

JLSC

ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | <http://jls-public.org>

Volume 3, Issue 2 (2015)

Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications

Philip Herold

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), eP1244. <http://dx.doi.org/10.7710/2162-3309.1244>

External Data or Supplements:

Herold, Philip, 2015, "Appendices to "Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications"", <http://dx.doi.org/10.7910/DVN/5BOGES>, Harvard Dataverse



© 2015 Herold. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications

Philip Herold

*Research & Learning Director for Agricultural, Biological, and Environmental Sciences,
University of Minnesota-Twin Cities*

INTRODUCTION Understanding the differing data management practices among academic disciplines is an important way to inform existing and emerging library research support and services. This paper reports findings from a study of data sharing practices among ecology, evolution, and natural resources scientists at the University of Minnesota. It examines data sharing rates, methods, and disciplinary differences and discusses the characteristics of researchers, data, methods, and aspects of data sharing across this group of disciplines. **METHODS** Data sharing practices are investigated by reviewing the two most recently published research articles ($n=155$) for each faculty member ($n=78$) in three departments at a single large research university. All mentions of data sharing in each publication were pursued in order to locate, analyze, and characterize shared data. **RESULTS** Seventy-two of 155 (46%) articles indicated that related research data was publicly shared by some method. The most prevalent method for data sharing was via journal websites, with 91% of data sharing articles using this method. Ecology, evolution, and behavior scientists shared data at the highest rate (70% of their articles), contrasting with fisheries, wildlife, and conservation biologists (18%), and forest resources (16%). **DISCUSSION** Differences between data sharing practices may be attributable to a range of influences: funder, journal, and institutional policies; disciplinary norms; and perceived or real rewards or incentives, as well as contrasting concerns, cost, or other barriers to sharing data. **CONCLUSION** Study results suggest differential approaches to data services outreach based on discipline and research type and support the need for education and influence on both scientist and journal practices.

Received: 02/27/2015 Accepted: 06/05/2015

Correspondence: Philip Herold, 94a Magrath Library, 1984 Buford Ave, St Paul, MN 55108, herol008@umn.edu



© 2015 Herold. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

IMPLICATIONS FOR PRACTICE

1. Data associated with published research articles was shared in almost half (46%) of the instances studied, indicating authors' willingness to share and opportunity for data-related education or influence.
2. The prevalent method for sharing data among ecology, evolution, and natural resources scientists is through journal websites (91%), often not adhering to preservation best practices; this method poses an open question about persistence.
3. There are important differences in data sharing rates related to research type, making disciplinary generalizations difficult, but those about research methods/types more useful.
4. Despite overlap between the types of science, there are data sharing practice differences between natural resources, and ecology and evolution scientists that suggest librarians take differential approaches to service and outreach to these disciplines.

INTRODUCTION

With an increasing number of public and private funding agency requirements for data management plans to be included as part of research grant proposals, there is an accompanying trend among libraries to develop library services that support research data management. Data management planning, data curation, data literacy, and metadata expertise are areas in which many academic libraries either have had, or are establishing, expert roles and services in support of their academic communities' research enterprise (Shaffer, 2013). Librarians are gaining expertise in aspects of research data and its management needs, and this necessarily includes understanding disciplinary differences in data practices, including data sharing practices (Thessen & Patterson, 2011).

The rationale for increased data sharing in the sciences was made prominent and explicit by the National Research Council (2003). This was followed by more specific recommendations that focused on data sharing in natural resources, ecology, and evolution in papers by Parr and Cummings (2005); Borgman, Wallis, and Enyedy (2007); Moore, McPeck, Rausher, Rieseberg, and Whitlock (2010); and more recently, Hampton et al. (2013). While these works discussed the need for, and potential beneficial impacts of, increased data sharing in these life sciences disciplines, few studies have produced evidence from measures of current practice. Hampton et al. (2013) analyzed sample of 100 ecology research papers associated with National Science Foundation grants and concluded that ecologists lack a culture of data sharing. But more study is needed to understand the disciplinary culture and practices around data sharing. This study provides insights into the data sharing practices of natural resources, ecology, and evolution scientists using quantitative and descriptive research methods.

The sciences include a broad range of applied and basic biological sciences research, and there are many differences in the cultures of the different disciplines that constitute the life sciences (Thessen & Patterson, 2011). Even within life and earth sciences, there are important cultural differences between disciplines in the realm of scholarly communications. Some disciplines, such as physics and astronomy, have active pre-print cultures, for example, with an established system of informal peer review for manuscripts that precedes formal peer review as part of the publishing process (Jamali & Nicholas, 2009). So it is true with data practices that disciplinary culture plays an important role in how researchers manage and share their data (Akers & Doty, 2013; Key Perspectives, 2010). Within each scientific discipline, the methods by which data are produced, the common types of data produced, data characteristics (e.g., size, format, complexity), metadata standards, attitudes towards data sharing and re-use, data repositories and other places where the data is shared or stored, are all important facets of data management practices that librarians will need to know as they work with researchers.

This paper reports on a study of data sharing practices among ecology, evolution, and natural resources scientists at the University of Minnesota. This was accomplished by examining a systematic selection of faculty research publication evidence. The study examines who shares their research data, how much they share, how and where it is shared, what is shared, and it discusses the characteristics of researchers, data, methods, and aspects of data sharing across this group of disciplines.

LITERATURE REVIEW

Among the key drivers that have helped move libraries toward new roles and services in support of data management and curation, the National Institutes of Health (NIH) (2003) and National Science Foundation (NSF) (2011) requirements for data management plans (DMPs) as part of grant proposals have paved the way for the other agencies to follow with their own requirements. DMP requirements have greatly increased the expectation that researchers create a plan for their data, which may include description, storage, sharing and accessibility that will permit study replication and data re-use. On many university campuses the requirements have helped to catalyze library data management programs.

The trend towards open science and open data preceded many funder requirements and was encouraged from many directions, including from the National Research Council (2003), which laid a foundation of ten recommendations for the life sciences community that encouraged sharing of publication-related data and materials. Parr and Cummings (2005) argued that increased data sharing in ecology, evolution, conservation biology, and animal behavior could speed scientific advancement in these fields, in addition to facilitating study replication and being economically sound practice.

