Data in Brief: Can a Mega-Journal for Data be Useful?[1]

As part of the current move towards open science, there is increasing pressure for scientists to share their research data. In support of this, several journals only publish descriptions of data generated from research: data papers. It is not clear whether this service encourages data reuse, however. This article assesses the prevalence and impact of the largest such journal, *Data in Brief*, comparing it with 24 other general or specialist data journals. The results show that *Data in Brief* became the largest data journal in 2016 and that its papers attracted over five Mendeley readers each, within a year of publication, as well as a non-trivial amount of citations. Its papers have been cited for relevance or facts contained in them in addition to acknowledging the reuse of associated datasets in about 1% of cases. Some papers describe electronic dataset whereas other papers embedded the tables or images that formed the shared data. Overall, the journal seems to make a positive contribution to science by enabling access to multiple types of data, even though its papers rarely lead to data reuse.
Keywords: Open science; data sharing; data journals; data papers; citation analysis; Mendeley.

Introduction

Sharing research data has at least two benefits for science. It makes research more transparent by supporting methods checks of articles, for method triangulation or to identify errors (Borgman, 2012). It also makes research efforts reusable by allowing others to re-analyse the data for new purposes, especially in fields with relatively standardised data, such as genomics (Field, Sansone, Collis, Booth, Dukes, Gregurick, & Millard, 2009), biodiversity (Khan, Thelwall, & Kousha, 2019) and parts of neuroscience (Poldrack & Gorgolewski, 2014). Because of the collective advantages for science, there is increasing pressure on authors to share data, including through funder mandates and journal policies. This sharing is possible with supplementary files or links to datasets in repositories elsewhere (Reilly, Schallier, Schrimpf, Smit, & Wilkinson, 2011). The disadvantages of data sharing are primarily individual: it requires effort to share data in an understandable and adequately documented format, and scientists may wish to fully exploit their own data to avoid the risk that others can publish before them (Tenopir, Allard, Douglass, Aydinoglu, Wu, Read, & Frame, 2011), perhaps especially if they are working in a resource-poor environment.

Announcing a dataset through a data paper (Callaghan, Donegan, Pepler, et al., 2012) may be an attractive option for authors since it publicises the data and gives them a record for their CV in return for providing a systematic description. Some journals have allowed data papers to be published since at least 2009 (Newman & Corke, 2009), and journals exclusively publishing dataset descriptions have existed since at least *Earth System Science Data* in 2009 (Pfeiffenberger & Carlson, 2011). A survey of journals publishing data papers in 2014 found 116, including six data journals (Candela, Castelli, Manghi, & Tani, 2015).

Scientists deciding whether to share data may write a data paper if they think that the benefits are likely to outweigh the likely financial and time costs. Data sharing does not

[1] Thelwall, M. (in press). Data in Brief: Can a mega-journal for data be useful? *Scientometrics*.

automatically translate to data reuse (Wallis, Rolando, & Borgman, 2013) and data citations (Silvello, 2018), although there is evidence from one field that authors sharing data may attract more citations to the associated article (Piwowar & Vision, 2013). Nevertheless, there is no evidence yet about the likely benefits of publishing data papers in data journals, so cost-benefit analyses must rely on guesswork. This article addresses this gap with an analysis of data journals at the end of 2019, focusing on the new mega-journal, *Data in Brief*.

Data in Brief is a journal launched by Elsevier in 2014 as a place for (a) articles describing data that may be of lasting value when the originator could *not* generate a publishable analysis of it, and (b) more extensive descriptions of data than could reasonably be accommodated in an article published with the data (Shaklee, 2014). Both have the aims of ensuring that data has lasting value to the academic community, irrespective of whether it is associated with a published article. Although no analysis is needed for a data paper, the authors must provide an explanation about why the data has value to the scientific community (Shaklee, 2014). *Data in Brief* is a gold open access journal with a publication fee of $600 (December, 2019: https://www.elsevier.com/journals/data-in-brief/2352-3409/open-access-journal). Articles are peer reviewed, presumably for coherence of the description and to check the value of the data. Publishing a *Data in Brief* article is nevertheless a relatively straightforward and cheap way to get an article in Scopus for inexperienced authors, so it may also be attractive as a publishing outlet in its own right. Associated datasets can be shared online in the *Data in Brief* Dataverse (dataverse.harvard.edu/dataverse/dib; 28 datasets by December 2019) or elsewhere.

