

Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets

Christopher W. Belter*

LAC Group, Central Library, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, United States of America

Abstract

Evaluation of scientific research is becoming increasingly reliant on publication-based bibliometric indicators, which may result in the devaluation of other scientific activities – such as data curation – that do not necessarily result in the production of scientific publications. This issue may undermine the movement to openly share and cite data sets in scientific publications because researchers are unlikely to devote the effort necessary to curate their research data if they are unlikely to receive credit for doing so. This analysis attempts to demonstrate the bibliometric impact of properly curated and openly accessible data sets by attempting to generate citation counts for three data sets archived at the National Oceanographic Data Center. My findings suggest that all three data sets are highly cited, with estimated citation counts in most cases higher than 99% of all the journal articles published in Oceanography during the same years. I also find that methods of citing and referring to these data sets in scientific publications are highly inconsistent, despite the fact that a formal citation format is suggested for each data set. These findings have important implications for developing a data citation format, encouraging researchers to properly curate their research data, and evaluating the bibliometric impact of individuals and institutions.

Citation: Belter CW (2014) Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. PLoS ONE 9(3): e92590. doi:10.1371/journal.pone.0092590

Editor: Howard I. Browman, Institute of Marine Research, Norway

Received: July 24, 2013; **Accepted:** February 25, 2014; **Published:** March 26, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: No funding beyond employment by NODC was received for this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: Chris.Belter@noaa.gov

Introduction

In recent years there has been increasing interest in, and use of, bibliometric indicators for the evaluation and ranking of research institutions. Bibliometric indicators feature prominently in global mixed-method ranking schemes such as the Academic Ranking of World Universities [1] and the Times Higher Education ranking [2]. They also feature in national mixed-method research assessment exercises in the UK, Brussels, Italy, and Australia. Other global ranking schemes are based solely on bibliometric indicators [3,4]. Bibliometric indicators are often recommended to supplement [5–9] or even replace [10,11] peer review in evaluating research institutions.

Partially in response to the growing importance of bibliometrics in research evaluation, and partially in response to other factors, there is also a growing movement focused on the development of a standard method of citing data sets in academic publications [12]. Reasons for developing a citation format for data sets include verification of published results, reuse of data sets for additional research purposes, and attribution to data collectors and archivists. Such suggestions have been made in bioinformatics [13–16], genetics [17], climate sciences [18], geochemistry [19–21], oceanography [22,23], earth sciences [24,25], and multidisciplinary sciences [26–28], among others.

Although there is widespread agreement within the movement that a minimum set of information is necessary to a complete data set citation, there seems to be two schools of thought as to how this ought to be accomplished. One school favors a direct citation to the data set as it resides in an established repository. This model

was first adopted for nucleotide sequence datasets in the formation of GenBank [29] and adapted for the marine [30] and earth [31,32] sciences before being more widely recommended [33–35] and implemented in various subject-specific and general data repositories such as the California Digital Library (<http://www.cdlib.org/>), DataONE (<http://www.dataone.org/>), the Dataverse Network (<http://thedata.org/>), Dryad (<http://datadryad.org/>), ICPSR (<http://www.icpsr.umich.edu/>), Pangaea (<http://www.pangaea.de/>), NOAA's climatic (<http://www.ncdc.noaa.gov/>), geophysical (<http://www.ngdc.noaa.gov/>), and oceanographic (<http://www.nodc.noaa.gov/>) data centers, etc. One of the fundamental components of this model is the creation and citation of an identifier that uniquely identifies the data set being cited. This identifier typically takes the form of a doi assigned through DataCite [36], although other identifiers may also be used.

This model is beginning to be incorporated into the products of commercial scientific information providers. In 2012, Thomson Reuters launched the Data Citation Index (http://wokinfo.com/products_tools/multidisciplinary/dci/), a database of data sets that provides suggested citation formats for each data set indexed in the database and attempts to generate citation linkages to articles indexed in its other Web of Science databases. More recently, Elsevier, in cooperation with DataCite and numerous data repositories, launched a similar project that attempts to link papers available in ScienceDirect to the data sets that they use or have deposited in repositories (<http://www.elsevier.com/about/content-innovation/database-linking>) through data set dois or other unique identifiers.

The other school of thought favors the citation of a ‘data paper’ or ‘data publication’ describing the data set. In this model, the metadata necessary for using a data set and a link to the data set is presented in a paper published either in a traditional journal or in a specialized data journal. Data papers differ from more traditional publications in that no analyses or findings resulting from the data set are required. Researchers wishing to cite the data set would then cite the data paper, rather than the data set. This model has been suggested in the neurosciences [37,38], genetic sciences [39], and bioinformatics [40] communities, and implemented in the geosciences community through the formation of data journals such as *Earth System Science Data* (<http://www.earth-system-science-data.net>) and *Geoscience Data Journal* (<http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%292049-6060>) and the publication of data papers in journals such as the *Quarterly Journal of the Royal Meteorological Society*, *Eos*, and *Oceanography*. Examples of recent data papers in the earth sciences include those describing the ERA-40 reanalysis in atmospheric sciences [41], the Argo profiling floats [42], and a database of iron enrichment experiment results [43] in oceanography.

Closely tied to the development of data citation standards is the growing awareness of the need to properly preserve, describe, and provide access to data sets, a collection of activities sometimes referred to as data curation. In order for a data set to be cited, it must first have been deposited in a repository, preserved in an interoperable format, adequately described by a formal set of metadata attached to the data set, and made available to other researchers for reuse. Although technical issues exist at each step in this process, the idea of sharing data sets with other researchers has proven to be the most controversial [44–47]. In addition to concerns over the idea of freely sharing research data, many researchers are reluctant to devote the time necessary to properly curate their research data, especially since many have not received training on how to do so. Although mandates for data preservation and sharing have been established by the National Science Foundation (<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>), the American Geophysical Union (<http://publications.agu.org/author-resource-center/publication-policies/data-policy/>), and the US Office of Science and Technology Policy (<http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>) among others, it is not yet clear whether these mandates will motivate researchers to do so in the future.

Although bibliometric indicators can be a useful compliment to peer review in the evaluation of scientific research [7–9], the growing reliance on publication-based indicators for research evaluation could potentially lead to the devaluation of activities that do not typically result in the publication of articles in scientific journals. Participation in workshops, policy formulation, peer-review of submitted manuscripts, public education, and mentoring are all critical to the advancement of scientific research and to the translation of that research into societal benefits, but few of these activities ‘count’ in bibliometric evaluations because they rarely result in formal publications. Since the incorporation of bibliometric indicators into research evaluation is known to affect the subsequent behavior of those being evaluated [48–51], it seems likely that the growing reliance on bibliometric indicators could result in a disincentive to engage in such activities.

