Data Practices Ten Years Later: A New Review of Selected Publications by Crop Sciences Faculty

Sarah C. Williams

The Version of Record of this manuscript has been published and is available in the *Journal of Agricultural & Food Information*, 2022, 10.1080/10496505.2021.2013850

Abstract

This study explores data reuse and data sharing based on a review of the two most recent publications of each faculty member in crop sciences at the University of Illinois at Urbana-Champaign. Following the methodology of a study conducted in 2011, this 2021 study reveals current practices and compares present-day findings with those of the original study and other literature on data practices. In particular, this work addresses the variety of data sources used by scientists, data citation practices, common data sharing methods, and the challenges of determining the effects of funder policies on data sharing.

Keywords

Agriculture, crop sciences, data citation, data policy, data reuse, data sharing

Introduction

Data reuse and sharing practices are continually evolving for a variety of reasons. In recent years, scientists have indicated more willingness to share and reuse data (Tenopir et al., 2015). In some cases, they change their practices to fulfill disciplinary expectations, promote transparency, or facilitate reuse (Cooper, 2021; Williams et al., 2019). External factors also lead to changes in scientists' practices. In the past decade, many publishers have established data policies, especially since 2014 when Public Library of Science (PLoS) instituted a policy requiring data availability as a condition of publication (Briney et al., 2017). Publisher policies vary widely, ranging from strict requirements to general recommendations. Since the early 2000s, scientists also have needed to respond to a growing number of funder data policies, starting with the National Institutes of Health (NIH) in 2003 and the National Science Foundation in 2011 (Briney et al., 2017). Then, in 2013, the White House Office of Science and Technology Policy (OSTP) released a memo directed at federal agencies with large research and development expenditures to increase public access to publications and data resulting from federally funded research (Holdren, 2013). By 2016, agency responses to the OSTP memo ranged from newly implemented requirements to preliminary plans for compliance (Briney et al., 2017; Kriesberg, et al., 2017).

The OSTP memo applies to the United States Department of Agriculture (USDA), which is a major funder of agricultural research at land-grant institutions and beyond. The National Institute of Food and Agriculture (NIFA) is the USDA's extramural research agency. The USDA also enables research through the Agricultural Research Service (ARS), which is the USDA's in-house research agency with scientists based throughout the country, and through Hatch funding, which supports research programs at state agricultural experiment stations around the U.S. The USDA released its public access implementation plan in early 2015 (Adler, 2015; United States Department of Agriculture, 2014). For data, the USDA and NIFA took a phased approach, which can be pieced together from several sources (Table 1). To further facilitate access to USDA-funded research data, the National Agricultural Library developed the Ag Data Commons (https://data.nal.usda.gov/), a data catalog and repository to which USDA researchers and grant recipients can submit data. A beta version of the Ag Data Commons launched in 2015. In 2019, the Ag Data Commons relaunched with a new interface and new functionality (United States Department of Agriculture, National Agricultural Library, 2019).

Amid these changes, many studies investigated scientists' data practices. Tenopir et al. (2011) conducted a foundational study, in which over 1000 scientists responded to a survey about their data sharing practices and perceptions. Since then, Tenopir et al. (2015, 2020) published the results of two large follow-up surveys. In their most recent article, Tenopir et al. (2020) noted the importance of understanding scientists' actual behaviors in order to support data management best practices. Kim (2017), and Kim and Yoon (2017) also conducted broad surveys investigating scientists' data sharing and data reuse behaviors, respectively. In both publications, the authors encouraged future research to examine actual data practices, such as evidence of data sharing from data repositories and journals.

Other investigations focused more specifically on the data practices of researchers in agricultural sciences and related disciplines. Diekmann (2012) published foundational research using semi-structured interviews to explore data practices of agricultural scientists. Herold (2015) and Kerby (2015) analyzed publications of researchers in ecology, evolution, and natural resources and veterinary medicine, respectively, to examine their data practices. More recently, Cooper (2021) conducted a mixed-method study using a survey and semi-structured interviews to investigate data sharing attitudes and practices of plant sciences faculty. Moore et al. (in press) discussed common practices in agricultural data management, and from two Driving Innovation through Data in Agriculture (https://www.nal.usda.gov/legacy/ks/didag2018) workshops, they summarized key recommendations that could advance data-intensive research in agriculture.

In an earlier study, Williams (2012) analyzed publications of crop sciences faculty to explore actual data reuse and data sharing practices in the literature, and to highlight resources for acquiring and sharing research data. That original study yielded valuable information about data practices prior to significant developments such as most publisher and funder data policies. Ten years later, this research uses a similar methodology to reveal current practices of data reuse and sharing in the crop sciences. This work also compares the current findings to those of the original study and other literature on data practices. As circumstances continue to change, a solid understanding of data practices can inform how information professionals support agricultural researchers, who notably were among the most likely to ask librarians and data managers for help in the most recent survey by Tenopir et al. (2020).

Methods

To investigate data reuse and sharing practices, this study involved a thorough review of two recent articles by each member of the faculty of the Department of Crop Sciences, College of Agricultural, Consumer, and Environmental Sciences at the University of Illinois at Urbana-Champaign. To identify the faculty members, the author used the Crop Sciences website directory (https://cropsciences.illinois.edu/people/faculty/) for faculty and faculty affiliates, which includes specialized research faculty, ARS researchers, and research affiliates. This study excluded specialized teaching faculty, adjunct and emeritus faculty members, and a campus administrator with an appointment in Crop Sciences. Fifty-one faculty members fit these criteria, resulting in a total of 102 articles reviewed.