McCain (1995) examined the role of journals as gatekeepers with important influence on data sharing practices in her review of 850 natural science, medical, and engineering journal policies, finding that 132 have some policy statement concerning aspects of data sharing. She noted sizeable gaps among journal policies and was not able to gauge the effectiveness of these policies.

Arzberger et al. (2004), members of an Organisation for Economic Co-operation and Development (OECD) group, advocated for greater data openness, defining key issues involved with data sharing and access and proposing a path forward that depends heavily on governmental policy development to advance increased data sharing.

Uhlir and Schröder (2007) focused on the need for open data to advance international science, describing many benefits of openness that include diversity of analysis and opinion, automated discovery, result verification, testing new hypotheses and methods of analysis, the creation of new data sets and knowledge from combined data, and the promotion of interdisciplinary research. They examined the policy frameworks that will promote greater openness, and described the then-emerging open access models.

Borgman et al. (2007) used habitat ecology as a case study to explore the role of data in a field moving from analog to digital science. They explored a range of data issues, including characteristics, sharing, policy, and architecture. With data sharing, they paid special interest to cultural aspects, such as the scientists' willingness and reluctance to share data. They found distinctions between experimental and contextual data, the latter more favorable for sharing, and that higher degrees of automation in data capture also influenced willingness to share favorably.

More recently, Moore et al.'s (2010) editorial, introducing a new policy shared by several key evolutionary biology journals, described the need for evolution scientists to share data, citing the increasing number of meta-analyses that rely on accessible data, and future, unanticipated data uses, such as providing baseline population data. The following year Whitlock (2011) addressed best practices in data archiving for ecology and evolution, explaining how data creators, users, and editors and publishers can responsibly preserve, use, and share scientific data sets using widely-accepted practices.

Several studies, using a variety of research methods, reported research findings about data sharing practices for different science disciplines. Campbell and Bendavid's (2002) survey about attitudes toward data sharing and withholding in medical sciences technology transfer showed strong support for post-publication sharing.

Swan and Brown's (2008) Research Information Network-commissioned study examined data practices of science researchers, offering comparative observations and discussing qualitative differences. The PARSE.Insight study (Kuipers & van der Hoeven, 2009) reported aggregated findings (all disciplines) that showed 25% of researchers shared data openly and an additional 11% shared data openly within a research discipline. Results from the Data Curation Profiles project (Cragin, Palmer, Carlson, & Witt, 2010) research demonstrated low rates of interest in sharing raw or normalized data, even with embargo periods, with small samples from 12 science disciplines that did not include ecologists or natural resources scientists. Tenopir et al.'s (2011) large survey study investigated data sharing practices across several broad science categories, including environmental sciences and ecology, and found significant differences among both data sharing practices and attitudes about data sharing. Diekmann (2012) interviewed agricultural researchers at Ohio State University about their data practices, a sample that included natural resources management and environmental scientists. He found general openness to the idea of sharing, but few researchers had actual experience sharing their data, and they cited a range of concerns about secondary use. Williams (2013) conducted interviews with crop sciences faculty who had already shared data to gain insight into reasons for sharing and to learn about potential roles for the library. Akers and Doty's (2013) survey of researchers at one institution indicated higher rates of data sharing in the basic sciences than in all other science and non-science areas, but did not distinguish among disciplines that constitute the basic sciences.

Among non-survey or interview studies, Savage and Vickers (2009) found low rates of data sharing when requested from the ten corresponding authors they contacted from studies in *PLoS Clinical Trials* and *PLoS Medicine*.

METHODS

The method for this study was modeled after Williams' (2012) study of data practices among crop sciences faculty at the University of Illinois. The two most recent publications of each faculty member in the Department of Ecology, Evolution, and Behavior (EEB), Department of Forest Resources (FR), and Department of Fisheries, Wildlife, and Conservation Biology (FWCB) at the University of Minnesota, Twin Cities, College of Food, Agricultural, and Natural Resource Sciences (CFANS) were thoroughly reviewed. The review was conducted from July 2014 to September 2014. Faculty names were identified through departmental website directories¹ and included all regular faculty (not adjunct or emeritus) with assistant,

¹ EEB: <http://www.cbs.umn.edu/explore/departments-institutes/ecology-evolution-and-behavior/contact-eeb/faculty-directory> ; FWCB: <http://fwcb.cfans.umn.edu/Faculty/index.htm> ; FR: <http://www.forestry.umn.edu/People/index.htm>

associate, or full professor rank. Seventy-nine faculty members were listed in the directories (35 EEB, 23 FR, 21 FWBC), with one person listed twice due a joint appointment in both FR and FWCB, for a total of 78 individuals. Two publications were reviewed for each individual, except for one person, for which only one article was discoverable, for a total of 155 publications.

Faculty names were searched individually in both Web of Science (WOS) and Scopus citation databases in order to determine the most recent publications. Author affiliation was included in searches in order to disambiguate author names. Only *articles* were included in this review, with all other publication types (e.g., reviews, letters, editorials, conference proceedings, book chapters) excluded. WOS and Scopus indexing was used to determine publication type. Articles in press were also excluded. In cases where one or both of a faculty member's publications were co-authored by another faculty member in one of the three departments, the next most recent publication(s) was used.

Each article was thoroughly reviewed to identify and record all of several possible elements present that indicated data sharing by the article author(s), as well as the type of study, location of shared data, type(s) of data shared in each location, and characteristics of the location, or host, for shared data. An instance of data sharing was counted when some form of data, including numeric, tabular, code, data images, gene sequences, and other data types was shared outside the published article. Supplementary or other external files that did not contain a form of data, such as those files that only contained an additional description of methods, figures, or appendices, were not counted as data sharing instances. It is accurate to describe this threshold for data sharing as minimally "shared some data," noting a similar definition for one criterion used by Hampton et al. (2013). While this definition casts a relatively wide net, as it may not include all data used and/or generated by, or necessary to replicate, a study, it is nonetheless data sharing, as in a case where all gene sequence data was shared in GenBank (though other data from the same study may not have been shared). A close approximation of this definition was used, making comparisons possible with a study conducted by Williams (2012), which used a similar definition. Furthermore, using a more prescriptive definition for data sharing is made difficult, and is beyond the scope of this paper, due to the complexity and heterogeneity of ecological data types, as described by Hampton et al. (2013).