One such activity that is likely to be devalued in this context is data curation. Despite its importance to the scientific community, data curation rarely results in the production of scientific journal articles, meaning that scientists and institutions devoting time and effort to data curation are unlikely to be rated favorably by bibliometric indicators in comparison with their more prolific peers. This is likely to undermine support for data curation efforts,

since scientists are unlikely to devote the time and effort required to properly curate their data sets if they are unlikely to be rewarded for doing so.

The purpose of this analysis is to combine these trends by attempting to show the value of data curation in bibliometric terms. Specifically, I attempt to generate citation counts for three oceanographic data sets curated by the National Oceanic and Atmospheric Administration (NOAA)’s National Oceanographic Data Center (NODC). In doing so, I hope to demonstrate the utility of data curation to scientific research, since these data sets could not have been cited without the curation activities performed by NODC and its partners. In the process, I also hope to inform the discussion surrounding the development of data citation standards by identifying how these data sets are currently cited and referenced in scientific articles. Such baseline information can be useful in identifying both the metadata that should be included in a data citation format and how such a format ought to be applied.

Although many articles advocating data citation standards mention the usefulness of such standards for bibliometric evaluation, efforts to actually generate citation counts for data sets are fairly rare. Chao [52] measured dataset reuse in the earth sciences and found that earth science data sets were primarily cited in physical science and multidisciplinary journals, suggesting that data sets generated in one discipline may also have applications in other disciplines. Parsons et al. [24] used Google Scholar to search for mentions of snow cover data sets archived at the National Snow and Ice Data Center and found between 100 and 600 mentions per year. Piwowar and colleagues [53,54] used a similar method, searching PubMed Central for mentions of data sets archived in the Gene Expression Omnibus (GEO) database and estimated that GEO data sets had been cited over 1,150 times by the end of 2010. The Inter-university Consortium for Political and Social Research (ICPSR) maintains a bibliography of several thousand publications that cite one or more data sets archived by ICPSR [55]. Finally, several studies suggest that articles with publically available data sets are more highly cited than articles that do not make their data publically available [e.g. 56,57].

Methods

In consultation with NODC, I selected three highly-used data sets for this analysis: the World Ocean Atlas and World Ocean Database (WOA/WOD), the Pathfinder Sea Surface Temperature (PSST) data set, and the Group for High Resolution Sea Surface Temperature (GHRSSST) data set. The World Ocean Atlas is a quality-controlled set of objectively analyzed global *in situ* observational data published in four volumes focused on the variables of temperature [58], salinity [59], oxygen [60], and nutrients [61]. Although NODC considers the World Ocean Atlas a data product, rather than a raw data set, because it is a compilation of many individual data sets gathered at various times and locations around the world and because of the quality control and analysis done on the underlying data, I consider it a data set for the purposes of this analysis. The World Ocean Database [62] is an interactive database of the data used to create the World Ocean Atlas. Since the Atlas and the Database utilize same underlying data, I will refer to them in combination as the WOA/WOD. The WOA/WOD was initially published in 1982 as the ‘Climatological Atlas of the World Ocean’ [63] and rereleased with updated data in 1994, 1998, 2002, 2006, and 2009–2010. The PSST data set [64] is a long-term set of global sea surface temperature data derived from the Advanced Very High Resolution Radiometer (AVHRR) sensor mounted on NOAA’s

polar-orbiting satellites. The GHRSSST data set [65,66] is a global set of combined satellite and in situ sea surface temperature data contributed by a number of institutions from around the world. GHRSSST data are initially collected from these institutions by the NASA Jet Propulsion Laboratory and then transferred 30 days after observation to NODC for long term preservation and access.

In the first phase of this analysis, conducted in March 2013, I attempt to generate citation counts for these data sets using three data sources: Web of Science, Science Citation Index Expanded (WoS), the full text search capabilities provided by various journal publishers' websites (Elsevier, Springer, Wiley, etc), and Google Scholar. I search these data sources to find citations to, or mentions of, these data sets in scientific publications and compile the number of results retrieved. In this analysis, I count mentions of these data sets as citations, a broader definition of 'citation' than is currently used, because scientific articles utilizing or discussing these data sets may or may not formally cite them. This definition is consistent with that employed by previous studies [24,53,54]. The search strings used in this phase are listed in Table 1. These search strings are deliberately restrictive to improve the precision of the retrieved results. As a result, although the resulting counts are likely to be fairly, although not entirely, accurate, they are also likely to be undercounts of the actual number of publications citing or mentioning these data sets.

To generate citation counts for each data set in Web of Science, I use the search strings to search the title, abstract, keywords, and funding text (or acknowledgements) fields and add all of the resulting records to my 'Marked List.' I then perform cited reference searches to identify citations to publications or reports associated with each data set and add all of the resulting records to my 'Marked List'. The final number of records in my 'Marked List' is then noted as the WoS citation count for each data set. Using the 'Marked List' in this way allows me to avoid potentially double counting records retrieved by multiple searches.

To generate citation counts for each data set using publishers' websites, I use the search strings to search the full text of all records on each site using the sites' internal search engine and note the number of results retrieved for each data set. I then combine these totals across websites to generate a final citation count for each data set. The publishers' sites searched were: ScienceDirect, SpringerLink, the Wiley Online Library, the American Meteorological Society's Online Journals page, the Nature Publishing Group website, the Science (AAAS) website, the PNAS website, Taylor and Francis Online, IEEE Xplore, the Public Library of Science website, the American Chemical Society website, and the Ecological Society of America's online journals site. Searching each of these sites individually was necessary because no formal full text database covering the oceanographic, marine, and geosciences is available.

To generate citation counts for each data set using Google Scholar, I use the search strings to search the database without restriction and then note the number of results retrieved as the

final citation count. Due to known indexing and metadata issues with Google Scholar [67,68], these counts are likely to be inflated, and include non-peer-reviewed publication types, but are also likely to provide reasonable estimates of how often these sets are used overall and to provide accurate rankings of these data sets relative to each other [69,70].

In the second phase of this analysis, conducted in January 2014, I attempt a more comprehensive cited reference search in WoS to generate citation counts to all editions of the WOA/WOD over time. Since WOA/WOD originally was, and still is, distributed as a print publication, it seems likely that formal citations to this data set would be more numerous than for the other data sets analyzed here. To allow for wide variance in citation formats, I search for the author(s) and publication year(s) of each edition of the WOA/WOD and then manually select the relevant search results. The search strategies used in this process are summarized in Table 2.

