Online Information Review

Data bibliometrics: metrics before norms

David Stuart,

Article information:

To cite this document:
David Stuart, (2017) "Data bibliometrics: metrics before norms", Online Information Review, Vol. 41 Issue: 3, pp.428-435, https://doi.org/10.1108/OIR-01-2017-0008
Permanent link to this document:
https://doi.org/10.1108/OIR-01-2017-0008

Data bibliometrics: metrics before norms

David Stuart
Statistical Cybermetrics Research Group, University of Wolverhampton, UK

Received 9 January 2017
Revised 9 January 2017
Accepted 20 January 2017

Abstract

Purpose – The purpose of this paper is to highlight the problem of establishing metrics for the impact of research data when norms of behaviour have not yet become established.

Design/methodology/approach – The paper considers existing research into data citation and explores the citation of data journals.

Findings – The paper finds that the diversity of data and its citation precludes the drawing of any simple conclusions about how to measure the impact of data, and an over emphasis on metrics before norms of behaviour have become established may adversely affect the data ecosystem.

Originality/value – The paper considers multiple different types of data citation, including for the first time the citation of data journals.

Keywords Scientometrics, Citation analysis, Open data, Data journals

Paper type Research paper

Introduction

As the scholarly publishing ecosystem changes so quickly, the gaps in our knowledge seem to grow faster than our understanding, and nowhere can this be more clearly seen than in the publishing and citing of research data. In recent years, research data have come to be seen as a public good and citations are seen as having the potential to both incentivize and reward the publishing of data. However, with limited understanding of the norms of data publishing and data citation, too early an emphasis on metrics may do damage to the nascent data ecosystem.

This viewpoint considers the heterogeneous nature of research data, the limited extent of current research into data publication and citation, and the risks of an over emphasis on metrics before norms are established. Particular attention is paid to the potential of data journals, one of the more recent and less studied areas of data publishing. Data journals are promoted as providing a recognisable and citable publication, "a bridge between traditional journal publication and data set submission" (Force *et al.*, 2016, p. 27), but as an initial analysis of data journals shows here: the huge diversity in the extent to which data journals are cited precludes the drawing of any simple conclusions.

Background

Evaluative citation analysis is based on relatively simple assumptions about the publishing and citation process: researchers publish research which builds upon the work of others, the intellectual debt is paid in the form of citations, and if the citations are aggregated they can provide useful indicators of the influence of a work. For these indicators to be meaningful, however, it is required that units of analysis are sufficiently similar to be aggregated, i.e., there are norms of citing behaviour, and sufficiently similar genres of document. The problem of aggregating units that are not sufficiently similar may be seen in criticisms of both journal-based metrics and citation analysis more generally: articles in the same journal are not necessarily of similar quality, research is cited for different reasons, and there are disciplinary differences in citation and publishing practices. Issues surrounding dissimilarity of units of analysis may be expected to be writ large when it comes to citation of research data.

Online Information Review
Vol. 41 No. 3, 2017
pp. 428-435

1468-4527
DOI 10.1108/OIR-01-2017-0008

The open data rationale

Borgman's (2012) suggested rationales for sharing research data include the potential for the verification of results, public access to publicly funded research, new questions being asked of existing data, and the advancement of research and innovation. However, such rationales give precedence to the advancement of science over the individual scientist, and the "dirty little secret" of open data promotion has been a lack of research data sharing (Borgman, 2012, p. 1059). There is a mismatch between what people believe should be done and what is actually being done (Tenopir *et al.*, 2015); a disparity between the interests of the academic community and individual academic's behaviour. Self-interested rationales are therefore more important for changing researcher behaviour; more self-interested rationales include permissiveness (Spires-Jones *et al.*, 2016), quality improvement (Fecher *et al.*, 2015), and recognition (Fecher *et al.*, 2015).

Researchers who will make their data publicly available are permitted to get funding from certain sources or publish in particular journals, however, open data requirements are not as strict as open access requirements (Spires-Jones *et al.*, 2016), and are not consistently enforced (Borgman, 2012). Publishing data provide the opportunity for work to be improved and new ideas to emerge from others exploring a researcher's data, although this comes with the risk of data misuse and misinterpretation, and the time necessary to support others reuse of the data (Fecher *et al.*, 2015). As neither permissiveness nor quality improvement are unadulterated drivers of open data there is a lot of interest in the role of recognition.

