JLSC

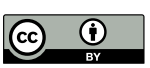
ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | http://jlsc-pub.org

Volume 5, General Issue (2017)

Open Data, [Open] Access: Linking Data Sharing and Article Sharing in the Earth Sciences

Samantha Teplitzky

Teplitzky, S. (2017). Open Data, [Open] Access: Linking Data Sharing and Article Sharing in the Earth Sciences. *Journal of Librarianship and Scholarly Communication, 5*(General Issue), eP2150. https://doi.org/10.7710/2162-3309.2150



ISSN 2162-3309 10.7710/2162-3309.2150

RESEARCH

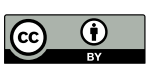
Open Data, [Open] Access: Linking Data Sharing and Article Sharing in the Earth Sciences

Samantha Teplitzky
Earth and Physical Sciences Librarian, University of California Berkeley

INTRODUCTION The norms of a research community influence practice, and norms of openness and sharing can be shaped to encourage researchers who share in one aspect of their research cycle to share in another. Different sets of mandates have evolved to require that research data be made public, but not necessarily articles resulting from that collected data. In this paper, I ask to what extent publications in the Earth Sciences are more likely to be open access (in all of its definitions) when researchers open their data through the Pangaea repository. **METHODS** Citations from Pangaea data sets were studied to determine the level of open access for each article. **RESULTS** This study finds that the proportion of gold open access articles linked to the repository increased 25% from 2010 to 2015 and 75% of articles were available from multiple open sources. **DISCUSSION** The context for increased preference for gold open access is considered, and future work linking researchers' decisions to open their work to the adoption of open access mandates is proposed.

Received: 08/30/2016 Accepted: 02/02/2017

Correspondence: Samantha Teplitzky, 351 LeConte Hall, University of California, Berkeley, CA, 94720-7300, steplitz@library.berkeley.edu



IMPLICATIONS FOR PRACTICE

1. Data repositories like Pangaea can be used as discrete data sources for examining discipline specific trends in the adoption of open science practices.
2. There is an increasing preference for gold open access journals among Earth Scientists who post data in Pangaea, but also a notable overlap in access types with articles also being shared in green repositories and through academic social networking sites.
3. A coupled understanding of open data and open publications may help librarians understand researcher preferences and provide opportunities to link the promotion of research data management and institutional open access policies.

INTRODUCTION

Practices for sharing data and sharing publications have evolved separately, and the decisions to make an article and its related data open are subject to different considerations. Researchers are being asked to make data sharing part of their workflows, both by funders who require the submission of data management plans and publishers who require researchers to make supplemental materials available. Access to articles has followed a different path, through the open access movement, through mandates to increase access to US federally funded research as well as more recent institutional open access policies that have encouraged green open access. Due to the way these two systems have evolved, there can be a disconnect between the article and its supplementary data. It is possible to fulfill data sharing obligations while publishing in a subscription-based journal. As the library liaison to the Department of Earth and Planetary Sciences, I strive to understand current research practices in my area of specialization and in this paper, I use the Earth Sciences literature as an example in order to consider to what extent data sharers are more likely to open their articles as well, and in what form.

The ways that articles are made open vary considerably. The Science Metrix report (Archambault et al., 2014), "Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels" introduces useful categories delineating these somewhat cloudy definitions of OA. These categories (p. 4-5) shed light onto the variable access points for articles in the Earth Sciences, including:

- Gold OA Articles published in full open access journals
- Gold OA Articles published in subscription or hybrid open access journals
- Green DOA (Delayed Open Access): open before or immediately after publication by author or institutional repository

- Gold DOA (Delayed Open Access): embargoed self-archiving and embargoed journals
- ROA (Robin Hood or Rogue Open Access): papers archived in non-institutional repositories or social networking sites such as ResearchGate and Academia.edu

