

# The Changing Influence of Journal Data Sharing Policies on Local RDM Practices

Dylan Dearborn  
University of Toronto

Steve Marks  
University of Toronto

Leanne Trimble  
University of Toronto

## Abstract

The purpose of this study was to examine changes in research data deposit policies of highly ranked journals in the physical and applied sciences between 2014 and 2016, as well as to develop an approach to examining the institutional impact of deposit requirements. Policies from the top ten journals (ranked by impact factor from the Journal Citation Reports) were examined in 2014 and again in 2016 in order to determine if data deposits were required or recommended, and which methods of deposit were listed as options. For all 2016 journals with a required data deposit policy, publication information (2009-2015) for the University of Toronto was pulled from Scopus and departmental affiliation was determined for each article. The results showed that the number of high-impact journals in the physical and applied sciences requiring data deposit is growing. In 2014, 71.2% of journals had no policy, 14.7% had a recommended policy, and 13.9% had a required policy (n=836). In contrast, in 2016, there were 58.5% with no policy, 19.4% with a recommended policy, and 22.0% with a required policy (n=880). It was also evident that U of T chemistry researchers are by far the most heavily affected by these journal data deposit requirements, having published 543 publications, representing 32.7% of all publications in the titles requiring data deposit in 2016. The Python scripts used to retrieve institutional publications based on a list of ISSNs have been released on GitHub so that other institutions can conduct similar research.

*Received* 20 October 2016 ~ *Revision Received* 23 February 2017 ~ *Accepted* 23 February 2017

Correspondence should be addressed to Dylan Dearborn, Map & Data Library, 130 St. George Street, 5th Floor, Toronto, ON, Canada M5S 1A5. Email: [dylanne.dearborn@utoronto.ca](mailto:dylanne.dearborn@utoronto.ca)

An earlier version of this paper was presented at the 12<sup>th</sup> International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



## Introduction

While in some countries, such as the US and the UK, funding agencies mandate activities such as data management, preservation and sharing (SHERPA/JULIET, 2017), but in Canada this is still an emerging area (Research Data Canada, 2008; Shearer, 2015). Mandates can provide structured guidance both for researchers and for data service providers responsible for designing institutional support. In the absence of such guidance, institutions must turn to other sources to understand the drivers of researcher practice. There have been efforts at the University of Toronto (U of T) to understand science and engineering researcher research data management (RDM) needs and practices, one of which was a survey of faculty and postdoctoral fellows (Sewerin et al., 2015). The results highlighted differing data practices between subject areas, as well as indicated that only 14.7% of responding researchers (n=95) are currently depositing data in repositories and 31.6% are depositing data with journals as supplementary files. We can learn more about the international community by examining journal policies as a potential driver for data deposit and sharing, since journal policies are typically representative of community established norms.

In 2014, in order to better understand the impact of publisher policy pressures on U of T researchers, we ran a journal policy analysis intended to discover disciplinary patterns in data sharing and deposit requirements. We approached the analysis with the assumption that researchers would be motivated to publish in high-impact journals, and that if highly ranked journals had data deposit requirements, then researchers would be motivated to comply. This led us to examine the author requirements for the highest ranked journals by impact factor using the Journal Citations Reports (JCR), 2012 science edition, for areas within the physical and applied sciences.

More than two years have passed since the original analysis and the broader research data landscape has changed significantly. In 2016, we ran this analysis again in order to determine whether there have been changes in journal data sharing policies among high impact titles in physical and applied science areas, and to re-assess the impact on U of T researchers.

## Literature Review

While research data management (RDM) has gained significant attention in the last decade, recognition of the importance of data sharing is by no means new. A pioneering study of journal policies was McCain's investigation of about 850 journals in medicine, engineering and the natural sciences (1995). At the time, only about 16% of all journals had a policy which made some mention of the deposit or sharing of research data.