The process of executing these search strategies for one edition of the WOA/WOD is as follows. First, I perform a Cited Reference Search using the search criteria listed in Table 2. In step 2 of the Cited Reference Search process, I select the relevant citation variants through manual inspection of the step 1 results, and manually count the number of cited reference variants selected. After retrieving the final list of citing articles, I then analyze the results using the tools provided by WoS to obtain citation counts per year, subject category, and country for that edition of the WOA/WOD. I then repeat this process for each subsequent edition of the WOA/WOD.

Results

The citation counts generated for WOA/WOD, PSST, and GHRSSST during the first phase of this analysis are summarized in Figure 1. Two consistent patterns seem to emerge in these counts. First, the total number of citations generated for each data set increases as the coverage of the data source increases. WoS is the most limited of the three data sources, since it only indexes article metadata, acknowledgements, and cited references. Publishers' sites have wider coverage, since they allow access to articles' full text, but I only searched a limited number of these sites. Google Scholar has the broadest coverage, in that it offers access to the full text of a broad range of publishers' websites as well as to conference proceedings, institutional repositories, and other websites. The citation counts generated using these data sources seem to follow this pattern, with citation counts generated from publishers' sites being nearly four times higher than those generated from WoS and counts generated from Google Scholar being nearly eight times higher than WoS.

Second, the data sets are consistently ranked relative to each other across the three data sources. The citation counts for WOA/WOD are higher than those for PSST, which are higher than those for GHRSSST. The magnitude of these differences also seems consistent, with WOA/WOD receiving approximately four times

Table 1. Search strings used to generate citation counts for three data sets in WoS, publishers' full text websites, and Google Scholar.

Data Set	Search String
WOA/WOD	"world ocean atlas" OR "world ocean database"
PSST	"pathfinder sst" OR "pathfinder sea surface" OR "avhrr sst"
GHRSSST	GHRSSST OR "group for high resolution" OR "goade high resolution"

doi:10.1371/journal.pone.0092590.t001

Table 2. Search strategies used to identify articles citing each edition of the WOA/WOD in WoS.

Edition	Search Strategy
1982	Cited Author=(levitus s*) AND Cited Year=1982
1994	Cited Author=(conkright m* OR levitus s* OR boyer t*) AND Cited Year=1994
1998	Cited Author=(antonov j* OR levitus s* OR boyer t* OR conkright m*) AND Cited Year=1998
2001	Cited Author=(stephens c* OR boyer t* OR conkright m* OR locarnini r* OR levitus s*) AND Cited Year=2002
2005	Cited Author=(boyer t* OR locarnini r* OR antonov j* OR garcia h*) AND Cited Year=2006
2009	Cited Author=(boyer t* OR antonov j* OR garcia h* OR locarnini r*) AND Cited Year=(2009 OR 2010)

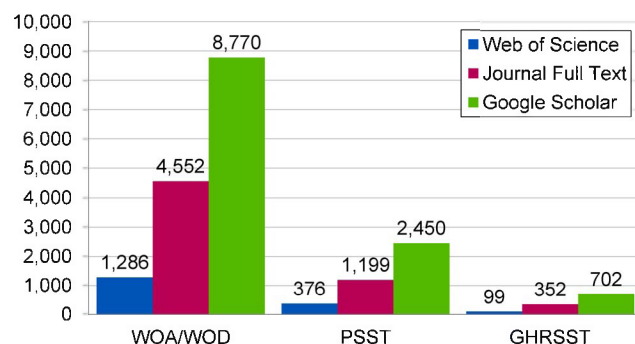
doi:10.1371/journal.pone.0092590.t002

more citations in each data source than PSST and PSST receiving approximately three times more citations than GHRSSST. The consistency of these patterns across data sets and data sources suggests that these findings are robust, although much additional work would be necessary to verify their accuracy.

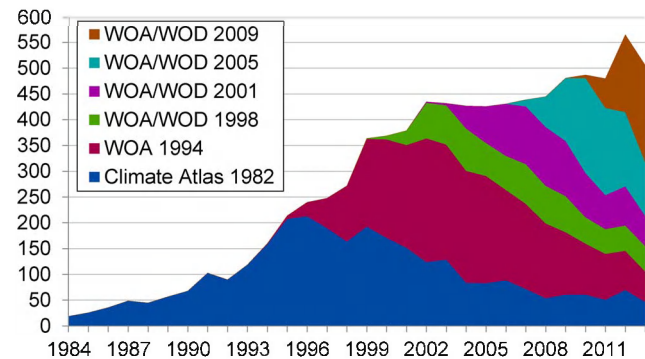
In compiling these citation counts, I also find a wide variety in the methods used to refer to these data sets. Examples of this variety are given in Figure 2. Some articles include formal citations to these data sets, but the format of these citations is highly variable, despite the fact that NODC provides a suggested citation format for each of these data sets. Many other articles simply mention the data set in the text of the article, although the format of such mentions is also highly variable. The data sets are referred to by various names (PSST alone is referred to as 'Pathfinder Sea Surface Temperature', 'Pathfinder SST', 'Advanced Very High Resolution Radiometer SST', 'AVHRR SST', etc) and a URL to the online source of the data is not always included.

In the second phase of this analysis, a more comprehensive Cited Reference Search in WoS for articles citing WOA/WOD, I find a total of 8,412 articles citing all six editions of the WOA/WOD from 1984 to 2013. The 1982 edition has been cited 2,987 times, the 1994 edition has been cited 2,577 times, the 1998 edition has been cited 810 times, the 2001 edition has been cited 842 times, the 2005 edition has been cited 795 times, and the 2009 edition has been cited 401 times. The distribution of articles citing WOA/WOD over editions and years is presented in Figure 2.

These distributions display a number of interesting features. First, versions of the WOA/WOD seem to require at least four, and up to 14, years after their initial release date to reach their peak citation rate. The time necessary for versions to reach their peak rate has declined with each version: the 1982 Climatological Atlas reached its peak 14 years after its initial publication, whereas

**Figure 1.** Citation counts retrieved for three oceanographic data sets from three data sources.

doi:10.1371/journal.pone.0092590.g001

**Figure 2.** Number of citations received per year by all editions of the WOA/WOD.

doi:10.1371/journal.pone.0092590.g002

the 1998 version required six years and the 2005 version required four. The amount of time necessary for the older versions of the WOA/WOD to reach their peak citation rate is longer than the 2–5 years required for most journal articles [71,72], although the 2005 version seems to have peaked within that timeframe.

This delay may be due to the media in which each successive version of WOA/WOD was distributed. The 1982 climatological atlas was distributed via magnetic tape and personal communication, the 1994 and 1998 editions were distributed via CD-ROM, the 2001 version was distributed via DVD, the 2005 was distributed via DVD and online access, and the 2009 version was distributed online. Each successive version made the data more accessible and usable, possibly leading to quicker incorporation of the data into scientific articles. In addition, updates began to be incorporated into the WOA/WOD every three months starting in 2008, allowing the WOA/WOD to be used for more timely investigations.

