This is a section of doi:10.7551/mitpress/12200.001.0001

The Open Handbook of Linguistic Data Management

Edited by: Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister

Citation:

The Open Handbook of Linguistic Data Management
Edited by: Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister
DOI: 10.7551/mitpress/12200.001.0001
ISBN (electronic): 9780262366076
Publisher: The MIT Press
Published: 2022

The MIT Press

12 Metrics for Evaluating the Impact of Data Sets

Robin Champieux and Heather L. Coates

1 Introduction

Research is a social activity, involving a complex array of resources, actors, activities, attitudes, and traditions (Sugimoto & Larivière 2018). There are many norms, including the sharing of new work in the form of books and journal articles and the use of citations and acknowledgments to recognize the influence of earlier work, but what it means to produce impactful scholarship is difficult to define and measure. The goals, methods, metrics, and utility of evaluating the impact of data sets are situated within this broader context of scholarly communication and evaluation. An understanding of the dynamic history, current practices, concepts, and critiques of measuring impact for and beyond research data sets can help researchers navigate the scholarly dissemination landscape more strategically and gain agency in regard to how they and their work are evaluated and described.

What is research impact? As Roemer and Borchardt (2015) describe, the concept involves two important ideas: the change a work influences and the strength of this effect. These effects can include, but are not limited to, advances in understanding and decision making, policy creation and change, economic development, and societal benefits. For example, rich documentation of an endangered language might lead to and support community and governmental revitalization efforts. However, the linkages between a specific scholarly product and its effects are rarely direct, there are disciplinary differences between how research is communicated and endorsed, and some outcomes take a very long time to manifest (Greenhalgh et al. 2016). This makes the assessment of research impact very labor intensive, even at a small scale, so researchers and decision makers often rely on data and metrics that are regarded as indicative of certain kinds of impact.

Many communities, particularly those outlined herein, have been interested in and have contributed to data-based methods for assessing research impact. Libraries and librarians have a long relationship with research evaluation. In 1927, Gross and Gross published an article in *Science* describing a method for counting and analyzing the citations among chemistry journals to guide library support for graduate level training. The journal impact factor (JIF), which will be discussed in more detail in section 2, is rooted in this history, as it was originally developed to guide decision making for journal indexing and library collection development. Similarly, as institutionalized support for research has grown, largely since the mid-twentieth century, governments, universities, and other organizations have sought to track and benchmark the progress and outcomes of their investments (Sugimoto & Larivière 2018). Why and how metrics are used by such evaluators will be discussed in section 4, but recent examples of this intense interest in impact evaluation include the creation and application of assessment frameworks, such as the Australian Research Council's Research Impact Principles and Framework (Australian Research Council 2019), intended to capture the complex routes by which knowledge is created, shared, and applied (Greenhalgh et al. 2016), and the development of new metrics, such as the Relative Citation Ratio (Hutchins et al. 2016).

The development of *scientometrics*, the quantitative study of scholarly literature, as a field is connected to and has influenced this history. The creation of the Science Citation Index (Garfield 1963) made it possible to access and use data about the citation relationships between publications at a scale allowing for quantitative investigation and measurement of the growth and impact of science. Most of the metrics used today for evaluation purposes, such as the h-index, an author-level, citation-based metric, are an output of such scientometric research (Sugimoto & Larivière 2018).

Finally, it is worth highlighting the roles and influence of individual researchers within this landscape. They are often the focus of impact questions with high-stakes consequences and can be both the objects of evaluation and consumers of metrics. As such, researchers are some of the most prominent advocates for improving the ways in which their work is assessed. The Declaration on Research Assessment (DORA), for example, was initiated at the 2012 Annual Meeting of the Society for Cell Biology and argues against evaluating the quality and impact of research based upon where it is published; DORA has since been signed by "more than sixteen thousand" individuals and organizations, including the Linguistic Society of America (San Francisco Declaration on Research Assessment [DORA] n.d.)

DORA and the ideas, concerns, and practices described within it reflect a changing scholarly communication system. How scholarship is created and shared, when and where it is discussed, and how it is endorsed is evolving. Most scholarly work is now published online, and technology changes have made it possible to share research products more immediately. Scholarly communication is also taking place in informal spaces, such as Twitter and research blogs, which can be tracked. Additionally, ideas about what should be considered valuable and citable research products have expanded to include such things as data sets, protocols, and software, along with an increasing interest in making these products more accessible and reusable. As Cronin and Sugimoto (2014:9) observe, "today the scholarly communication system is less linear, less rigid, and less opaque than before; both the process and end products are being transformed."

These developments have influenced the ways in which research impact is understood and evaluated. Data about research, particularly citation-based data, are increasingly used to administer science and scholarship, from hiring and promotion decisions to grant funding and university rankings (Aksnes, Langfeldt, & Wouters 2019). However, while the strength and validity of citation-based metrics are regularly examined and debated within the scientometric community, evaluation-based users of metrics are often unaware of an indicator's limitations, which can lead to misuse (Hicks et al. 2015). Several initiatives and resources have emerged to draw attention to and address this issue, such as the Leiden Manifesto (Hicks et al. 2015), which describes best practices for guiding metrics-based research evaluation, and the Metrics Toolkit (https://www.metrics-toolkit.org/), which uses a standardized schema to provide evidence-based information about metrics, including appropriate and inappropriate use cases. The availability of data that trace online and informal attention to and engagement with research products, and discussions about the types of impact these data reflect, are also increasingly a part of the modern impact landscape. Many data sources for online engagement have emerged and been integrated within traditional publishing and indexing platforms. Finally, the call to treat research data and other kinds of scholarly products as first-class research products has grown alongside the availability of new data sources for tracking and quantifying their impact.

Research impact is a complicated concept that exists within a highly dynamic scientific and scholarly communication ecosystem, and although this volume specifically deals with data and data management, metrics for data are best understood as components of research impact writ large. Understanding and participating in what can seem like a constant cycle of impact evaluation can be overwhelming; this chapter is designed to introduce foundational concepts for the understanding of impact and metrics, then specify how to apply these concepts to data. In section 2, we address fundamental metrics concepts, including the use of metrics for personal career advancement and by evaluators. Following that, we describe the practices that apply and are unique to research data and describe how to make use of them during data collection, analysis, and sharing. Our aims are (1) to help individual researchers develop metrics and evaluation literacy to make strategic decisions about how to share and disseminate research data and other scholarly products and (2) to equip them to be proactive participants in the research evaluation process who can effectively advocate for the value of their work.