Piwowar and Chapman built upon this work with their 2008 study of journal policies related to gene microarray data. Things had changed significantly since the 1990s: they found that 76% of the 70 identified journals had a policy making some mention of data sharing, and about 43% of these policies were considered "strong" with respect to gene microarray data (meaning that an accession number from the NCBI GEO database was required prior to publication) (Piwowar and Chapman, 2008). Strong policies were more likely from academic rather than commercial publishers, and from journals with high impact factors.

Outside of the medical sciences, progress has been slower. For example, a 2010 study of 307 journal policies in the environmental sciences found only 14% which either “requested” or “required” the archiving of data (Weber, Piwowar and Vision, 2010). A study of sociology journals found only 5% with an explicit data policy, although 67% did refer to a common policy of the Association of Learned and Professional Society Publishers – an organization to which the journals belong and which does encourage data sharing (Zenk-Moltgen and Laphien, 2014). The JoRD project conducted a broader study of nearly 400 journals in all fields of study (Sturges et al., 2014; 2015). In this study, about half of the journals had no policy, and of those that did, only about 24% were considered strong, using Piwowar and Chapman’s definition. Only 15% of policies named a specific repository which researchers should use.

Few studies to date have focused on the impact of journal policies on researchers at specific institutions. Researchers at the University of Rochester’s River Campus Libraries conducted a study assessing whether their researchers were complying with the data sharing policies of the journals in which they published (Fear, 2015). They identified the 109 journals that Rochester researchers published in most frequently in 2014, then reviewed journal policies, ultimately narrowing the sample to 161 articles from 13 journals which required data sharing. They learned that for half of these articles, the researchers had not shared their data.

Fear’s work assessing publishing patterns and policy compliance at the institutional level demonstrates the kind of research libraries can undertake to help direct the development of focused outreach efforts – for example, conducting workshops specifically on compliance with particular publisher or journal requirements.

## Methods

### Part 1: Identify High-Impact Journals and Review Their Data Sharing Policies

The methods employed the first time this analysis was undertaken, are described briefly in a poster presented at the 2015 IDCC conference (Dearborn and Marks, 2015). Here we describe the methodology in more detail as well as note minor changes made for the 2016 run, including the automation of parts of the process through scripting.

The top ten journal policies (ranked by impact factor) from 114 categories in the physical and applied sciences were exported from the 2015 science edition of Journal Citation Reports, resulting in 1,140 ISSN records. This dataset included 880 unique journal ISSNs, since some journals appeared in more than one category. The categories were the same ones used in 2014 (from the 2012 science edition of the Journal Citation Reports), with the exception of seven categories which were not included in the 2014 study – six categories in areas U of T does not focus in, and one new category that did not exist in 2012. These categories were included in the 2016 study to make the coverage more comprehensive.

For each journal, the policies/author guidelines were located and read. This included analysing title-level policies as well as any publisher-level policies that were explicitly linked from the author guidelines of the journal. The policies were coded according to their data deposit policy: required, recommended, or no policy. This was a challenging process, requiring a careful reading of the policy, considering the words used in context; for example, the use of the word “should” sometimes connoted a requirement, and

sometimes only a recommendation. If the journal's policy mentioned data, but did not clearly recommend that data be deposited in a public repository, it was classed as "no policy". Information was recorded on what modes of data deposit were mentioned in the policy, including sharing via the journal itself (i.e. as "supplementary information"), via institutional repositories, or via subject repositories (which we defined to include references to "appropriate" or "publicly available" data repositories). Any specific repositories mentioned as suitable for deposit were recorded.

Many policies contained both a recommendation and a requirement for data deposit. For example, a journal might encourage data deposit for all data, but specify that specific data types must be deposited as a condition of publication. For example, BioMed Central's policy is worded: "strongly encourages that all datasets on which the conclusions of the paper rely should be available to readers, and where there is a community established norm for data sharing, BioMed Central mandates data deposition". In these cases, the journal was coded as "deposit required", even if only the deposit of one data type was required. This skews the policy coding slightly, as certain types of data (e.g. DNA and RNA sequences, microarray data, crystallographic data, etc.) have established norms of deposit. This skewing particularly affected publisher-level policies, where a data deposit policy exists for all journals under one publisher and therefore, for example, a civil engineering journal which would not contain DNA sequences was still classified as a "deposit required" in the coding.