Scholarly journals dedicated to data have existed for many decades within specialist areas. An early example, *Nuclear Data Sheets*, is a scholarly journal in the sense that it has an editorial board of academics, uses peer reviewers and contains articles written by, and targeted at, academics. Articles report data and evaluate data of a specific type. For example, since 1975 it has published at least six articles with the same title, "Nuclear data sheets for A=155", with the latest reporting, "This work represents an update of the previous evaluation of the nuclear data on the A=155 nuclides" (Nica, 2019). This could be viewed as a traditional journal with an emphasis on data or a data journal with an emphasis on data evaluation.

Data journals perhaps supplement repositories that hold specific types of data for communities, such as genomics data, biodiversity information, telescope recordings, and protein structures (e.g., Robertson, Döring, Guralnick, et al., 2014; White, Carrier, Thompson, Greenberg, & Scherle, 2008), as well as generic data websites, such as FigShare (Thelwall & Kousha, 2016). For datasets that do not match specialist databases, a textual description is probably necessary to explain what the data is and how it was collected. Some publication platforms, such as F1000Research, also publish descriptions of data, which are described as data papers. By December 2019, Scopus had indexed 5,232 documents that it categorised as Data Papers. These were mainly (91%: 4,781) from *Data In Brief*, with other main journals including *Chemical Data Collections* (299) and *Geoscience Data Journal* (44), all of which are data journals. Over the past decade, several general data journals have begun publication, and there have been at least 100 journals or platforms that were willing to publish, or exclusively published, data papers (Candela, Castelli, Manghi, & Tani, 2015). Data journals, sometimes for a fee, publish data and associated descriptions as articles. Some of these journals are associated with publishers so that authors, when they are

encouraged or mandated to share data associated with an article, may be offered the option to publish a data paper in the publisher's journal.

The aim of this article is to assess whether *Data in Brief* makes a valuable contribution to scholarship through a descriptive analysis of its articles. A simple descriptive analysis is appropriate since this journal has not been systematically analysed before and it is of a relatively new type. It is therefore useful to assess its basic properties initially. Citation counts can be used as a proxy for attention here since the focus is on the academic community. Although data can be used in many ways that do not lead to citations, such as for methods training and validation of the analyses in associated papers, data creators can use citations as tangible evidence of the value of their work. Thus, a citation is both proof that data sharing has been useful (unless it is a self-citation) and a reward for its creator. As a secondary analysis, Mendeley was queried to assess whether counts of readers could give useful insights into the value of recent papers for which citations would not be available. In theory, people that use datasets for purposes other than generating new knowledge might register the associated papers in Mendeley, making it potentially a better source of impact evidence than citation counts.

Methods

The overall research design was to obtain citation and Mendeley readership information for *Data in Brief* papers and to contextualise it against similar information from other data journals.

A list of data journals was created by searching the web for pages listing them and searching Scopus and the Web of Science for journals with "data" in their names, manually filtering out periodicals that matched the query but were not data journals (e.g., *Data Science*). A data journal was defined as a journal that advertised itself as primarily publishing articles about data in some form, whether or not that data was usually associated with other papers or saved in a separate dataset.

The journals found were queried in Scopus to obtain a list of outputs of all types from them. Editorials were filtered out from the results. The journals were also queried in Dimensions.ai for triangulation. Scopus results can be incomplete because it does not aim for exhaustive coverage of the academic literature whereas Dimensions results can be incomplete if the information is not available from publishers or it has not found it (Thelwall, 2018b). The results were also compared with the Web of Science, but this did not produce additional matches so these results were not reported.

The Scopus and Dimensions APIs were used with journal name searches to identify papers in relevant journals, with the results manually checked and filtered for false matches. The following data journals not in Scopus were downloaded from Dimensions.ai instead: Journal of Open Archaeology Data; Open Health Data; Journal of Open Psychology Data; Dataset Papers in Science; Research Data Journal for the Humanities and Social Sciences.