Second, although all releases of the WOA/WOD are highly cited, some versions are clearly more highly cited than others. The 1982, 1994, and 2009 versions all received over 200 citations in a single year, whereas the 1998 version never received more than 82 citations in a single year and the 2001 version never received more than 114. Since the 1998 and 2001 versions presented similar data, were prepared using similar methods, and compiled by many of the same authors as the 1994 and 2009 versions, it is unclear what conclusions to draw from these trends. In addition, the 2005 version seems to have been the most rapidly cited version of the data set, accumulating 795 citations in the eight years since it was published, although the 2009 version seems to be following a similar trajectory, having received 401 citations in the five years following its publication.

Finally, all versions of the WOA/WOD continue to be highly cited well beyond their publication date, even when one or several newer versions of the data set are available. The 1982, 1994, 1998, and 2001 versions all received between 50 and 60 citations in 2013. It is unclear whether researchers continue to use the older versions of the WOA/WOD out of habit, convenience, unawareness of newer versions, or other reasons. For whatever reason, researchers continue to cite these data sets well beyond the cited article half-life of 9.1 years recorded for Oceanography in the 2011 edition of *Journal Citation Reports*, suggesting that all versions of the WOA/WOD continue to be valuable resources for scientific research.

Analysis of the articles citing all versions of the WOA/WOD also reveals some interesting features. An analysis of the WoS subject categories of these citing articles, presented in Figure 3, shows that although the WOA/WOD is predominantly cited by articles in Oceanography, it is also cited by other related fields. The high number of citations from the Meteorology & Atmospheric Sciences category suggests that the WOA/WOD is frequently used by studies examining the effects of the ocean on weather and climate. Its number of citations in Paleontology, primarily by articles published in the journal *Paleoceanography*, suggests that it is used in studies of the prehistoric ocean as well as those of the modern ocean. Finally, its use in the Environmental Sciences, Marine and Freshwater Biology, and Ecology subject categories suggest that it is being used by studies examining the effects of ocean conditions on marine biota.

Figure 4 presents an analysis of these citing articles by country. Citing articles by authors from multiple countries are counted as whole citations for each country, rather than fractionally. Creation of the WOA/WOD is an international project in that the WOA/WOD consists of data sets contributed by researchers from numerous countries around the world. Figure 4 suggests that the international nature of its creation is reflected by the international scope of its use. The WOA/WOD is not only highly cited by the traditionally prolific scientific countries such as the United States, France, Germany, and the United Kingdom, but also by rising scientific countries such as China, India, and Brazil. This suggests that although NODC is an institution of the US government, its work to archive, quality control, and freely provide the data comprising the WOA/WOD is useful to the global scientific community.

Finally, as with phase 1 of this analysis, I find the format of citations to the WOA/WOD to be highly inconsistent. I found 377 variant methods of citing the 1982 version, 305 variants of the 1998 version, 221 variants of the 2001 version, 200 variants of the 2005 version, and 77 variants of the 2009 version, for a total of

1,180 variant methods of citing all versions of the WOA/WOD captured in WoS as of early 2013. See Figure 5 for a sample of the citation variants to the 2005 version. Because I did not attempt to search for erroneous citations (citations to the wrong year of publication, misspelled author names, etc), these figures are likely to underestimate the actual number of variant methods of citing the WOA/WOD in WoS. In addition, since phase 1 of this analysis suggests that articles are more likely to reference data sets in their text than in their cited references lists, the actual number of methods that articles use to refer to the WOA/WOD is likely to be substantially higher than I estimate here.

Discussion

These results seem to have a number of implications for data curation and data citation initiatives. First, my results indicate that all three of these data sets are highly cited. My phase 1 results suggest that, if they were counted as journal articles in WoS, both the WOA/WOD and the PSST data sets would have citation counts higher than 99% of all articles in Oceanography in WoS from any single publication year from 1995 to the present. Using the more expansive journal full-text method, each of the three data sets would be ranked in the top 1% for citation counts of all articles published in Oceanography during the same year, while the WOA/WOD and PSST data sets would be ranked in the top 0.1%. My phase 2 results indicate that each version of the WOA/WOD would be ranked in the top 0.1% of articles in Oceanography that were published during the same year and the 1982 and 1994 versions have been cited more than twice as often as the most highly cited article in Oceanography published in 1982 and 1994. Percentile values and article citation counts for journal articles in Oceanography were obtained by using the search string “WC = oceanography AND PY = 1995” and sorting the results by “Times Cited – highest to lowest.” This string was then repeated for the other publication years. Because of the limitations of my search methods noted above, these citation counts are likely to be underestimates of the actual totals for each data set.

These high citation counts are surprising in light of the fact that previous studies [24,52–54] have reported more modest citation counts to individual data sets. I speculate that the high citation counts reported here could result from the unique features of the particular data sets that I analyzed. First, each data set is freely and publicly available and has been described in enough detail to permit its reuse. Second, the GHRSSST and WOA/WOD data sets are composites of multiple smaller data sets contributed by multiple researchers to form a more comprehensive, global data set. Third, each data set has been available from a consistent source for an extended period of time. Finally, each data set is quality controlled to ensure the consistency and accuracy of the data contained in each set. Each of these features adds value to the original data sets, making the final data sets more useful to the oceanographic community. It may be that the high citation counts to these data sets, and particularly to the WOA/WOD, reflect the somewhat unique nature of these data sets.

If this is the case, it suggests a potential path forward for data repositories and data sharing in other disciplines. A single data set in isolation may have limited applications because of the methodology and parameters of its collection, but if that data set is quality controlled, adjusted, and merged with other similar data sets, it can be used to create a more comprehensive, overarching data product that can be queried for analysis at local, regional, or global disciplinary scales. The more data incorporated into such a product, the more useful the product becomes. Data repositories

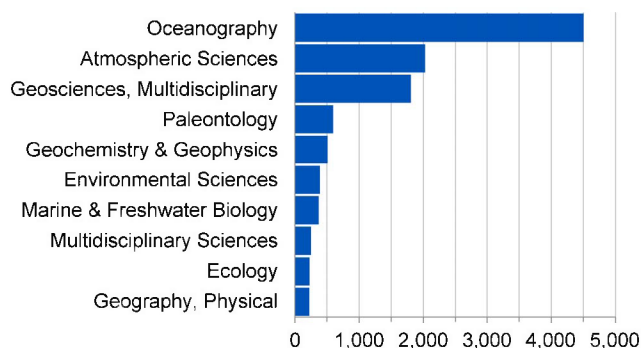


Figure 3. Number of articles citing all versions of the WOA/WOD per WoS-defined subject category.
doi:10.1371/journal.pone.0092590.g003