Candidate publications were identified via Illinois Experts (https://experts.illinois.edu/), a portal of University of Illinois faculty research and publications primarily populated from Elsevier's *Scopus* database. This study focused on scholarly research articles. Publications like editorials, book chapters, or plant registrations were not included. Two unique, recent publications were identified for each faculty member. If an article already was recorded for a co-author, the next most recent article was noted. Identification of the publications and the review process took place from April through June 2021.

The review process involved reading the entire article to find any mention of data reuse or sharing, including in the main text, data availability statement, supplementary materials, and references. Data reuse was defined as the use of data from an external source in the research project, not simply a citation to a source for background information or in a literature review. In each case, the author recorded the data source and reuse characteristics (e.g., use of a data citation, citation of a publication, or mention of the data source in the text of the article). For this study, data reuse did not include code, computer systems, software packages, physical specimens, or germplasm. Data sharing was defined as providing additional information beyond what was published in the article, not including code, computer systems, software packages, physical specimens, or germplasm. In each case, the author recorded the data sharing method and sharing characteristics, including where data sharing was described in the article and the file formats of any supplementary materials. The author also assigned each publication one or more research/publication type(s) to explore the data practices of different research areas. The research/publication types were: field, genetic, greenhouse/laboratory, model/method, and review. To be assigned a type, the reported research must include the type as a major component of the project based upon the author's reading of the article.

To provide context, the author noted the access status of the article on the journal website: Subscription, Gold Open Access (OA), Gold OA Hybrid (OA article in a subscription journal), and Bronze OA (freely available, not OA, article in a subscription journal). Piwowar et al. (2018) described the category Bronze OA to encompass publications that are free (not licensed OA) on publishers' websites. In addition, if an article provided funding information, the specific funding sources (e.g., USDA-NIFA, Gates Foundation) and general funding categories (e.g., academic institution, soybean checkoff) were captured. If a publication mentioned the year(s) during which the research occurred, that time frame was noted as well. Lastly, any other potentially pertinent information was collected in a notes field.

Results

A wide variety of journals published the 102 articles reviewed, and many of the articles were OA and published recently. The number of articles by publication year was: 2021 (46), 2020 (43), 2019 (10), 2018 (2), and 2009 (1). A majority of the publications (56%) were OA, with 37 Gold OA and 20 Gold OA hybrid. The articles appeared in 66 different journals (Table 2). These journals were the most frequent outlets: *Agronomy Journal* (6), *Science of the Total Environment* (5), *Crop Science* (4), *Plant Disease* (4), *Frontiers in Genetics* (3), *GCB Bioenergy* (3), *Pest Management Science* (3), and *Plant Physiology* (3). Interestingly, 11 of the 66 journals (17%) started publication in 2011 or after, including six MDPI titles and two Nature titles. Nearly a quarter of the journals (24%) started publication in 2010 or after, including eight from MDPI, three from Nature, and three from Frontiers.

The research/publication types and funding sources also varied widely. There were 15 different combinations of research/publication types, with the most common being greenhouse/laboratory research projects (19), genetic and greenhouse/laboratory projects (15), and model/method projects (14). Table 3 provides a full breakdown of the research/publication type combinations. A vast majority of the articles (89%) provided information about research funding, and 41 publications (40% of all) listed more than one funding source. Funding sources included U.S. federal agencies, government agencies outside of the U.S., academic institutions, state-level organizations, and the Gates Foundation. The most common funder was the USDA, comprised of NIFA grants and ARS and Hatch funding. Fifteen articles acknowledged U.S. Department of Energy (DOE) funding, and nine of these publications listed the same major DOE grant.

Data Reuse

Slightly more than half of the articles (55) reused data from an external source. Of these 55, 64% were Gold OA or Gold OA Hybrid, appearing in 39 different journals (Table 2). Analyzing data reuse by research/publication type, 73% of the research projects (22 of 30) with a major model/method component reused data, followed by 60% (25 of 42) with a major genetic component, and 53% (17 of 32) with a major field component. While 60% (3 of 5) of the review articles reused data, the sample size was small in comparison with the other research/publication types. Forty-three percent (22 of 51) of publications with a major greenhouse/laboratory component reused data, but of the articles with only a greenhouse/laboratory component (i.e., not in combination with another type), about a quarter (5 of 19) did so.

The 55 articles cited or mentioned dozens of data sources (Table 4). The most common source of data was research literature (23), but this took many forms. In some cases, the data clearly came from a published article, and at least one article specifically acknowledged and thanked authors for publishing their data. In other cases, researchers reused data and included a citation to a publication. However, the source of the data was unclear as to whether the reused data came directly from the cited publication and its supplementary materials, or from a data repository deposit connected to the publication. Another category of data sources was that of disciplinary data repositories, such as DDBJ (DNA DataBank of Japan), FungiDB (fungal and oomycete informatics resources), GenBank (National Center for Biotechnology [NCBI] genetic sequence database), GRASSIUS (grass regulatory information server), Panzea (maize diversity project), Phytozome (plant comparative genomics portal), Rice Genome Annotation Project, and SoyBase (soybean genetics and genomics portal). Another common category was that of sources for weather and climate data, which were used most frequently by field research projects and model/method research projects. The weather and climate data sources ranged from global (e.g., ERA5 from the European Centre for Medium-Range Weather Forecasts; Global Precipitation Climatology Project from the National Center for Atmospheric Research), to national (e.g., U.S. Climate Normals from the National Oceanic and Atmospheric Administration; Bangladesh Meteorological Department), to state (e.g., Illinois Climate Network from the Illinois State Water Survey; Climate Data for Minnesota from the Minnesota Department of Natural Resources), down to local (e.g., an established weather station near a research plot).