Recognition of the contribution a data set has made to later research can come in many forms, ranging from acknowledgement, to citation, and co-authorship (Fecher *et al.*, 2015). The interest in data citations is understandable, and formal citation or acknowledgement is considered "essential" (Tenopir *et al.*, 2015) for data reuse. Citations are the most common currency of traditional scholarly publishing (Cronin, 1984) and are increasingly used as an indicator of impact (Hicks *et al.*, 2015). But data citation is not directly analogous to paper citations, and we should be wary about overburdening data citations before the norms of behaviour have emerged.

Data are heterogeneous

Scholarly data differ far more as a unit of analysis than the research article; there is not just one type of data, but a great diversity between data (Borgman, 2015). Data can be as big at the petabytes spewing forth from the Large Hadron Collider or as small as an exploratory social science survey; it can be as public as the amount of traffic on the street or as private as health records in a medical trial; it can be as time sensitive as the sequencing of a virus during an pandemic or as slow moving as some areas of astrophysics; it can be as reproducible as the findings from a chemical experiment or as irreproducible as historic carbon dioxide levels in the air; it can be published according to open standards on the web or be hidden within proprietary software on a personal hard drive.

All such factors may be expected to influence the impact of data, but from the perspective of how we operationalise a bibliometric study, where the data are published and the nature of the recognition are of paramount importance. Data may be published in public repositories, on public websites, or as supplementary materials (Borrego and Garcia, 2013), it may be associated with a traditional research paper, a data paper, or without an accompanying publication. Data recognition include intratextual citations of data sets at different levels of granularity, acknowledgements, co-authorship, and formal citation of associated data and research papers.

Evaluative data citation studies

The great diversity in "data" and its potential citation has not yet been met with similar quantities of research. The bibliometric studies that give us some insights into the impact of scholarly data on the web can be broadly split into two: those that consider the impact of open data on the citation of the traditional research paper, and those that have considered the citing of data directly.

Open data citation advantage. A number of studies have found signs of a citation advantage for research articles from sharing data. Piwowar *et al.* (2007) found cancer microarray clinical trial publications with publicly available data had significantly more citations. It was found again in a later study of microarray data, albeit with a smaller advantage than previously estimated (Piwowar and Vision, 2013). Leitner *et al.* (2016) identified citation advantages in neuroscience and molecular biology for those papers indexed in the Web of Science with data-related MeSH terms (e.g. atlas or database). Gordon *et al.* (2016) showed that articles in the *Journal of Molecular Spectrometry* with supplementary material were cited more often, and Drachen *et al.* (2016) found an increase in the number of citations for papers linking to data in three astrophysics journals.

The association between data sharing and increased citations does not imply causation (Piwowar *et al.*, 2007), and it may be that additional factors account for the data citation advantage. For example, high quality papers may be both more likely to publicly share data and receive more citations. However, Piwowar and Vision's (2013) study included a classification of the reasons for citation, and found that at least some of the citations, albeit only 6 per cent, reflected the reuse of data. The impact of data may vary over time, however, with changes to data and tool availability, and indeed Piwowar and Vision (2013) found whereas nearly all the microarray data sets published in 2001 were reused at least once, this fell to about 20 per cent in later years.

Data set citation. Data sets can also be cited directly, although citation norms are less established for the citing of data, and the lack of proper recognition for the contribution of data is seen as a significant barrier to the publishing and sharing of data (Tenopir *et al.*, 2011; De Castro *et al.*, 2013; Helbig *et al.*, 2015). There are also practical problems with the identification of the citations that are made, especially as many fields continue to favour intratextual data citations (Mayo *et al.*, 2016).

Nonetheless, some studies have explored the direct citation of data sets using the Data Citation Index (DCI), an index of data repositories, their data sets, and citations. Robinson-García *et al.* (2016) found that the vast majority of data sets and data studies (88.1 per cent of all DCI records) were not cited, with 43 repositories receiving no citations at all. Peters *et al.* (2016) combined the DCI with altmetrics to also find that most of the data sets (over 85 per cent) are not cited, and those data sets that are cited are not particularly visible in altmetrics.