Options for making data open vary as well. Many journals in the Earth Sciences have made the Pangaea repository a recommended repository for their supplemental materials (including high impact publications like *Nature Geoscience* and *Geochimica et Cosmochimica Acta*) ("Supported data repositories," 2016). Pangaea began in 1993 as a network for geological and environmental data and was designed to offer a flexible approach to project data management with considerations of the needs of the Earth Sciences, including heterogenous data types and geocoding functionality (Diepenbroek et al., 2002). Although including references to publications was not the main objective of the repository, articles that were linked to open data sets in Pangaea are useful proxies to illustrate the connection between sharing data and sharing articles and to demonstrate how the preferred mode of open access is shifting along with specific journal preferences among Earth Science researchers. For researchers, Pangaea offers linked data displays that georeference datasets for display on a related Google map; coordinated linking and uploading with several publishers; long-term archiving and open access to data; and easy sharing, citing, and identification using a Digital Object Identifier (DOI) ("About Pangaea Linked Data," 2016). For this study, Pangaea offered the benefit of covering different subfields of Earth Sciences, an interface that offered easy searching based on publication year and reference and a system for quick downloading of large sets of results.

LITERATURE REVIEW

Much has been written about recent interest in open science and the general movement toward openness in all aspects of research, both for resulting publications and their underlying data, but few have looked at the extent to which sharing in one domain influences the other.

Several studies have considered the preferences of scholars in the realm of data sharing that are relevant for the field of Earth Science. Herold (2015) studied the data sharing practices of Ecology, Evolution, and Natural Science researchers at the University of Minnesota and found high levels of sharing (46% of articles shared data), with most (91%) sharing data through journal websites. Cragin, Palmer, Carlson, and Witt (2010) studied data sharing in small science and found variety and complexity in how these subdisciplines created, curated, and shared data, often dependent on the practices and resources

of the lead PI or small network of collaborators. Schmidt, Gemeinholzer, and Treloar (2016) and Hsu, Martin, McElroy, Litwin-Miller, and Kim (2015) both address discipline specific preferences, expectations, and challenges in data sharing, Schmidt et al. regarding open data in global environmental research and Hsu et al., experimental geomorphology. They discuss the availability and growth of open data specific publications, but not open access publications in general.

Other studies and reports have tracked changes in article sharing and evaluated open access policies and mandates. Most studies follow the trajectory of the various policies that ask researchers to open their funded research to the public. In the US, the most prominent example comes from the NIH policy requiring deposit of articles whose research was funded by the NIH within 12 months of publication. The initial voluntary policy in 2005 saw compliance rates of 19%. Poynder (2012) explains that the mandatory policy (introduced in April 2008) saw compliance grow to 49% by the end of 2008 and 70% by the end of 2009. Van Noorden (2013) cites a compliance rate close to 80%. Gargouri, Larivière, Gingras, Carr, and Harnad (2012) also found that "mandates almost immediately triple the baseline Green OA self-archiving rate." Institutional mandates have seen slower uptake. Swan, Gargouri, Hunt, and Harnad (2015) compare deposits in mandated and non-mandated institutional repositories by discipline and show that for Earth Sciences, only about 12% of articles were deposited for mandated institutions and barely 2% for non-mandated institutions.

Another set of studies has focused on examining the advantages of publishing OA or the motivations of researchers, either increased readership (Davis, Lewenstein, Simon, Booth, & Connolly, 2008) or increased citations (Gargouri et al., 2010), as well as the implication of being able to find a large proportion of articles available online without fees. Archambault et al. (2014) reported that 57.8% of earth and environmental sciences articles published between 2011-2013 were gold, green, or other open access at the time of the study, and concluded:

> The fact that more than 50% of the papers published in peer-reviewed journals can now be downloaded for free by users who do not have to register to use a web site or to pay, that is, papers available in OA … certainly has important implications for academia, for university librarians, and perhaps even more so for the scientific, technical and medical publishing industry. (p. 28)

Piwowar (2011) examined who shares research data and found that among other factors including prior experience sharing or funding (by NIH), authors were likely to share data if their study was "published in an open access journal or a journal with a relatively strong

data sharing policy." She notes that more research is required to understand this association. This paper aims to delve more into this association between open data and open publications.