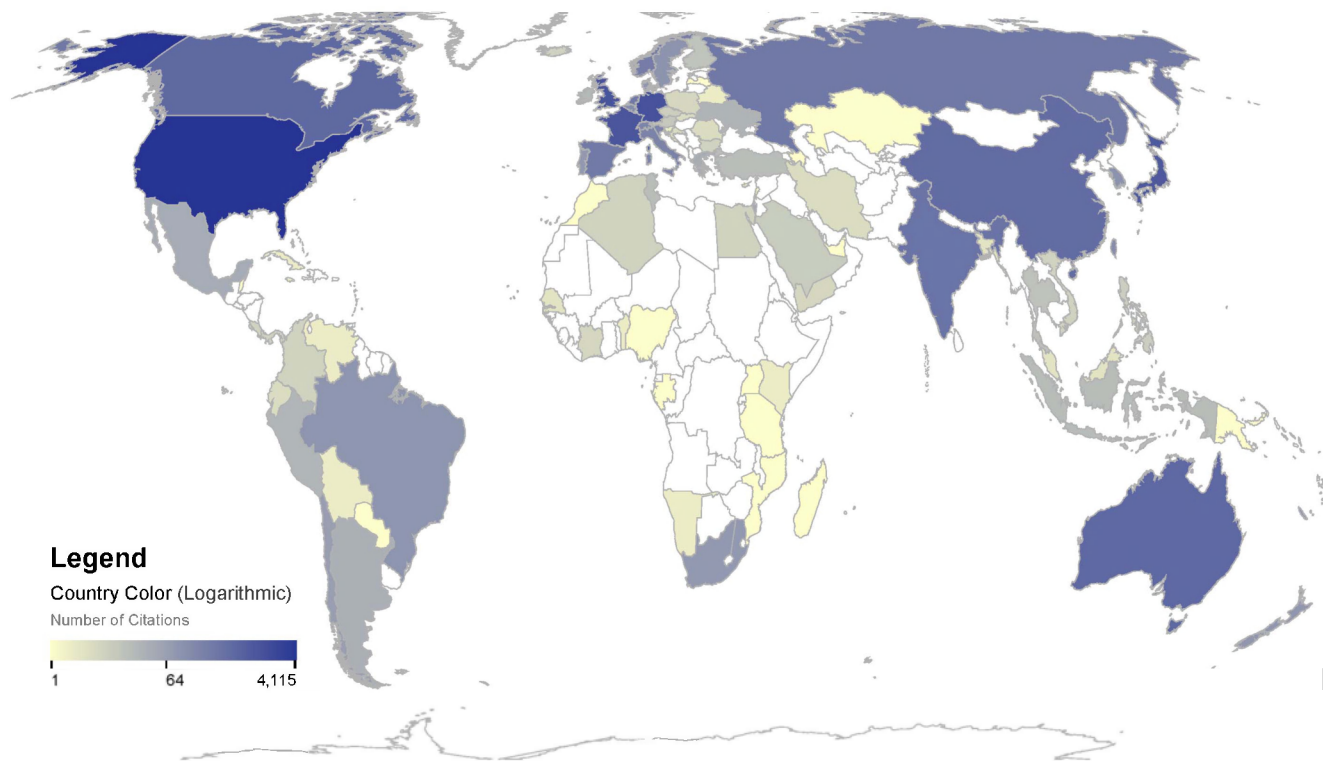


Figure 4. Number of articles citing all versions of the WOA/WOD per country.
doi:10.1371/journal.pone.0092590.g004

creating such products could then become central hubs for disciplinary, and potentially interdisciplinary, research, leveraging the limited research funding available in each discipline to ensure that individual pieces of research performed in that discipline eventually benefits the entire disciplinary community. In a sense, such a model could be considered a quality-controlled Wikipedia of data – the combination of individual pieces of expertise to create a resource larger and more comprehensive than anything that could be achieved individually. Obviously there are significant social, technological, and political barriers to implementing such a model, but the examples in oceanography of NODC and the Intergovernmental Oceanographic Commission [73] show that such barriers can be overcome.

Second, my results suggest that the majority of references to these data sets occur in the full text of articles, rather than in the title, abstract, keywords, acknowledgements, or cited references sections of these articles. The citation counts retrieved from full text sources—publishers’ websites and Google Scholar—are consistently and substantially higher than those retrieved from WoS. The only exception is that the publishers’ website total for WOA/WOD is lower than my phase 2 results, but this may be due to the large number of reference variants I found during phase 2. This pattern suggests that most articles do not refer to these data sets in a section indexed by WoS, calling into question the appropriateness of citation-indexing databases for compiling citation counts for these data sets.

Third, I find wide disparities in the methods used to cite or refer to these data sets, despite the fact that a formal citation format is suggested for each. This suggests that although a suggested citation format exists, researchers are not, for whatever reason, using it consistently to refer to these data sets. This finding is consistent with that of Part 2 of Mooney and Newton [28] and with many anecdotal accounts of citation practices among authors. It is likely

that the multiple points of access to these data sets may account for some of this inconsistency. PSST, for example, is also available from NASA, leading some to refer to the data set as the ‘NASA Pathfinder SST’ data set. The implication of this trend for the data citation community seems to be that although the development of a standard citation format is necessary, that format by itself is not sufficient to guarantee consistent citation of data sets. It seems that in addition to developing this format, it will be necessary to encourage researchers to use the format and, perhaps more importantly, to obtain commitments from journal editors, reviewers, and publishers to ensure that it is used.

More consistent adoption and usage of a doi to refer to a data set, either by directly assigning a doi to a data set or through the publication of a data paper with a doi, has the potential to considerably reduce the issues resulting from this inconsistency. From a purely bibliometric perspective, the format and content of a reference or citation is irrelevant as long as a doi is present. The consistent use of a doi to refer to a data set would enable a researcher to search full-text or citation-indexing databases for that doi to retrieve a reasonably accurate set of articles citing that data set. Again, however, such consistency requires both data providers to assign dois to their data sets and authors to include these dois in their papers. NODC has begun to lay the groundwork for such consistency in oceanography, having recently assigned its first doi to a version of the PSST data set and developing a process for assigning dois to its other data sets.

