviews seems inconsistent: the *Cochrane Handbook for Systematic Reviews of Interventions* lists Google Scholar among sources for searching, but none of the five "highlighted reviews" on the Cochrane web site at the time of this article's writing used Google Scholar in their methodologies. The UK organization National Institute for Health and Care Excellence's manual (National Institute for Health and Care Excellence (NICE)) only mentions Google Scholar in an appendix of search sources under "Conference Abstracts."

A study by Gehanno et al. (2013) found Google Scholar contained 100% of the references from 29 systematic reviews, and suggested Google Scholar could be the first choice for systematic reviews or meta-analyses. This finding prompted a slew of follow-up studies in the next three years. An

immediate response by Giustini and Boulos (2013) pointed out that systematic reviews are not performed by searching for article titles as with Gehanno et al.'s method, but through search strategies. When they tried to replicate a systematic review's topical search strategy in Google Scholar, the citations were not easily discovered. In addition the authors were not able to find all the papers from a given systematic review even by title searching. Haddaway et al. also found imperfect coverage: for one of the seven reviews examined, 31.5% of citations could not be found (2015, 11). Haddaway also noted that special characters and fonts (as with chemical symbols) can cause poor matching when such characters are part of article titles. Recent literature concurs that it is still necessary to search multiple databases when conducting a systematic review, including abstracts and indexes, no matter how good Google Scholar's coverage seems to be. No one database's coverage is complete, including Google Scholar (Thielen et al. 2016), and practical recall of Google Scholar is exceptionally low due to the 1,000 result limit, yet at the same time, Google Scholar's lack of precision is costly in terms of researchers' time (Bramer, Giustini, and Kramer 2016b; Haddaway et al. 2015).

The challenges limiting study of Google Scholar's coverage also bedevil those wishing to use it for reviews, especially the 1,000 result retrieval limit, lack of batch export, and lack of exported abstracts (Levay et al. 2016). Additionally, Google Scholar's changing content, unknown algorithm and updating practices, search inconsistencies, limited Boolean functions, and 256-character query limit prevent the tool from accommodating the detailed, reproducible search methodologies required by systematic reviews (Bonato 2016; Haddaway et al. 2015; Giustini and Boulos 2013). Bonato noted Google Scholar retrieved different results with Advanced and Basic searches; could not determine the format of items (e.g. conference papers); and found other inconsistent results.[14] Bonato also lamented the lack of any kind of document type limit.

Despite the limitations and logistical challenges, practitioners and scholars are finding solid reasons for including academic web search engines as part of most systematic review methodologies (Cals and Kotz 2016). Stansfield et al. noted that "relevant literature for low- and middle-income countries, such as working and policy papers, is often not included in databases," and that Google Scholar finds additional journal articles and grey literature not indexed in databases (2016, 191). For eight systematic reviews by EPPI-Center, "over a quarter of relevant citations were found from websites and internet search engines" (Stansfield, Dickson, and Bangpan 2016, 2).

Specific tools and practices have been recommended when using search engines within the context of systematic reviews. Software is available to record search strategies and results (Harzing and Alakangas 2016b; Haddaway 2015). Haddaway suggests the use of snapshot tools (Haddaway 2015) to record the first 1,000 Google Scholar records rather than the typical assessment of the first 50 search results as had been done in the past: "This change in practice

[14] Bonato (2016) found zero hits for conference papers when limiting by year 2015-2016, but found two papers presented at a 2015 meeting.

could significantly improve both the transparency and coverage of systematic reviews, especially with respect to their grey literature components." (Haddaway et al. 2015, 15). Both Haddaway (2015) and Cochrane recommend that review authors print or save locally electronic copies of the full text or relevant details rather than bookmarking web sites, "in case the record of the trial is removed or altered at a later stage" (Higgins and Green 2011). New methods for searching, downloading, and integrating academic search engine results into review procedures using free software to increase transparency, repeatability, and efficiency have been proposed by Haddaway and his colleagues (2015).

Google Scholar Citations and Metrics

Google Scholar Citations and Metrics are not academic search engines, but this article included them because these products are interwoven into the fabric of the Google Scholar database. Google Scholar Citations, launched in late 2011 (Martín-Martín et al. 2016b, 12) groups citations by author, while Google Metrics (launch date uncertain) provides similar data for articles and journals. Readers interested in an in-depth literature review of Google Scholar Citations for earlier years (2005-2012) are directed to (Thelwall and Kousha 2015b).

In his comprehensive review of more recent literature about using Google Scholar Citations for citation analysis, Waltman (2016) described several themes. Google Scholar's coverage of many fields is significantly broader than Web of Science and Scopus, and this seems to be continuing to improve over time. However studies regularly report Google Scholar's inaccuracies, content gaps, phantom data, easily manipulatable citation counts, lack of transparency, and limitations for empirical bibliometric studies.

As discussed in the coverage section, Google Scholar's citation database is competitive with other major databases such as Web of Science and has been growing dramatically in the last few years (Winter, Zadpoor, and Dodou 2014; Harzing and Alakangas 2016b; Harzing 2014) but has recently stabilized (Harzing and Alakangas 2016b). More and more studies are concluding that Google Scholar will report more comprehensive information about citation impact than Web of Science or Scopus. Across a sample of articles from many years of one science journal, Trapp (2016) found the proportion of articles with zero citations was 37% for Web of Science, 29% for Scopus, and 19% for Google Scholar. Some of Google Scholar's superiority for citation analysis in the social sciences and humanities is due to its inclusion of book content, software, and additional journals (Prins et al. 2016; Bornmann et al. 2016). Bornmann et al. (2016) noted citations to all ten of a research institute's ten books published in 2009 were found in Google Scholar, whereas Web of Science found citations for only two books. Furthermore they found data in Google Scholar for 55 of the total of 71 of the institute's book chapters. For the four conference proceedings they could identify in Google Scholar, there were 100 citations, of which 65 could be found in Google Scholar. The comparative success of Google Scholar for citation impact varies by discipline, however: (Levay et al. 2016) found Web of Science to be more reliable than Google Scholar, quicker for

downloading results, and better for retrieving 100% of the most important publications in public health.

Despite Google Scholar's growth, using all three major tools (Scopus, Web of Science, and Google Scholar) still seems to be necessary for evaluating researcher productivity. Rothfus (2016) compared Web of Science, Scopus, and Google Scholar citation counts for evaluating the impact of the Canadian Network of Observational Drug Effect Studies (CNODES), as represented by a sample of 222 citations from five articles. Attempting to determine citation metrics for the CNODES research team yielded different results for every article when using the three tools. They found that "using three tools (Web of Science, Scopus, Google Scholar) to determine citation metrics as indicators of research performance and impact provided varying results, with poor overall agreement among the three" (237). Major academic libraries' web sites often explain how to find one's h-index in all three (Suiter and Moulaison 2015).