2 Foundational metrics concepts

Here, we provide brief descriptions for the terms used throughout the rest of this chapter.

Activities: The ways in which users (scholars or machines) can interact with a scholarly product.

Altmetrics: The metrics and qualitative data that are calculated or derived from ways in which people interact online with scholarly content. These are often considered complementary or supplementary to traditional, citation-based metrics.

Awareness: The degree to which scholars within a particular group know about a product. This is strongly connected to the visibility and discovery of specific products.

Citation-based metrics: The metrics derived from citations. They are influenced by the content selected for indexing.

Dissemination: The process of distributing a scholarly product to relevant audiences.

Impact: The way(s) in which a scholarly product (or body of work) has affected the world. Often, this happens on a longer time scale than evaluation processes.

Indicators: An indicator is a quantified way of measuring a concept (e.g., impact, quality). The usefulness of an indicator depends on its explicit linkage to a concept and sufficient evidence that it is a valid measure for the concept (Gringas 2016). Gringas also suggests that indicators of different concepts should not be combined into a single composite indicator.

Normalization: A process for modifying data that are measured on different scales to produce values that share a common scale. This is necessary for valid comparisons of scholarly products.

Outcomes: The consequences or changes in a target that directly result from a research intervention, program, or discovery.

Quality: This is a broad concept that is operationalized differently across fields of research and which includes different elements such as novelty, creativity, and integrity depending on the field of research and methods employed. Data quality is often assessed based on its utility for a specified purpose. Similarly, the utility of a metric is influenced by the data on which it is calculated. No publication metric is a direct measure for quality because the standards for determining quality vary by individual, organization, discipline, setting, and many other factors. Despite its flaws, bibliometrics scholars (e.g., Gringas 2016; Waltman 2018) agree that there is no substitute for expert judgment via peer review in evaluating the quality and integrity of individual scholarly products or an author's body of work.

Scholarly product: The object the metric describes or applies to (e.g., journal, article, author).

Usage: The ways in which scholars use scholarly products. This encompasses viewing abstracts, skimming articles to filter for relevant information, reading online to scan for new ideas, downloading to save for later reference or citation, finding related articles and resources, storing and organizing into citation management databases for long-term use, text and data mining, and more.

Prior to the widespread adoption of network technologies for disseminating scholarly publications, metrics centered on analyzing the connections contained within citation data that exist between documents, authors, and journals. Many of the metrics not based on citations are simple counts of the ways in which users engage with products in an online environment. These are commonly referred to as altmetrics (short for "alternative metrics"). For example, in our current networked scholarly environment, a user can find, access, download, and export data about an item to their citation management tool. Each of these activities leaves a digital trace. Thus, we can see how many people on Twitter or how many media sources have mentioned the findings reported in a particular article. Altmetrics providers such as Altmetric and PlumX harvest and gather these digital traces for users to access and use. Other metrics are derived from more complex formulas; these include the JIF, the Eigenfactor Article Influence Score, and Field Normalized Citation Impact metrics. No single metric can reflect the full impact of an article, data set, or other product. In general, metrics are indirect indicators of impact, rather than direct measures.

A variety of terms are used to describe metrics about research. The term *publication metrics* is typically used to describe metrics for journal articles, books, and book chapters. As the processes for conducting research and disseminating the resulting knowledge have shifted online, it has become possible to capture interactions with a broader range of scholarly products. The field of *bibliometrics* uses statistical and network analysis techniques to analyze data related to scholarly products. This body of research often describes activities related to or relationships between three core objects—the scholarly product, authors, and the venue for dissemination, such as a journal or press. Many publication metrics are constructed from a few key attributes—citation, authorship, and institutional affiliation—and the relationships between them. *Altmetrics* is a more recent term coined in 2011 to describe a broader range of metrics derived from engagement with scholarly resources online and via social media (Priem et al. 2010). Altmetrics are often counts of specific activities that take place online, including views, downloads, endorsements (e.g., shares or likes, which vary by platform), and the use/reuse that takes place outside of the publication ecosystem described by journal indexes and citation databases. They are commonly seen as a supplement to citation metrics (i.e., metrics derived from citation data). The altmetrics manifesto (Priem et al. 2010) reflects frustration with the limitations of citation-based metrics and how they are used, particularly the JIF. Altmetrics are often siloed by social media platform; there is no metric, citation-based or altmetric, that is a reliable indicator of usage or engagement across all platforms. *Article-level metrics* are distinct from altmetrics; this term is used to describe both citation metrics and altmetrics about a particular article, as opposed to metrics about journals, such as the JIF. Finally, scientometrics is the "study of science, technology, and innovation from a quantitative perspective" (Leydesdorff & Milojević 2012), whereas *research impact metrics* is an umbrella term used to describe a range of metrics that speak to the impact of research. The rest of this chapter focuses on metrics about scholarly outputs or products, rather than metrics related to the inputs and outcomes of research.

Metrics are indicators, that is, indirect measures, rather than direct measures of impact. The many metrics just described exist because the data are available, not because they were designed to measure a particular concept for evaluation purposes (e.g., quality, impact, reputation). The strength of the relationship between a bibliometric indicator and the corresponding concept is crucial to the validity of the indicator (Sugimoto & Larivière 2018). Put simply, for an indicator to be valid, it must measure the concept that people expect it to and do it reasonably well. In the case of bibliometrics and altmetrics, the validity of the metric is often unclear to users, partly due to disciplinary differences in scholarship, publication, and the ways in which quality and impact are understood and valued. There is often a mismatch between the criteria that evaluators use (e.g., quality) and the metrics (e.g., citations). For instance, some administrators equate high impact research with publishing in a journal that has a comparatively high JIF. The JIF is "a measure reflecting the annual average (mean) number of citations to recent articles published in that journal" (Metrics Toolkit Editorial Board 2019). The JIF is problematic for a number of reasons. First, it is a measure about a journal as a whole, rather than individual articles. Second, it is not a valid predictor for individual articles because the distribution of citations is highly skewed. Thus, the mean of a highly skewed range of citation counts is not representative of most of the individual citation counts. Most of the articles published in a journal will not receive the number of citations suggested by the JIF. In some cases, like that of the journal *Nature*, 75% of articles receive fewer citations than the average citations reflected in the JIF (Larivière et al. 2016).