Conclusion

In this analysis, I attempted to generate citation counts for three oceanographic data sets curated by NODC by searching WoS, publishers’ websites, and Google Scholar for mentions of these data sets in the bibliographic information or full text of scientific

CITED REFERENCE INDEX
References: 1 - 50 of 85

Page 1 of 2 Go

<div> <div>Select Page</div> <div>Select All*</div> <div>Clear All</div> <div>Finish Search</div> </div>									
Select	Cited Author	Cited Work [SHOW EXPANDED TITLES]	Year	Volume	Issue	Page	Identifier	Citing Articles **	View Record
<input type="checkbox"/>	Antonov J [Show all authors]	NOAA Atlas NESDIS	2006					1	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	NOAA ATLAS NESDIS	2006	62		1		3	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	SALINITY	2006	2				14	
<input type="checkbox"/>	Antonov J I [Show all authors]	NOAA Atlas NESDIS 62	2006			182		7	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	WORLD OCEAN ATLAS 20	2006	62				5	
<input type="checkbox"/>	Antonov, J. I [Show all authors]	WORLD OCEAN ATLAS	2006	62				7	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	NOAA ATLAS NESDIS	2006	2				8	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	WORLD OCEAN ATLAS 20	2006	2				69	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	NOAA ATLAS NESDIS	2006	4		396		8	
<input type="checkbox"/>	Antonov, J. I. [Show all authors]	WORLD OCEAN DATABASE	2006	60				14	
<input type="checkbox"/>	ANTONOV JI	NOAA ATLAS NESDIS	2006	62		2		1	
<input type="checkbox"/>	ANTONOV JI	NOAA ATLAS NESDIS	2006	61				3	
<input type="checkbox"/>	Antonov, JI [Show all authors]	NOAA ATLAS NESDIS	2006	6		182		1	
<input type="checkbox"/>	ANTONOV JI	NOAA ATLAS NESDIS 62	2006	2		182		1	
<input type="checkbox"/>	ANTONOV JI	NOAA ATLAS NESDIS 62	2006	2		3021		3	
<input type="checkbox"/>	ANTONOV JI	NOAA ATLAS NESDIS 62	2006					1	
<input type="checkbox"/>	ANTONOV JI	NOAA ATLAS NFSDIS	2006	62				1	
<input type="checkbox"/>	ANTONOV JI	OFFICE	2006	62		182		3	
<input type="checkbox"/>	ANTONOV JI	SALINITY	2006	2		65		1	
<input type="checkbox"/>	ANTONOV JI	SALINITY	2006	2		17712		1	
<input type="checkbox"/>	ANTONOV JI	SALINITY NOAA ATLAS	2006	62				5	
<input type="checkbox"/>	ANTONOV JI	SALINITY NOAA ATLAS	2006	62		62		1	
<input type="checkbox"/>	ANTONOV JI	SALINITY NOAA ATLAS	2006	2				1	
<input type="checkbox"/>	ANTONOV JI	WORLD OCEAN ATLAS 20	2006	2		62		2	
<input type="checkbox"/>	Antonov, JI [Show all authors]	WORLD OCEAN ATLAS 20	2006					2	
<input type="checkbox"/>	Antonov, J.I. [Show all authors]	NOAA ATLAS NESDIS	2006	62		182		93	
<input type="checkbox"/>	Antonov, JI [Show all authors]	NOAA ATLAS NESDIS	2006	62				33	
<input type="checkbox"/>	Antonov, J.I. [Show all authors]	WORLD OCEAN ATLAS 20	2006	2		182		7	
<input type="checkbox"/>	ANTONOV JI	62 NOAA	2006	2				1	

Figure 5. Variant methods of citing the 2005 version of the WOA/WOD in WoS.
doi:10.1371/journal.pone.0092590.g005

articles. I found that although there were substantial differences in the citation counts derived from each source, all three data sets were highly cited in all sources. The WOA/WOD was particularly highly cited, with all versions of the data set having received over 8,000 citations since its first release in 1982. My results suggest that scientific articles are more likely to mention these data sets in the text than in the acknowledgements or cited references sections. I also find wide discrepancies in the methods used to refer to these data sets, both in the full text and in the cited references sections. I

found 377 variant methods of citing different versions of a single data set, WOA/WOD, suggesting that researchers are not consistently using the citation formats provided for these data sets.

Although I limited this analysis to oceanographic data sets in an attempt to control for potential differences in citation practices among fields [74], it seems likely that the findings and issues identified here may be similar for data sets in other disciplines. Previous studies have suggested that data sets in other disciplines are highly cited [54,56] and that references to data sets retrieved

from full text sources are higher than formal citation counts [24], both of which are consistent with the findings presented here. In addition, inconsistent referral to data sets in scientific papers is often raised as a motivator for the development of a data citation standard, suggesting that the large number of reference variants identified here is likely to be an issue with data sets in other disciplines as well.

However, although the findings of this analysis may be broadly applicable to data sets in other disciplines, much additional research would be needed to determine what, if any, differences exist in data citation and referencing patterns among disciplines. Since it is known that citation practices for journal articles differ among disciplines [74], it is likely that such differences also exist for data sets. Such differences may be compounded by differences in disciplinary collaboration rates, the existence and utilization of discipline-specific data repositories, or other factors.

In addition to citation counts, future research on the impact of data sets and data curation activities might focus on alternative metrics such as download counts, social network discussion, or social bookmarking to measure other forms of engagement with these data sets beyond formal citation [13,75], or on comparing such altmetric indicators with traditional cited reference counts. Although download counts could be easily obtained, other altmetric indicators might be more problematic to obtain due to the inconsistency with which data sets are cited. Unique identifiers such as dois might alleviate this issue somewhat, but only for data sets that have been assigned such identifiers and to the degree that researchers include these identifiers in their publications.

Finally, this analysis demonstrates that individuals and institutions can make substantial contributions to scientific research

without producing formal publications. My results suggest that these data sets are often used in the production of original research in oceanography. This use is possible because researchers posted their data sets to oceanographic data repositories and because these data sets were properly archived, described, and made available to the scientific research community. The high citation counts identified here suggest that these data sets – and, by extension, the curation activities necessary for their use in scientific articles – are at least as important to the advancement of oceanographic research as the findings presented in the vast majority of journal articles published in the field. Future evaluations of NODC and other organizations that curate scientific data ought to take such considerations into account.

Acknowledgments

I thank Kenneth S. Casey and Sydney Levitus of NODC for discussions about data citation and bibliometrics which ultimately led to the analyses presented here. Kenneth Casey and Timothy Boyer of NODC and two anonymous reviewers provided excellent suggestions that improved the manuscript considerably. Opinions presented in this paper are solely my own and do not necessarily represent those of NOAA, the Department of Commerce, or the US Government. Use of products in this analysis does not constitute a recommendation of these products by NOAA or the US Government.

Author Contributions

Conceived and designed the experiments: CWB. Performed the experiments: CWB. Analyzed the data: CWB. Wrote the paper: CWB.