Data citation practices varied among, and in some cases within, the publications. In some articles, a data citation appeared for one source, while another data source only would be mentioned in the text; in other cases, a data source would be cited, while another would have a citation to a publication connected to its source. Data citations for weather and climate data appeared frequently. Another frequently cited source was the USDA Natural Resources Conservation Service Web Soil Survey (https://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm). It appeared in eight articles, with seven of those including a citation to the Web Soil Survey; the eighth instance was just a mention in the text of the article.

Despite including only 102 articles, this study uncovered some interesting data reuse situations. One research project utilized publicly available Long Term Ecological Research (LTER) arthropod data. However, the LTER data had never been gathered into a single dataset and inconsistencies existed between datasets. The researchers had to harmonize the datasets to facilitate analysis. Two other articles cited programs that digitize or extract data – WebPlotDigitizer and Data Thief – to create a more usable form for the researchers.

Data Sharing

Sixty-nine articles (68%) shared data beyond that published in the article. Of these, about two-thirds (64%) were OA (i.e., Gold OA or Gold OA Hybrid). The data sharing articles appeared in 50 different journals (Table 2). For almost all research/publication types, more than half of the articles shared data; the only exception was that of review articles, in which 20% (1 of 5) shared data. Otherwise, there were data shared in 86% (36 of 42) of articles with a major genetic component, 69% (35 of 51) of articles with a major greenhouse/laboratory component, 66% (21 of 32) of articles with a major field component, and 63% (19 of 30) of articles with a major model/method component.

There were three categories of data sharing methods: supplementary materials found on the journal website, data deposited in a repository, and by request. Some articles used more than one method. Data sharing via supplementary materials was most common, with 88% (61 of 69) of the publications reflecting this method. The vast majority of supplementary material files were Word documents (.docx), PDF files (.pdf), or Excel spreadsheets (.xlsx), with Word and PDF nearly even as most common format. Nearly 40% (27) of the articles shared data via a repository. Twelve articles (17%) shared data by request, and three articles (4%) shared data only by request. Over two dozen articles had a data availability statement, which commonly noted that data were available in a repository or by request. In addition to this statement, some Nature articles also included a "Nature Research Reporting Summary" at the end of the article to "improve the reproducibility of the work that we [Nature] publish" (https://www.nature.com/documents/nr-reporting-summary-flat.pdf). The summary includes a standardized statistics checklist, code and data availability, and information in study-specific fields (e.g., study sample size, sampling strategy, timing and spatial scale, replication).

Of the 27 articles that shared data via a repository, slightly more than half had a major genetic component. Field, greenhouse/laboratory, and model/method components were nearly equally represented (around 30% each) in these 27 articles. Authors noted this repository data sharing in various sections of the articles. Two sections were most commonly used: a data availability statement in 13 publications (10 with a stand-alone statement, and 3 with the data availability statement in the materials and methods section), and supplementary materials in nine articles. Notations regarding data sharing via repository appeared less frequently in the following sections: article information (1 article), materials and methods with no data availability statement (3), results (1), discussion (1), acknowledgements (2), and references (2).

There were a limited number of repositories utilized for data sharing (Table 4). More than one publication used the following repositories: FigShare (12), NCBI Sequence Read

Archive (5), GenBank (4), Mendeley Data Repository (3), Dryad Digital Repository (2), GitHub (2), and Illinois Data Bank (University of Illinois' data repository) (2). The use of FigShare was tied mainly to publisher partnerships; for eight articles, supplementary materials were available on the journal website but were linked to, and also downloadable from, FigShare. This connection between supplementary materials and FigShare explains why supplementary materials were such a common place to indicate data sharing, as described in the previous paragraph. While many of the repositories accept data from a range of disciplines, some authors used repositories with a clear agricultural focus, specifically CIMMYT (International Maize and Wheat Improvement Center) Dataverse and MaizeGDB. None of the articles mentioned sharing data via Ag Data Commons, the National Agricultural Library's catalog and repository for research data generated by USDA funding. To explore this further, for every article that noted USDA NIFA or ARS funding, the author searched the article DOI in Ag Data Commons to see if it might be a related article in any data records. None of the article DOIs yielded a result in Ag Data Commons.

In an analysis of some funder information as it relates to data sharing, nearly all (94%) of the data sharing articles listed at least one funding source. Of the 33 articles that did not share data beyond that published in the article, 79% listed at least one funding source. For some key funders, the percentage of articles in this study that shared data was: 100% of NSF funded articles (7 of 7), 87% of DOE funded articles (13 of 15), 80% of industry funded articles (4 of 5), 75% of USDA-NIFA funded articles (18 of 24), 67% of Gates Foundation funded articles (4 of 6), and 65% of USDA-ARS funded articles (11 of 17). However, the percentages for some funders are based on only a handful of articles.