Research types were categorized into 13 broad categories: case study, essay, field, focus group, genetic, genomic, phylogenetic, interview, laboratory, method, model, policy, and survey in order to determine if there were significant differences in data sharing practices that might correlate with research and data types. As Williams (2012) notes, many studies do not fit neatly into a single category of research type. In this study, research types were defined during the article review process. These were later grouped into broad categories in order to make

associations with data sharing. Many studies are not of a single type, but are hybrid, such as study that both uses modeling/simulation to create a new research method (i.e., model/method). Several phylogenetic studies collected field specimen data and then performed gene sequencing (i.e., field/genetic). These were recorded appropriately as hybrid type studies.

Data sharing in this study is defined as publicly sharing research data for use by others. There are many different ways to share data by this definition, for example supplementary data can be shared via a journal website, or data can be placed in a disciplinary data repository, such as Dryad or GenBank. When any method of data sharing was indicated anywhere in the article, or supplementary data was present on the publishing journal's website, that mention was traced to one or more data sharing destinations. It is possible that in some cases data was shared without mention in the related article or the journal website. It is also possible that authors are willing to share data upon request, but this was not measured in this study. Additional information related to data sharing, beyond whether or not it was shared and where it was shared, was also recorded in this study. Whether the data was openly accessible, or had use restrictions, or was available only to journal subscribers, was noted. Characteristics of shared data, including data types (e.g., data tables, figures, computer code), number of data files, and file formats was also collected.

RESULTS

For the 78 faculty members, 155 articles were reviewed in this study. Seventy-two articles (46%) indicated that related research data was publicly shared by some method. By a wide margin the most common method for sharing data was using the journal website to host supplemental data files, with 66 of 72 data sharing articles (91%) using this method (see Table 1, following page). Thirteen (18%) of the data sharing articles placed data in the Dryad repository. Nine (12%) of the data sharing articles placed DNA sequence data in GenBank. Eleven (15%) more of the data sharing articles used one of eight additional repositories or websites to share data. These ranged considerably in size and nature, from the Gene Expression Omnibus (GEO) database hosted by the National Center for Biotechnology Information (NCBI) to a regional Fish & Wildlife Research Unit website.

Twenty-one (29%) of the data sharing articles indicated that data was shared in more than one location. In most of these instances, different data from the same study was shared in separate locations where there was some logical alignment of data with a disciplinary repository, such as DNA sequence data in GenBank. For example, for one study DNA sequence data was deposited in GenBank, phylogenetic trees and amino acid alignments were deposited in Dryad, and additional supplementary figures and tables were shared on the journal website. In all 21 instances one of the multiple data sharing locations was the journal website.

Data sharing location	Number of data sharing articles* (n=72)	Percentage of data sharing articles
Journal website	66	91
Dryad	13	18
GenBank	9	12
TreeBase	2	3
European Nucleotide Sequence Read Archive	2	3
Other**	10	14

*Twenty-one (29%) of the data sharing articles shared data in more than one location.

**Other locations included, followed by number of instances: Cedar Creek Ecosystem Science Reserve at University of Minnesota, Twin Cities (1); Computation of Microbial Ecosystems in Time and Space (COMETS) (1); Gene Expression Omnibus Database (GEO) (1); Minnesota Cooperative Fish and Wildlife Research Unit (1); National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (1); SILVIS Lab at University of Wisconsin, Madison (1)

Table 1. Data Sharing Locations

Discipline / Department	Rank						Total data sharing articles n=72
	Asst. Professor n=13		Assoc. Professor n=19		Full Professor n=46		
	Shared	Shared	Shared	Shared	Shared	Shared	
Ecology, Evolution & Behavior	8	8	11	11	16	13	49
Fisheries, Wildlife & Conservation Biology	4	4	2	2	15	8	16
Forest Resources	2	1	6	2	15	2	7

Table 2. Data Sharing Articles by Faculty Discipline and Rank

Differences between data sharing practice by faculty departmental affiliation was also investigated (see Table 2). Faculty in EEB shared data at the highest rate, with 49 of 70 articles (70%) indicating data sharing. Twenty of the 21 instances (95%) in which data were shared in multiple locations occurred with articles from EEB faculty. Only three of 35 EEB faculty members (8%) did not share data associated with at least one of the two reviewed articles (92% did share data).

In contrast, FR faculty members shared data from just 7 of 43 articles (16%) and accounted for no instances where data from an article were shared in multiple places. Only 4 of 22 FR faculty members (18%) shared data for either of the two articles reviewed (82% did not share data). The only type of location for data sharing among FR faculty articles was via the publishing journal's website.

Articles from FWCB faculty members indicated data sharing in 16 of 42 (38%) instances. Just one of 19 instances in which data was shared in multiple locations came from an FWBC article. The primary means for data sharing was through the publishing journal's website (13 of 16 articles), but three articles indicated data sharing in Dryad, one in GenBank, and one through a regional U.S. Fish & Wildlife Unit website.

Each article was categorized into one or more of 13 research types (see Table 3 and Figure 1, following pages). Research types were then cross tabulated with article data sharing results. Sample sizes for some research types were quite small, but results indicated higher prevalence of data sharing for some research types. For example, there were 31 *field* study articles; 12 (39%) of these were also data sharing articles. Twenty-six additional articles had a *field* component combined with one or more other components (genetic, laboratory, method, model, and/or phylogenetic). Twenty-six (45%) of the combined 57 articles that were pure *field* studies or had a *field* component shared data. *Genetic* (5 out of 5 (100%)), *genomic* (3 out of 3 (100%)), and *phylogenetic* (13 out of 14 (93%)) studies had very high rates of data sharing. Combined, these three research article types shared data at a 95% rate (see Figure 2, page 11).

In contrast, articles that had a *case study* (2 of 6 (33%)), *essay* (0 of 3 (0%)), *focus group* (0 of 2 (0%)), *interview* (0 of 2 (0%)), *policy* (0 of 3 (0%)), or *survey* (1 of 12 (8%)) component had a comparatively low data sharing frequency. For these combined 28 articles that used qualitative methods, only 3 (11%) of these research types had related shared data.