Citation count is a common example of a metric that appears to be simple and consistent on the surface, but which has a tremendous amount of hidden variance. There is variance in meaning; scholars cite for many reasons—to acknowledge awareness of existing literature, to contradict or argue with previous findings, to support their own methodological decisions, and more (Bornmann & Daniel 2008). Neither the fact that article A cited article B nor the count of citations for article B inherently conveys meaning about the quality of article B, nor the quality of the research that led to its creation. Cultural norms and practices for when to cite also differ by discipline or research field. There is also substantial variance in the data available for different disciplines, publication languages, and regions of the world (Sugimoto & Larivière 2018). However, citations and derived metrics can be construed as an indicator for impact on a scholarly audience, given the appropriate context (Bornmann & Daniel 2008). Context is a tremendously important factor to consider when interpreting and using metrics (Paul-Hus et al. 2017). Over the last 50 years, we have learned much about the value of publication metrics as well as the dangers in using them to evaluate individual scholars. It is important to keep those lessons in mind as we construct and implement metrics about data sets so that we do not recreate the same biases, limitations, and challenges.

3 Understanding metrics for career advancement

Scholars are the subject of constant evaluation for hiring, publishing, funding proposal reviews, annual reviews, promotion and tenure, and recognition through awards. As candidates for promotion and tenure, scholars are expected to tell a compelling story about how their scholarship has affected the world, in addition to meeting disciplinary and institutional standards for quality, productivity, and often funding. The ways in which institutions and the schools and departments contained within them operationalize concepts such as quality, prestige, and impact vary widely. Early career scholars are expected to navigate these systems and construct their research programs to meet these expectations. In some cases, the expectations of institutions diverge from those of peer institutions; thus, scholars face the additional challenge of building a research program that is successful within their current institution and still viewed favorably by those at other institutions. Evaluation of scholars for hiring, promotion and tenure, and awards is often centered on notions of prestige, reputation, excellence, and impact. Often, these concepts are not sufficiently operationalized for the evaluation task, leaving evaluators to interpret and apply them within the limits of their own knowledge, expertise, and available time. At best, this results in inconsistent evaluation; at worst, it significantly disadvantages scholars whose work does not fit the expectations of the evaluators.

A scholar's ability to build reputation and gain access to resources is crucial to their career success. Successful scholars are effective at crafting a compelling narrative about the value of their work in grant proposals, reports, publications, and their dossier. Though these narratives may incorporate concepts such as "high-impact" and "prestige" that are used and valued by their institutions and disciplines, they frequently do so in very vague terms and without evidence to support the claims. There is a significant disconnect between the rhetoric used by both scholars and evaluators in describing scholarship and the evidence available to support those claims.

Uncertainty and the challenges of evaluating scholars as just described have contributed to an overreliance on easily available publication metrics, such as the JIF. Though not intentional, metrics have increasingly been used as proxies for values such as productivity, impact, and quality. This is problematic for two reasons. First, many metrics can be gamed; over time, people change their behavior to improve their performance according to the metric. The behaviors that result in metric improvement may compete with and win out over behaviors leading to better quality research and more reproducible results. Second, many metrics are not fit to be used in evaluating scholars. A prime example is the JIF, which was initially designed to improve information retrieval and support collection development activities by librarians before the availability of electronic databases and the Internet. The JIF was not designed to be used in the evaluation of individual scholars, yet it frequently is.

3.1 Best practices for increasing the visibility of your scholarship

As the publishing ecosystem has changed, strategies for disseminating and discovering scholarship have also changed. These practices can enhance or constrain the metrics available for particular scholarly products. In the last decade, the following recommended practices have emerged to maximize dissemination of scholarly products and enable availability of publication metrics that support career advancement.

3.1.1 Own your scholarly profile First, scholars must create an online profile, or digital identity, that is independent from their institution. Institutional websites frequently change, and faculty may not be able to easily update or customize those sites. The core platforms recommended are ORCID (Open Researcher and Contributor ID; https://orcid.org/) and Google Scholar (https://scholar.google.com/). A structured profile, such as ORCID, can be exported to streamline and export bibliographic data for biosketches and other proposal and reporting requirements. Scholars who need a more robust digital identity or who engage with public and community audiences may want to consider creating a hosted site that helps their audiences understand and find their work. Core elements for these digital profiles are educational background, scholarly products, awards, and digital projects. It is crucial to provide clear and accessible information about scholarship that is geared toward a public audience. A digital profile can also increase the visibility of community-based service-learning projects or community-based participatory research. Links to scholarly profiles should be included in e-mail signatures, social media accounts, and other professional and institutional web pages to maximize discovery and visibility.

3.1.2 Share freely Once scholars have a digital profile that reflects their scholarly identity, the next step is to make the scholarly products as openly available as possible. Such products include journal articles as well as presentations, posters, book chapters, books, digital projects, data sets, and code or models, among other scholarly products. Because readers prefer one-click access, publications that are available openly are discovered and retrieved first. Greater access often leads to more citations (Piwowar et al. 2018; Colavizza et al. 2019; SPARC Europe, n.d.). Sharing freely does not necessarily require scholars to change their publishing practices; instead, they can utilize the expertise of librarians to identify which products can be shared openly and to deposit them into an appropriate repository or platform. The benefits offered by many institutional repositories include rich metadata and digital object identifier (DOI) registration to increase discovery, as well as a commitment to maintaining long-term access through archiving or preservation strategies. These services are typically available to affiliated scholars at no cost, as they are often subsidized by the library or research offices within the institution. The benefits of increased visibility and access take time to accrue, so creating a scholarly profile and sharing scholarly products freely are best completed as early as possible in a scholar's career trajectory.

3.1.3 Gather evidence When preparing a dossier for promotion and/or tenure, a funding proposal, or an award package, scholars need evidence of the impact and influence of their work (Alperin et al., chapter 13, this volume). For some scholars, these points of need may be the first time they think about research impact metrics. However, the availability of rich evidence in support of their case may depend on proactive planning and strategies for disseminating their scholarship. While a substantial proportion of the Western, English-language literature in the physical and biomedical sciences is represented in citation indexes, this is less true for the social science, humanities, and interdisciplinary literature. Scholars should aim to gather a range of publication metrics (e.g., citation counts, normalized citation metrics, JIF) as well as altmetrics (e.g., Tweets, blog mentions, policy mentions) as potential evidence. With altmetrics, it is often the case that the stories behind the numbers are more compelling than the numbers themselves. When possible, normalized metrics are recommended. Field normalized metrics are a "ratio between the actual citations received by a publication and the average number of citations received by all other similar publications" (Metrics Toolkit Editorial Board 2018). Field normalized metrics in particular enable easy comparison by evaluators because they are standardized according to the time of publication and research field. By intentionally disseminating their work, scholars can maximize its visibility, access, and use to generate evidence of attention, use, and impact.