The Mendeley API was used to identify reader counts of all articles found by Scopus or Dimensions. Both DOI searches and metadata searches (e.g., title:A reanalysis dataset of the South China Sea AND author:Zeng AND year:2014) were combined by Webometric Analyst, with filtering to remove false matches (http://lexiurl.wlv.ac.uk/searcher/mendeley.htm; for full details, see: Thelwall & Wilson, 2016) to obtain the most complete Mendeley reader counts (Zahedi, Haustein, & Bowman, 2014). Mendeley provides earlier impact evidence than citation counts from Scopus

(Thelwall, 2017b, 2018a), which is useful here since *Data in Brief* is relatively young, and is useful for all academic fields (Thelwall, 2017a).

A content analysis was used to categorise citation contexts for a random set of 200 Data in Brief articles, selected with a random number generator. The citations were chosen from Google Scholar. Three independent experienced content analysis coders decided whether (a) each article had at least one non-self-citation, and (b) the reason for the citation, using the categories described in the results section. The Cohen Kappas for the two categories were 0.987 and 0.566, which are both adequate for reporting. The author resolved cases of disagreement.

Results

Background information about data journals is given to contextualise the position of *Data in Brief* before reporting a citation and readership analysis.

Data journals

There is a long tradition of data journals in some areas of science, including chemistry, engineering, physics, and health (Table 1). As far as indexed by Scopus, the largest data journal appears to be the American Chemical Society's *Journal of Chemical & Engineering Data*. This journal publishes, "articles containing data on the phase behavior and the physical, thermodynamic, and transport properties of well-defined materials, including complex mixtures of known compositions" (ACS, 2019). In addition, "Articles should present a significant amount of experimental or computational data on properties of systems of technological or theoretical interest that are not available in the original literature, that have lower uncertainty than those published, or that help resolve conflicts in previously published values" (ACS, 2018). These articles seem to be a hybrid between purely descriptions of data and a more traditional article with a wider analytic component.

Whilst the table confirms that data journals have existed for a long time in specialist areas, from 2012 there has been a growth in the number of data journals, including four with general remits: *Data; Data in Brief; Dataset Papers in Science; Scientific Data*.

Table 1. Sizes and first publication years of the data journals investigated.

Data Journal	First year	Last year	Papers
J Chemical & Engineering Data	1956		16294
Radiological Health Data & Reports	1963	1971	767
Vital & Health Statistics, Series 10	1963	2013	134
Engineering Sciences Data Unit	1965	1989	677
Nuclear Data Sheets	1965		1472
Atomic Data & Nuclear Data Tables	1969		1124
J Physical & Chemical Reference Data	1972		969
Radiation Data & Reports	1972	1974	318
Vital & Health Statistics, Series 13	1977	2011	90
Advance Data	1981	2008	251
Capitation Rates & Data	1997	2008	474
NCHS Data Brief	2007		339
Earth System Science Data	2009		452
J Open Archaeology Data	2012		31
Biodiversity Data J	2013		580
Genomics Data	2013	2017	660
J Open Psychology Data	2013		32
Open Health Data	2013		23
Data in Brief	2014		4701
Scientific Data	2014		1108
Dataset Papers in Science	2015		23
Geoscience Data J	2015		46
Chemical Data Collections	2016		291
Data	2016		257
Research Data J Hum & Soc Sci	2016		21

In terms of evolution over time, *Journal of Chemical & Engineering Data* published the most articles annually until 2016, when *Data in Brief* surpassed it (Figure 1). *Scientific Data* became the third largest data journal in 2018, also experiencing rapid growth. Thus, recent years have seen two general data journals become successful for the first time.