Laboratory, *method*, and *model* were among the most common research types. Forty-one articles used study methods involving a *laboratory* component; 27 of these had shared data (66%). Fourteen studies involved the development of a new research *method*; 6 of these had shared data (43%). Thirty-five articles involved some sort of *modeling*; 17 of these had shared data (48%).

Faculty rank was cross tabulated with data sharing to evaluate differences between data sharing rates (see Table 2). Faculty members with full professor rank shared data for at least one of two reviewed articles at a 50 percent rate (23 of 46). Faculty at associate professor rank shared at a 79 percent rate (15 of 19), and those with assistant professor rank shared at a 93 percent rate (13 of 14).

Research type	Total number of articles (n=155)	Number of data sharing articles (n=72)	Percentage of data sharing articles for research type
Case Study	4	0	0
Case Study/Method	1	1	100
Case Study/Model	1	1	100
Essay	3	0	0
Field	31	12	39
Field/Genetic/Laboratory	1	1	100
Field/Laboratory	12	5	42
Field/Laboratory/Method	1	1	100
Field/Laboratory/Phylogenetic	3	3	100
Field/Method	1	1	100
Field/Model	8	4	50
Field/Phylogenetic	1	0	0
Focus Groups	2	0	0
Genetic	2	2	100
Genetic/Laboratory	1	1	100
Genetic/Laboratory/Model	1	1	100
Genomic	2	2	100
Genomic/Method	1	1	100
Interview	1	0	0
Interview/Survey	1	0	0
Laboratory	17	10	59
Laboratory/Method/Model	1	1	100
Laboratory/Model	2	2	100
Laboratory/Phylogenetic	2	2	100
Method	5	0	0
Method/Model	4	1	25
Method/Phylogenetic	1	1	100
Model	25	12	48
Model/Phylogenetic	4	4	100
Model/Survey	2	0	0
Phylogenetic	3	3	100
Policy	3	0	0
Survey	9	1	11

Table 3. Number of Articles by Research Type

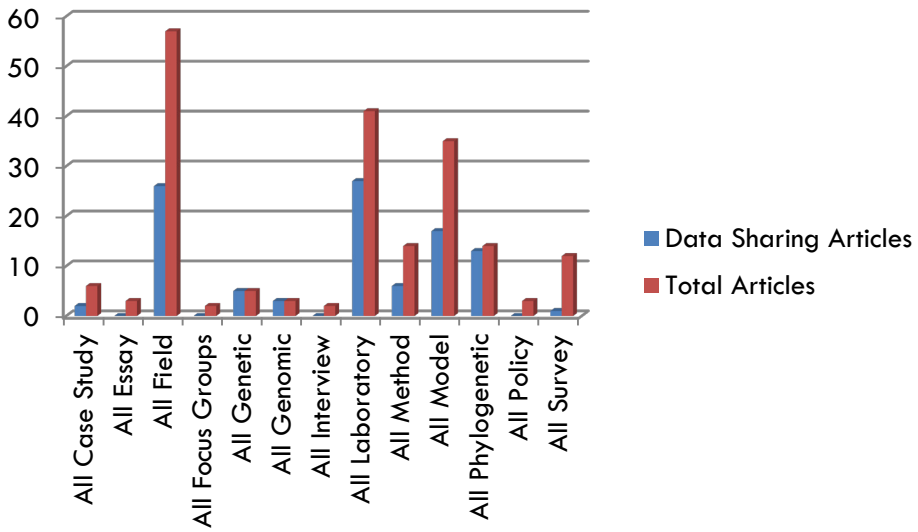


Figure 1. Data Sharing Articles by Article Type

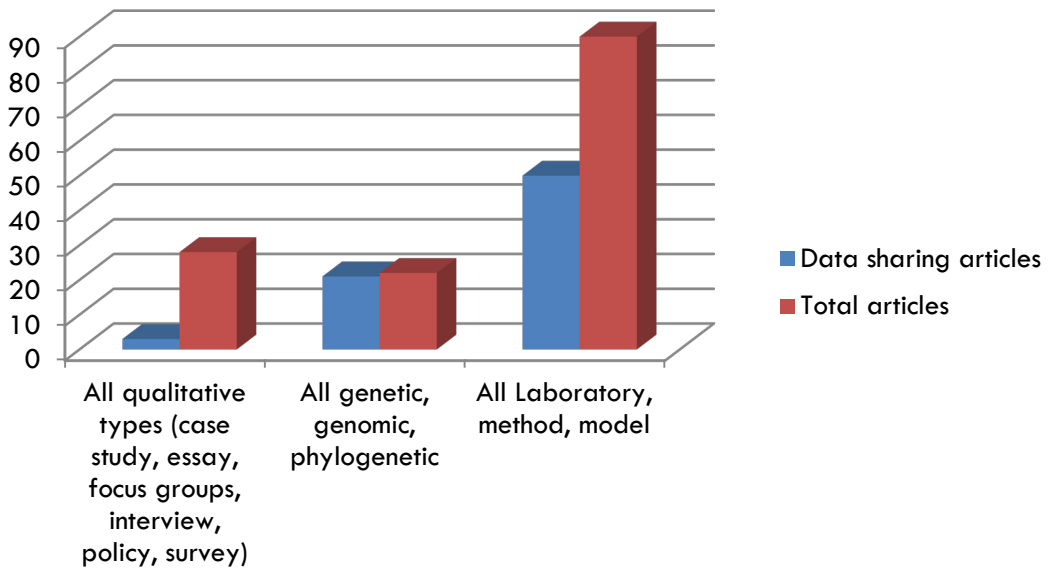


Figure 2. Data Sharing Articles by Research Type Groupings

The 155 articles reviewed in this study were published in 103 different journals (see Appendix A). Two or more articles appeared in only 29 (28%) of the 103 journals. The greatest number of articles published in the same journal was five, and this occurred in five journals: *Evolution*, *Northern Journal of Applied Forestry*, *PloS One*, *Proceedings of the Royal Society of London Series B-Biological Sciences*, and *Proceedings of the National Academy of Sciences (PNAS)*. All articles were published between 1999 and 2014, with 116 (75%) of the articles published in either 2013 or 2014 and only three articles published prior to 2008 (one each in 2005, 2000, and 1999) (see Table 4).