Researchers have also noted the disadvantages of Google Scholar for citation impact studies. Google Scholar is costly in terms of researcher time. Levay et al. (2016) estimated the cost of "administering results" from Web of Science to be 4 hours versus 75 hours for Google Scholar. Administering results includes using the search tool to search, download, and add records to bibliographic citation software, and removing duplicate citations. Duplicate citations are often mentioned as a problem (Prins et al. 2016), although Moed (2016) suggested the double counting by Google Scholar would occur only if the level of analysis is on target sources, not if it is on target articles.[15]

Downloaded citation samples can still suffer from double counts, however: Harzing and Alakangas described how cleaning "a fairly extreme case" in their study reduced the number of papers from 244 to 106 (2016b). Google Scholar also does not identify self-citations, which can dramatically influence the meaning of results (Prins et al. 2016). Furthermore, researchers have shown it is possible to corrupt Google Scholar Citations by uploading obviously false documents (Delgado López-Cózar, Robinson-García, and Torres-Salinas 2014).While the researchers noted traditional citation indexes can also be defrauded, Google's products are less transparent and abuses may not be easily detected. Google did not respond to the research team when contacted and simply deleted the false documents to which it had been alerted without reporting the situation to the affected authors, and the researchers concluded: "This lack of transparency is the main obstacle when considering Google Scholar and its by-products for research evaluation purposes" (453).

Because these disadvantages do not outweigh Google Scholar's seemingly broader coverage, many articles investigate workarounds for using Google Scholar more effectively when evaluating

[15] "if a document is, for instance, first published in ArXiv, and a next version later in a journal J, citations to the two versions are aggregated. In Google Scholar Metrics, in which ArXiv is included as a source, this document (assuming that its citation count exceed the h5 value of ArXiv and journal J) is listed *both* under ArXiv *and* under journal J, with *the same*, aggregate citation count (Moed 2016, 29).

research impact. Harzing and Alakangas (2016b) recommend the hI index[16], which is corrected for career length and co-authorship patterns, as the citation metric of choice for a fair comparison of Google Scholar with other tools. Bornmann et al. (2016) investigated a method to normalize data and reduce errors when using Google Scholar data to evaluate citations in the social sciences and humanities.

Researcher profiles can also be used to find other scholars by topic. In a 2014 survey of researchers (n=8,554), Dagienė and Krapavickaitė found that 22% used a third-party service such as Google Scholar or Microsoft Academic to produce lists of their scholarly activities and 63% reported their scholarly record was freely available on the Web (2016, 158, 161). Google Scholar ranked only second to Microsoft Word as the most frequently used software to maintain academic activity records (160). Martín-Martín et al. (2016b) examined 814 authors in the field of bibliometrics using Google Scholar Citations, ResearcherID, ResearchGate, Mendeley, and Twitter. Google Scholar was the most used social research sharing platform, followed by ResearchGate, with ResearcherID gaining wider acceptance among authors deemed "core" to the field. Only about one-third of the authors created a Twitter profile, and many Mendeley and ResearcherID profiles were found empty. The study found Google Scholar academic profiles' distinctive advantages to be automatic updates and its high growth rate, with disadvantages of scarce quality control, inherited metadata mistakes from Google Scholar, and its manipulatability. Overall, Martin-Martin and colleagues concluded that Google Scholar "should be the preferred source for relational and comparative analyses in which the emphasis is put on author clusters" (57).

Google Scholar Metrics provides citation information for articles and journals. In a sample of 1,000 journals, Orduña-Malea and Delgado López-Cózar found that "despite all the technical and methodological problems," Google Scholar Metrics provides sound and reliable journal rankings (2014, 2365). Google Scholar Metrics seems to be an annual publication; the 2016 edition contains 5,734 publications and 12 language rankings. Russian, Korean, Polish, Ukranian, and Indionesian were added this year, while Italian and Dutch were removed for unknown reasons (Martín-Martín et al. 2016a). Researchers also found that many discussion papers and working papers were removed in 2016. English-language publications are broken into subject areas and disciplines. Google Scholar Metrics often, but not always creates separate entries for each language in which a journal is published. Bibliometricians call for Google Scholar Metrics to display the total number of documents published in the publications indexed and the total number of citations received: "These are the two essential parameters that make it possible to assess the reliability and accuracy of any bibliometric indicator" (13). Adding country and language of publication and self-citation rates are among the other improvements listed by Lopez-Cozar and colleagues.

[16] Harzing and Alakangas (2016b) define the hIa as the hI norm/academic age. Academic age refers to the number of years elapsed since first publication. To calculate hI norm, one divides the number of citations by the number of authors for that paper, and then calculates the h-index of the normalized citation count.

Informing Practice

The glaring lack of research related to the coverage of arts and humanities scholarship, limited research on book coverage, and relaunch of Microsoft Academic make it impossible to form a general recommendation regarding the use of academic web search engines for serious research. Until the ambiguity of arts and humanities coverage is clarified, and until academic web search engines are transparent and stable, traditional bibliographic databases still seem essential for systematic reviews, citation analysis, and other rigorous literature search purposes. Discipline-specific databases also have features such as controlled vocabulary, industry classification codes, and peer review indicators that make scholars more efficient and effective. Nevertheless, the increasing relevance of academic search engines and solid coverage of sciences and social sciences make it essential for librarians to become as expert with Google Scholar, Google Books, and Microsoft Academic. For some scholarly tasks, academic search engines may be superior: for example, when looking up doi numbers for this paper's bibliography, the most efficient process seemed to be a Google search on the article title plus the term "doi," and the most likely site to display in the results was ResearchGate.[17] Librarians and scholars should champion these tools as an important part of an efficient, effective scholarly research process (Walsh 2015), while also acknowledging the gaps in coverage, biases, metadata issues and missing features available in other databases. Academic web search engines could form the centerpiece for instruction sessions surrounding the scholarly network, as shown by "cited by" features, author profiles, and full-text sources. Traditional abstracts and indexes could then be presented on the basis of their strengths. At some point, explaining how to access full text will likely no longer focus on the link resolver but on the many possible document versions a user might encounter (e.g. pre-prints or editions of books) and how to make an informed choice. In the meantime, even though web search engines and repositories may retrieve copious full text outside library subscriptions, college students should still be made aware of the library's collections and services such as interlibrary loan.

When considering Google Scholar's weaknesses, it's important to keep in mind Chen's observation that we may not have a tool available that does any better (Antell et al. 2013). While Google Scholar may be biased toward English-language publications, so are many bibliographic databases. Overall, Google Scholar seems to have increased the visibility of international research (Bartol and Mackiewicz-Talarczyk 2015). While Google Scholar's coverage of grey literature has been shown to be somewhat uneven (Bonato 2016; Haddaway et al. 2015), it seems to include more diversity among relevant document types than many abstracts and indexes (Ştirbu et al. 2015; Bartol and Mackiewicz-Talarczyk 2015). Although the rigors of systematic reviews may contraindicate the tool's use as a single source, it adds value to search results from other databases (Bramer, Giustini, and Kramer 2016a). User preferences and priorities should also be taken into account; Google

[17] Because the authority of ResearchGate is ambiguous, in such cases I then looked up the doi using Google to find the publisher's version. In some cases, the doi was not displayed on the publisher's result page (e.g., https://muse.jhu.edu/article/197091).

Scholar results have been said to contain "clutter," but many researchers have found the noise in Google Scholar tolerable given its other benefits (Ştirbu et al. 2015).

Google Books purportedly contains about 30 million items, focused on U.S.-published and English-language books. But its coverage is hit-or-miss, surprising Mays (2015) with an unexpected wealth of primary sources but disappointing Harper (2016) with limited coverage of academic health sciences books. Recent court decisions have enabled Google to continue progressing toward their goal of full-text indexing and making snippet views available for the Google-estimated universe of 130 million books, which suggests its utility may increase. Google Books is not integrated with link resolvers or discovery tools but has been found useful for providing information about scholarly research impact, especially for the arts, humanities, and social sciences.