Data journals. Data journals are scholarly publications that promote data papers, descriptions of data sets that have been made available online (Candela *et al.*, 2015). They provide a solution to some of the limitations of data publishing and citation: they can make data easier to find, cite, and reuse, provide a level of credibility, and show the value of data separate to the value of the research article. For Kratz and Strasser (2015, p. 3) there is a clear-cut appeal to data papers: "they are unquestionably peer-reviewed papers, so academia knows how (if perhaps not how much) to value them".

As might be expected, in their data citation study Robinson-García *et al.* (2016) found that data studies, data sets with a description of the experiment, received more citations on average than data sets.

The citation of data journals

There has been a relatively short window of opportunity to explore the impact of data journals, with many of the principle journals only emerging in the last few years. However, as they are positioned to play a pivotal role in the data ecosystem, it is necessary to have at least some crude understanding of the role they have. As such this viewpoint is supplemented with a simple analysis, exploring whether the data journals are being cited.

Methodology

A list of data journals was compiled for analysis through the Web of Science. As is appropriate for an exploratory paper on open data, this was based on previously published data sets. Candela *et al.*'s (2015) list of 116 data journals was supplemented by five journals identified in Kratz and Strasser's (2015) survey of recognised data journals, and seven additional journals identified by a web-based inventory study.

There is a great diversity in the journals that may be considered "data journals". All data journals must necessarily be willing to publish data papers, but the proportion of a journal that are made up by data papers can differ significantly, from those that are still primarily focussed on research articles but will publish data papers to those that are primarily focussed on data papers with the occasional editorial. Data journals also differ significantly in terms of longevity, as well as a number of new data journals (some of which have already been and gone), there are also data journals that predate the web (e.g. *Journal of Physical and Chemical Reference Data*). Each of the 128 journals was checked to exclude those that had ceased publication, those that were not primarily data journals, those that preceded the recent emphasis on open data on the web, and those that were not indexed by the Web of Science.

The remaining eight journals were then analysed more closely to determine the proportion of papers that were cited. More specifically the proportion of papers published in 2014, both including and excluding self-citations (where the citing paper shares at least one author with the cited data paper). Papers published in 2014 have at least two years for citations to be made and indexed (as the data were gathered in the last week of December 2016), and seven of the eight journals were indexed in the Web of Science by this point.

Findings

Table I shows the proportion of papers published in 2014 that had been cited by the end of 2016. There can be seen to be great variation between the journals, from those where a very small proportion of the journals' papers have been cited (e.g. *Biodiversity Data Journal* and *Scientific Data*) to those where the vast majority of the papers have been cited (e.g. *Earth System Science Data* and *Geoscience Data Journal*). The small proportion of data papers that were cited in *Scientific Data* is particularly interesting considering that it was the most recognised data journal in Kratz and Strasser's (2015) survey.

Self-citations are always an important issue in bibliometrics, but especially with regards to data papers as data papers may be expected to have fewer citations than research papers and authors of data papers can be expected to make use of their own data. Of the 134 papers

Journal	First year indexed by WoS	Number of papers published in 2014	Number of papers cited incl. self-citations	Number of papers cited excl. self-citations
Biodiversity Data Journal	2013	93	3 (3%)	3 (3%)
Data in Brief	2014	19	5 (26%)	4 (21%)
Earth System Science Data	2012	27	25 (93%)	20 (74%)
Genomics Data	2013	106	48 (45%)	34 (32%)
Geoscience Data Journal	2014	17	15 (88%)	15 (88%)
GigaScience	2012	37	33 (89%)	31 (84%)
Journal of Open Archaeology Data	2015	na	na	na
Scientific Data	2014	53	5 (9%)	4 (8%)
Total		352	134 (38%)	111 (32%)

Table I. Proportion of data papers published in 2014 cited by the end of 2016 that were cited at least once, 111 (83 per cent) were cited at least once after shared authorship was discounted. Self-citation is a particularly complex issue in data citation as recognition may be in the form of authorship.

At 38 per cent, the proportion of papers that were cited was higher than that found by Robinson-García *et al.* (2016) (88.1 per cent uncited), even for data studies (81.74 per cent uncited).This may be partly accounted for by the inclusion of other types of papers by using the journal as the starting point, although as the half-life of the citation of data journals may be expected to be longer than for that of research journals, it nonetheless suggests that data journals are a product that are making a real contribution to the scholarly publishing ecosystem. The difference with which the papers in the different journals are cited, however, emphasises the problem of drawing too many broad conclusions about data citation too quickly.