METHODS

The data repository Pangaea was consulted to obtain a set of articles whose data had been shared. Using the Advanced Retrieval Tool (ART) in Pangaea, the citations of publications from 2010-2015 (with related data sets) were downloaded in May 2016. Grobe, Diepenbroek, Dittert, Reinke, and Sieger (2006) explain that the Advanced Retrieval Tool "provides full access to all tables of the relational system and enables the user to retrieve individually configured subsets of data from the inventory." The ART was queried for "reference" and "year," and then article citations were sorted by their journal titles. Reports, theses, and book chapters were removed from consideration.

Gold open access trends

For each year, from 2010 through 2015, all cited articles were downloaded and sorted according to journal title. The status of journals as gold open access was checked on journal websites, as well as DOAJ (doaj.org) and the Geoscience Information Society list of open access journals relevant to the Earth Sciences (http://www.geoinfo.org/geooajour.html). Journal titles were examined and designated as gold open access when appropriate. The number of articles published in each of the gold open access journals was counted for each year.

Case studies: 2010 and 2015 article availability

Following the definitions of Archambault et al. (2014), each article for the end years 2010 and 2015 was searched in Google Scholar (using Google Scholar button installed in the Chrome Browser) for its availability as:

- Gold Open Access: published in a **Full** Open Access Journal
- Gold Open Access/Hybrid: published as an Open Access Article in a **Hybrid** Journal
- Green Open Access: available through an institutional or subject repository, or personal website
- Rogue Open Access: posted on Researchgate

Researchgate is an academic social media site used for collaboration, networking, as well as article sharing. It was chosen to represent rogue access because at time of these case studies, it hosted more papers and was visited more often than Academia.edu or Mendeley, two other well known sites that offer similar functionality (Van Noorden 2014).
In addition to the four categories above, the status of rolling, delayed, or embargoed access was noted. In many cases, articles were available from multiple sources, and this overlapping access was recorded as well.

RESULTS

Gold open access trends

Pangaea data depositors demonstrated an increased preference for full gold open access from 2010 through 2015 (Figure 1). In 2010, 9.7% (72), and in 2011 9.6% (72) of articles linked to Pangaea data sets were published in full gold open access journals. From 2011 forward, the percentage of articles in gold open access journals began to increase, to 19% (131) in 2012, 25% (158) in 2013, 26.9% (132) in 2014, and 30.9% (149) in 2015.

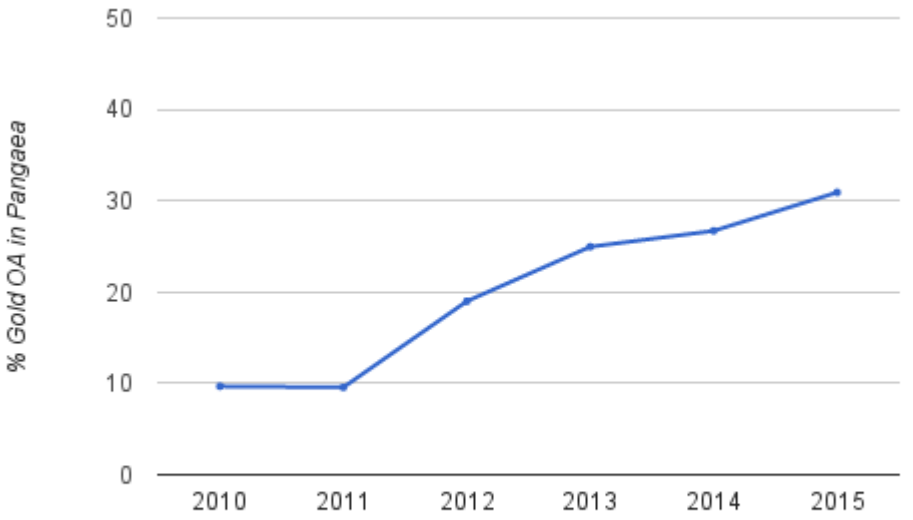


Figure 1. Percentage of Articles in Gold Open Access Journals linked to Pangaea Data Sets