Once the coding was complete, the policy data from the 2016 (2015 JCR ISSNs) and 2014 (2012 JCR ISSNs) policy reviews were merged into one data file which was then analysed using SPSS. The merged data file contains 2,209 records (1,140 from the 2016 review and 1,069 from the 2014 review). There were 880 unique ISSNs in the 2016 review, and 836 in 2014. In the merged dataset, this resulted in 610 ISSNs which appeared in both years and could be used to analyse changes to specific policies over time.

## **Part 2: Determine which Journal Policies Most Impact University of Toronto Researchers**

For each of the journals coded with a deposit required in the 2016 analysis (194 ISSNs), we retrieved a list of all publications with an author from U of T for the years 2009 through 2015. We selected a time window that was wide enough to result in a fairly large number of U of T publications in the selected journals. Our intention was to gain a broad understanding of the publication patterns at U of T in these areas, not to determine the impact of specific policies, as the policy review is so recent.

We then identified a subset of articles where the primary author was affiliated with U of T, using the "Corresponding Author" field in Scopus. This work was done through Python scripts which make calls to the Scopus Search and Abstracts APIs, and the data were cleaned and clustered using OpenRefine. The scripts are available on GitHub.<sup>1</sup>

The Scopus search query for all articles with a U of T affiliated author returned 3,487 results. The query used to determine whether the article's corresponding author was affiliated with the U of T refined this set to 1,672 articles. We chose to limit to corresponding author as they are typically the person responsible for the study and therefore adhering to any conditions of submission. This subset of articles was then analysed for departmental affiliation.

---

<sup>1</sup> Scripts available on GitHub: <https://github.com/sbmarks/scopus-queries>

## Results

### Journal Policy Analysis

The analysis was undertaken in two parts: journal policies, and U of T publications. The journal policy data was analysed in terms of broad shifts in policy requirements by subject area, as well as changes in individual policies between the two study years. Finally, policy mentions of specific repositories and/or institutional repositories was investigated.

#### 2016 journal policies by JCR subject category

For each JCR subject category included in our analysis, we pulled the ten journal ISSNs with the highest JCR impact factor. Table 1 shows the specific subjects within the physical and applied sciences that were most likely to have data sharing requirements in their high-impact journals.

Several subject categories were observed to have had a particularly large change in the number of policies requiring data deposit since 2014. These included: Biochemical research methods (jumped from 4/10 in 2014 to 8/10 in 2016); Chemistry, multidisciplinary (jumped from 2/10 to 6/10); Energy & fuels (jumped from 1/10 to 5/10); Meteorology & atmospheric sciences (jumped from 3/10 to 7/10); and Mycology (jumped from 3/10 to 7/10). Some of this could be explained by the fact that different journals may appear in the “top ten” list from year to year. However, given the overall increase in data sharing requirements, it seems likely that it reflects, at least to some extent, changes within these subject areas. Even though the specific journals on the “top ten” list change, the list remains generally representative of researcher requirements and disciplinary trends, since there is a desire to publish in whatever journals have the highest impact at the time.

**Table 1.** Top JCR subject categories for deposit requirements in 2016.

JCR category	Number of ‘deposit required’ policies (out of a possible 10)
Biochemical research methods	8
Biochemistry & molecular biology	8
Cell biology	8
Crystallography	8
Evolutionary biology	8
Ecology	7
Meteorology & atmospheric sciences	7
Multidisciplinary sciences	7
Mycology	7
Chemistry, multidisciplinary	6
Chemistry, organic	6
Chemistry, physical	6

JCR category	Number of 'deposit required' policies (out of a possible 10)
Mathematical & computational biology	6
Plant sciences	6
Agricultural economics policy	5
Agronomy	5
Energy & fuels	5
Entomology	5
Materials science, multidisciplinary	5
Nanoscience & nanotechnology	5
Physics, applied	5
Physics, condensed matter	5

There were also some categories that do not have any data deposit requirements or recommendations. Categories that did not have policy in both 2014 and 2016 include: Computer science, theory & methods; Electrochemistry; Engineering, industrial; Engineering, marine; Logic; and Physics, particles & fields.