3.1.4 Tell your story Metrics should not drive the story. Scholars should develop their story first and use relevant metrics to support their case. Using a mix of qualitative and quantitative evidence to support the case is more effective than relying on a single type or source of evidence. Few candidates are well served by limiting the metrics used to those available in platforms such as Dimensions, Journal Citation Reports, Scopus, and Web of Science. In the case of tenure and promotion, the short pretenure time frame combined with the time it takes for citations to accrue means that it is rare for quantitative publication metrics alone to make the case. Selected metrics should describe the scholar's work (e.g., article level metrics, altmetrics), rather than the merits of the publisher or journal. Ultimately, the scholar's case and the evaluation of it are about more than the metrics. It is up to the scholar to tell a compelling story and demonstrate that they have met the requirements. Likewise, evaluation consists of more than comparing numbers. The expertise of evaluators is a crucial component of research evaluation. Neither the quality of scholarship nor its impact can be summed up in a metric.

4 How metrics are used by evaluators

Within academic and research institutions, the use of research impact metrics has increased for both hiring and promotion decisions. Their use in promotion and tenure decisions tends to be more formalized and documented than in other processes such as hiring, perhaps because the policies and processes for promotion and tenure are more formalized. However, these metrics do not hold absolute meaning. Rather, their meaning is socially constructed within the context of institutional practices (Leydesdorff, Wouters, & Bornmann 2016). Faculty reviewers serving on promotion and tenure committees are tasked with evaluating their peers according to the standards and processes of their institution. They bring to this work implicit assumptions and expectations about the meaning of prestige, high-quality scholarship, and impact. These assumptions often place higher value on products such as journal articles, which are vetted through the trusted (but imperfect) process of peer review, with correspondingly less value placed on presentations, posters, data, software, and such. (Alperin et al., chapter 13, this volume). There are also disciplinary or professional beliefs about the relative value of specific journals, conferences, and presses. These beliefs often favor historically significant or long-lived organizations but are not necessarily informed by evidence or a clear understanding of the scholarly ecosystem in which they are created, disseminated, assessed, valued, and consumed. Such implicit expectations and beliefs become problematic when they conflict with the disciplinary culture of the scholar under review.

Reviewers use research metrics for a variety of reasons. Anecdotally, reliance on metrics appears to be heaviest when time constraints are a factor and reviewers are unfamiliar with the field or discipline of the scholar under review. At large research institutions where committees review a hundred or more promotion cases per year, it is not possible for reviewers to read a sample of the publications produced by a candidate during the review period; neither are they experts in each candidate's field. Scholarly norms such as authorship conventions and the specific products that are valued can vary widely, even within disciplines. Additionally, reviewers are generally not aware of the limitations of and biases in the data used to generate citation metrics and altmetrics (Leydesdorff, Wouters, & Bornmann 2016). Despite insufficient training and support for this work, reviewers must do the best they can with limited information and resources. Changing the metrics used and how they are understood requires confronting complex and implicit beliefs held by each scholar. Beliefs about the value of scholarly products, ways of disseminating those products, and the metrics that serve as indicators of their visibility, use, and impact on the world influence evaluation decisions. Change will also require stakeholders to acknowledge the value systems enforced by their institutional policies and procedures for hiring and promotion. Finally, reviewers must recognize their own knowledge gaps and be willing to seek out expertise and support.

5 How data are shared and used

Having covered the foundational concepts and uses of metrics, we turn now to metrics for data. Importantly, data metrics highly depend on how they are shared. Dissemination of data occurs along a spectrum, ranging from the informal, such as posting data on a lab website, to the formal, such as peer-reviewed data papers. Callaghan et al. (2012) describe two models of data sharing: *publication* and *Publication*. The former refers to when data are shared without a formal commitment to discovery, reuse, or long-term availability. An example in linguistics is the Corpus of Regional African American Language (CORAAL) project website. In contrast, *Published* data are stored, described, and organized in a way that enables potential users to find, access, understand, use, and cite them. Mature examples of this model include subject repositories such as the Inter-university Consortium for Political and Social Research or The Tromsø Repository of Language and Linguistics (TROLLing) archive. An emerging model is that of data journals such as *Scientific Data* and *GigaScience*, in which data are described with rich metadata, associated with a persistent ID, usually a DOI, and supported by an organizational commitment to maintain access for some period of time into the future. Some journals in linguistics have sections devoted to data sharing, such as *Phonological Data and Analysis* and *Language Documentation and Conservation,* though these may not necessarily publish data sets on their own.

Many options for sharing data produce high value for both data producers and consumers (see Callaghan's model in Van den Eynden 2011), including depositing data to specialized data centers or open-access repositories. Sharing data on a research group website is a low-barrier option for the research group, but generally does not support effective discovery and reuse. Other researchers are not likely to find the data, unless they know that the group has this practice of sharing. Thus, neither the producer nor the consumer gains much value from data shared in this way. Similarly, decisions about sharing have substantial impact on the availability of metrics for the data set. The available metrics often depend on a number of factors related to the platform selected for sharing. Does the platform use server logs or web analytics tools to capture usage? Does the platform provide rich descriptive information that can be harvested by search engines such as Google Dataset Search or scholarly indexes such as DataCite (e.g., structured metadata)? Is a persistent identifier provided to ensure access, even as platforms and web links change? Is there a commitment to data retention? Is there a commitment to data preservation? Can a user access the usage statistics for their data sets?

Though many platforms are beginning to address some of these issues, most are focused on easily implemented solutions that provide immediate value without addressing long-term access and preservation. One such example is the proliferation of DOIs for data sets. While they are crucial in our current ecosystem for discovery and citation, they do not ensure that the data they point to will be available in ten years, even if the DOI and its associated metadata are maintained. For a more in-depth discussion of archiving linguistic data, see chapter 7 (Andreassen, this volume).