Challenges of bibliometric indicators for data

The lack of citations

The uncomfortable truth that accompanies Borgman's (2012) dirty little secret is that much of the data that are published are not cited anyway (Torres-Salinas *et al.*, 2014; Robinson-García *et al.*, 2016; Peters *et al.*, 2016). Although data papers and data studies have been found to have a greater proportion of papers cited, the diversity in the extent data journals are cited means we should beware of overly simplistic pronouncements. Data in data journals get cited more often is appealing, but we may find that we drive forward the idea that the value of open data is in its citation as well as building false expectations about data journals.

As Candela *et al.* (2015) put it: "data journals are not the ultimate and complete solution for all data-sharing and reuse issues". The diversity of data means no single catch-all publishing solution can exist, and early unqualified pronouncements could easily lead researchers to sub-optimal publishing solutions. Whilst the quality of data documentation has been found to correlate positively with satisfaction with data reuse (Faniel *et al.*, 2016) and peer-reviewed data papers are deemed the most valuable type of data publication (Kratz and Strasser, 2015, p. 13), it is nonetheless difficult to generalise such positions to all data. In some instances such publication may be superfluous, duplicating information that may be unnecessary due to widely applied methodologies, and standardized data structures.

It is also important to recognise, as Piwowar and Vision (2013, p. 20) put it "Many important instances of data reuse leave no trace in the published literature". Data may be used for teaching students or checking research findings. If we emphasise the value of citations alone, a lack of citations may adversely affect the perceived value of researchers publishing data: it's not worth anything if it's not cited.

Data citations will also differ considerably in terms of the valued contribution they represent. Whereas this has always been the case with citations, it may be deemed even more extreme with data citations: one research paper may cite another because they disagree with the paper, or because it is a seminal paper in the field; a data paper may be cited because it queries the data collection methodology, or because a citing research paper is based solely on reusing the cited paper's associated data set.

The current interest in metrics for evaluating scientific outputs means we risk putting the cart before the horse. Products such as the DCI and the inclusion of data journals in the Web of Science will undoubtedly find a willing audience, even if it is hard to imagine some of the journals being so quickly indexed if it was not for their novelty.

Alternative metrics

Although citations are currently viewed as the most useful potential measure of impact (Kratz and Strasser, 2015), the possibility of alternative metrics – or altmetrics – has been widely acknowledged (e.g. Piwowar and Vision, 2013; Peters *et al.*, 2016) even though initial findings have not been positive: even those data sets that are cited are not particularly visible in altmetrics (Peters *et al.*, 2016), whilst data sets were the least viewed and the second least shared type of resource on Figshare (Thelwall and Kousha, 2016).

When considering the diversity of data that can be shared and the work that needs to be done before any data metric provides reliable insights that are accepted, there is nonetheless some surprising optimism about their potential: "[…] it may well be that in a few years' time young neuroscientists are being awarded grants and positions based at least partly on their articles' or data sets' download statistics, or something more imaginative and informative" (Spires-Jones *et al.*, 2016). Whilst Kratz and Strasser (2015, p. 16) note that "only one third of respondents found them [altmetrics] even somewhat useful in assessing impact", in comparison to the time it has taken for citations to be considered valuable for research assessment, and the huge gaps in the potential of altmetrics for research papers, that a third should consider them even somewhat useful in assessing impact could be considered quite an optimistic position.

Such optimism may be misplaced however. It may be that there is no particular metric suitable for understanding the importance or impact of a data set, alternative or otherwise. Whereas a seminal paper may be cited numerous times, it is quite plausible that a data set is reused for one seminal study, but is then never referenced again. Rather than developing metrics of reuse, it may be that the focus should be on the potential for reuse: the ability of a data set to form the basis of further studies, repeatability, metadata, and documentation.

Cultural change

It is important to remember that evaluative citation analysis is not the primary purpose for sharing research data (Piwowar and Vision, 2013), but rather it may be seen as one possible tool for encouraging a data sharing culture, the lack of which has been described as the main obstacle to academic data sharing (Fecher *et al.*, 2015).