As re-launched in 2016, Microsoft Academic shows real potential to compete with Google Scholar in coverage and utility for finding journal articles. As of February 2017 its index contains 120 million citations. In contrast to the mystery of Google Scholar's black-box algorithms and restrictive limitations, Microsoft Academic uses an open-system approach and offers an API. Microsoft Academic appears to have less coverage of books and grey literature compared with Google Scholar. Research is badly needed about the coverage and utility of both Google Books and Microsoft Academic.

Google Scholar continues to evolve, launching a new algorithm for known-item searching in 2016[18] that appears to work very well. Google Scholar does not reveal how many items it searches but studies have suggested 160 million documents have been indexed. Studies have shown the Google Scholar relevance algorithm to be heavily influenced by citation counts and language of publication. Google Scholar has been so heavily researched and is such a "black box" that more attention would seem to have diminishing returns, except in the area of coverage of and utility for arts and humanities research. Librarians may find these takeaways useful for working with or teaching Google Scholar:

- Little is known about coverage of arts and humanities by Google Scholar.
- Recent studies repeatedly find that in the sciences and social sciences Google Scholar covers as much if not more than library databases, has more recent coverage, and frequently provides access to full text without the need for library subscriptions.
- Although the number of studies is limited, Google Scholar seems excellent at retrieving known scholarly items compared with discovery tools.
- Using proper accent marks in the title when searching for non-English language items appears to be important.

[18] Google Scholar's blog notes that in January 2016, a change was made so "Scholar now automatically identifies queries that are likely to be looking for a specific paper" Technically speaking, "it tries hard to find the intended paper and a version that that particular user is able to read" https://scholar.googleblog.com/.

- Finding full text for non-English journal articles may require searching Google Scholar in the original language.
- While Google Scholar may include results from Google Books, it appears both tools should be used rather than assuming Google Books will appear in Google Scholar.
- While Google Scholar does include grey literature, these results do not usually rank highly.
- Google Scholar and Google must both be used to effectively search across institutional repository content.
- Free full text may be buried underneath the "All X versions" links because the publisher's web site is usually the dominant version presented to the user. The right-hand column links may help ameliorate this situation, but not reliably.
- Google Scholar is well-known in most academic communities and used regularly; however, it is seldom the only tool used, with scholars continuing to use other web search tools, library abstracts and indexes, and published web sites as well.
- Experts in writing systematic reviews recommend Google Scholar be included as a search tool along with traditional abstracts and indexes, using software to record the search process and results.
- For evaluating research impact, Google Scholar may be superior to Web of Science or Scopus, but using all three tools still seems necessary.
- As with any database, citation metadata should be verified against the publisher's data; with Google Scholar, publication dates should receive deliberate attention.
- When Google Scholar covers some of a major publisher's content, that does not imply it covers all of that publisher's content.
- Google Scholar Metrics appears to provide reliable journal rankings.

Research Agenda

This review of the literature also provides direction for future research concerning academic web search engines. Because this review focused on 2014-2016, researchers may need to review studies from earlier periods for methodological ideas and previous findings, noting that dramatic changes in search engine coverage and behavior can occur within only a few years.[19]

Across the studies, some general best practices were observed. When comparing the coverage of academic web search engines, their utility for establishing research impact, or other bibliometric studies, researchers should strongly consider using software such as Publish or Perish, and to design their research approach with previous methodologies in mind. Information scientists have charted a set of clear disciplinary methods; there is no need to start from scratch. Even when

[19] For example Ştirbu found that Google Scholar overlapped GeoRef by 57% and 62% (Ştirbu et al. 2015, 328), compared with a finding by Neuhaus in 2006 where Scholar overlapped with GeoRef by 26% (2006, 133).

performing a large-scale quantitative assessment such as (Kousha and Thelwall 2015), manually examining and discussing a subset of the sample seems helpful for checking assumptions and for enhancing the meaning of the findings to the reader. Some researchers examined the "top 20" or "top 10" results qualitatively (Kousha and Thelwall 2015), while others took a random sample from within their large-study sample (Kousha, Thelwall, and Rezaie 2011).

Academic search engines for arts and humanities research

Research into the use of academic web search engines within arts and humanities fields is sorely needed. Surveys show humanities scholars use both Google and Google Scholar (Inger and Gardner 2016; Kemman, Kleppe, and Scagliola 2013; Van Noorden 2014). During interviews of 20 historians by Martin and Quan-Haase (2016) concerning serendipity, five mentioned Google Books and Google Scholar as important for recreating serendipity of the physical library online. Almost all arts and humanities scholars search the Internet for researchers and their activities, and commonly expressed the belief that having a complete list of research activities online improves public awareness (Dagienė and Krapavickaitė 2016). Mays's (2015) practical advice and the few recent studies on citation impact of Google Books for these disciplines point to the enormous potential for this tool's use. Articles describing opportunities for new online searching habits of humanities scholars have not always included Google Scholar (Huistra and Mellink 2016).

Wu and Chen's interviews with humanities graduate students suggested their behavior and preferences were different from science and technology students, doing more known-item searching and struggling with "semantically ambiguous keywords" that retrieved irrelevant results (2014, 381). Platform preferences seem to have a disciplinary aspect: Hammarfelt's (2014) investigation of altmetrics in the humanities suggests Mendeley and Twitter should be included along with Google Scholar when examining citation impact of humanities research, while a 2014 Nature survey suggests ResearchGate is much less popular in the social sciences and humanities than in the sciences (Van Noorden 2014). In summary, arts and humanities scholars are active users of academic web search engines and related tools, but their preferences and behavior, and the relative success of Google Scholar as a research tool cannot be inferred from the vast literature focused on the sciences. Advice from librarians and scholars about the strengths and limitations of academic web search engines in these fields would be incredibly useful.

Specific examples of needed research, and related studies to reference for methodological ideas:

- Similar to the studies that have been done in the sciences, how well do academic search engines cover the arts and humanities? An emphasis on formats important to the discipline would be important (Prins et al. 2016).
- How does the quality of search results compare between academic search engines and traditional library databases for arts and humanities topics? To what extent can the user usefully accomplish her task? (Ruppel 2009)?

- To what extent do academic search engines support the research process for scholarship distinctive to arts and humanities disciplines (e.g. historiographies, review essays)?
- In academic search engines, how visible is the arts and humanities literature found in institutional repositories (Pitol and De Groote 2014)?