6 Metrics about data sets

6.1 Data citation

Citation has a pivotal role in the scholarly ecosystem, from both sociocultural and technical perspectives. In large part, citation has value because scholars believe that it has value. It represents our notion of credit and attribution within the research community (Kurtz & Bollen 2010). Like the citation of publications, citation of data serves a variety of needs. Data citation is a keystone practice for data sharing, discovery, access, and rewarding data creators for their work (Altman et al. 2015). The recent position paper by Berez-Kroeker et al. (2018) emphasizes the importance of data citation in supporting reproducible research in the field of linguistics. Developing and implementing standards for data citation is one way to address the lack of rewards for the important, but largely invisible, preprocessing work of preparing data for use. Citations are a familiar incentive for data creators to share their data because they are already embedded in hiring practices, career advancement, and awards and funding processes. A wide range of scholarly infrastructure and cultural norms influence data citation practices due to the many functions they have. Conzett and De Smedt (chapter 11, this volume) describe more fully the practices and principles of data citation. Here, we will focus on the role of data citation in attribution, discovery, and access as it shapes the availability of metrics.

Systematic data citation is a key element of the social and scholarly infrastructure that fosters trust in research (Funk et al. 2019). The linkages created between publications and data sets, particularly by machine actionable citation metadata and DOIs, are critical for enabling verification and validation of research conclusions (CODATA-ICSTI Task Group on Data Citation Standards and Practices 2013). Effective reuse depends on the data sets being usable digital objects (Borgman 2012). However, this additional value comes at a relatively high cost to the original research team and their institution; for data to be reused, the data must be managed, described, disseminated, and preserved. Researchers must choose how to invest their limited time in activities that are rewarded, such as data collection, analysis, and writing for publication, or in activities that are not rewarded, such as preparation of shareable data. For those who do not consider this work a core component of their research workflows, providing credit and recognition in the form of citations to data sets may be a motivating reward. Such rewards are crucial for changing the overall culture of research to one that values quality over quantity, reproducibility over novelty.

6.1.1 Discovery Though discovery is enhanced through data sharing and citation, the act of sharing data or making it available does not necessarily make it visible and discoverable to potential users. Borgman (2012) describes the discoverability of data as the ability to "determine the existence of a set of data objects with specified attributes." Such attributes commonly include data creator, date of creation, method of creation, description of contents, and its representation. One challenge is that those attributes deemed relevant and the terminology used to describe them differs across research disciplines and communities. Maximizing the visibility and discovery of data requires that they be well described in a way that both machines and humans can understand; data must also be identified uniquely and must be persistent and available in usable open formats (CODATA-ICSTI Task Group on Data Citation Standards and Practices 2013). By taking into account some of the important aspects of data citation outlined here and in Conzett and De Smedt (chapter 11, this volume), discovery may be enhanced and the relevant metrics that can show impact will be easier to collect and more meaningful.

The following examples illustrate how discovery can be affected by decisions related to description, registration, and citation.

Example 1: CORAAL

> Citation: Kendall, Tyler, and Charlie Farrington. 2018. *The Corpus of Regional African American Language*. Version 2018.10.06. Eugene, OR: The Online Resources for African American Language Project. http://oraal.uoregon.edu/coraal.

Though this data set does not have a persistent identifier, the name is relatively unique. Thus, we can conduct a quick test of discovery. We performed phrase searching of "Corpus of Regional African American Language" in Dimensions, Google Scholar, Scopus, and Web of Science, as well as the subject database ProQuest Linguistics and Language Behavior Abstracts (LLBA). Our search results are in table 12.1.

The relevance of the results varies. In the subject database (LLBA), the phrase was found in nine items. Because the LLBA database does not include reference metadata, it is not possible to determine whether any of the items cite the CORAAL data set. In Dimensions, the results include several items that contain the phrase within the article title or in the references, many of which are not the data set itself. Google Scholar retrieved thirty-three items, of which twenty-four include the phrase in the title or full text. Both citation databases (Scopus and Web of Science) retrieved fewer results but with higher precision than Google Scholar. Without examining the articles, it is difficult to determine for all searches whether the retrieved items contain analyses based on the data set or simply discuss it. In this case, the creation of a DOI would uniquely identify the data set separately from articles citing or mentioning it. While the searches are not a perfect way to identify citation counts for a data source, it does lead us to those publications that might. Unfortunately, at least one article in the journal *American Speech* does analyze data from the CORAAL data set but does not cite it in the references. This may be an important use of the data set that is missed with some citation counting practices.

> Example 2: TROLLing data set "Replication Data for: Seeing from without, seeing from within: Aspectual differences between Spanish and Russian"

Table 12.1
Results of phrase searches for "Corpus of Regional African American Language"

Source	Items retrieved
Dimensions	14
Google Scholar	33
ProQuest Linguistics and Language Behavior Abstracts (LLBA)	9
Scopus	11
phrase search References	7
phrase search TITLE-ABS-KEY-AUTH	
Web of Science	5

> Citation: Janda, Laura A., and Antonio Fábregas. 2018. Replication Data for: Seeing from without, seeing from within: aspectual differences between Spanish and Russian. https://doi.org/10.18710/WR4Y0Q. DataverseNO, V1, UNF:6:v5Lkz2Vq1VjqBSIUTbLvrA== [fileUNF].

In example 2, we see that the repository TROLLing from DataverseNO network (Norway) has implemented a number of best practices that enable the creation of usage and citation metrics. The record contains rich metadata about the source, contact information, version information, and temporal coverage of the data, including terminology specific to the linguistic research community. A suggested citation is visible at the top of the record, which also includes a Universal Numerical Fingerprint (UNF) identifier. The UNF, similar to a checksum, allows users to confirm that the data set downloaded is the data set that was deposited. Users can also see the number of downloads for all files in this record.

The availability of features such as those described in the previous paragraph depends on the repository system selected and the decisions of those who manage it. There are many repository systems with different emphases on functionality (e.g., preservation, sharing), content (e.g., preprints, publications, research data), and users. Each system has limitations and strengths that can affect how discoverable data are by search engines and scholarly indexes. Generally, scholars need to be aware of the key features for discovery and ask questions of the repository managers about these features; these discovery features may help a scholar decide how to best increase the impact of their data and how to measure that impact. For more detail, see section 7.