Research cultures do not change overnight, and as Tenopir *et al.* (2011) point out, building the infrastructure for data sharing can be easier than changing a culture. Some have expressed concerns about the system being taken over by "research parasites" (Longo and Drazen, 2016) or "data vultures" (Kratz and Strasser, 2015), but the prevalence and expected working practices depend heavily on the field of research, and in areas such as the social sciences data reuse is a common phenomenon with no associated stigma (Faniel *et al.*, 2016, p. 1413).

One of the principal reasons given for not sharing data is the time it takes to prepare and document the data (Kratz and Strasser; 2015; Tenopir *et al.*, 2011), as well as supporting others trying to reuse the data (Fecher *et al.*, 2015). Although the proportion of researchers who identified insufficient time as a barrier to data sharing had fallen in Tenopir *et al.*'s (2015) follow up survey, it was nonetheless identified by 38.6 per cent of respondents, second only to the need for researchers to publish their research first, which unlike a lack of time should not be considered a permanent barrier. When asking early career neuroscientists how long they thought it would take to make their data available for sharing the mean answer was nine days (Spires-Jones *et al.*, 2016); it seems unlikely that one or two citations will do much to persuade someone to spend nine days sharing their data.

Conclusions

At the early stages of an emerging data infrastructure it is important that the data ecosystem is not adversely affected by overburdening the metrics that are available. Whilst there is a place for bibliometric analysis of research data sets, any such use in the assigning of grants and positions should be a long way off. The sort of conservatism that has met traditional citation analysis should be welcome in data metrics until far more research is done.

Even then we should not be surprised to discover that data metrics offer few insights into the impact of data, or are insufficient to entice researchers to publish more data.

The need for more research into the citation of research data is widely recognised (e.g. Robinson-García *et al.*, 2016; Peters *et al.*, 2016), but bibliometrics suffers from a wealth of potential resources for analysis. The regular emergence of sites and services for public conversation inevitably draws attention from the more staid areas.

For now the promotion of open data may be best served by focusing on the other self-interested rationales: more clearly demonstrating the improvements in quality that can be made by researchers sharing data, as well as implementing and more strictly enforcing data sharing policies. Bibliometric investigations of data sets should be limited to gaining insights into the ecosystem rather than moving too quickly to evaluative applications in the real world.

References

Borgman, C.L. (2012), "The conundrum of sharing research data", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 6, pp. 1059-1078.

Borgman, C.L. (2015), *Big Data, Little Data, No Data: Scholarship in the Networked World*, MIT Press, Cambridge, MA.

Borrego, Á. and Garcia, F. (2013), "Provision of supplementary materials in library and information science scholarly journals", *ASLIB Proceedings*, Vol. 65 No. 5, pp. 503-514.

Candela, L., Castelli, D., Manghi, P. and Tanim, A. (2015), "Data journals: a survey", *Journal of the American Society for Information Science and Technology*, Vol. 66 No. 9, pp. 1747-1762.

Cronin, B. (1984), *The Citation Process*, Taylor Graham, Oxford.

De Castro, P., Calzolari, A., Napolitani, F., Rossi, A.M., Mabile, Cambon-Thomsen, A. and Bravo, E. (2013), "Open data sharing in the context of bioresources", *Acta Informatica Medica*, Vol. 21 No. 4, pp. 291-292.

Drachen, T.M., Ellegaard, O., Larsen, A.V. and Dorch, S.B.F. (2016), "Sharing data increases citations", *LIBER Quarterly*, Vol. 26 No. 2, pp. 67-82.

Faniel, I.M., Kriesberg, A. and Yakel, E. (2016), "Social scientists' satisfaction with data reuse", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 6, pp. 1404-1416.

Fecher, B., Friesike, S. and Hebing, M. (2015), "What drives academic data sharing?", *PLoS ONE*, Vol. 10 No. 2, p. e0118053, available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053 (accessed 9 January 2017).

Force, M., Robinson, N., Matthews, M., Auld, D. and Boletta, M. (2016), "Research data in journals and repositories in the web of science: developments and recommendations", *Bulletin of IEEE Technical Committee on Digital Libraries*, Vol. 12 No. 1, pp. 27-30, available at: www.ieee-tcdl.org/Bulletin/v12n1/papers/IEEE-TCDL-DC-2016_paper_3.pdf (accessed 9 January 2017).