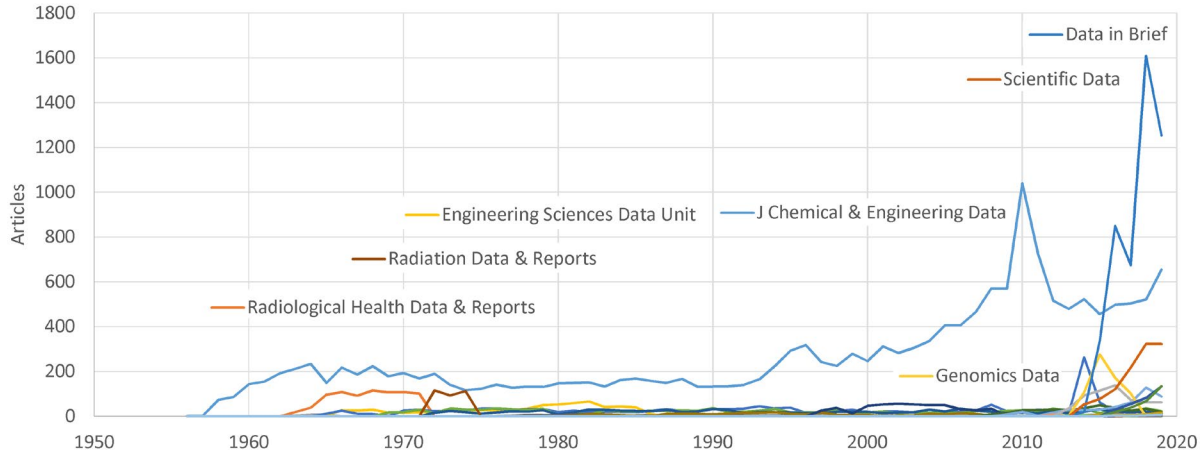

Figure 1. Number of documents in each of the 25 data journals, with selected larger journals labelled. Full graph available in the supplementary materials. This shows that *Data in Brief* became the largest data journal in 2018.

Of the recent set of journals, *Scientific Data* articles attracted the most average attention from Mendeley readers, but *Data in Brief* articles also attracted a substantial amount of attention (more than 5 readers per article until 2018) (Figure 2). Citation graphs and Mendeley reader graphs are available in the supplementary material for all journals.

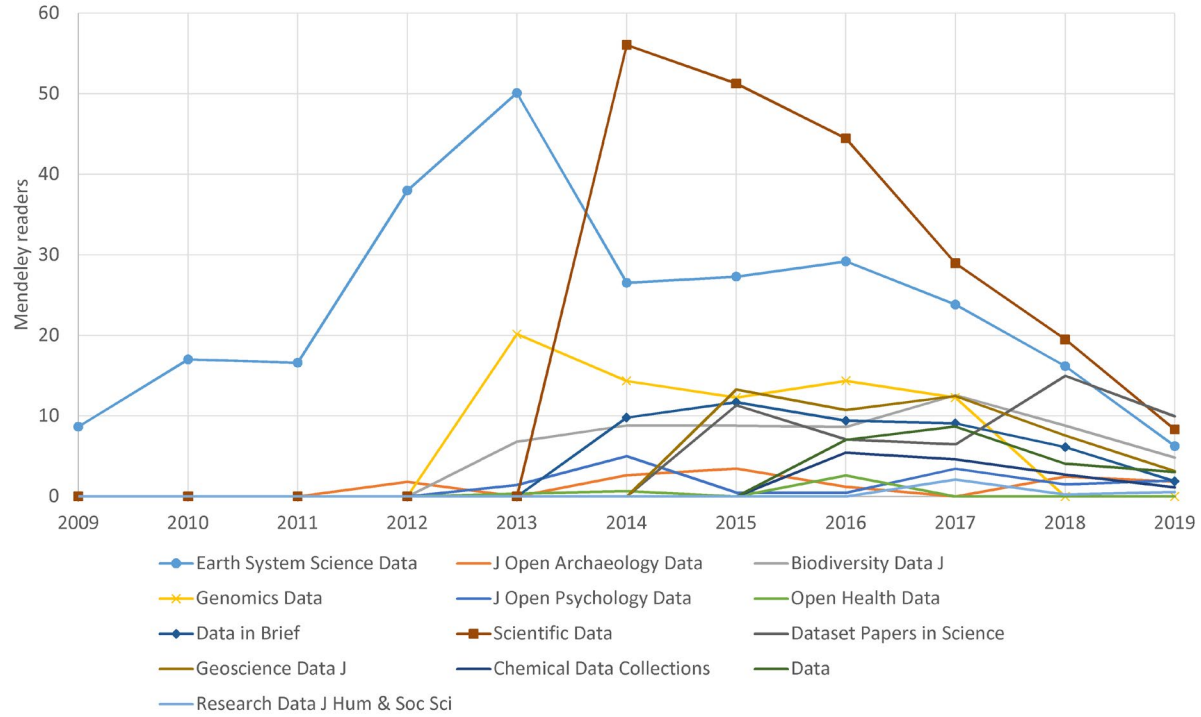

Figure 2. Average (geometric mean) number of readers per article for data journals that started publishing after 2008. This shows that *Data in Brief* attracts a moderate amount of Mendeley readers compared to contemporary data journals. Citation count and Mendeley reader count graphs for all journals are available in the supplementary materials: http://doi.org/10.6084/m9.figshare.11793876.