6.1.2 Citation metrics Data citation as a practice is not new. The social sciences have had some tradition of data citation, at least since Dodd (1979) wrote about citing data from the General Social Survey. Developments in infrastructure and changing expectations for the dissemination of research products have created an environment in which data citation is possible and the benefits are clearer than ever before, as digital infrastructure makes counting citations and presenting metrics easier. So far, most data citation efforts are focused on prospective data citation, rather than looking backward to improve the record of data citations for prior publications (Mayernik et al. 2017). As such, sufficiently comprehensive studies of data citations are not yet possible, which limits the ability to put data metrics into context and develop normalized scores, for now. However, as data citation practices improve, so will the study of metrics examining data citation as a method of demonstrating impact.

Citation metrics for data are currently limited to raw citation counts; field normalized citation metrics are not yet possible for data sets. As with citations to publications, the context of the citation can be compelling, but it is not fully represented by metrics. The current scholarly communication infrastructure can capture the reuse of data in the form of citations when the references are contained in indexed publications such as journal articles, books, and book chapters, or in items shared via institutional or domain repositories. A central challenge for the visibility of data citations is the lack of a universal way to exchange information about links between publications and data. The availability of metadata about the linkages between publications and data varies widely; there is little exchange of usage data between systems at this point. The Scholix initiative seeks to improve the exchange of data citations between global aggregators such as DataCite, Crossref, and OpenAIRE to ensure that that metrics are accurate. Thus far, the initiative has developed an information model and a metadata schema for links. Scholix is developing exchange protocols and working to reach consensus across the community of publishers, data centers, and service providers to improve the exchange of links between data and literature. However, many other forms of knowledge sharing are not captured, including course syllabi, blogs, community and cultural heritage efforts, and projects that do not have a web presence. An ecosystem that excludes many of the ways in which data are reused limits our perspective on impact and tends to reward and reinforce the uses that we can easily measure, rather than what we value most.

> Example 3: TROLLing data set "Replication Data for: Automatic parsing as an efficient pre-annotation tool for historical texts"
>
> > Citation: Eckhoff, Hanne, and Aleksandrs Berdicevskis. 2016. Replication Data for: Automatic parsing as an efficient pre-annotation tool for historical texts. https://doi.org/10.18710/FERT42. DataverseNO, V1.

In example 3, we can see that the data set is related to a paper presented at the Workshop on Language Technology Resources and Tools for Digital Humanities. However, because the authors did not cite their own data in the references section of the paper and the proceedings are not indexed in any citation databases, the citation linkage is only visible on the TROLLing item record. Additionally, searching Google Scholar for the title retrieves the paper, rather than the data set. This potential confusion is one reason that some advocate for creating distinct titles for data sets and the publications based on them; this distinction would allow potential readers and data users to more easily find the item they are looking for, and for authors to more clearly understand who is citing their work and how.

From a scientometric perspective, citations between articles are an incomplete representation of the relationships between researchers (e.g., coauthorship) and research programs. Despite the importance of authorship in the discourse, no widely adopted metrics capture coauthorship contributions in a quantitative way; in part, this is due to the tremendous variety of authorship norms across disciplines and even journals. Some large research institutions deal with this issue by requiring authors to specify their contribution in percentages and getting confirmation of that effort from coauthors. There is a body of literature in scientometrics that describes coauthorship patterns and trends, but not for the purposes of evaluating individuals.

6.2 Altmetrics for data

The acceptance and use of altmetrics for data are highly localized to a few communities. A survey by Kratz and Strasser (2015) found that respondents valued citation and download counts over altmetrics and search rank. Indeed, the available metrics for data are thin and inconsistent, though expanding rapidly. Sharing data, whether via open or controlled mechanisms, is a prerequisite for data metrics to accrue (Costas et al. 2013). The altmetrics data are not yet reliable and consistent enough for scholars to trust that altmetrics for data sets can provide a reasonably accurate representation of their impact. Despite the widespread implementation of Altmetric and PlumX products by publishers, altmetrics are highly decentralized and generally dependent on the assignment of persistent identifiers such as DOIs to products.

The usage data, which the Kratz and Strasser (2015) survey indicates are of more interest to their respondents, are not typically captured by altmetrics vendors. These data, typically consisting of views, downloads, and user characteristics such as location, are siloed by the platforms on which they are created, requiring users to go to great lengths to gather and make sense of them. Even when such data are aggregated, it is currently impossible to compare or normalize them. In spite of these issues, usage metrics for data hold great potential for demonstrating engagement and use outside the scholarly environment. This potential depends on the development and adoption of common standards, data exchange or aggregation, and trust in the providers.

Key challenges for the adoption of altmetrics for data include (1) a lack of shared definitions and standards for gathering and reporting the data and (2) the decentralized storage of the data. In 2013–2016, the National Information Standards Organization led an initiative that explored a range of metrics beyond citations, including usage-based metrics, social media activities, and network behavioral analysis. The work resulted in a report, "Outputs of the NISO Alternative Assessment Metrics Project" (National Information Standards Organization 2016). The report identifies three themes or functions of metrics—showcase achievements, research evaluation, and discovery—and describes eight use cases that inform the recommendations. The report offers a helpful and relatively accessible overview and key recommendations related to data metrics.

7 A call to action

This chapter has described the complicated cultural and technical ecosystem in which data sharing and the evaluation of data sets as a scholarly product take place. This ecosystem includes many stakeholders whose knowledge and actions influence not only the success of individual researchers, but also the value of their scholarship. Thus, in conclusion, we have outlined practices that scholars, evaluators, academic institutions, publishers, and data repository managers can implement to facilitate good data sharing and its rewards.