References

- ShanghaiRanking Consultancy (2012) Academic Ranking of World Universities. Shanghai Jiao Tong University.
- Times Higher Education (2013) The Times Higher Education World University Rankings.
- Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, et al. (2012) The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology* 63: 2419–2432. doi:10.1002/Asi.22708.
- SCImago Research Group (2012) SCImago Institutions Ranking: SIR World Report 2012: Global Ranking. 98 p.
- Allen L, Jones C, Dolby K, Lynn D, Walport M (2009) Looking for Landmarks: The Role of Expert Review and Bibliometric Analysis in Evaluating Scientific Publication Outputs. *PLoS ONE* 4: e5910. doi:10.1371/journal.pone.0005910.
- Bornmann L, Mutz R, Neuhaus C, Daniel HD (2008) Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics* 8: 93–102. doi:10.3354/esep00084.
- Haefliger-Cavallion N, Graillot-Gak C (2009) The use of bibliometric indicators to help peer-review assessment. *Archivum Immunologiae Et Therapiae Experimentalis* 57: 33–38. doi:10.1007/s00005-009-0004-2.
- Moed H (2007) The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy* 34: 575–583. doi:10.3152/030234207x255179.
- van Raan AFJ (1996) Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics* 36: 397–420.
- Abramo G, Cicero T, D'Angelo CA (2012) National peer-review research assessment exercises for the hard sciences can be a complete waste of money: the Italian case. *Scientometrics* 95: 311–324. doi:10.1007/s1192-012-0875-6.
- Abramo G, D'Angelo C, Di Costa F (2011) National research assessment exercises: a comparison of peer review and bibliometrics rankings. *Scientometrics* 89: 929–941. doi:10.1007/s1192-011-0459-x.
- Uhlir PF (2012) For Attribution — Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop: The National Academies Press. 219 p.
- Chavan VS, Ingwersen P (2009) Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *Bmc Bioinformatics* 10: S2. doi:10.1186/1471-2105-10-s14-s2.
- Costello MJ (2009) Motivating Online Publication of Data. *Bioscience* 59: 418–427. doi:10.1525/bio.2009.59.5.9.
- Moritz T, Krishnan S, Roberts D, Ingwersen P, Agosti D, et al. (2011) Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group. *Bmc Bioinformatics* 12: S1. doi:10.1186/1471-2105-12-s15-s1.
- Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology and Evolution* 28: 454–461.
- Mons B, van Haagen H, Chichester C, t Hoen PB, den Dunnen JT, et al. (2011) The value of data. *Nature Genetics* 43: 281–283. doi:10.1038/ng0411-281.
- Chandler RE, Thorne P, Lawrimore J, Willett K (2012) Building trust in climate science: data products for the 21st century. *Environmetrics* 23: 373–381. doi:10.1002/env.2141.
- Helly J, Staudigel H, Koppers A (2003) Scalable models of data sharing in Earth sciences. *Geochemistry Geophysics Geosystems* 4: 1010. doi:10.1029/2002gc000318.
- Staudigel H, Albarede F, Anderson DL, Derry L, McDonough B, et al. (2001) Electronic data publication in geochemistry: A plea for “full disclosure”. *Geochemistry Geophysics Geosystems* 2: 2001GC000234. doi:10.1029/2001GC000234.
- Staudigel H, Helly J, Koppers AAP, Shaw HF, McDonough WF, et al. (2003) Electronic data publication in geochemistry. *Geochemistry Geophysics Geosystems* 4: 8004. doi:10.1029/2002gc000314.
- Conway EM (2006) Drowning in data: Satellite oceanography and information overload in the Earth sciences. *Historical Studies in the Physical and Biological Sciences* 37: 127–151. doi:10.1525/hsp.2006.37.1.127.
- Hofmann EE, Gross E (2010) IOC Contributions to Science Synthesis. *Oceanography* 23: 152–159. doi:10.5670/oceanog.2010.30.
- Parsons MA, Duerr R, Minster J-B (2010) Data Citation and Peer Review. *Eos, Transactions American Geophysical Union* 91: 297–298. doi:10.1029/2010eo340001.
- Walton DWH (2010) Data Citation - Moving to New Norms. *Antarctic Science* 22: 333–333. doi:10.1017/s0954102010000520.
- Science editors (2011) Dealing with Data. *Science* 331: 692–729.
- Klump J, Bertelmann R, Brase J, Diepenbroek M, Grobe H, et al. (2006) Data publication in the open access initiative. *Data Science Journal* 5: 79–83. doi:10.2481/dsj.5.79.
- Mooney H, Newton MP (2012) The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication* 1: eP1035. doi:10.7710/2162-3309.1035.