Discussion

Findings from this study revealed some notable data reuse practices. Publications often were cited as a data source. Some data clearly came from a published article, but in other cases, it was unclear if the data came directly from the publication and its supplementary materials or from a data repository deposit connected to the publication. As Park and Wolfram (2017) noted, scientists may regularly cite publications instead of directly citing datasets. Yet, in a study by Imker et al. (2021), survey respondents reported obtaining data mostly from publications (i.e., copied from tables, figures, etc.), followed by supplementary materials, and then data repositories. Based on the findings of this current study, crop scientists appear to have similar practices – reusing data from publications, supplementary materials, and data repositories, but not always clearly citing these sources, which is a potential area of improvement.

Scientists seem regularly to reuse data from PDF files despite limitations. Publications were typically in PDF format, and as found in this and other studies (Herold, 2015; Kenyon et al., 2016; Williams, 2012), supplementary materials were often PDF files. Studies have articulated concerns about PDF files limiting data reuse (Kenyon et al., 2016; Williams, 2016). In a DATA (Discoverability, Accessibility, Transparency, and Actionability) assessment of data sharing, Van Tuyl and Whitmire (2016) gave a zero score for the Actionability criteria if data required more than minimal reprocessing or reformatting to be analyzed (e.g., PDF files). Nevertheless, scientists find ways to reuse data from PDF files. Imker et al. (2021) noted "the results of this study suggest that data within PDFs are more readily usable than commonly accepted, at least in certain circumstances" (p. 9). In a study by Kim and Yoon (2017), "perceived usefulness was found to be the most important factor influencing data reuse intentions, whereas the perceived effort of data reuse is not associated with scientists' intentions to reuse data" (p. 2715). In this study of researchers in the crop sciences, the publications citing programs to digitize or extract data illustrate researchers' willingness to reformat data deemed useful for their research. While data could be shared in formats that more easily facilitate reuse, data reuse from PDF files is still possible and perhaps even common. With existing research as a foundation (Imker et al., 2021; Kim & Yoon, 2017), this will be an important topic to continue exploring.

Not surprisingly, data reuse differed between research/publication types. As Pasquetto et al. (2017) wrote, "Data integration and reuse are much more difficult in areas where standards are unavailable or premature" (p. 5). In the original crop sciences study of ten years ago, genetic research articles most frequently reused data, especially from well-established data repositories like GenBank, SoyBase, and MaizeGDB (Williams, 2012). In the current study, the research/publication types that most frequently reused data were model/method research, followed by genetic research and review articles. Curty et al. (2017) found a large positive effect on data reuse behavior with researchers who use models and remote-sensing data, which seems analogous to some of the weather/climate models and data in this crop sciences study. Among review articles in this study, more than half reused data, and Imker et al. (2021) reported half of the review articles in their study contained data. While the number of review articles in both studies was small, these findings suggest review articles might be good candidates for future research on data reuse and sharing practices.

In both the 2011 and current studies, there were dozens of resources from which data were reused. The variety and scope of the data sources were remarkable, but especially notable was the lack of overlap between the data sources between the two studies. Only seven resources/organizations were a source of data in both studies: the Illinois State Water Survey, Phytozome, RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank, SoyBase, unspecified NCBI databases, but also specifically NCBI GenBank and NCBI RefSeq. The lists of data sources in both studies help illustrate the vast landscape of potential data resources and the diversity of research undertaken by crop sciences faculty.

Data citation practices are a potential area of improvement, as noted earlier and in previous studies (Curty et al., 2017; Dallmeier-Tiessen et al., 2014). The current study revealed inconsistencies, even within an article – a mix of formal data citations, data sources cited with a publication, and data sources only mentioned in the article text. Other authors (Park & Wolfram, 2017; Parsons et al., 2019; Pasquetto et al., 2017) also noted inconsistent data citation practices. Establishing clear and effective data citation practices is especially critical given the broad belief among scientists in the importance of data citation (Sayogo & Pardo, 2013; Tenopir et al., 2015; Tenopir et al., 2020). To help inform this effort, future research could examine closely the inconsistencies in data citation, to understand why scientists cite (or do not cite) data in certain ways. Publishers also have a significant role in data citation practices (Hrynaszkiewicz, 2019; Parsons et al., 2019).

Shifting to how data sharing is indicated in articles, the practices appear to be more, but not completely, consistent. Data availability statements are common, especially with certain publishers. In the 2011 study, no articles had a data availability statement, even articles published by PLoS, because the original study predated PLoS's 2014 data policy (Williams, 2012). Ten years later, this study identified over two dozen articles with a data availability statement. Compared to earlier studies (Park & Wolfram, 2017; Williams, 2016), the existence of supplementary materials was clearly indicated in the PDF version of the articles in this study. Despite these improvements, some data sharing information still was dispersed throughout the text of the publications, as found in other studies (Park & Wolfram, 2017; Van Tuyl & Whitmire, 2016). While progress clearly has been made, efforts to further standardize data sharing information in journal articles should continue.