Specific aspects of academic search engine coverage

This review suggests that broad studies of academic search engine coverage may have reached a saturation point. However, specific aspects of coverage need additional investigation:

- **Grey literature:** Although Google Scholar's inclusion of grey literature is frequently mentioned as valuable, empirical studies evaluating its coverage are scarce. Additional research following the methodology of Haddaway (2015) could investigate the bibliographies of literature other than systematic reviews, investigate various disciplines, or use a sample of valuable known items (similar to Kousha, Thelwall, and Rezaie's (2011) methodology for books).
- **Non-Western, non-English language literature:** For further investigation of the repeated finding of non-Western, non-English language bias (Abrizah and Thelwall 2014; Cavacini 2015), comparisons to library abstracts and indexes would be helpful for providing context. To what extent is this bias present in traditional research tools? Hilbert et al. found the coverage of their sample increased for English language in both Web of Science and Scopus, and "to a lesser extent" in Google Scholar (2015, 260).
- **Books:** Any investigations of book coverage in Microsoft Academic and Google Scholar would be welcome. Very few 2014-2016 studies focused on books in Google Scholar, and even looking in earlier years turned up little research. Georgas (2015) compared Google with a federated search tool for finding books, so her study may be a useful reference. Kousha et al. (2011) found three times as many citations in Google Scholar than in Scopus to a sample of 1,000 academic books. The authors concluded "there are substantial numbers of citations to academic books from Google Books and Google Scholar, and it therefore may be possible to use these potential sources to help evaluate research in book-oriented disciplines" (Kousha, Thelwall, and Rezaie 2011, 2157).
- **Institutional Repositories:** Yang (2016) recommended that "librarians of digital resources conduct research on their local digital repositories, as the indexing effects and discovery rates on metadata or associated text files may be different case by case," and the studies found 2014-2016 show that IR platform and metadata schema dramatically affect discovery, with some IRs nearly invisible (Weideman 2015; Chen 2014; Orduña-Malea and López-Cózar 2015; Yang 2016) and others somewhat findable by Google Scholar (Lee et al. 2015; Obrien et al. 2016). Askey and Arlitsch (2015) have explained how Google Scholar's decisions regarding metadata schema can dramatically affect results.[20] Libraries who

[20] For example, Google's rejection of Dublin Core.

would like their institutional repositories to serve as social sharing platforms for research should consider conducting a study similar to (Martín-Martín et al. 2016b). Finally, a study of IR journal article visibility in academic web search engines could be extremely informative.

- **Full-text retrieval**: The indexing coverage of academic search engines relates to the retrieval of full text, which is another area ripe for more research studies, especially in light of the impressive quantity of full text that can be retrieved without user authentication. Johnson and Simonsen (2015) found that more of the engineering students they surveyed obtained scholarly articles from a free download or getting a PDF from a colleague at another institution than used the library's subscription. Meanwhile, libraries continue to pay for costly subscription resources. Monitoring this situation is essential for strategic decision-making. Quint (2016) and Karlsson (2014) have suggested strategies for libraries and vendors to support broader access to subscription full text through creative licensing and per-item fee approaches. Institutional repositories have had mixed results in changing scholars' habits (both contributors and searchers) but are demonstrably contributing to the presence of full text in the academic search engine experience. When will academic users find a good-enough selection of full-text articles that they no longer need the expanded full text paid for by their institutions?

Google Books

Similarly to Microsoft Academic, Google Books as a search tool also needs dedicated research from librarians and information scientists about its coverage, utility, and/or adoption. A purposeful comparison with other large digital repositories such as HathiTrust (https://www.hathitrust.org) would be a boon to practitioners and the public. While HathiTrust is transparent about its coverage (https://www.hathitrust.org/statistics_visualizations), specific areas of Google Books' coverage have been called into question. Weiss (2016) suggested a gap in Google Books exists from about 1915-1965 "because many publishers either have let it fall out of print, or the book is orphaned and no one wants to go through the trouble of tracking down the copyright owners" and found that copies in Google Books "will likely be locked down and thus unreadable, or visible only as a snippet, at best" (303). Has this situation changed since the court rulings concerning the legality of snippet view? Longitudinal studies in the growth of Google Books similar to (Harzing 2014) could illuminate this and other questions about Google Books's ability to deliver content. Uneven coverage of content types, geography, and language should be investigated. Mays noted a possible geographical imbalance within the United States (Mays 2015, 26). Others noted significant language and international imbalances, and large disciplinary differences (Weiss 2016; Abrizah and Thelwall 2014; Kousha and Thelwall 2015). Weiss and others suggest the implications of Google Books' coverage imbalance have enormous social implications: "Google and other [massive digital libraries] have essentially canonized the books they have scanned and contribute to the marginalization of those left unscanned" (301). Therefore more holistic quantitative investigations of the types of information in Google Books and possible skewness

would be welcome. Finally, Chen's study (2012) comparing the coverage of Google Books and WorldCat could be repeated to provide longitudinal information.

The utility of Google Books for research purposes also needs further investigation. Books are far more prevalently cited in Wikipedia than are research articles (Thelwall and Kousha 2015a). Examining samples of Wikipedia articles' citation lists for the prevalence of Google Books could reveal how dominant a force Google Books has become in that space. On a more philosophical level, investigating the ways Google Books might transform scholarly processes would be useful. Szpiech (2014) considered how the Google Books version of a medieval manuscript transformed his relationship with texts, causing a rupture "produced by my new power to extract words and information from a text without being subject to its order, scale, or authority" (78). He hypothesized readers approach Google Books texts as consumers, rather than learners, whereby "the critical sense of the gestalt" is at risk of being forgotten" (84). Have other researchers in experienced what he describes?

Microsoft Academic

Given the stated openness of Microsoft's new academic web search engine,[21] the closed nature of Google Scholar, and the promising findings of bibliometricians (Harzing 2016b; Harzing and Alakangas 2016a), librarians and information scientists should embark on a thorough review of Microsoft Academic with similar enthusiasm to which they approached Google Scholar. The search engine's coverage, utility for research, and suitability for bibliometric analysis[22] all need to be examined. Microsoft Academic's abilities for supporting scholarly social networking would also be of interest, perhaps using Ward et al. (2015) as a theoretical groundwork. The tool's coverage and utility for various disciplines and research purposes is a wide-open field for highly useful research.

Professional and Instructional Approaches Based on User Research

To inform instructional approaches, more study on user behavior is needed, perhaps repeating Herrera's (2011) study with Google Scholar and Microsoft Academic. In light of the recent focus on graduate students, research concerning the use of academic web search engines by undergraduates, community college students, high school students, and other groups would be welcome. Using an interview or focus group generates exploratory findings that could be tested through surveys with a larger, more representative sample of the population of interest. Studying searching behaviors has been common; can librarians design creative studies to investigate reading, engagement, and reflection when web search engines are used as part of the process? Is there a way to study whether the "Matthew Effect" (Antell et al. 2013, 281), the aging citation

[21] Microsoft's FAQ says the company is "adopting an open approach in developing the service, and we invite community participation. We like to think what we have developed is a community property. As such, we are opening up our academic knowledge as a downloadable dataset" and offers the Academic Knowledge API (https://www.microsoft.com/cognitive-services/en-us/academic-knowledge-api).

[22] See Jacsó (2011) for methodology.

phenomenon (Verstak et al. 2014; Martín-Martín et al. 2016a; Davis and Cochran 2015), or other epistemological hypotheses are influencing scholarship patterns? A bold study could be performed to examine differences in quality outcomes between samples of students using primarily academic search engines versus traditional library search tools. Exploratory studies in this area could begin by surveying students about their use of search tools for research methods courses or asking them to record their research process in a journal, and correlating the findings with their grades on the final research product.

Three specific areas of user research needed are the use of scholarly social network platforms, researcher profiles, and the influence of these on scholarly collaboration and research (Ward, Bejarano, and Dudás 2015, 178); the performance of Google's relatively new known-item search[23] (compared with Microsoft Academic's known-item search abilities), and searching in non-English languages. Regarding the latter, Albarillo's (2016) method which he applied to library databases could be repeated with Google Scholar, Microsoft Academic, and Google Books.