Publication year	Total number of articles (n=155)	Number of data sharing articles (n=72)
2014	70	39
2013	46	22
2012	16	8
2011	14	2
2010	2	1
2009	3	0
2008	1	0
2005	1	0
2000	1	0
1999	1	0

Table 4. Number of Articles by Publication Year

The 72 data sharing articles were published in 46 journals (see Appendix A). Thirteen of these journals contained two or more data sharing articles. The journals containing the highest number of data sharing articles were *Evolution* and *PLoS One*, with 5 each, followed by *Molecular Phylogenetics and Evolution* and *Proceedings of the Royal Society B*, which each contained four data sharing articles. Most (86%) data sharing articles were published in 2013 (23) or 2014 (39). No data sharing articles were published prior to 2010; only 3 prior to 2012.

The journal websites containing data sharing articles were a frequent location for hosting some or all of the data that was shared, accounting for 91% of data sharing instances. There were 66 data sharing articles for which some data or all shared data was located on a journal website; forty-four journals hosted data from one or more of these data sharing

articles. *Evolution* (5), *Molecular Phylogenetics and Evolution* (4), and *PLoS One* hosted data for the highest number of data sharing articles.

Supplemental data hosted on journal websites was most frequently available only to subscribers. Supplemental data from 48 (73%) of the 66 data sharing articles that shared via journal websites were available only to journal subscribers. Data from 18 articles (27%) were available open access, including one where it was explicit that the author paid to make the article and data open access in an otherwise non-open access journal.

File format types were identified and recorded for all shared data (see Appendix B). In many cases more than one file format was used to share different supplemental data for a single article. The most commonly-used file format for data sharing on journal websites (n=66) was PDF (portable document format) with 39 instances. Second most common were Microsoft Word .doc and .docx formats, with a combined 25 instances. Microsoft Excel was the next most common format, with six instances. Archive (.zip and .gz) files were used 5 times to package groups of files. Plain-text files (.txt) were used 5 times. A wide variety of other file formats were used to shared data, none more than three times. They included: .html (hypertext mark-up language), .jpg (image), .tif (image), .mov (video), EPS (Encapsulated PostScript), .mp3 (audio), .csv (comma-separated values), .r (R statistical software code), .c (C programming code), .nex (NEXUS phylogenetic format), and .tre (tree age decision tree format).

Among the thirteen articles with shared data in the Dryad repository, .zip files were used frequently to archive/package groups of files or to compress large files. Plain-text files (.txt) were the most often used format to share text-based data, and comma-separated value (.csv) files the most often used format to share numerical or text tabular data. There were few instances of PDF, Microsoft Excel, or Word formatted files in Dryad. Data were also shared in a variety of plain text file formats for use with specific software packages, such as NEXUS and R.

DISCUSSION

While few studies report specifically on the data sharing practices of science disciplines, even fewer use methods that examine observable evidence to form conclusions, in other words, study what researchers do rather than what they say they do or would do. This study used the same method as Williams' (2012) study of crop sciences researchers' data practices. Williams' study is most directly comparable to this study because of the shared method and relative nearness of the disciplines studied. Williams found that 24% of the instances in her sample shared some form of data beyond that contained in the journal

articles she examined. While this study follows three years later and reports a 46% sharing rate (72 of 155 instances), the higher rate may be attributable to any of several factors, including a trend towards increased data sharing that is influenced by funding agency policies, disciplinary cultural differences, and/or an increase in the availability of publisher or disciplinary platforms for data sharing.

Hampton et al. (2013) also used a similar method to analyze data sharing from publications resulting from NSF-supported ecology studies. Among data-producing studies, they found 28% shared all data, 15% shared some data, and 57% did not share any associated data.

It is difficult to make comparisons between the results of this study and the results of other data practices or data sharing studies that used qualitative methods, primarily surveys and interviews. Most of the qualitative studies cover a wider range of disciplines and many also extend to other areas of data practices, including data re-use, data citation, or other practices. While this study evaluates aspects of data sharing, such as rates, disciplinary differences, and methods, other factors that may influence an individual researcher's decision to share data, or not share, cannot be discerned by reviewing published articles. Many of the survey and interview studies are better able to probe questions about motivations and perceptions among researchers and disciplines (Akers & Doty, 2013; Campbell & Bendavid, 2002; Cragin, Palmer, Carlson, & Witt, 2010; Diekmann, 2012; Kuipers & van der Hoeven, 2009; Swan & Brown, 2008; Tenopir et al., 2011; Williams, 2013). This study of natural resources, ecology, evolution and behavior scientists does, however, shed some light on these disciplinary cultures—there are some obvious norms and expectations widely shared among the broader community of researchers. For example, ecology and evolution scientists, as well as those from many other disciplines, that generate DNA sequence data most often follow the norm of depositing sequence data in a shared repository, like GenBank (Whitlock, 2011). But pro data sharing norms are not well developed in some disciplines, as suggested by the data for Forest Resources faculty, nor are they established for all research types, suggested by the lack of sharing for data generated in most field and qualitative studies. Further study into data sharing practices of different scientific disciplines, using both qualitative and quantitative approaches, will help researchers and the professionals who support them understand this changing landscape.

Journal and publisher policies appear to influence data sharing practices. Piwowar and Chapman's (2008) study of journal data sharing policies and their impacts on gene expression microarray data sharing found a positive association between strength of the policies and rates of sharing. Some journal and publisher policies require that the data related to a manuscript be shared as a condition of publication. *Science*, *Nature* (for certain data research types), and *PLoS* journals are prominent examples of this, providing clear

statements to authors about data sharing (nature.com, 2015; Public Library of Science, 2015; Science Magazine, 2015). Policies such as these provide an important incentive to share data, but questions about author compliance are raised in Savage and Vickers' (2009) study of data sharing in *PLoS* journals. While admittedly a very small sample size, nine of the 10 papers reviewed in the present study published in *Science*, *Nature*, or *PLoS* journals shared the data underlying those publications.