29. Cinkosky MJ, Fickett JW, Gilna P, Burks C (1991) ELECTRONIC DATA PUBLISHING AND GENBANK. *Science* 252: 1273–1277. doi:10.1126/science.1925538.
30. Dodge C, Majewski F, Marx B, Pfeifferberger H, Reinke M (1996) Providing global access to marine data via the World Wide Web. *Journal of Visualization and Computer Animation* 7: 159–168. doi:10.1002/(sici)1099-1778(199607)7:3<159::aid-vis149>3.0.co;2-t.
31. Brase J (2004) Using digital library techniques - Registration of scientific primary data. In: Heery R, Lyon L, editors. *Research and Advanced Technology for Digital Libraries*. pp. 488–494.
32. Schindler U, Brase J, Diepenbroek M (2005) Webservices infrastructure for the registration of scientific primary data. In: Rauber A, Christodoulakis S, Tjoa AM, editors. *Research and Advanced Technology for Digital Libraries*. pp. 128–138.
33. Altman M, King G (2007) A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib Magazine* 13. doi:10.1045/march2007-altman.
34. Goodman L, Lawrence R, Ashley K (2012) Data-set visibility: Cite links to data in reference lists. *Nature* 492: 356–356. doi:10.1038/492356d.
35. Thorisson GA (2009) Accreditation and attribution in data sharing. *Nature Biotechnology* 27: 984–985. doi:10.1038/nbt1109-984b.
36. Starr J, Ashton J, Brase J, Bracke P, Gastl A, et al. (2011) DataCite Metadata Schema for the Publication and Citation of Research Data (Version 2.1). *DataCite*. 27 p. doi:10.5438/0003.
37. De Schutter E (2010) Data Publishing and Scientific Journals: The Future of the Scientific Paper in a World of Shared Data. *Neuroinformatics* 8: 151–153. doi:10.1007/s12021-010-9084-8.
38. Gorgolewski K, Margulies DS, Milham MP (2013) Making data sharing count: a publication-based solution. *Frontiers in Neuroscience* 7: 9. doi:10.3389/fnins.2013.00009.
39. Peterson J, Campbell J (2010) Marker papers and data citation. *Nature Genetics* 42: 919–919. doi:10.1038/ng1110-919.
40. Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *Bmc Bioinformatics* 12: S2. doi:10.1186/1471-2105-12-s15-s2.
41. Uppala SM, Kallberg PW, Simmons AJ, Andrae U, Bechtold VD, et al. (2005) The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society* 131: 2961–3012. doi:10.1256/qj.04.176.
42. Gould J, Roemmich D, Wijffels S, Freeland H, Ignaszewsky M, et al. (2004) Argo profiling floats bring new era of in situ ocean observations. *Eos, Transactions of the American Geophysical Union* 85: 185–191. doi:10.1029/2004EO190002.
43. Boyd PW, Bakker DCE, Chandler C (2012) A New Database to Explore the Findings from Large-Scale Ocean Iron Enrichments Experiments. *Oceanography* 25: 64–71.
44. Borgman CL (2012) The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63: 1059–1078. doi:10.1002/asi.22634.
45. Nature editors (2009) Data Sharing. *Nature* 461.
46. Sedransk N, Young LJ, Kehler KL, Moffitt RA, Thakar A, et al. (2010) Make Research Data Public?—Not Always so Simple: A Dialogue for Statisticians and Science Editors. *Statistical Science* 25: 41–50. doi:10.1214/10-sts320.
47. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6: e21101. doi:10.1371/journal.pone.0021101.
48. Kostoff RN, Geisler E (2007) The unintended consequences of metrics in technology evaluation. *Journal of Informetrics* 1: 103–114. doi:10.1016/j.joi.2007.02.002.
49. Moed H (2008) UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics* 74: 153–161. doi:10.1007/s11192-008-0108-1.
50. van Dalen HP, Henkens K (2012) Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey. *Journal of the American Society for Information Science and Technology* 63: 1282–1293. doi:10.1002/asi.22636.
51. Weingart P (2005) Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics* 62: 117–131. doi:10.1007/s11192-005-0007-7.
52. Chao TC (2011) Disciplinary reach: Investigating the impact of dataset reuse in the earth sciences. *Proceedings of the American Society for Information Science and Technology* 48: 1–8. doi:10.1002/meet.2011.14504801125.
53. Piwowar HA, Carlson JD, Vision TJ (2011) Beginning to track 1000 datasets from public repositories into the published literature. *Proceedings of the American Society for Information Science and Technology* 48: 1–4. doi:10.1002/meet.2011.14504801337.
54. Piwowar HA, Vision TJ, Whitlock MC (2011) Data archiving is a good investment. *Nature* 473: 285–285. doi:10.1038/473285a.
55. ICPSR (2011) ICPSR Bibliography of Data-Related Literature. Inter-university Consortium for Political and Social Research: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/citations/>.
56. Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2: e308. doi:10.1371/journal.pone.0000308.
57. Piwowar HA, Vision TJ (2013) Data reuse and the open data citation advantage. *PeerJ* 1: e175. doi:10.7717/peerj.175.
58. Locarnini RA, Mishnov AV, Antonov JJ, Boyer TP, Garcia HE, et al. (2010) World Ocean Atlas 2009, Volume 1: Temperature. In: Levitus S, editor. *NOAA Atlas NESDIS 68*. Washington, DC: U.S. Government Printing Office. pp. 184.
59. Antonov JJ, Seidov ID, Boyer TP, Locarnini RA, Mishnov AV, et al. (2010) World Ocean Atlas 2009, Volume 2: Salinity. In: Levitus S, editor. *NOAA Atlas NESDIS 69*. Washington, D.C.: U.S. Government Printing Office. pp. 184.
60. Garcia HE, Locarnini RA, Boyer TP, Antonov JJ, Baranova OK, et al. (2010) World Ocean Atlas 2009, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. In: Levitus S, editor. *NOAA Atlas NESDIS 70*. Washington, D.C.: U.S. Government Printing Office. pp. 344.
61. Garcia HE, Locarnini RA, Boyer TP, Antonov JJ, Zweng MM, et al. (2010) World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, silicate). In: Levitus S, editor. *NOAA Atlas NESDIS 71*. Washington, D.C.: U.S. Government Printing Office. pp. 398.
62. Boyer TP, Antonov JJ, Baranova OK, Garcia HE, Johnson DR, et al. (2009) World Ocean Database 2009. In: Levitus S, editor. *NOAA Atlas NESDIS 66*. Washington, D. C.: U.S. Government Printing Office. pp. 216.
63. Levitus S (1982) *Climatological Atlas of the World Ocean*. Washington DC, US Government Printing Office: NOAA Professional Paper No. 13. 173 p.
64. Casey KS, Brandon TB, Cornillon P, Evans R (2010) The Past, Present and Future of the AVHRR Pathfinder SST Program. In: Barale V, Gower JFR, Alberotanza L, editors. *Oceanography from Space: Revisited*. Springer Verlag. doi:10.1007/978-90-481-8681-5_16.
65. Donlon C, Robinson I, Casey KS, Vazquez-Cuervo J, Armstrong E, et al. (2007) The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society* 88: 1197–1213. doi:10.1175/bams-88-8-1197.
66. Donlon CJ, Casey KS, Robinson IS, Gentemann CL, Reynolds RW, et al. (2009) The GODAE high-resolution sea surface temperature pilot project. *Oceanography* 22: 34–45.
67. Aguillo IF (2011) Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics* 91: 343–351. doi:10.1007/s11192-011-0582-8.
68. Jászó P (2010) Metadata mega mess in Google Scholar. *Online Information Review* 34: 175–191. doi:10.1108/14684521011024191.
69. Franceschet M (2010) A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics* 83: 243–258. doi:10.1007/s11192-009-0021-2.
70. Kousha K, Thelwall M, Rezaie S (2010) Using the Web for research evaluation: The Integrated Online Impact indicator. *Journal of Informetrics* 4: 124–135. doi:10.1016/j.joi.2009.10.003.
71. Costas R, van Leeuwen TN, van Raan AF (2011) The “Mendel syndrome” in science: durability of scientific literature and its effects on bibliometric analysis of individual scientists. *Scientometrics* 89: 177–205. doi:10.1007/s11192-011-0436-4.
72. Eom YH, Fortunato S (2011) Characterizing and Modeling Citation Dynamics. *PLoS ONE* 6: e24926. doi:10.1371/journal.pone.0024926.
73. Glover DM, Wiebe PH, Chandler CL, Levitus S (2010) IOC Contributions to International, Interdisciplinary Open Data Sharing. *Oceanography* 23: 140–151. doi:10.5670/oceanog.2010.29.
74. Garfield E (1979) *Citation Indexing. Its theory and application in science, technology and humanities*. New York: Wiley.
75. Priem J, Hemminger BM (2010) *Scientometrics 2.0: New metrics of scholarly impact on the social Web*. *First Monday* 15: 2874.