Finally, to continue their strong track record as experts in navigating the landscape of digital scholarship, librarians need to research assumptions regarding best practices for scholarly logistics. For example, searching Google for article titles plus the term "doi," then scanning the results list for ResearchGate was found by this study's author to most efficiently provide doi numbers: but is this a reliable approach? Does ResearchGate have sufficient accuracy to be recommended as the optimal tool for this task? What is the most efficient way for a scholar to locate full text for a citation? Are academic search engines' bibliographic citation management software export tools competitive with third-party commercial tools such as RefWorks? Another area needing investigation is the visibility of links to free full text in Google Scholar. Pitol and DeGroote found that 70% percent of the items in their study had at least one free full-text version available through a "hidden" Google Scholar version (2014, 603), and this author's work on this review article indicates this problem still exists – but to what extent? Also, when free full text exists in multiple repositories (e.g. ResearchGate, Digital Commons, Academic.edu), which are the most trustworthy and practically useful for scholars? Librarians should discuss the answers to these questions and be ready to provide expert advice to users.

CONCLUSION

With so many users opting to use academic web search engines for research, librarians need to investigate the performance of Microsoft Academic, Google Books, and of Google Scholar for the arts and humanities, and to re-think library services and collections in light of these tools' strengths and limitations. The evolution of web indexing and increasing free access to full text should be monitored in conjunction with library collection development. To remain relevant to

[23] Google Scholar's blog notes that in January 2016, a change was made so "Scholar now automatically identifies queries that are likely to be looking for a specific paper" Technically speaking, "it tries hard to find the intended paper and a version that that particular user is able to read" https://scholar.googleblog.com/.

modern researchers, librarians should continue to strengthen their knowledge of and expertise with public academic web search engines, full-text repositories, and scholarly networks.

BIBLIOGRAPHY

Abrizah, A., and Mike Thelwall. 2014. "Can the Impact of Non- Western Academic Books be Measured? An Investigation of Google Books and Google Scholar for Malaysia." *Journal of the Association for Information Science & Technology* 65 (12): 2498-2508. https://doi.org/10.1002/asi.23145.

Albarillo, Frans. 2016. "Evaluating Language Functionality in Library Databases." *International Information & Library Review* 48 (1): 1-10. https://doi.org/10.1080/10572317.2016.1146036.

Antell, Karen, Molly Strothmann, Xiaotian Chen, and Kevin O'Kelly. 2013. "Cross-Examining Google Scholar." *Reference & User Services Quarterly* 52 (4): 279-282. https://doi.org/10.5860/rusq.52n4.279.

Asher, Andrew D., Lynda M. Duke, and Suzanne Wilson. 2012. "Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources." *College & Research Libraries* 74(5):464-488. https://doi.org/10.5860/crl-374.

Askey, Dale, and Kenning Arlitsch. 2015. "Heeding the Signals: Applying Web Best Practices When Google Recommends." *Journal of Library Administration* 55 (1): 49-59. https://doi.org/10.1080/01930826.2014.978685.

Authors Guild. "Authors Guild v. Google." Accessed January 1, 2016, https://www.authorsguild.org/where-we-stand/authors-guild-v-google/.

Bartol, Tomaž, and Maria Mackiewicz-Talarczyk. 2015. "Bibliometric Analysis of Publishing Trends in Fiber Crops in Google Scholar, Scopus, and Web of Science." *Journal of Natural Fibers* 12 (6): 531. https://doi.org/10.1080/15440478.2014.972000.

Boeker, Martin, Werner Vach, and Edith Motschall. 2013. "Google Scholar as Replacement for Systematic Literature Searches: Good Relative Recall and Precision Are Not Enough." *BMC Medical Research Methodology* 13 (1): 1.

Bonato, Sarah. 2016. "Google Scholar and Scopus for Finding Gray Literature Publications." *Journal of the Medical Library Association* 104 (3): 252-254. https://doi.org/10.3163/1536-5050.104.3.021.

Bornmann, Lutz, Andreas Thor, Werner Marx, and Hermann Schier. 2016. "The Application of Bibliometrics to Research Evaluation in the Humanities and Social Sciences: An Exploratory Study using Normalized Google Scholar Data for the Publications of a Research Institute."

Journal of the Association for Information Science & Technology 67 (11): 2778-2789. https://doi.org/10.1002/asi.23627.

Boumenot, Diane. "Printing a Book from Google Books." One Rhode Island Family. Last modified December 3, 2015, accessed January 1, 2017. https://onerhodeislandfamily.com/2015/12/03/printing-a-book-from-google-books/.

Bøyum, Idunn, and Svanhild Aabø. 2015. "The Information Practices of Business PhD Students." *New Library World* 116 (3): 187-200. https://doi.org/10.1108/NLW-06-2014-0073.

Bramer, Wichor M., Dean Giustini, and Bianca M. R. Kramer. 2016. "Comparing the Coverage, Recall, and Precision of Searches for 120 Systematic Reviews in Embase, MEDLINE, and Google Scholar: A Prospective Study." *Systematic Reviews* 5(39):1-7. https://doi.org/10.1186/s13643-016-0215-7.

Cals, J. W., and D. Kotz. 2016. "Literature Review in Biomedical Research: Useful Search Engines Beyond PubMed." *Journal of Clinical Epidemiology* 71: 115-117. https://doi.org/10.1016/j.jclinepi.2015.10.012.

Carlson, Scott. 2006. "Challenging Google, Microsoft Unveils a Search Tool for Scholarly Articles." *Chronicle of Higher Education* 52 (33).

Cavacini, Antonio. 2015. "What is the Best Database for Computer Science Journal Articles?" *Scientometrics* 102 (3): 2059-2071. https://doi.org/10.1007/s11192-014-1506-1.

Chen, Xiaotian. 2012. "Google Books and WorldCat: A Comparison of their Content." *Online Information Review* 36 (4): 507-516. https://doi.org/10.1108/14684521211254031.

———. 2014. "Open Access in 2013: Reaching the 50% Milestone." *Serials Review* 40 (1): 21-27. https://doi.org/10.1080/00987913.2014.895556.

Choong, Miew Keen, Filippo Galgani, Adam G. Dunn, and Guy Tsafnat. 2014. "Automatic Evidence Retrieval for Systematic Reviews." *Journal of Medical Internet Research* 16 (10): 1-1. https://doi.org/10.2196/jmir.3369.

Ciccone, Karen, and John Vickery. 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. https://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/23845.

Conrad, Lettie Y., Elisabeth Leonard, and Mary M. Somerville. 2015. "New Pathways in Scholarly Discovery: Understanding the Next Generation of Researcher Tools." Paper presented at the Association of College and Research Libraries annual conference, March 25-27, Portland, OR. https://pdfs.semanticscholar.org/3cb1/315476ccf9b443c01eb9b1d175ae3b0a5b4e.pdf.

Dagienė, Eleonora, and Danutė Krapavickaitė. 2016. "How Researchers Manage their Academic Activities." *Learned Publishing* 29(3):155-163. https://doi.org/10.1002/leap.1030.

Davis, Philip M., and Angela Cochran. 2015. "Cited Half-Life of the Journal Literature." *arXiv Preprint arXiv:1504.07479*. https://arxiv.org/abs/1504.07479.

Delgado López-Cózar, Emilio, Nicolás Robinson-García, and Daniel Torres-Salinas. 2014. "The Google Scholar Experiment: How to Index False Papers and Manipulate Bibliometric Indicators." *Journal of the Association for Information Science & Technology* 65 (3): 446-454. https://doi.org/10.1002/asi.23056.

Erb, Brian, and Rob Sica. 2015. "Flagship Database for Literature Searching Or Flelpful Auxiliary?" *Charleston Advisor* 17 (2): 47-50. https://doi.org/10.5260/chara.17.2.47.