Of the three data sharing categories in this study – in supplementary materials, via a repository, and by request, data shared in supplementary materials was most common, with 88% of the data sharing articles using this method. In the 2011 study, this method also was most prevalent, used by 96% of the data sharing articles (Williams, 2012). In the study by Tenopir et al. (2020), survey respondents representing agriculture and natural resources were among those who reported the most difficulty finding a suitable repository, which might help explain the heavy reliance on supplementary materials in crop sciences research; it will be interesting to see if the availability of the Ag Data Commons has an impact over time. Other studies of scientists' data sharing practices have also revealed supplementary materials as a common data sharing method (Federer, et al., 2018; Herold, 2015; Imker et al., 2021; Kerby, 2015). Frequent use of supplementary materials might be attributed in part to Kim's (2017) finding that "scientists perceive more career benefits when they share their data through journal supplements" (p. 880), as compared to sharing data via repositories or personal communications. Similarly, Pham-Kanter et al. (2014) found the net benefit (i.e., a combined measure of respondents' experience as data users and contributors) higher with supplementary materials than with repositories.

No articles in this study indicated data sharing via a faculty or departmental website, which is a positive finding. In the original study, only two articles mentioned sharing data this way, but in both cases, the data was not discoverable (Williams, 2012). Similarly, Van Tuyl and Whitmire (2016) identified 10 data management plans that stated data would be shared via a principal investigator or project website, but they discovered the data in only two of those instances. Given the small sample size of this study, no generalizations can be made about how frequently individual or institutional websites are used for data sharing. With known significant problems with persistence, hopefully this practice is waning. Future research could investigate the use of individual or institutional websites for data sharing over time.

The effects of funder requirements on data sharing are challenging to determine. Funding agencies announced data sharing and management policies over many years, starting with NIH in 2003 and NSF in 2011, but NIH and NSF are not major funders of agricultural research. Articles in this study commonly noted USDA-NIFA and DOE as funders; while the 2013 OSTP memo on public access applied to both agencies, the agencies' implementation timelines differed. The USDA data policy was phased in around 2015-2019 (United States Department of Agriculture, 2014), and the DOE data policy around 2014-2015 (United States Department of Energy, 2014). To get a sense of when the USDA-NIFA and DOE research projects in this study started, the author gathered the USDA-NIFA initial award fiscal year, which ranged from 2009 to 2020, and the DOE initial project start year, which ranged from 2016-2020. For further context, Tenopir et al. (2020) ran a survey in 2017-2018 that asked whether a scientists' primary funding agency required a data management plan, and the agricultural and natural resources participants responded: yes (39.3%), no (52.5%), and don't know (7.9%). Across all survey participants, 48.6% responded yes.

Some studies have relied on the passage of time to help ensure funder data policies would apply to the publications analyzed. Imker et al. (2021) studied articles published in 2015 partially because it was several years after the NIH and NSF requirements. Van Tuyl and Whitmire (2016) assessed articles published from 2012-2014 assuming they represented outputs of research funded after the 2011 NSF requirement. However, they noted this assumption of a quick turnaround was a limitation, since articles can be published years after research funding is received. To explore this, the author recorded any mention of the year(s) during which the research was conducted. The information is limited, as specific mentions of years were typical only in publications with a field research component. Nevertheless, among articles published between 2018-2021, the reported research years ranged from 2002 to 2020, which illustrates years can pass from a research project start date until its results are published.

Considering all of this, funder data requirements would have applied to some of the published research analyzed in this study, but certainly not all of it. Another complication is that 40% of all of the articles listed multiple funders, which means an article might need to follow multiple funder, and perhaps publisher, requirements regarding data sharing. Kerby (2015) also recognized the issue of multiple funders. This complex situation could benefit from additional research.

The current study has two main limitations. First, the sample size is relatively small, so the results may not be generalizable, especially when analyzing sample subsets. For example, some funders had a high percentage of data sharing articles (e.g., NSF, industry, Gates Foundation), but those percentages were based only on a handful of articles. Second, the author is not a domain expert in all the research areas represented in the articles. Despite a thorough review of each article, scholarly research is complex, and some authors write generally about data they reuse or share. As a result, the author may have miscategorized some articles as to their research/publication type(s), or may have misunderstood some references to data reuse or sharing. Van Tuyl and Whitmire (2016) reported this challenge in their research too.

This 2021 study reveals current practices of actual data reuse and sharing in publications of crop sciences faculty and provides a valuable comparison to a similar study conducted in 2011

(Williams, 2012). Articles noted data reused from dozens of data sources, with little overlap with the data sources in the original study. Data citation practices in both studies were very inconsistent, which suggests an area of future research and improvement. Compared to the original study, indications of data sharing were typically more apparent, especially with data availability statements, although some data sharing information was spread throughout the text of the articles. Data most commonly were shared via supplementary materials, which mirrored the original study and other studies (Federer, et al., 2018; Herold, 2015; Imker et al., 2021; Kerby, 2015). This study begins to explore the effects of funder policies on data sharing, but this complex issue will require future research. The passage of time also will help ensure that funder policies apply to published research, given phased policies and the sometimes lengthy time frame for publication of the research. Areas of future research abound, and time will tell if another review in 10 years might again be valuable.

Acknowledgements

I would like to thank Peg Burnette for providing feedback on a draft of this article, and Rodney Vance for clarifying some aspects of USDA-NIFA's data policy implementation.