Gordon, I.E., Potterbusch, M.R., Bouquin, D., Erdmann, C.C., Wilzewski, J.S. and Rothman, L.S. (2016), "Are your spectroscopic data being used?", *Journal of Molecular Spectroscopy*, Vol. 327, pp. 232-238.

Helbig, K., Hausstein, B. and Toepfer, R. (2015), "Supporting data citation: experiences and best practices of a DOI allocation agency for social sciences", *Journal of Librarianship and Scholarly Communication*, Vol. 3 No. 2, available at: http://doi.org/10.7710/2162-3309.1220 (accessed 9 January 2017).

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. (2015), "Bibliometrics: the Leiden Manifesto for research metrics", *Nature*, Vol. 520 No. 7548, pp. 429-431.

Kratz, J.E. and Strasser, C. (2015), "Researcher perspectives on publication and peer review of data", *PLoS ONE*, Vol. 10 No. 2, p. e0117619, available at: http://doi.org/10.1371/journal.pone.0117619 (accessed 9 January 2017).

Leitner, F., Bielza, C., Hill, S.L. and Larrañaga, P. (2016), "Data publications correlate with citation impact", *Frontiers in Neuroscience*, Vol. 10, September, Article 419, available at: http://journal.frontiersin.org/article/10.3389/fnins.2016.00419/full (accessed 9 January 2017).

Longo, D.L. and Drazen, J.M. (2016), "Data sharing", *New England Journal of Medicine*, Vol. 374 No. 3, pp. 276-277.

Mayo, C., Vision, T.J. and Hull, E.A. (2016), "The location of the citation: changing practices in how publications cite original data in the Dryad Digital Repository", paper presented at the 11th International Digital Curation Conference, Amsterdam, 22-25 February, available at: http://doi.org/10.5281/zenodo.32412 (accessed 9 January 2017).

Peters, I., Kraker, P., Lex, E., Gumpenberger, C. and Gorraiz, J. (2016), "Research data explored: an extended analysis of citations and altmetrics", *Scientometrics*, Vol. 107 No. 2, pp. 723-744.

Piwowar, H.A. and Vision, T.J. (2013), "Data reuse and the open data citation advantage", *PeerJ*, Vol. 1, p. e175, available at: https://doi.org/10.7717/peerj.175 (accessed 9 January 2017).

Piwowar, H.A., Day, R.S. and Fridsma, D.B. (2007), "Sharing detailed research data is associated with increased citation rate", *PLoS ONE*, Vol. 2 No. 3, p. e308, available at: http://dx.doi.org/10.1371/journal.pone.0000308 (accessed 9 January 2017).

Robinson-García, N., Jiménez-Contreras, E. and Torres-Salinas, D. (2016), "Analyzing data citation practices using the data citation index", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 12, pp. 2964-2975.

Spires-Jones, T.L., Poirazi, P. and Grubb, M.S. (2016), "Opening up: open access publishing, data sharing, and how they can influence your neuroscience career", *European Journal of Neuroscience*, Vol. 43 No. 11, pp. 1413-1419.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. and Frame, M. (2011), "Data sharing by scientists: practices and perceptions", *PLoS ONE*, Vol. 6 No. 6, p. e21101, available at: http://dx.doi.org/10.1371/journal.pone.0021101 (accessed 9 January 2017).

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D. and Dorsett, K. (2015), "Changes in data sharing and data reuse practices and perceptions among scientists worldwide", *PLoS ONE*, Vol. 10 No. 8, p. e0134826, available at: http://dx.doi.org/10.1371/journal.pone.0134826 (accessed 9 January 2017).

Thelwall, M. and Kousha, K. (2016), "Figshare: a universal repository for academic resource sharing?", *Online Information Review*, Vol. 40 No. 3, pp. 333-346.

Torres-Salinas, D., Jiménez-Contreras, E. and Robinson-García, N. (2014), "How many citations are there in the data citation index?", presented at the STI Conference, Leiden, 3-5 September, available at: https://arxiv.org/ftp/arxiv/papers/1409/1409.0753.pdf (accessed 9 January 2017).

Corresponding author
David Stuart can be contacted at: dp_stuart@hotmail.com

This article has been cited by:

1. Gianmaria Silvello. 2018. Theory and practice of data citation. *Journal of the Association for Information Science and Technology* **69**:1, 6-20. [Crossref]