#### **Policy changes from 2014 to 2016**

Because journals could appear in multiple subject categories in the same year, we de-duplicated the ISSNs before conducting further analysis. As mentioned above, there were 880 unique ISSNs in our 2016 dataset (ISSNs pulled from 2015 JCR impact factor rankings). In the 2014 dataset (ISSNs pulled from 2012 JCR impact factor rankings) there were 836.

As seen in Table 2, in 2014, 71.2% of journals had no policy, 14.7% had a recommended policy, and 13.9% had a required policy (n=836). In contrast, in 2016, there were 58.5% with no policy, 19.4% with a recommended policy, and 22.0% with a required policy (n=880). Though the datasets for the two years did not contain identical ISSNs (because there were changes in which journals were considered high impact between the years), this gives a general understanding of shifts in the policy landscape of high impact titles.

**Table 2.** Overall changes in data deposit policies, 2014 to 2016.

	Year			
	2014		2016	
	Count	Percent	Count	Percent
No Policy	597	71.4%	515	58.5%
Recommended Policy	123	14.7%	171	19.4%
Required Policy	116	13.9%	194	22.0%
Total Policies	836		880	

### Policy shifts within individual journals

For the next stage of our analysis, we specifically examined the 610 unique ISSNs that appeared in both years. We then looked closely at how individual policies were changing.

Of these 610 ISSNs, 158 (26.0%) had undergone a policy shift, with 130 (82.3%) of these shifts towards greater data sharing. Overall, in 2016 there were 30 more ISSNs with a recommended policy than in 2014 (a 32.5% increase), and 49 more ISSNs with a required policy (a 54.4% increase). See Table 3 for a detailed breakdown of the nature of these changes.

**Table 3.** Breakdown of policy changes for journal policies reviewed in both 2014 and 2016. A total of 158 out of 610 total policies changed.

2014 policies → 2016 policies	Journals with policy changes
No policy → Recommended	72
No policy → Required	30
Recommended → Required	28
Required → Recommended	5
Recommended → No policy	19
Required → No policy	4
Total policy changes	158

### Recognition of institutional solutions

One of the factors of interest was whether journals recognised institution-based data solutions as acceptable options for the sharing and preservation of data. We coded each policy with a “1” if the policy mentioned institutional repositories or other institutional data solutions, and a “0” if they did not or if they explicitly stated that it was not an acceptable solution. There has been a noticeable increase in the number of policies that address institutional solutions, going up from 6 in 2014 (0.7%, n=836) to 78 in 2016 (8.9%, n=880). Anecdotally, changes to publisher-level policies may account for some of this shift, as policies within titles from the Nature Publishing Group, the Royal



Society of Chemistry, BioMed Central, Wiley-Blackwell and the American Meteorological Society account for the most mentions of institutional solutions.

**Table 4.** Institutional data solutions as an option for data deposit, by year.

	Year			
	2014		2016	
	Count	Percent	Count	Percent
Institutional repository not mentioned	830	99.3%	802	91.1%
Institutional repository explicitly mentioned	6	0.7%	78	8.9%
Total Policies	836		880	

### Recommended repositories

Also recorded were any specific data repositories mentioned as recommended or required options in the title-level or publisher-level policies. In 2014, 239 policies made mention of repositories a total of 3,233 times and in 2016, 280 policies made mention of repositories a total of 7,343 times. Overall, 287 unique repositories were named over both years. In some cases these were consortia or overhead bodies. For example, if the International Molecular Exchange Consortium (IMEx) was named, that could be referring to any number (or all) of their 16 partner repositories, which include DIP and IntAct.