Fagan, Jody Condit, and David Gaines. 2016. "Take Charge of EDS: Vet Your Content." Presentation to the EBSCO Users' Group, Boston, MA, May 10-11.

Gehanno, Jean-François, Laetitia Rollin, and Stefan Darmoni. 2013. "Is the Coverage of Google Scholar Enough to be Used Alone for Systematic Reviews." *BMC Medical Informatics and Decision Making* 13 (1): 1. https://doi.org/10.1186/1472-6947-13-7.

Georgas, Helen. 2015. "Google vs. the Library (Part III): Assessing the Quality of Sources found by Undergraduates." *portal: Libraries and the Academy* 15 (1): 133-161. https://doi.org/10.1353/pla.2015.0012.

Giustini, Dean, and Maged N. Kamel Boulos. 2013. "Google Scholar is Not Enough to be Used Alone for Systematic Reviews." *Online Journal of Public Health Informatics* 5 (2). https://doi.org/10.5210/ojphi.v5i2.4623.

Gray, Jerry E., Michelle C. Hamilton, Alexandra Hauser, Margaret M. Janz, Justin P. Peters, and Fiona Taggart. 2012. "Scholarish: Google Scholar and its Value to the Sciences." *Issues in Science and Technology Librarianship* 70 (Summer). https://doi.org/10.1002/asi.21372/full.

Haddaway, Neal R. 2015. "The Use of Web-Scraping Software in Searching for Grey Literature." *Grey Journal* 11 (3): 186-190.

Haddaway, Neal Robert, Alexandra Mary Collins, Deborah Coughlin, and Stuart Kirk. 2015. "The Role of Google Scholar in Evidence Reviews and its Applicability to Grey Literature Searching." *PloS One* 10 (9): e0138237. https://doi.org/10.1371/journal.pone.0138237.

Hammarfelt, Björn. 2014. "Using Altmetrics for Assessing Research Impact in the Humanities." *Scientometrics* 101 (2): 1419-1430. https://doi.org/10.1007/s11192-014-1261-3.

Hands, Africa. 2012. "Microsoft Academic Search – http://academic.research.microsoft.com." *Technical Services Quarterly* 29 (3): 251-252. https://doi.org/10.1080/07317131.2012.682026.

Harper, Sarah Fletcher. 2016. "Google Books Review." *Journal of Electronic Resources in Medical Libraries* 13 (1): 2-7. https://doi.org/10.1080/15424065.2016.1142835.

Harzing, Anne-Wil. 2013. "A Preliminary Test of Google Scholar as a Source for Citation Data: A Longitudinal Study of Nobel Prize Winners." *Scientometrics* 94 (3): 1057-1075. https://doi.org/10.1007/s11192-012-0777-7.

———. 2014. "A Longitudinal Study of Google Scholar Coverage between 2012 and 2013." *Scientometrics* 98 (1): 565-575. https://doi.org/10.1007/s11192-013-0975-y.

———. 2016a. *Publish Or Perish*. Vol. 5. http://www.harzing.com/resources/publish-or-perish.

———. 2016b. "Microsoft Academic (Search): A Phoenix Arisen from the Ashes?" *Scientometrics* 108 (3): 1637-1647.https://doi.org/10.1007/s11192-016-2026-y.

Harzing, Anne-Wil, and Satu Alakangas. 2016a. "Microsoft Academic: Is the Phoenix Getting Wings?" *Scientometrics*: 1-13.

Harzing, Anne-Wil, and Satu Alakangas. 2016b. "Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison." *Scientometrics* 106 (2): 787-804. https://doi.org/10.1007/s11192-015-1798-9.

Herrera, Gail. 2011. "Google Scholar Users and User Behaviors: An Exploratory Study." *College & Research Libraries* 72 (4): 316-331. https://doi.org/10.5860/crl-125rl.

Higgins, Julian, and S. Green, eds. 2011. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 ed.: The Cochrane Collaboration. http://handbook.cochrane.org/.

Hilbert, Fee, Julia Barth, Julia Gremm, Daniel Gros, Jessica Haiter, Maria Henkel, Wilhelm Reinhardt, and Wolfgang G. Stock. 2015. "Coverage of Academic Citation Databases Compared with Coverage of Scientific Social Media." *Online Information Review* 39 (2): 255-264. https://doi.org/10.1108/OIR-07-2014-0159.

Hoffmann, Anna Lauren. 2014. "Google Books as Infrastructure of in/Justice: Towards a Sociotechnical Account of Rawlsian Justice, Information, and Technology." *Theses and Dissertations*. Paper 530. http://dc.uwm.edu/etd/530/.

———. 2016. "Google Books, Libraries, and Self-Respect: Information Justice Beyond Distributions." *The Library* 86 (1). https://doi.org/10.1086/684141.

Horrigan, John B. "Lifelong Learning and Technology." Pew Research Center, last modified March 22, 2016, accessed February 7, 2017, http://www.pewinternet.org/2016/03/22/lifelong-learning-and-technology/.

Hug, Sven E., Michael Ochsner, and Martin P. Braendle. 2016. "Citation Analysis with Microsoft Academic." *arXiv Preprint arXiv:1609.05354*.https://arxiv.org/abs/1609.05354.

Huistra, Hieke, and Bram Mellink. 2016. "Phrasing History: Selecting Sources in Digital Repositories." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 49 (4): 220-229. https://doi.org/10.1093/llc/fqw002.

Inger, Simon, and Tracy Gardner. 2016. "How Readers Discover Content in Scholarly Publications." *Information Services & Use* 36 (1): 81-97. https://doi.org/10.3233/ISU-160800.

Jackson, Joab. 2010. "Google: 129 Million Different Books have been Published." *PC World*, August 6, 2010. http://www.pcworld.com/article/202803/google_129_million_different_books_have_been_published.html.

Jacsó, P. 2008. "Live Search Academic." Peter's Digital Reference Shelf, April.

Jacsó, Péter. 2011. "The Pros and Cons of Microsoft Academic Search from a Bibliometric Perspective." *Online Information Review* 35 (6): 983-997. https://doi.org/10.1108/14684521111210788.

Jamali, Hamid R., and Majid Nabavi. 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.

Johnson, Paula C., and Jennifer E. Simonsen. 2015. "Do Engineering Master's Students Know What They Don't Know?" *Library Review* 64 (1): 36-57. https://doi.org/10.1108/LR-05-2014-0052.

Jones, Edgar. 2010. "Google Books as a General Research Collection." *Library Resources & Technical Services* 54 (2): 77-89. https://doi.org/10.5860/lrts.54n2.77.

Karlsson, Niklas. 2014. "The Crossroads of Academic Electronic Availability: How Well does Google Scholar Measure Up Against a University-Based Metadata System in 2014?" *Current Science* 107 (10): 1661-1665. http://www.currentscience.ac.in/Volumes/107/10/1661.pdf.

Kemman, Max, Martijn Kleppe, and Stef Scagliola. 2013. "Just Google It-Digital Research Practices of Humanities Scholars." *arXiv Preprint arXiv:1309.2434*. https://arxiv.org/abs/1309.2434.

Khabsa, Madian, and C. Lee Giles. 2014. "The Number of Scholarly Documents on the Public Web." *PloS One* 9 (5): https://doi.org/10.1371/journal.pone.0093949

Kirkwood Jr., Hal, and Monica C. Kirkwood. 2011. "Historical Research." *Online* 35 (4): 28-32.