Within ecology and evolution, journals of the Ecological Society of America are among those that have specific data sharing expectations (Ecological Society of America, 2013). *American Naturalist*, *Evolution*, *Journal of Evolutionary Biology*, *Molecular Ecology*, and *Heredity* share a policy requiring data sharing, stating that underlying data should be shared in a stable, public repository, such as Dryad, GenBank, TreeBASE, or the Knowledge Network for Biocomplexity (Whitlock, McPeck, Rausher, Rieseberg, & Moore, 2010). While the individual data-archiving policies of these five journals all require DNA sequences to be deposited in GenBank or another public sequence database, for other data types they vary in some important details. They all require other types of underlying data to be shared, but *Evolution and Molecular Ecology* (both Wiley-published journals) allow authors to share non-sequence underlying data as supplementary files on the journal website, which may effectively remove them from access except to journal subscribers (*Evolution*, 2015) (*Molecular Ecology*, 2015). *Heredity* requires non-sequence data to be shared via Dryad (*Heredity*, 2015). *American Naturalist* requires associated data to be shared in a public archive, but does not require a specific location except for gene sequence data (*The American Naturalist*, 2015).

In this study, articles published in these journals combine to form a very small sample, but all eight articles analyzed from among the five journals shared underlying data in compliance with specific journal requirements. Of the eight articles, five were published in *Evolution*, and one each in *American Naturalist*, *Heredity*, and *Molecular Ecology*. Authors of four of these articles shared data in Dryad, which included computer code, gene sequence assemblies, a range of raw experimental data, both original and re-used from other studies, phylogenetic tree figures and other data. Three of the articles generated gene sequences shared in GenBank, once in combination with Dryad and once in combination with the journal website. Authors of six of the articles shared data on the journal website, three times as the only data sharing location, twice in combination with Dryad, and once in combination with GenBank. All six instances in which the journal website was used to share data were for articles in *Evolution* or *Molecular Ecology*, which allow this practice.

Funding agency mandates also may influence the decision to share research data. The National Institutes of Health (NIH) and National Science Foundation (NSF) now require

data management plans that should include plans for data sharing to be submitted with grant proposals (National Institutes of Health (NIH), 2003; National Science Foundation (NSF), 2011). A 2013 memorandum from the Federal Office of Science and Technology Policy (OSTP) mandated federal agencies expending more than \$100 million in research and development expenditures to develop plans that will increase public access to federally-funded research outputs, including publications and research data (Holdren, 2013). Some large private science research funders are also now requiring that the data underlying published research be made openly accessible, such as the new Bill & Melinda Gates Foundation Open Access Policy (2015). While this study did not explore the connection between funding sources and data sharing, Pham-Kanter, Zinner, and Campbell (2014) found that recent data sharing policies and infrastructures have a “sizable effect” on data sharing practice, with some variability in degree of influence for different types of policies.

There is an increasing number of institutional data sharing policies at U.S. universities that may also influence data sharing practices. This study was conducted before the adoption of a data management policy at the University of Minnesota in January 2015 (University of Minnesota, 2014). While this new institutional policy did not influence data sharing practices in this study, it states clear responsibilities for data ownership, stewardship, preservation, storage, security, transfer, and access and reinforces the expectation that principal investigators will comply with funder and publisher mandates regarding data management and sharing.

Measuring increased citation rate, impact, or other perceived or real rewards from data sharing were not the focus of this study, but may also have influenced data sharing rates. Piwowar, Day, and Fridsma’s (2007) study of cancer microarray clinical trial publications found a positive association between public data sharing and increased citation rate. Another Piwowar (2011) study of open data sharing among gene expression microarray research found associations affecting the probability of data sharing that included “an experiment’s topic, impact, funding, publishing, institutional, and authorship environments” (p. e18657). Impact, as measured by increased citation rate, may be a motivating factor for some researchers to share data and is certainly a compelling set of findings that librarians can use to raise awareness of the link between data sharing and impact, though Doty (2013) concluded from a small survey study focused on publications (not data) of science faculty that the open access citation effect is unlikely to convert those “who are not already convinced of the benefits of OA [open access]” (p. 1).

With many factors that may serve to encourage natural resources and ecology scientists to share their research data, there are as many reasons why they may choose not to share. Foremost among these may be sociological and cultural barriers, which may be the most

difficult to overcome. Reichman et al. (2011) argue that historically ecologists have not had a data sharing culture, and sharing was neither rewarded nor seen as essential to doing science. Findings in the present study marginally support the suggestion that senior faculty, those at the full professor rank, are less likely to share their data than those at the associate and assistant professor rank though this is a questionable conclusion due to small sample size.

Sharing data requires extra work to document, prepare, and organize, requiring time and resources that may be scarce for many scientists. Scientists may fear being scooped by others if they share their data with the broader community before its value has been exhausted; they may generally feel protective of data that was costly and time consuming to collect. Data may not be shared because it is related to patentable discovery and/or due to intellectual property concerns, or because it involves human subjects. Misinterpretation or misuse of a scientist's data is a frequently cited reason for not sharing data. There may not be an obvious or appropriate place to put a research data set. There may be privacy concerns about data, particularly when human subjects are involved. Several studies have surfaced these reasons, and others, explaining why scientists choose not to share their data (Campbell & Bendavid, 2002; Cragin et al., 2010; Diekmann, 2012; Piwowar, 2011; Reichman et al., 2011; Savage & Vickers, 2009; Swan & Brown, 2008; Tenopir et al., 2011).

Journals and publishers clearly have an important role in encouraging and supporting data sharing among natural resources and ecology scientists. Journal policies may influence the decision to share one's data, and in this study journal websites were the most common method for sharing supplementary data (see Table 1). Thus, access to shared data is largely controlled by journals. Large publishers often each have their own common way to host supplementary data, applied consistently in each of their many journal titles. Notable considerations among journal practices include the fact that some subscription-based journals from large publishers, such as Springer, make supplementary data files openly accessible, even when the associated articles are accessible only to paid subscribers. Others keep data and articles accessible only to subscribers. Some journals will make data and articles openly available after an embargo period, such as the *Journal of Bacteriology* (six months), or *Journal of the Royal Society Interface* (one year). The Ecological Society of America maintains a separate website for data files which is linked to articles on journal websites. This site and all its data are open access. *Oikos* has a separate website for supplemental data and appendices which is also open access.