Appendix A, details the top 20 repositories mentioned for both the 2014 and 2016 policy analysis. NCBI GenBank was the repository named most often in both 2016 and 2014, with 148 and 121 mentions respectively. One factor that may have influenced the list of repositories mentioned was the existence of multiple publisher-level lists of recommended repositories. These lists, from publishers such as Springer Nature or the American Geophysical Union, can be lengthy and cover many types of data and journals. A specific journal may link to one of these publisher lists, but in practice its articles would not likely involve data relevant to all of the repositories mentioned. In addition to exhaustive lists of options, other journals provided only a few examples of appropriate repository solutions. Some policies mentioned that using the listed repository for a particular type of data was the only acceptable deposit location.

### Other noted differences

Though data was not systematically collected on facets other than deposit recommendations and requirements, a few other changes in the author guidelines between 2014 and the end of 2016 were noted.

### Data Papers

In 2016, we noted an increase in data papers being encouraged as an additional option to showcase data mentioned in the article submission. Additionally, some journals, such as *Annals of Forest Science*, now include data papers as a type of submission to the journal itself alongside review and research articles (instead of a separate journal dedicated to the publication of data papers).



### ***Supplementary Files***

In 2016, most journals continue to offer the option of the submission of supplementary files. It was also noted that there are now partnerships between some publishers and data repositories for the hosting of supplementary information. For example, OSA Publishing, as well as some Royal Society and Taylor & Francis Group titles, have partnered with figshare. Similarly, certain Elsevier titles offer a Mendeley Data deposit option. The supplemental files submitted to these journals are then hosted on the partner repository instead of by the journal or publisher. This action is facilitated by the journal or publisher upon paper submission or acceptance, instead of the author uploading the files directly to the repository.

### ***Data linking and data citation***

The 2016 policies included a number which discussed data linking (appearing both as a recommendation and a requirement) for referenced data, as well as original datasets discussed in papers. Many journals also had detailed information on how to reference datasets and whether it was acceptable to reference unpublished data. Although data was not gathered on data linking and data citation, this is another area that implies authors must think about the need to make referenced data (theirs or others) findable and available. Statements of data availability were also mentioned in many policies, which allow for both publicly available datasets as well as those only available through contact with the author.

### **Impact on University of Toronto Authors**

The top physical and applied science journals requiring data deposit among U of T researchers are listed in Table 5. All 1,662 publications were spread among a total of 134 separate journals.

**Table 5.** ‘Deposit required’ journals with the highest number of University of Toronto publications.

Journal Title	U of T Publications
Journal of the American Chemical Society	141
Environmental Science and Technology	111
Proceedings of the National Academy of Sciences of the United States of America	106
Dalton Transactions	63
Organic Letters	60
Lab on a Chip - Miniaturisation for Chemistry and Biology	55
ACS Nano	48
Organometallics	45
Bioinformatics	39
Journal of Organic Chemistry	37
Nature Communications	37
Advanced Materials	35

Journal Title	U of T Publications
Science	34
Nature	29
Journal of Climate	25
Physical Chemistry Chemical Physics	25
Scientific Reports	24
Inorganic Chemistry	22
Journal of Biomolecular NMR	21
Small	20

The U of T is a large, diverse university, and there were 62 separate departments represented in the dataset of articles published in high-impact physical and applied science journals. Table 5 displays the departments with the largest number of publications in these journals from 2009 to 2015. The Department of Chemistry was by far the most prolific, with 543 publications (32.7% of all publications). Authors working in departments related to the areas of biology, medicine, and engineering also regularly publish in journals requiring data deposit. It was interesting that departments in the Faculty of Medicine were so well represented in publications in the physical and applied sciences, given that health and medical science categories were not included in our list of high-impact journals.

**Table 6.** University of Toronto departments with the highest number of publications in “deposit required” journals high impact journals.