Data in Brief

Data in Brief papers were investigated in more detail to find out who was publishing in the journal. The national affiliations of authors broadly reflect active research publishing nations (Table 2), but Nigeria is an exception. The list of most publishing institutions is led by a new evangelical Christian university from Nigeria, with several of its faculty publishing many articles in the journal (Table 3). Other institutions in the list are not amongst the most active in the world, suggesting that some institutions, or active individuals within them, have systematically adopted *Data in Brief* as a publishing outlet – perhaps because they produce useful data or for policy reasons.

Table 2. Number of *Data in Brief* articles published by December 2019 by country.

Country	**Papers**
United States	910
China	379
Italy	321
Germany	317
India	315
United Kingdom	290
Iran	263
Nigeria	259
Japan	241
Canada	200

Table 3. Number of *Data in Brief* articles published by December 2019 by author affiliation.

Affiliation	**Papers**
Covenant University	228
Tehran University of Medical Sciences	82
CNRS Centre National de la Recherche Scientifique	59
Chinese Academy of Sciences	47
Ministry of Education China	39
Kermanshah University of Medical Sciences	38
Russian Academy of Sciences	38
Universidade de Sao Paulo - USP	36
Bushehr University of Medical Sciences	36
Consiglio Nazionale delle Ricerche	35

Documents citing Data in Brief Data Papers from 2017

Whilst the Mendeley reader counts reported above suggest that *Data in Brief* papers are attracting attention, they do not reveal why people have used them. Citation information can give some insights into this. The Scopus citations to *Data in Brief* papers from 2017 were mostly from standard articles, although a substantial minority were from Data Papers or Reviews. Authors of *Data in Brief* papers sometimes cite other *Data in Brief* papers by others or themselves to give context to their contribution.

Table 4. Documents citing *Data in Brief* articles from 2017, by document type, as of December 2019.

Document type	**Papers**	**Percentage**
Article	965	69.1%
Data Paper	152	10.9%
Review	140	10.0%
Conference Paper	95	6.8%
Book Chapter	27	1.9%
Letter	4	0.3%
Editorial	3	0.2%
Note	3	0.2%
Book	2	0.1%
Undefined	6	0.4%

Some of the institutions most citing *Data in Brief* articles from 2017 (Table 5) are the same as those most publishing *Data in Brief* articles (Table 3), which is presumably due to self-citations. This is unsurprising since *Data in Brief* papers can be associated with full articles by the same authors that report an analysis of the data.

Table 5. Documents citing *Data in Brief* articles from 2017, by affiliation, as of December 2019.

Affiliation	**Papers**
Covenant University	95
Tehran University of Medical Sciences	41
Ahvaz Jundishapur University of Medical Sciences	35
Hindustan Institute of Technology and Science	27
Università degli Studi di Catania	26
Bushehr University of Medical Sciences	24
Islamic Azad University, Bushehr Branch	22
CNRS Centre National de la Recherche Scientifique	19
Neyshabur University of Medical Sciences	18
Abadan University of Medical Sciences	15

A wide range of Scopus broad subjects cite *Data in Brief* articles (Table 6), confirming that it is a multidisciplinary journal.

Table 6. Documents citing *Data in Brief* articles from 2017, by Scopus broad subject, as of December 2019. Documents can be multiple classified, so percentages add up to more than 100%.