Journals owned by large publishers, such as Elsevier, Springer, and Wiley have mechanisms in place for authors to submit supporting data in a variety of file formats and also have mechanisms for sharing this supplementary data, openly or to subscribers only, on journal websites hosting the related published article. Some of the smaller society-owned journals,

such as those from the Society of American Foresters, have neither the mechanism for authors to submit supplementary data nor a way to share such data on the journal website. When examining disciplinary differences in data sharing practices, this is an important consideration and one that can influence interpretation. Journals that do not support data sharing are an obvious barrier and may be seen as opportunities for both scientific authors and librarians to influence change that will support funder mandates, data re-use, research replicability, and a range of additional benefits.

This investigation found distinctions between the data sharing practices of faculty in the different disciplines studied (see Table 2) and between rates of data sharing associated with different research types/methods (see Table 3). Similar to the findings of Weller and Monroe-Gulick (2014), who studied disciplinary differences in data practices of academic researchers (but not data sharing specifically), implications for library or other services supporting data management and sharing, such as data curation services, are several. Given wide differences in data sharing practices between the authors of studies that use qualitative methods (e.g., case studies, interviews, surveys, etc.) and those using other scientific methods (e.g., laboratory, genetic), targeted services may be most effective at responding to different individuals. For example, many of the qualitative studies reviewed in this study involved human subjects, and authors may have been unable to or reluctant to share data that would need to be de-identified. Expertise in methods for de-identifying data is a service that could be targeted to researchers conducting studies involving subjects. None of the genetic/genomic/phylogenetic studies reviewed in this study involved human subjects, and the high rate of data sharing among these studies may also reflect the cultural norms surrounding this type of science, as well as the availability of stable and reputable data repositories, such as GenBank and TreeBase. Researchers leading these studies may be targets for services around their supplementary data that typically is shared via a journal website, rather than the DNA sequence or phylogenetic tree data that is shared via GenBank and TreeBase.

One area for concern among findings of this study is the common practice of sharing data through journal sites, and the preservation and persistence of this data. Similar concerns are raised by the Williams (2013) and the PARSE.Insight (Kuipers & van der Hoeven, 2009) studies. While all data mentioned in the reviewed data sharing articles was discoverable and available in this study, the majority were recently published. Further investigation into journal preservation practices and guidelines is needed, but it is clear from the large number of data sets in a wide range of proprietary formats identified in this study that many journals either do not recommend or require digital preservation best practices, such as file format normalization for office documents, tabular data, media, or computer code, or these practices are not followed by authors. There is risk both of broken links (non-persistence): that shared supplementary data will disappear as journals migrate to new platforms or are

acquired by new publishers; and, that files will be rendered obsolete over time as their formats fall out of favor and use. These are both areas where libraries have well established expertise and can play both an educational, transformative role, and a behind-the-scenes operational role supporting data preservation and persistence locally, and advocating for change with publishers and scientist practices.

CONCLUSION

Understanding current practices of scientists in different disciplines is an imperative in order to develop data management and curation services that are useful, anticipatory, and responsive to the needs of scientists. This study offers insight into the data sharing practices of natural resources, ecology, and evolution and behavior scientists that informs the development of support and services in libraries and other campus service providers. It identifies disciplinary differences in data sharing practices within these sciences as well as distinctions between data sharing rates based on the scientific methods used and types of data generated and considers some implications of these differences. It examines a sample of journals and publishers, focusing on data policies and practices supported by them. It also discusses a range of external influences and possible motivations for scientists to share, or not share, their data.

This study differs from most other studies that examine data sharing practices of science researchers. It examines data sharing practices using existing evidence, and reports based on what researchers actually do, not what they say they do, or would do. Among studies that also report findings based on existing evidence, this study is unique in its focus on natural resources, ecology, and evolution researchers.

The results of this study advance the collective understanding of research data sharing practices among ecology, evolution, and natural resources science faculty. Results show clear differences in how frequently ecology and evolution scientists share their data compared to natural resources scientists; both sharing levels can be compared with other disciplines and be used to inform library approaches to data services. These findings suggest that data support service providers should not follow a one-size-fits-all approach to outreach. The low rate of data sharing among natural resources scientists can be seen as an opportunity to influence change through education about the benefits of sharing and re-use, by removing barriers to data sharing, building accessible platforms that encourage sharing, and advocating for the use of best practices as the disciplinary culture evolves its data sharing norms.

The high rate of data sharing among ecology, evolution, and behavior scientists is not a recommendation for library data services outreach complacency. Although there are helpful

exceptions, most of the current data sharing happens through journal websites that lack standardization, provide few guidelines, and do not follow preservation best practices. The library community can look to the few journals with forward looking preservation and sustainable policies, such as those from the Ecological Society of America, and shared repositories, such as Dryad, as positive examples that may be used to influence other journals and publishers, or to shape the development of local or other data curation and repository services. Compelling data sharing alternatives will be needed to change the current sharing practice of sharing supplemental data via journal websites.

ACKNOWLEDGEMENTS

This research was made possible by a University of Minnesota Libraries professional development leave.

REFERENCES

- Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5–26. <http://dx.doi.org/10.2218/ijdc.v8i2.263>
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., & Moorman, D. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, 3(29 November), 135–152. <http://dx.doi.org/10.2481/dsj.3.135>
- Bill & Melinda Gates Foundation. (2015). *Bill & Melinda Gates Foundation open access policy*. Retrieved December 30, 2014, from <http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>
- Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, 7(1-2), 17–30. <http://dx.doi.org/10.1007/s00799-007-0022-9>
- Campbell, E., & Bendavid, E. (2002). Data-sharing and data-withholding in genetics and the life sciences: Results of a national survey of technology transfer officers. *Journal of Health Care Law and Policy*, 6(2), 241–255. Retrieved from http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/hclwpo6§ion=19
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 368(1926), 4023–4038. <http://dx.doi.org/10.1098/rsta.2010.0165>

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

Doty, R. C. (2013). Tenure-track science faculty and the “Open access citation effect.” *Journal of Librarianship and Scholarly Communication*, 1(3), eP1052. <http://dx.doi.org/10.7710/2162-3309.1052>