Best practices for scholars:

- Develop a proactive strategy for disseminating your work to the most important audiences and systematically gather both qualitative and anecdotal evidence of use and impact.
- Use a combination of quantitative and qualitative evidence to support claims made in your case/story; do not let bibliometric data drive your case.
- Be aware of the limitations of the metrics used in evaluating you and your work, including your data (see the Metrics Toolkit; http://metrics-toolkit.org/).
- Use normalized metrics, statistically adjusted for publication date and field, when they are available. Examples include the Field Normalized Citation Impact (Field Weighted Citation Impact from Elsevier Scopus, Field Citation Ratio from Dimensions, or the Relative Citation Ratio for National Institutes of Health–funded work).
- Include the source for all metrics used (e.g., Scopus, Web of Science, Dimensions).
- Ask the editors of journals in which you publish to adopt standards and practices that support data citation, including contributing "Cited-by" metadata to Crossref.
  - Talk to the publishing arm of your professional society, if one exists, to ask the same and ask that they contribute to the Initiative for Open Citations (https://i4oc.org/).
- Get involved with groups such as the Research Data Alliance (RDA) Working Groups. RDA Working Groups typically have smaller committees that deal with data citation and measurement issues. For example, under the umbrella of the Joint RDA/World Data System Publishing Data Interest Group (https://rd-alliance.org/groups/rdawds-publishing-data-ig.html), several working groups have produced recommendations and guidance for issues related to the links between publications and data as well as metrics for publications and data. Two current working groups—Data Usage Metrics WG and the Data Citation WG—are especially relevant to this conversation. The Data Usage Metrics WG works closely with the Make Data Count project. Within RDA are several discipline-specific interest groups, including one for linguistics (https://rd-alliance.org/rda-disciplines/rda-and-linguistics), which works to apply data citation recommendations specifically to the field.

Best practices for evaluators:

- Be transparent and explicit regarding the assumptions about how indicators are valued and used within your institution and field—write them down, share them, and discuss them so that they are remembered during evaluation processes.

- Consider the value of scholarly products beyond journal articles, books or book chapters, and conference presentations. Refer to disciplinary guidance on the evaluation of these products, such as the Linguistic Society of America's Statement on Evaluation of Language Documentation for Hiring, Tenure, and Promotion (https://www.linguisticsociety.org/resource/statement-evaluation-language-documentation-hiring-tenure-and-promotion).
- Use evidence such as metrics to inform expert decision making, rather than using metrics to rank or compare individual scholars across disciplines or institutions.
- Consider evidence within the context of the candidate's case for advancement and disciplinary norms for scholarship.
- If expected to evaluate dissemination venues, choose the relevant data, put metrics into context, and use normalized metrics when possible.
- Balance consideration of dissemination venues with item- or article-level metrics.
- Ask your institution to provide training on research metrics for evaluation.
- In your evaluation toolkit/resource package, include information or point to resources that describe the appropriate uses, limitations, and biases of commonly used metrics (see the Metrics Toolkit).

Best practices for academic institutions:

- Provide relevant training to evaluators and sufficient support for hiring and review processes.
- Foster discussions about how institutional values should show up in evaluation processes; use data gathered from those discussions to inform guidance.
- Recognize that the university has shared responsibility for the long-term storage of scholarly records so that they remain available and accessible (Smith 2012).
- Commit sufficient resources to upholding the responsibility for ensuring the infrastructure that affiliated scholars need to do their work is maintained and well-functioning (Smith 2012).
- Make specific and explicit statements about the value of scholarly work such as data collection and data management in evaluation, hiring, and promotion guidance.

Best practices for journal editors and publishers:

- Assign unique identifiers, such as a DOI, to your articles.
- Adopt clear policies and standards regarding data availability for research published in your journal. If possible, require authors to share their data, either openly or via controlled mechanisms.
- Adopt policies and standards that promote data citation in alignment with the Austin Principles of Data Citation in Linguistics (https://site.uit.no/linguisticsdatacitation/austinprinciples/).
- Contribute Cited-by metadata to Crossref (https://www.crossref.org/services/cited-by/).

Best practices for managers of data repositories/archives/centers:

- Make it easier for users to judge relevance, accessibility, and reusability from the search summary. Some ways to do this include highlighting search terms, making the data availability and license clear, enabling a preview of the data, and displaying usage statistics (Wu et al. 2019).
- Make individual metadata records both readable and analyzable (Wu et al. 2019).
- Enable sharing and downloading of bibliographic references to data sets (Wu et al. 2019).
- Identify and/or aggregate duplicate metadata records for the same object (Wu et al. 2019).
- Make metadata indexed and searchable in ways compatible with Schema.org (Wu et al. 2019).

References

Aksnes, Dag W., Liv Langfeldt, and Paul Wouters. 2019. Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open* 9 (1): 2158244019829575. https://doi.org/10.1177/2158244019829575.

Altman, Micah, Christine Borgman, Mercè Crosas, and Maryann Martone. 2015. An introduction to the joint principles for data citation. *Bulletin of the Association for Information Science and Technology* 41 (3): 43–45. https://scholar.harvard.edu/mercecrosas/publications/introduction-joint-principles-data-citation.

Australian Research Council. 2019. Research impact principles and framework. *Australian Research Council*. March 27, 2019.

https://www.arc.gov.au/policies-strategies/strategy/research-impact-principles-framework.

Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, et al. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56 (1): 1–18. https://doi.org/10.1515/ling-2017-0032.

Borgman, Christine L. 2012. Why are the attribution and citation of scientific data important? In *Report from Developing Data Attribution and Citation Practices and Standards: An International Symposium and Workshop. National Academy of Sciences' Board on Research Data and Information*, ed. P. F. Uhlir, 1–8. Washington, DC: The National Academies Press. https://escholarship.org/uc/item/65b51130.

Bornmann, Lutz, and Hans-Dieter Daniel. 2008. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* 64 (1): 45–80. https://doi.org/10.1108/00220410810844150.

Callaghan, Sarah. 2012. Data citation in the earth and physical sciences. In *For Attribution—Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*, 49–54. Washington, DC: The National Academies Press. https://doi.org/10.17226/13564.

CODATA-ICSTI Task Group on Data Citation Standards and Practices. 2013. Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data. *Data Science Journal* 12:CIDCR1–75. https://doi.org/10.2481/dsj.OSOM13-043.

Colavizza, Giovanni, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, and Barbara McGillivray. 2019. The citation advantage of linking publications to research data, version 1. *ArXiv: 1907.02565 [cs.DL]*, July. http://arxiv.org/abs/1907.02565.

Costas, R., I. Meijer, Z. Zahedi, and P. F. Wouters. 2013. The value of research data: Metrics for data sets from a cultural and technical point of view. A Knowledge Exchange report. http://hdl.handle.net/1887/23586.

Cronin, Blaise, and Cassidy R Sugimoto. 2014. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge: MIT Press.

Dodd, Sue A. 1979. Bibliographic references for numeric social science data files: Suggested guidelines. *Journal of the American Society for Information Science* 30 (2): 77–82. https://doi.org/10.1002/asi.4630300203.