Subject	**Papers**	**Percentage**
Biochemistry, Genetics and Molecular Biology	310	22.2%
Medicine	272	19.5%
Engineering	254	18.2%
Environmental Science	194	13.9%
Multidisciplinary	174	12.5%
Chemistry	148	10.6%
Agricultural and Biological Sciences	144	10.3%
Computer Science	143	10.2%
Materials Science	135	9.7%
Chemical Engineering	97	6.9%
Pharmacology, Toxicology and Pharmaceutics	96	6.9%
Neuroscience	75	5.4%
Social Sciences	73	5.2%
Energy	69	4.9%
Physics and Astronomy	68	4.9%
Immunology and Microbiology	63	4.5%
Earth and Planetary Sciences	46	3.3%
Mathematics	40	2.9%
Health Professions	22	1.6%
Psychology	19	1.4%
Arts and Humanities	16	1.1%
Business, Management and Accounting	15	1.1%
Decision Sciences	13	0.9%
Economics, Econometrics and Finance	10	0.7%
Nursing	10	0.7%
Veterinary	7	0.5%
Dentistry	2	0.1%

The registered Mendeley readers of *Data in Brief* articles are mainly junior researchers or students rather than senior researchers (Table 7). This is partly because Mendeley users tend to be more junior (Mohammadi, Thelwall, Haustein, & Larivière, 2015).

Table 7. Declared occupations of Mendeley readers of *Data in Brief* articles from all years, as of December 2019.

Occupation	Total	Percentage
Professor	1727	4.9%
Associate Professor	1923	5.5%
Senior Lecturer	456	1.3%
Lecturer	1236	3.5%
Researcher	5695	16.2%
PhD Student	6685	19.0%
Doctoral Student	2543	7.2%
Postgraduate	1435	4.1%
Master Student	6945	19.8%
Bachelor Student	4338	12.4%
Librarian	366	1.0%
Other	1771	5.0%
Total	**35120**	**100.0%**

Examples of Data in Brief articles

A random sample of 200 *Data in Brief* articles from 2017 were investigated to assess why they had been cited, using Google Scholar to identify citations, and excluding self-citations (Table 8). Most (60%) articles had at least one non-self-citation. When the non-self-citation could be checked, data reuse occurred in only 2% of cases, with the most common citation contexts being a discussion of related work (45%) or the invocation of a specific fact (52%). Data reuse is impossible for the many *Data in Brief* articles that contain their data in the paper in the form of tables or figures, rather than as an electronic dataset. An example is the figures shared in the paper, “Biphasic calcium phosphates (BCP) of hydroxyapatite (HA) and tricalcium phosphate (TCP) as bone substitutes: Importance of physicochemical characterizations in biomaterials studies”. For *Data in Brief* articles without external datasets, citing a fact in them equates to fully using the data in the article. The key categories are discussed below.

- **Related**. Articles may be cited within lists of references to show awareness of related papers without explicitly referring to their contents. For example, “Mass spectrometry data from label-free quantitative proteomic analysis of harmless and pathogenic strains of infectious microalgae, Prototheca spp.” was cited in a list, “Despite the development of proteomics platforms, not many approaches have been developed to study microalgae proteomes [6,29–51].” (Fajardo, Amil-Ruiz, Fuentes-Almagro, et al., 2019). Articles may also be cited as background information to an article rather than directly using the information presented to inform the methods or interpretation of the results of the citing article.
- **Data reuse**. The data described in the paper may be re-used by other researchers for follow-up research (i.e., the standard use-case for data sharing). For example, “Tashkeela: Novel corpus of Arabic vocalized texts, data for auto-diacritization systems” was cited as follows, “To build the Arabic frequency dictionary, we used the freely available Tashkeela5 corpus [reference] composed of around 70 million diacriticized Arabic words.” (Nassiri, Lakhouaja, & Cavalli-Sforza, 2017).
- **Cite for fact.** The data paper may be cited for a fact within it. For example, “Data for Korean college students’ anxious and avoidant attachment, self-compassion, anxiety

and depression" was cited by an MSc, "findings support the mediating role of self compassion in attachment style and emotional distress in college students [reference]" (Dökmeci, 2017). Another example is, "Dataset on the absorption of PCDTBT:PC70BM layers and the electro-optical characteristics of air-stable, large-area PCDTBT:PC70BM-based polymer solar cell modules, deposited with a custom built slot-die coater" cited as follows, "This material and its derivatives have been selected as the electron acceptors to improve light absorption in the visible light range, which leads to more excitons in the active layer. [reference]" (Mhamdi, Sweii, & Bouazizi, 2019). The cited facts could be complex and described as data, as in the following example, "For the analysis of this section, the regular ground is chosen to be the surface of CIRC case 2. PV panels are assumed to be Si pillar solar cells (pillar 3) with spectral reflectance data given by Ref. [8]" (Li, 2018), which refers to figures in the Data in Brief paper (which does not have a separate dataset).