Ecological Society of America. (2013). *ESA Data Registry*. Retrieved November 25, 2014, from <http://esapubs.org/esapubs/DataReg.htm>

Evolution. (2015). *Evolution* - Author Guidelines - Wiley Online Library. Retrieved May 26, 2015, from [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1558-5646/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1558-5646/homepage/ForAuthors.html)

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., ... Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11(3), 156–162. <http://dx.doi.org/10.1890/120103>

Heredity. (2015). *Heredity* - Guide for Authors. Retrieved May 26, 2015, from http://mts-hdy.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions#datapolicy

Holdren, J. P. (2013). *Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research*. Executive Office of the President, Office of Science and Technology Policy. Retrieved from http://scholar.google.com/scholar?q=Memorandum+for+the+Heads+of+Executive+Departments+and+Agencies+2013&btnG=&hl=en&as_sdt=0,24#0

Jamali, H. R., & Nicholas, D. (2009). E-print depositing behavior of physicists and astronomers: An intradisciplinary study. *Journal of Academic Librarianship*, 35, 117–125. <http://dx.doi.org/10.1016/j.acalib.2009.01.005>

Key Perspectives Ltd. (2010). *Data dimensions: Disciplinary differences in research data sharing, reuse and long term viability. A comparative review based on sixteen case studies*. Edinburgh, Scotland: Digital Curate Centre.

Kuipers, T., & van der Hoeven, J. (2009). *Insight into digital preservation of research output in Europe* (p. 83). PARSE.Insight.

McCain, K. W. (1995). Mandating sharing: Journal policies in the natural sciences. *Science Communication*, 16(4), 403–431. <http://dx.doi.org/10.1177/1075547095016004003>

Molecular Ecology. (2015). *Molecular Ecology* - Author Guidelines - Wiley Online Library. Retrieved May 26, 2015, from [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1365-294X/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-294X/homepage/ForAuthors.html)

Moore, A. J., McPeck, M. A., Rausher, M. D., Rieseberg, L., & Whitlock, M. C. (2010). The need for archiving data in evolutionary biology. *Journal of Evolutionary Biology*, 23, 659–660. <http://dx.doi.org/10.1111/j.1420-9101.2010.01937.x>

National Institutes of Health (NIH). (2003). *NOT-OD-03-032: Final NIH statement on sharing research data*. Retrieved December 30, 2014, from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>

National Research Council. (2003). *Sharing publication-related data and materials: Responsibilities of authorship in the life sciences*. The National Academies Press.

National Science Foundation (NSF). (2011). *Grant proposal guide, Chapter II.C.2.j*. Retrieved December 30, 2014, from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#IIC2j

nature.com. (2015). *Availability of data, material and methods*. Retrieved from <http://www.nature.com/authors/policies/availability.html>

Parr, C. S., & Cummings, M. P. (2005). Data sharing in ecology and evolution. *Trends in Ecology & Evolution*, 20(7), 361–2. <http://dx.doi.org/10.1016/j.tree.2005.04.006>

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

Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE*, 6(7), e18657. <http://dx.doi.org/10.1371/journal.pone.0018657>

Piwowar, H. A., & Chapman, W. W. (2008). A review of journal policies for sharing research data. In *Open scholarship: Authority, community, and sustainability in the age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing, ELPUB 2008* (pp. 1–14). Retrieved from http://elpub.scix.net/cgi-bin/works/Show?001_elpub2008

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, 2(3), e308. <http://dx.doi.org/10.1371/journal.pone.0000308>

Public Library of Science. (2015). *PLoS ONE manuscript guidelines*. Retrieved February 27, 2015, from <http://www.plosone.org/static/guidelines.action#datareporting>

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331(6018), 703–5. <http://dx.doi.org/10.1126/science.1197962>

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One*, 4(9), e7078. <http://dx.doi.org/10.1371/journal.pone.0007078>

Science Magazine. (2015). *General information for authors*. Retrieved from http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml

Shaffer, C. (2013). The role of the library in the research enterprise. *Journal of eScience Librarianship*, 2(1), 8–15. <http://dx.doi.org/10.7191/jeslib.2013.1043>

Swan, A., & Brown, S. (2008). *To share or not to share: Publication and quality assurance of research data outputs*. A report commissioned by the Research Information Network, (June). Retrieved from <http://eprints.soton.ac.uk/266742/>

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS One*, 6(6), e21101. <http://dx.doi.org/10.1371/journal.pone.0021101>

The American Naturalist. (2015). *The American Naturalist* - Instructions for authors. Retrieved May 26, 2015, from <http://press.uchicago.edu/journals/an/instruct.html?journal=an>

Thessen, A. E., & Patterson, D. J. (2011). Data issues in the life sciences. *ZooKeys*, *150*, 15–51. <http://dx.doi.org/10.3897/zookeys.150.1766>

Uhlir, P. F., & Schröder, P. (2007). Open data for global science. *Data Science Journal*, *6*(June), 36–53. <http://dx.doi.org/10.2481/dsj.6.od36>

University of Minnesota, Office of the Vice President for Research. (2014). Research data management: Archiving, ownership, retention, security, storage, and transfer. Retrieved February 18, 2015, from <http://www.policy.umn.edu/Policies/Research/RESEARCHDATA.html>

Weller, T., & Monroe-Gulick, A. (2014). Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, *32*(3), 467–482. <http://dx.doi.org/10.1108/LHT-02-2014-0021>

Whitlock, M. C. (2011). Data archiving in ecology and evolution: best practices. *Trends in Ecology & Evolution*, *26*(2), 61–5. <http://dx.doi.org/10.1016/j.tree.2010.11.006>

Whitlock, M. C., McPeck, M. A., Rausher, M. D., Rieseberg, L., & Moore, A. J. (2010). Data archiving. *The American Naturalist*, *175*(2), 145–6. <http://dx.doi.org/10.1086/650340>

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. <http://dx.doi.org/10.1080/10496505.2012.717846>

Williams, S. C. (2013). Data sharing interviews with crop sciences faculty: Why they share data and how the library can help. *Issues in Science and Technology Librarianship*, Spring. <http://dx.doi.org/10.5062/F4T151M8>