Funk, Cary, Meg Hefferon, Brian Kennedy, and Courtney Johnson. 2019. Trust and mistrust in Americans' views of scientific experts. *Pew Research Center*. August 2, 2019. https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/.

Garfield, Eugene. 1963. Science Citation Index. http://garfield.library.upenn.edu/papers/80.pdf.

Greenhalgh, Trisha, James Raftery, Steve Hanney, and Matthew Glover. 2016. Research impact: A narrative review. *BMC Medicine* 14 (1): 1–16. https://doi.org/10.1186/s12916-016-0620-8.

Gringas, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Abuses*. History and Foundations of Information Science. Cambridge: MIT Press. https://mitpress.mit.edu/books/bibliometrics-and-research-evaluation.

Gross, P. L. K., and E. M. Gross. 1927. College libraries and chemical education. *Science* 66 (1713): 385–389. https://doi.org/10.1126/science.66.1713.385.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. Bibliometrics: The Leiden manifesto for research metrics. *Nature News* 520 (7548): 429. https://doi.org/10.1038/520429a.

Hutchins, B. Ian, Xin Yuan, James M. Anderson, and George M. Santangelo. 2016. Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLOS Biology* 14 (9): e1002541. https://doi.org/10.1371/journal.pbio.1002541.

Kratz, John Ernest, and Carly Strasser. 2015. Researcher perspectives on publication and peer review of data. *PLOS One* 10 (2): e0117619. https://doi.org/10.1371/journal.pone.0117619.

Kurtz, Michael J., and Johan Bollen. 2010. Usage bibliometrics. *Annual Review of Information Science and Technology* 44 (1): 1–64. https://doi.org/10.1002/aris.2010.1440440108.

Larivière, Vincent, Véronique Kiermer, Catriona J. MacCallum, Marcia McNutt, Mark Patterson, Bernd Pulverer, Sowmya Swaminathan, Stuart Taylor, and Stephen Curry. 2016. A simple proposal for the publication of journal citation distributions. *BioRxiv* (September). https://doi.org/10.1101/062109.

Leydesdorff, Loet, and Staša Milojević. 2012. Scientometrics, version 1. *ArXiv:1208.4566 [cs.DL]*, August. http://arxiv.org/abs/1208.4566.

Leydesdorff, Loet, Paul Wouters, and Lutz Bornmann. 2016. Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators—a state-of-the-art report. *Scientometrics* 109 (3): 2129–2150. https://doi.org/10.1007/s11192-016-2150-8.

Mayernik, Matthew S., David L. Hart, Keith E. Maull, and Nicholas M. Weber. 2017. Assessing and tracing the outcomes and impact of research infrastructures. *Journal of the Association for Information Science and Technology* 68 (6): 1341–1359. https://doi.org/10.1002/asi.23721.

Metrics Toolkit Editorial Board. 2018. Field Normalized Citation Impact. *Metrics Toolkit*. https://www.metrics-toolkit.org/field-normalized-citation-impact/.

Metrics Toolkit Editorial Board. 2019. Journal Impact Factor. *Metrics Toolkit*. https://www.metrics-toolkit.org/journal-impact-factor/.

National Information Standards Organization. 2016. Outputs of the NISO Alternative Assessment Metrics Project. *National Information Standards Organization*. https://www.niso.org/publications/rp-25-2016-altmetrics.

Paul-Hus, Adèle, Nadine Desrochers, Sarah de Rijcke, and Alexander D. Rushforth. 2017. The reward system of science. *Aslib Journal of Information Management* 69 (5): 478–485. https://doi.org/10.1108/AJIM-07-2017-0168.

Piwowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. 2018. The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ* 6 (February): e4375. https://doi.org/10.7717/peerj.4375.

Priem, Jason, D. Taborelli, P. Groth, and Cameron Neylon. 2010. Altmetrics: A manifesto. October 26, 2010. *Altmetrics*. http://altmetrics.org/manifesto/.

Roemer, Robin Chin, and Rachel Borchardt. 2015. *Meaningful Metrics: A 21st Century Librarian's Guide to Bibliometrics, Altmetrics, and Research Impact*. Chicago: The Association of College and Research Libraries, a division of the American Library Association.

San Francisco Declaration on Research Assessment (DORA). n.d. *DORA*. https://sfdora.org/. Accessed August 12, 2019.

SCHOLIX. n.d. SCHOLIX: A Framework for Scholarly Link eXchange. http://www.scholix.org/. Accessed December 3, 2019.

Smith, Mackenzie. 2012. Institutional perspective on credit systems for research data. In *For Attribution—Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop*, ed. National Research Council. Washington, DC: National Academies Press. http://www.nap.edu/catalog/13564.

SPARC Europe. n.d. Open Access Citation Advantage (OACA) list. https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/oaca-list/. Accessed August 3, 2019.

Sugimoto, Cassidy R., and Vincent Larivière. 2018. *Measuring Research: What Everyone Needs to Know*. New York: Oxford University Press.

Van den Eynden, Veerle. 2011. *Managing and Sharing Data: Best Practice for Researchers*. Colchester: UK Data Archive.

Waltman, Ludo. 2018. Responsible metrics: One size doesn't fit all. Slides presented at the 23rd International Conference on Science and Technology, Leiden, the Netherlands. https://www.slideshare.net/LudoWaltman/responsible-metrics-one-size-doesnt-fit-all.

Wu, Mingfang, Fotis Psomopoulos, Siri Jodha Khalsa, and Anita de Waard. 2019. Data discovery paradigms: User requirements and recommendations for data repositories. *Data Science Journal* 18 (1): 3. https://doi.org/10.5334/dsj-2019-003.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Berez-Kroeker, Andrea L., editor. | McDonnell, Bradley James, editor. | Koller, Eve, editor. | Collister, Lauren B., editor.
Title: The open handbook of linguistic data management / edited by Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller and Lauren B. Collister.
Description: Cambridge, Massachusetts : The MIT Press, [2021] | Series: Open handbooks in linguistics series | Includes bibliographical references and index.
Identifiers: LCCN 2020044363 | ISBN 9780262045261 (hardcover)
Subjects: LCSH: Computational linguistics. | Natural language processing (Computer science) | Data mining.
Classification: LCC P98 .O64 2021 | DDC 410.285—dc23
LC record available at https://lccn.loc.gov/2020044363