Table 8. Context of citations to a random sample of *Data in Brief* articles from 2017, as checked in December 2019. Contexts are based on one randomly-selected non-self-citation for each qualifying article, using Google Scholar data.

Citation type	**Total**	**Percentage**
Reference without in-text citation	1	1%
Data reuse	2	2%
Related work (cited in list or as part of review)	46	45%
Fact (cited for a specific fact reported, not part of a review)	53	52%
Documents with non-self-citations checked	**102**	**100%**
Citing document paywalled	19	
No citations	40	20%
Self-citations only	39	20%
Total	**200**	**100%**

Discussion

Several shortcomings should be considered when interpreting the findings. The results are limited by the choice of journals. It was not possible to check the tens of thousands of academic journals individually and there may be journals not examined that could reasonably be classified as data journals. For example, many papers in *Zookeys* (covering, "zoological taxonomy, phylogeny, and biogeography") seem to be descriptive and this journal could conceivably be described as mainly a data journal. The Mendeley data may be misleading if it includes automated "readers". This could not be checked since the site does not reveal readers' identities for a paper.

The results show that *Data in Brief* is numerically the most successful current data journal of any type, in terms of attracting papers. They also suggest that *Data in Brief* papers attract attention after publication (both citations and Mendeley readers), adding value to academia. Concerns have previously been raised that shared data may typically be ignored (Wallis, Rolando, & Borgman, 2013), but the *Data in Brief* results are more positive than previously found for data sharing sites (Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2016; Thelwall & Kousha, 2016). This seems to be due to *Data in Brief* papers sometimes containing citable facts in an easily accessible open access format so that they are more like a brief communication short paper than a dataset with metadata. Underlying this difference, data sharing clearly means different things to different specialties. For some,

the data to be shared might be a set of tables, graphs or pictures that can be embedded in a data paper, whereas for others it might be huge databases of information, with the data paper describing their characteristics and provenance.

The origins of a large section of papers from relatively young universities suggests that *Data in Brief* may also serve as a viable publishing outlet for scholars without access to the mentoring that may be necessary to produce a standard journal article. This seems positive if it serves as a pathway to eventual publishing of full articles or a method to publish information that might otherwise be unpublished or overlooked, if the host university or funder can afford it.

An interesting facet of *Data in Brief* is that many of its papers embed the data that they are describing through tables or images rather than storing it elsewhere and linking to it, as might be expected (Callaghan, Donegan, Pepler, et al., 2012). This possibility was not mentioned by a previous survey of journals publishing data papers (Candela, Castelli, Manghi, & Tani, 2015). Such papers might be thought of as presenting facts or unanalysed information rather than describing a dataset, but, most importantly, this seems to be a useful service.

Conclusions

The results suggest that *Data in Brief* fills a useful role in the scholarly publishing landscape by providing a relatively cheap open access format for researchers to publish data in the form of tables, figures or descriptions of datasets. The data papers in the journal seems to have attracted a reasonable amount of interest and may be cited by others for facts, data re-use or awareness of related work. The only caveat is that papers rarely lead to formal data reuse in the form of citations from academic documents.

References

ACS (2018). Author guide, Version: December 13, 2018. http://pubsapp.acs.org/paragonplus/submission/jceaax/jceaax_authguide.pdf?

ACS (2019). About. https://pubs.acs.org/page/jceaax/about.html

Borgman, C. L. (2012). The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, 63(6), 1059-1078.

Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., & Wright, D. (2012). Making data a first class scientific output: Data citation and publication by NERCs environmental data centres. International Journal of Digital Curation, 7(1), 107-113. doi: 10.2218/ijdc.v7i1.218

Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. Journal of the Association for Information Science and Technology, 66(9), 1747-1762.

Dökmeci, M. (2017). Self-compassion as a mediator between parental acceptance-Rejection and emotional reactivity among university students (Master's thesis, Boğaziçi University).