Department	Number of Articles
Department of Chemistry	543
Department of Ecology and Evolutionary Biology	109
Department of Molecular Genetics	103
Department of Electrical and Computer Engineering	70
Department of Cell and Systems Biology	65
Department of Chemical Engineering and Applied Chemistry	60
Institute of Biomaterials and Biomedical Engineering	56
Department of Physics	45
Department of Physical and Environmental Sciences, UTSC	43
Donnelly Centre for Cellular and Biomolecular Research	42
Department of Mechanical and Industrial Engineering	41
Banting and Best Department of Medical Research	37
Department of Biology, UTM	35
Department of Medical Biophysics	34
Department of Biochemistry	33
Department of Computer Science	29

Department	Number of Articles
Department of Laboratory Medicine and Pathobiology	24
Department of Physiology	23
Department of Biological Sciences, UTSC	22
Department of Civil Engineering	18
Department of Materials Science and Engineering	18
Department of Medicine	18

## Discussion

In 2016, 41.5% of the journals we reviewed had a data deposit policy, with a roughly even split between recommended versus required wording. This indicates that data deposit requirements in the physical and applied sciences are fairly common, but not as common as in the field of medicine, where nearly a decade ago policies were already widespread (Piwowar and Chapman, 2008). In addition, our study shows a strong general trend towards an increase in the number of data deposit policies in physical and applied science journals between 2014 and 2016. This is something we will continue to monitor.

Despite the increasing prominence of data deposit requirements, journals have by no means adopted a consistent approach to handling the matter. Wording varied widely and requirements were often ambiguously or inconsistently stated, which made coding a challenge. For example, the journal *FEMS Microbiology Reviews* mentions data in its author instructions, but only as supplementary data; however, in its journal policies it strongly recommends the deposit of organism, virus, and vector data into publicly available repositories. In some cases, individual journals would have their own policies, but would also link to a broader policy adopted by the publisher and intended to apply to many journals (often with only one policy for a wide range of subject areas). This was particularly common in cases where the publisher had recently made a change to their policies, but individual journals may have not all caught up (e.g. Nature Publishing Group/Springer Nature).

We identified that U of T chemistry researchers are by far the most heavily affected by journal data deposit requirements for the titles we examined, followed by biology, engineering, and medicine (though medical researchers were not in the original target group of physical and applied sciences researchers). Future work will involve a compliance review to identify whether these researchers are, in fact, complying with data deposit requirements in particular journals. It may be that the library can provide targeted training and support to those researchers who are not currently complying.

Looking at it from another perspective, U of T departments in subject areas related to chemistry, biology, medicine and some facets of engineering, may be very well prepared to handle funder data sharing requirements when they emerge in Canada. Departments which publish less regularly in journals requiring data deposit, may be more in need of training and support. At U of T, this would potentially include forestry, geology, mathematics, and public health, as well as areas that are traditionally considered social sciences (but which sometimes publish in physical science journals) such as geography and management.

Over the past two years there has also been a large increase in policies identifying institutional repositories as an option for data deposits. Some policies did specify that these institutional solutions should, or must, be able to provide a DOI for the deposited data. This is useful information for planning institutional services and outreach.

Also useful for outreach is the list of repositories mentioned in the policies. These can reveal highly recommended repositories for specific areas, which can help us tailor our training and discussions with researchers. They can provide guidance as to where to refer researchers for deposit, or where we might locate U of T data in the absence of funding body recommendations.

We intend to continue to run this analysis on a regular basis, to monitor the changing journal policy landscape, and to continue to build knowledge about local institutional needs and practices.

Please contact the authors to inquire about data sharing. The Python script generated during the current study is available on the GitHub repository.<sup>2</sup>