References

Adler, P. (2015, February 20). AHRQ, NASA, USDA release plans for public access to funded research. *Association of Research Libraries Community Updates*. https://www.arl.org/news/ahrq-nasa-usda-release-plans-for-public-access-to-funded-research/

Briney, K., Goben, A., & Zilinski, L. (2017). Institutional, funder, and journal data policies. In L. R. Johnston (Ed.), *Curating research data, volume 1: Practical strategies for your digital repository* (pp. 61-78). Association of College and Research Libraries.

Cooper, K. A. (2021). Data sharing attitudes and practices in the plant sciences: Results from a mixed method study. *Journal of Agricultural & Food Information*, *22*(1-2), 37-58. https://doi.org/10.1080/10496505.2021.1891923

Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. *PLoS ONE*, *12*(12), Article e0189288. https://doi.org/10.1371/journal.pone.0189288

Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Matthews, B., Mele, S., Suhonen, J., & Wilson, M. (2014). Enabling sharing and reuse of scientific data. *New Review of Information Networking*, *19*(1), 16-43. https://doi.org/10.1080/13614576.2014.883936

Diekmann, F. (2012). Data practices of agricultural scientists: Results from an exploratory study. *Journal of Agricultural & Food Information*, *13*(1), 14-34. https://doi.org/10.1080/10496505.2012.636005

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. *PLoS ONE*, *13*(5), Article 0194768. https://doi.org/10.1371/journal.pone.0194768

Herold, P. (2015). Data sharing among ecology, evolution, and natural resources scientists: An analysis of selected publications. *Journal of Librarianship and Scholarly Communication*, *3*(2), Article eP1244. https://doi.org/10.7710/2162-3309.1244

Holdren, J. P. (2013). *Increasing access to the results of federally funded scientific research*. Office of Science and Technology Policy.

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Hrynaszkiewicz, I. (2019). Publishers' responsibilities in promoting data quality and reproducibility. In A. Bespalov, M. C. Michel, & T. Steckler (Eds.), *Good research practice in non-clinical pharmacology and biomedicine* (pp. 319-348). Springer.

Imker, H. J., Luong, H., Mischo, W. H., Schlembach, M. C., & Wiley, C. (2021). An examination of data reuse practices within highly cited articles of faculty at a research university. *Journal of Academic Librarianship*, *47*(4), Article 102369. https://doi.org/10.1016/j.acalib.2021.102369

Kenyon, J., Sprague, N., & Flathers, E. (2016). The journal article as a means to share data: A content analysis of supplementary materials from two disciplines. *Journal of Librarianship and Scholarly Communication*, *4*(0), Article p.eP2112. https://doi.org/10.7710/2162-3309.2112

Kerby, E. E. (2015). Research data practices in veterinary medicine: A case study. *Journal of eScience Librarianship*, *4*(1), Article e1073. https://doi.org/10.7191/jeslib.2015.1073

Kim, Y. (2017). Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods. *Information Processing and Management*, *53*(4), 871-885. https://doi.org/10.1016/j.ipm.2017.03.003

Kim, Y., & Yoon, A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *68*(12), 2709-2719. https://doi.org/10.1002/asi.23892

Kriesberg, A., Huller, K., Punzalan, R., & Parr, C. (2017). An analysis of federal policy on public access to scientific research data. *Data Science Journal*, *16*, Article 27. http://doi.org/10.5334/dsj-2017-027

Moore, E. K., Kriesberg, A., Schroeder, S., Geil, K., Haugen, I., Barford, C., Johns, E. M., Arthur, D., Sheffield, M., Ritchie, S. M., Jackson, C., & Parr, C. (in press). Agricultural data management and sharing: Best practices and case study. *Agronomy Journal*. https://doi.org/10.1002/agj2.20639

Park, H., & Wolfram, D. (2017). An examination of research data sharing and re-use: Implications for data citation practice. *Scientometrics*, *111*, 443-461. https://doi.org/10.1007/s11192-017-2240-2

Parsons, M. A., Duerr, R. E., & Jones, M. B. (2019). The history and future of data citation in practice. *Data Science Journal*, *18*, Article 52. http://doi.org/10.5334/dsj-2019-052

Pasquetto, I. V. Randles, B. M., & Borgman, C. L. (2017). On the reuse of scientific data. *Data Science Journal*, *16*, Article 8. http://doi.org/10.5334/dsj-2017-008

Pham-Kanter, G., Zinner, D. E., & Campbell, E. G. (2014). Codifying collegiality: Recent developments in data sharing policy in life sciences. *PLoS ONE*, *9*(9), Article e108451. https://doi.org/10.1371/journal.pone.0108451

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ*, *6*, Article e4375. https://doi.org/10.7717/peerj.4375

Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, *30*(Supplement 1). S19-S31. https://doi.org/10.1016/j.giq.2012.06.011

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, El., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, *10*(8), e0134826. https://doi.org/10.1371/journal.pone.0134826

Tenopir, C., Rice, N. M. Allard, S., Baird, L., Borycz, J.., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientist worldwide. *PLoS ONE*, *15*(3), Article e0229003. https://doi.org/10.1371/journal.pone.0229003

United States Department of Agriculture. (2014, November 7). Implementation plan to increase public access to results of USDA-funded scientific research. https://www.usda.gov/sites/default/files/documents/USDA-Public-Access-Implementation-Plan.pdf

United States Department of Agriculture, National Institute of Food and Agriculture. (2015, April). Data management plan for NIFA-funded research projects. https://nifa.usda.gov/sites/default/files/resource/Data_Managagement_Plan_NIFA_research_Apr_2015.PDF