Döring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., & Desmet, P. (2014). The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. PloS One, 9(8), e102623.

Fajardo, C., Amil-Ruiz, F., Fuentes-Almagro, C., De Donato, M., Martinez-Rodriguez, G., Escobar-Niño, A., & Fernandez-Acero, F. J. (2019). An “omic” approach to Pyrocystis

lunula: New insights related with this bioluminescent dinoflagellate. Journal of Proteomics, 209, 103502

Field, D., Sansone, S. A., Collis, A., Booth, T., Dukes, P., Gregurick, S. K., & Millard, S. (2009). 'Omics data sharing. Science, 326(5950), 234-236.

Khan, N., Thelwall, M., & Kousha, K. (2019). Data citation and reuse practice in biodiversity: Challenges of adopting a standard citation model. Proceedings of the 17th International Conference on Scientometrics & Informetrics, Rome, Italy (pp. 1220-1225).

Li, M. (2018). Spectral modeling of solar and atmospheric radiation for solar power integration. University of California.

Mhamdi, A., Sweii, F. B. S., & Bouazizi, A. (2019). Effect of Thermal Annealing on the Electrical Properties of Inverted Organic Solar Cells Based on PCDTBT: PC 70 BM Nanocomposites. Journal of Electronic Materials, 48(1), 352-357.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. Journal of the Association for Information Science and Technology, 66(9), 1832-1846. doi:10.1002/asi.23286

Nassiri, N., Lakhouaja, A., & Cavalli-Sforza, V. (2017). Modern standard Arabic readability prediction. In International Conference on Arabic Language Processing (pp. 120-133). Berlin, Germany: Springer.

Newman, P., & Corke, P. (2009). Data paper - Peer reviewed publication of high quality data. The International Journal of Robotics Research, 28(5), 587.

Nica, N. (2019). Nuclear data sheets for A= 155. Nuclear Data Sheets, 160(1), 1-404.

Pfeiffenberger, H. & Carlson, D. (2011). “Earth System Science Data” (ESSD): A peer reviewed journal for publication of data. D-Lib Magazine, 17(1/2). doi: 10.1045/january2011-pfeiffenberger

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175.

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. Nature Neuroscience, 17(11), 1510.

Reilly, S., Schallier, W., Schrimpf, S., Smit, E., & Wilkinson, M. (2011). Report on integration of data and publications (Tech. Rep.). Alfred-Wegener-Institut, http://epic.awi.de/31397/1/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. Journal of the Association for Information Science and Technology, 67(12), 2964-2975.

Shaklee, P. M. (2014). Data in brief–Making your data count. Data in Brief, 1, 5-6.

Silvello, G. (2018). Theory and practice of data citation. Journal of the Association for Information Science and Technology, 69(1), 6-20.

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.

Thelwall, M., & Kousha, K. (2016). Figshare: a universal repository for academic resource sharing? Online Information Review, 40(3), 333-346.

Thelwall, M., & Wilson, P. (2016). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. Journal of the Association for Information Science and Technology, 67(8), 1962-1972.

Thelwall, M. (2017a). Are Mendeley reader counts useful impact indicators in all fields? Scientometrics, 113(3), 1721–1731.

Thelwall, M. (2017b). Are Mendeley reader counts high enough for research evaluations when articles are published? Aslib Journal of Information Management, 69(2), 174-183.

Thelwall, M. (2018a). Early Mendeley readers correlate with later citation counts. Scientometrics, 115(3), 1231–1240.

Thelwall, M. (2018b). Dimensions: A competitor to Scopus and the Web of Science? Journal of Informetrics, 12(2), 430–435.

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PloS One, 8(7), e67332.

White, H., Carrier, S., Thompson, A., Greenberg, J., & Scherle, R. (2008). The Dryad Data Repository: A Singapore framework metadata architecture in a DSpace environment. In Dublin Core Conference (pp. 157-162). https://dcpapers.dublincore.org/pubs/article/view/928

Zahedi, Z., Haustein, S. & Bowman, T. (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. Presentation at SIGMET Metrics 2014 workshop. https://crc.ebsi.umontreal.ca/en/publications/exploring-data-quality-and-retrieval-strategies-for-mendeley-reader-counts/