Koler-Povh, Teja, Primož Južnic, and Goran Turk. 2014. "Impact of Open Access on Citation of Scholarly Publications in the Field of Civil Engineering." *Scientometrics* 98 (2): 1033-1045. https://doi.org/10.1007/s11192-013-1101-x.

Kousha, Kayvan, Mike Thelwall, and Somayeh Rezaie. 2011. "Assessing the Citation Impact of Books: The Role of Google Books, Google Scholar, and Scopus." *Journal of the American Society*

for Information Science and Technology 62 (11): 2147-2164. https://doi.org/10.1002/asi.21608.

Kousha, Kayvan, and Mike Thelwall. 2017. "Are Wikipedia Citations Important Evidence of the Impact of Scholarly Articles and Books?" *Journal of the Association for Information Science and Technology*. 68(3):762-779. https://doi.org/10.1002/asi.23694.

Kousha, Kayvan, and Mike Thelwall. 2015. "An Automatic Method for Extracting Citations from Google Books." *Journal of the Association for Information Science & Technology* 66 (2): 309-320. https://doi.org/10.1002/asi.23170.

Lee, Jongwook, Gary Burnett, Micah Vandegrift, Hoon Baeg Jung, and Richard Morris. 2015. "Availability and Accessibility in an Open Access Institutional Repository: A Case Study." *Information Research* 20 (1): 334-349.

Levay, Paul, Nicola Ainsworth, Rachel Kettle, and Antony Morgan. 2016. "Identifying Evidence for Public Health Guidance: A Comparison of Citation Searching with Web of Science and Google Scholar." *Research Synthesis Methods* 7 (1): 34-45. https://doi.org/10.1002/jrsm.1158.

Levy, Steven. "Making the World's Problem Solvers 10% More Efficient." Backchannel. Last modified October 17, 2014, accessed January 14, 2016, https://medium.com/backchannel/the-gentleman-who-made-scholar-d71289d9a82d.

Los Angeles Times. 2016. "Google, Books and 'Fair Use'." *Los Angeles Times*, April 19, 2016. http://www.latimes.com/opinion/editorials/la-ed-google-book-search-20160419-story.html

Martin, Kim, and Anabel Quan-Haase. 2016. "The Role of Agency in Historians' Experiences of Serendipity in Physical and Digital Information Environments." *Journal of Documentation* 72 (6): 1008-1026. https://doi.org/10.1108/JD-11-2015-0144.

Martín-Martín, Alberto, Juan Manuel Ayllón, Enrique Orduña-Malea, and Emilio Delgado López-Cózar. 2016a. "2016 Google Scholar Metrics Released: A Matter of Languages... and Something Else." *arXiv Preprint arXiv:1607.06260*. https://arxiv.org/abs/1607.06260.

Martín-Martín, Alberto, Enrique Orduña-Malea, Juan M. Ayllón, and Emilio Delgado López-Cózar. 2016b. "The Counting House: Measuring those Who Count. Presence of Bibliometrics, Scientometrics, Informetrics, Webometrics and Altmetrics in the Google Scholar Citations, ResearcherID, ResearchGate, Mendeley & Twitter." *arXiv Preprint arXiv:1602.02412*. https://arxiv.org/abs/1602.02412.

Martín-Martín, Alberto, Enrique Orduña-Malea, Juan Manuel Ayllón, and Emilio Delgado López-Cózar. 2014. "Does Google Scholar Contain All Highly Cited Documents (1950-2013)?" *arXiv Preprint arXiv:1410.8464*. https://arxiv.org/abs/1410.8464.

Martín-Martín, Alberto, Enrique Orduña-Malea, Juan Ayllón, and Emilio Delgado López-Cózar. 2016c. "Back to the Past: On the Shoulders of an Academic Search Engine Giant." *Scientometrics* 107 (3): 1477-1487. https://doi.org/10.1007/s11192-016-1917-2.

Martín-Martín, Alberto, Enrique Orduña-Malea, Anne-Wil Harzing, and Emilio Delgado López-Cózar. 2017. "Can we Use Google Scholar to Identify Highly-Cited Documents?" *Journal of Informetrics* 11 (1): 152-163. https://doi.org/10.1016/j.joi.2016.11.008.

Mays, Dorothy A. 2015. "Google Books: Far More Than Just Books." *Public Libraries* 54 (5): 23-26. http://publiclibrariesonline.org/2015/10/far-more-than-just-books/

Meier, John J., and Thomas W. Conkling. 2008. "Google Scholar's Coverage of the Engineering Literature: An Empirical Study." *The Journal of Academic Librarianship* 34 (3): 196-201. https://doi.org/10.1016/j.acalib.2008.03.002.

Moed, Henk F., Judit Bar-Ilan, and Gali Halevi. 2016. "A New Methodology for Comparing Google Scholar and Scopus." *arXiv Preprint arXiv:1512.05741*.https://arxiv.org/abs/1512.05741.

Namei, Elizabeth, and Christal A. Young. 2015. "Measuring our Relevancy: Comparing Results in a Web-Scale Discovery Tool, Google & Google Scholar." Paper presented at the Association of College and Research Libraries annual conference, March 25-27, Portland, OR. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2015/Namei_Young.pdf

National Institute for Health and Care Excellence (NICE). "Developing NICE Guidelines: The Manual." Last modified April 2016, accessed November 27, 2016. https://www.nice.org.uk/process/pmg20.

Neuhaus, Chris, Ellen Neuhaus, Alan Asher, and Clint Wrede. 2006. "The Depth and Breadth of Google Scholar: An Empirical Study." *portal: Libraries and the Academy* 6 (2): 127-141. https://doi.org/10.1353/pla.2006.0026.

Obrien, Patrick, Kenning Arlitsch, Leila Sterman, Jeff Mixter, Jonathan Wheeler, and Susan Borda. 2016. "Undercounting File Downloads from Institutional Repositories." *Journal of Library Administration* 56 (7): 854-874. https://doi.org/10.1080/01930826.2016.1216224.

Orduña-Malea, Enrique, and Emilio Delgado López-Cózar. 2014. "Google Scholar Metrics Evolution: An Analysis According to Languages." *Scientometrics* 98 (3): 2353-2367. https://doi.org/10.1007/s11192-013-1164-8.

Orduña-Malea, Enrique, and Emilio Delgado López-Cózar. 2015. "The Dark Side of Open Access in Google and Google Scholar: The Case of Latin-American Repositories." *Scientometrics* 102 (1): 829-846. https://doi.org/10.1007/s11192-014-1369-5.

Orduña-Malea, Enrique, Alberto Martín-Martín, Juan M. Ayllon, and Emilio Delgado López-Cózar. 2014. "The Silent Fading of an Academic Search Engine: The Case of Microsoft Academic

Search." *Online Information Review* 38(7):936-953. https://doi.org/10.1108/OIR-07-2014-0169.

Ortega, José Luis. 2015. "Relationship between Altmetric and Bibliometric Indicators Across Academic Social Sites: The Case of CSIC's Members." *Journal of Informetrics* 9 (1): 39-49. https://doi.org/10.1016/j.joi.2014.11.004.

Ortega, José Luis, and Isidro F. Aguillo. 2014. "Microsoft Academic Search and Google Scholar Citations: Comparative Analysis of Author Profiles." *Journal of the Association for Information Science & Technology* 65 (6): 1149-1156. https://doi.org/10.1002/asi.23036.