United States Department of Agriculture, National Agricultural Library. (2019, August 5). Ag Data Commons relaunch – new look, feel, and function. *Ag Data Commons News*. https://data.nal.usda.gov/ag-data-commons-relaunch-new-look-feel-and-function

United States Department of Agriculture, National Institute of Food and Agriculture. (2019a, September 18). *Implementation of data management plan* [Video]. YouTube. https://www.youtube.com/watch?v=uQpdHdkiD5Q

United States Department of Agriculture, National Institute of Food and Agriculture. (2019b, October 1). *Research terms and conditions*. https://nifa.usda.gov/sites/default/files/resource/NIFA-19-012-NIFA-Research-Agency-Specific.pdf

United States Department of Agriculture, National Institute of Food and Agriculture. (2021, October 29). *Data management plan for NIFA-funded research, education, and extension projects*. https://nifa.usda.gov/resource/data-management-plan-nifa-funded-research-projects

United States Department of Energy. (2014, July 24). Public access plan. https://www.energy.gov/sites/default/files/2014/08/f18/DOE_Public_Access%20Plan_FINAL.pdf

Van Tuyl, S., & Whitmire, A. L. (2016). Water, water, everywhere: Defining and assessing data sharing in academia. *PLoS ONE*, *11*(2), Article e0147942. https://doi.org/10.1371/journal.pone.0147942

Williams, S. C. (2012). Data practices in the crop sciences: A review of selected faculty publications. *Journal of Agricultural & Food Information*, *13*(4), 308-325. https://doi.org/10.1080/10496505.2012.717846

Williams, S. C. (2016). Practices, policies, and persistence: A study of supplementary materials in crop science journals. *Journal of Agricultural & Food Information*, *17*(1), 11-22. https://doi.org/10.1080/10496505.2015.1120213

Williams, S. C., Farrell, S. L., Kerby, E. E., & Kocher, M. (2019). Agricultural researchers' attitudes toward open access and data sharing. *Issues in Science and Technology Librarianship*, 91. https://doi.org/10.29173/istl4

Table 1. USDA-NIFA phased approach to public access to data

Time Period	**Event**	**Source(s)**
Fiscal Year 2015	NIFA posted its first data management plan (DMP) guidance for a DMP pilot with several research and integrated programs	(United States Department of Agriculture, National Institute of Food and Agriculture, 2015)
Fiscal Year 2019	NIFA requested DMPs for all competitive grant programs	(United States Department of Agriculture, National Institute of Food and Agriculture, 2019a, 2021)
Fiscal Year 2020	NIFA Research Terms and Conditions updated to require investigators to make their data available as stated in their DMPs	(United States Department of Agriculture, National Institute of Food and Agriculture, 2019b)

Table 2. Journals included in this study

Journal Title	**Total Number of Articles (n = 102)**	**Number of Data Reuse Articles (n = 55)**	**Number of Data Sharing Articles (n = 69)**
Acta Horticulturae	1	0	0
Agronomy	2	1	2
Agronomy Journal	6	4	3
Animals	1	0	1
Applied Engineering in Agriculture	1	0	0
Applied Soil Ecology	1	1	1
Aquacultural Engineering	1	1	1
Biofuels, Bioproducts and Biorefining	1	1	0
Biological Control	1	0	0
Biomass and Bioenergy	1	0	1
BMC Genomics	1	1	1
Canadian Journal of Plant Pathology	1	0	1
Crop Protection	1	1	0
Crop Science	4	3	3
Current Plant Biology	1	1	1
Ecological Informatics	1	1	1
Environmental Microbiology	1	1	1

Environmental Science and Technology	1	1	1
Evolutionary Bioinformatics	1	1	1
Field Crops Research	1	1	0
Florida Entomologist	1	0	0
Frontiers in Environmental Science	1	0	1
Frontiers in Genetics	3	3	3
Frontiers in Plant Science	2	0	2
G3: Genes, Genomes, Genetics	2	2	2
GCB Bioenergy	3	3	3
Genes	2	2	2
Geoderma	1	0	1
Insects	1	1	0
International Journal of Food Microbiology	1	0	0
Journal of Agricultural and Food Chemistry	1	0	1
Journal of Chemical Ecology	1	0	1
Journal of Cleaner Production	1	1	0
Journal of Ecology	1	0	1
Journal of Economic Entomology	2	0	1
Journal of Experimental Botany	2	1	1
Journal of General Plant Pathology	1	1	0

Journal of General Virology	1	1	1
Life	1	1	1
Microorganisms	1	1	1
Molecules	1	0	1
Nature Communications	1	1	1
Nature Ecology and Evolution	1	1	1
New Phytologist	2	2	2
Nucleic Acids Research	1	0	1
Pathogens	1	0	1
Pest Management Science	3	1	1
Plant Biotechnology Journal	1	1	1
Plant Breeding	2	0	1
Plant, Cell & Environment	1	0	1
Plant Disease	4	1	1
Plant Journal	2	2	2
Plant Physiology	3	1	1
Plant Physiology and Biochemistry	1	1	1
Plants	1	1	1
PLoS Genetics	1	0	1
PLoS ONE	2	1	2
Precision Agriculture	1	0	0
Remote Sensing	1	0	0
Sarhad Journal of Agriculture	1	0	0