## References

- Dearborn, D. & Marks, S. (2015). Local assessment of science RDM practices as examined through journal policies. Poster presented at the 10<sup>th</sup> International Digital Curation Conference, London, UK. Retrieved from [http://www.dcc.ac.uk/sites/default/files/documents/IDCC15/210\\_rdmpractices.pdf](http://www.dcc.ac.uk/sites/default/files/documents/IDCC15/210_rdmpractices.pdf)
- Fear, K. (2015). Building outreach on assessment: Researcher compliance with journal policies for data sharing. *Bulletin of the Association for Information Science & Technology*, 41(6): 18-21. Retrieved from <https://www.asist.org/publications/bulletin/aug-2015/building-outreach-on-assessment/>
- McCain, K. (1995) Mandating sharing: Journal policies in the natural sciences. *Science Communication*, 16(4): 403-431.
- Piwovar, H.A. & Chapman, W.W. (2008). A review of journal policies for sharing research data. *Proceedings of the ELPUB Conference on Electronic Publishing*, Toronto, Canada. Retrieved from <https://elpub.architexturez.net/doc/oai-elpub-id-001-elpub2008>
- Research Data Canada – Research Data Strategy Working Group. (2008). Stewardship of research data in Canada: A gap analysis. Government of Canada. Retrieved from <https://www.rdc-drc.ca/?wpdmdl=679>
- Sewerin, C., Dearborn, D., Henshilwood, A., & Spence, M. (2015). Report on research data management faculty and postdoctoral survey. University of Toronto. Retrieved from <http://hdl.handle.net/1807/69948>
- Shearer, K. (2015). Comprehensive brief on research data management policies. Unpublished report. Retrieved from <http://docplayer.net/17594465-Comprehensive-brief-on-research-data-management-policies.html>

---

<sup>2</sup> Scripts available on GitHub: <https://github.com/sbmarks/scopus-queries>

- SHERPA/JULIET. (2017). List of all research funders. Website. Retrieved from <http://www.sherpa.ac.uk/juliet/index.php?la=en&mode=advanced&page=browse>
- Sturges, P., Bamkin, M. Anders, J. & Hussain A. (2014). Access to research data: Addressing the problem through journal data sharing policies. *2014 IATUL Proceedings*. Retrieved from <http://docs.lib.purdue.edu/iatul/2014/openaccess/3/>
- Sturges, P., Bamkin, M., Anders, J. Hubbard, B., Hussain, A. & Heeley, M. (2015). Research data sharing: Developing a stakeholder-driven model for journal policies. *ASIS&T Conference Proceedings*. doi:10.1002/asi.23336
- Weber, N., Piwovar, H. & Vision, T. (2010). Evaluating data citation and sharing policies in the environmental sciences. *2010 ASIST Conference*. Retrieved from [https://www.asis.org/asist2010/proceedings/proceedings/ASIST\\_AM10/submissions/445\\_Final\\_Submission.pdf](https://www.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/445_Final_Submission.pdf)
- Zenk-Möltgen, W. & Lepthien, G. (2014). Data sharing in sociology journals. *Online Information Review*, 38(6): 709-722. doi:10.1108/OIR-05-2014-0119

## Appendix A

**Appendix A.** Top 20 named repositories for data deposit by number of times mentioned in the 2014 and 2016 journal policies. Repositories listed as they appeared in the policy, which can include entries for consortia and overhead bodies.

Rank	Repository	Times mentioned	
		2014	2016
1	NCBI GenBank	121	148
2	Cambridge Crystallographic Data Centre (CCDC)	105	144
3	NCBI Gene Expression Omnibus (GEO)	94	136
4	PANGAEA	93	119
5	RCSB Protein Data Bank (PDB)	93	115
6	Dryad Digital Repository	86	113
7	DNA Data Bank of Japan (DDBJ)	77	109
8	ArrayExpress (EMBL-EBI)	68	90
9	Molecular Interaction database (MINT)	59	90
10	PubChem	47	81
11	HepData (High Energy Physics Data Repository)	44	81
12	NCBI Taxonomy database	44	80
13	SIMBAD Astronomical Database	44	80
14	European Nucleotide Archive (ENA EMBL-EBI)	41	80
15	PRoteomics IDentifications database (PRIDE EMBL-EBI)	41	80
16	NCBI Sequence Read Archive (SRA)	40	79
17	Database of Interacting Proteins (DIP)	40	79
18	Universal Protein Resource (UniProt)	39	79
19	Biological Magnetic Resonance Bank (BMRB)	37	76
20	IntAct Molecular Interaction Database (EMBL-EBI)	37	76