Pitol, Scott P., and Sandra L. De Groote. 2014. "Google Scholar Versions: Do More Versions of an Article Mean Greater Impact?" *Library Hi Tech* 32 (4): 594-611. https://doi.org/0.1108/LHT-05-2014-0039.

Prins, Ad A. M., Rodrigo Costas, Thed N. van Leeuwen, and Paul F. Wouters. 2016. "Using Google Scholar in Research Evaluation of Humanities and Social Science Programs: A Comparison with Web of Science Data." *Research Evaluation* 25 (3): 264-270. https://doi.org/10.1093/reseval/rvv049.

Quint, Barbara. 2016. "Find and Fetch: Completing the Course." *Information Today* 33 (3): 17-17.

Rothfus, Melissa, Ingrid S. Sketris, Robyn Traynor, Melissa Helwig, and Samuel A. Stewart. 2016. "Measuring Knowledge Translation Uptake using Citation Metrics: A Case Study of a Pan-Canadian Network of Pharmacoepidemiology Researchers." *Science & Technology Libraries* 35 (3): 228-240. https://doi.org/10.1080/0194262X.2016.1192008.

Ruppel, Margie. 2009. "Google Scholar, Social Work Abstracts (EBSCO), and PsycINFO (EBSCO)." *Charleston Advisor* 10 (3): 5-11.

Shultz, M. 2007. "Comparing Test Searches in PubMed and Google Scholar." *Journal of the Medical Library Association : JMLA* 95 (4): 442-445. https://doi.org/10.3163/1536-5050.95.4.442.

Stansfield, Claire, Kelly Dickson, and Mukdarut Bangpan. 2016. "Exploring Issues in the Conduct of Website Searching and Other Online Sources for Systematic Reviews: How Can We be Systematic?" *Systematic Reviews* 5 (1): 191. https://doi.org/10.1186/s13643-016-0371-9.

Ştirbu, Simona, Paul Thirion, Serge Schmitz, Gentiane Haesbroeck, and Ninfa Greco. 2015. "The Utility of Google Scholar when Searching Geographical Literature: Comparison with Three Commercial Bibliographic Databases." *The Journal of Academic Librarianship* 41 (3): 322-329. https://doi.org/10.1016/j.acalib.2015.02.013.

Suiter, Amy M., and Heather Lea Moulaison. 2015. "Supporting Scholars: An Analysis of Academic Library Websites' Documentation on Metrics and Impact." *The Journal of Academic Librarianship* 41 (6): 814-820. https://doi.org/10.1016/j.acalib.2015.09.004.

Szpiech, Ryan. 2014. "Cracking the Code: Reflections on Manuscripts in the Age of Digital Books." *Digital Philology: A Journal of Medieval Cultures* 3(1): 75-100. https://doi.org/10.1353/dph.2014.0010.

Testa, Matthew. 2016. "Availability and Discoverability of Open-Access Journals in Music." *Music Reference Services Quarterly* 19 (1): 1-17. https://doi.org/10.1080/10588167.2016.1130386.

Thelwall, Mike, and Kayvan Kousha. 2015b. "Web Indicators for Research Evaluation. Part 1: Citations and Links to Academic Articles from the Web." *El Profesional De La Información* 24 (5): 587-606.https://doi.org/10.3145/epi.2015.sep.08.

Thielen, Frederick W., Ghislaine van Mastrigt, L. T. Burgers, Wichor M. Bramer, Marian H. J. M. Majoie, Sylvia M. A. A. Evers, and Jos Kleijnen. 2016. "How to Prepare a Systematic Review of Economic Evaluations for Clinical Practice Guidelines: Database Selection and Search Strategy Development (Part 2/3)." *Expert Review of Pharmacoeconomics & Outcomes Research*: 1-17. https://doi.org/10.1080/14737167.2016.1246962.

Trapp, Jamie. 2016. "Web of Science, Scopus, and Google Scholar Citation Rates: A Case Study of Medical Physics and Biomedical Engineering: What Gets Cited and What Doesn't?" *Australasian Physical & Engineering Sciences in Medicine*. 39(4): 817-823. https://doi.org/10.1007/s13246-016-0478-2.

Van Noorden, R. 2014. "Online Collaboration: Scientists and the Social Network." *Nature* 512 (7513): 126-129. https://doi.org/10.1038/512126a.

Varshney, Lav R. 2012. "The Google Effect in Doctoral Theses." *Scientometrics* 92 (3): 785-793. https://doi.org/10.1007/s11192-012-0654-4.

Verstak, Alex, Anurag Acharya, Helder Suzuki, Sean Henderson, Mikhail Iakhiaev, Cliff Chiung Yu Lin, and Namit Shetty. 2014. "On the Shoulders of Giants: The Growing Impact of Older Articles." *arXiv Preprint arXiv:1411.0275*. https://arxiv.org/abs/1411.0275.

Walsh, Andrew. 2015. "Beyond "Good" and "Bad": Google as a Crucial Component of Information Literacy." In *The Complete Guide to Using Google in Libraries*, edited by Carol Smallwood, 3-12. New York: Rowman & Littlefield.

Waltman, Ludo. 2016. "A Review of the Literature on Citation Impact Indicators." *Journal of Informetrics* 10 (2): 365-391. https://doi.org/10.1016/j.joi.2016.02.007.

Ward, Judit, William Bejarano, and Anikó Dudás. 2015. "Scholarly Social Media Profiles and Libraries: A Review." *Liber Quarterly* 24 (4): 174–204.https://doi.org/10.18352/lq.9958.

Weideman, Melius. 2015. "ETD Visibility: A Study on the Exposure of Indian ETDs to the Google Scholar Crawler." Paper presented at ETD 2015: 18th International Symposium on Electronic Theses and Dissertations, New Delhi, India, November 4-6. http://www.web-

visibility.co.za/0168-conference-paper-2015-weideman-etd-theses-dissertation-india-google-scholar-crawler.pdf.

Weiss, Andrew. 2016. "Examining Massive Digital Libraries (MDLs) and their Impact on Reference Services." *Reference Librarian* 57 (4): 286-306. https://doi.org/10.1080/02763877.2016.1145614.

Whitmer, Susan. 2015. "Google Books: Shamed by Snobs, a Resource for the Rest of Us." In *The Complete Guide to using Google in Libraries*, edited by Carol Smallwood, 241-250. New York: Rowman & Littlefield.

Wildgaard, Lorna. 2015. "A Comparison of 17 Author-Level Bibliometric Indicators for Researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar." *Scientometrics* 104 (3): 873-906. https://doi.org/10.1007/s11192-015-1608-4.

Winter, Joost, Amir Zadpoor, and Dimitra Dodou. 2014. "The Expansion of Google Scholar Versus Web of Science: A Longitudinal Study." *Scientometrics* 98 (2): 1547-1565. https://doi.org/10.1007/s11192-013-1089-2.

Wu, Tim. 2015. "Whatever Happened to Google Books?" *The New Yorker*, September 11, 2015.

Wu, Ming-der, and Shih-chuan Chen. 2014. "Graduate Students Appreciate Google Scholar, but Still Find use for Libraries." *Electronic Library* 32 (3): 375-389. https://doi.org/10.1108/EL-08-2012-0102.

Yang, Le. 2016. "Making Search Engines Notice: An Exploratory Study on Discoverability of DSpace Metadata and PDF Files." *Journal of Web Librarianship* 10 (3): 147-160. https://doi.org/10.1080/19322909.2016.1172539.