Science of the Total Environment	5	4	4
Scientific Reports	2	0	2
Soil and Tillage Research	1	1	1
Viruses	1	1	1
Weed Science	2	0	0
Weed Technology	2	0	0

Table 3. Number of articles by research/publication type

Research/Publication Type	**Total Number of Articles (n = 102)**	**Number of Data Reuse Articles (n = 55)**	**Number of Data Sharing Articles (n = 69)**
Field	10	4	6
Genetic	10	6	10
Greenhouse/Laboratory	19	5	11
Model/Method	14	13	8
Review	5	3	1
Field & Greenhouse/Laboratory	5	4	4
Field & Genetic	3	2	2
Field & Model/Method	5	1	3
Genetic & Greenhouse/Laboratory	15	7	13
Genetic & Model/Method	3	3	3
Greenhouse/Laboratory & Model/Method	2	0	0
Field & Genetic & Greenhouse/Laboratory	5	2	3
Field & Genetic & Model/Method	1	1	1

Genetic & Greenhouse/Laboratory & Model/Method	2	1	2
Field & Genetic & Greenhouse/Laboratory & Model/Method	3	3	2

Table 4. Data repositories, systems, and organizational websites included in this study

Data repositories, systems, and organizational websites	**Number of articles that reused data from these resources**	**Number of articles that shared data via these resources**
APSIM (Agricultural Production Systems Simulator)	1	
ARG-ANNOT (Antibiotic Resistance Gene Annotation)	1	
Argonne National Laboratory Repository (Argonne Scientific Publications)		1
Bangladesh Meteorological Department	1	
BETYdb (Biofuel Ecophysiological Traits and Yields Database)	1	
CIMMYT (International Maize and Wheat Improvement Center) Research Data & Software Repository Network		1
Climate Data for Minnesota	1	
CoGe (Comparative Genomics)	1	
Colorado Climate Center	1	
Corn Nitrogen Rate Calculator	1	
CyVerse Discovery Environment		1
DDBJ (DNA Data Bank of Japan)	1	

DOE (Department of Energy) ESS-DIVE (Environmental Systems Science Data Infrastructure for a Virtual Ecosystem		1
EBI (European Bioinformatics Institute) UniProtKB/Swiss-Prot	1	
ECMWF (European Centre for Medium-Range Weather Forecasts) ERA5	2	
Ensembl Genome Browser	1	
EPA (Environmental Protection Agency) National Recommended Water Quality Criteria Tables	1	
Dryad		2
FigShare		12
FungiDB (Fungal & Oomycete Informatics Resources)	1	
Genomes to Fields	1	
GISAID (global health)	1	
GitHub	1	2
Global Human Footprint Index	1	
GRASSIUS (Grass Regulatory Information Server)	1	
Illinois Data Bank		2

Illinois State Water Survey/Illinois Climate Network	6	
International Wheat Genome Sequencing Consortium	1	
Iowa State University Soil Moisture Network	1	
MaizeGDB (Maize Genetics and Genomics Database)		1
Mendeley Data		3
NASA (National Aeronautics and Space Administration) Global Land Data Assimilation System	1	
NASA MODIS (Moderate Resolution Imaging Spectroradiometer) Terra+Aqua Combined Land Cover	1	
NCAR (National Center for Atmospheric Research) Global Precipitation Climatology Project	1	
NCBI (National Center for Biotechnology Information) unspecified databases	4	
NCBI GenBank	6	4
NCBI GEO (Gene Expression Omnibus)		1

NCBI RefSeq (Reference Sequence Database)	1	
NCBI SRA (Sequence Read Archive)		5
NOAA (National Oceanic and Atmospheric Adminstration) U.S. Climate Normals	1	
NSF (National Science Foundataion) LTER (Long Term Ecological Research) data	1	
ORNL (Oak Ridge National Laboratory) Daymet (Daily Surface Weather and Climatological Summaries)	1	
ORNL ISLSCP II (International Satellite Land Surface Climatology Project)	1	
Panzea (maize diversity project)	2	
Percent Tree Cover, global version	1	
Phyllis (database for biomass and waste)	1	
Phytozome (plant comparative genomics portal)	4	1
Random Forest Modelling of the Lake Erie Microbial Community	1	

RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank	2	
Rice Genome Annotation Project	1	
SILVA Ribosomal RNA Database	1	
SoyBase (Soybean Genetics and Genomics Database)	4	
Sugarcane Genome Hub	1	
TRY Plant Trait Database	1	
University of Illinois Herbicide Evaluation Program database	1	
U.S. Climate Data [usclimatedata.com]	1	
USDA (United States Department of Agriculture) AMS (Agricultural Marketing Service) Plant Variety Protection Office	1	
USDA NASS (National Agricultural Statistics Service) Crop Frequency Layers	1	
USDA NASS Cropland Data Layer	1	
USDA NASS QuickStats/data	3	
USDA NRCS (Natural Resources Conservation Service) Gridded National Soil Survey Geographic Database	1	

USDA NRCS Web Soil Survey	8	
USGS (U.S. Geological Survey) National Water Information System	1	
Weather Company [ibm.com/weather]	1	
WHO (World Health Organization) Global Health Observatory	1	
WorldClim [worldclim.org]	1	
Worldometer [worldometers.info]	1	