The open access publishers, European Geosciences Union (EGU) and PLOS represented an increased share of articles by the end of the study period. PLOS showed a small increase from .5% to 4%. EGU attracted a growing share of publications in Pangaea, more than doubling from 7.1% to 15.4% over these five years. The top ten gold open access journals are shown in Figure 2. Throughout the 2010-2015 study period, Biogeosciences, an EGU journal, published 170 of the gold open access Pangaea articles, the largest number of articles of any gold OA journal cited and the largest number in 4 out of 6 of the study years (2010, 2012, 2013, 2015). Another notable change in this time period were the increased submissions to discussion journals which represented 1% of the full open access journals cited in 2010 and 17% in 2015.

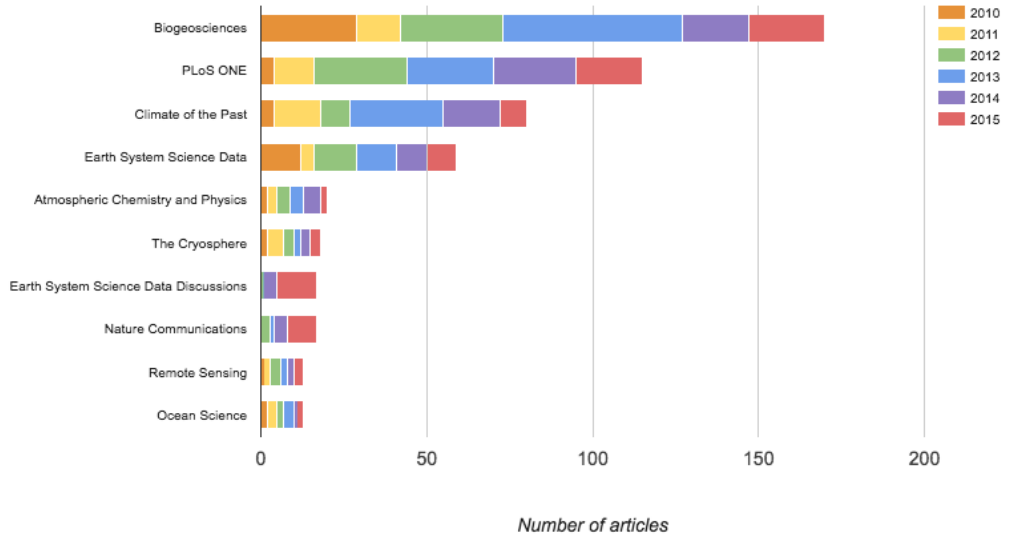


Figure 2. Top ten gold open access journals in Pangaea

Case studies: 2010 and 2015 article availability

To gain a deeper understanding of the preferred forms of open access represented by Pangaea articles, the years 2010 and 2015 were chosen for in depth case studies. Each article from 2010 and 2015 was considered for its type of open or free availability. In total, 75% (561) of 2010 articles are available freely, findable through Google Scholar, and 72% (348) of 2015 articles are similarly findable. Table 1 includes results by type of open access.

	2010		2015	
Gold/Full OA	72	9.7%	149	30.9%
Gold/Hybrid	50	6.7%	50	10.4%
Green	341	45.8%	163	33.8%
Researchgate	482	64.8%	238	49.4%
Embargo/Delay	113	15.2%	51	10.6%
TOTAL ARTICLES	**744**	**100%***	**482**	**100%***
Freely Available	561	75.4%	348	72.2%
Subscription access only	183	24.6%	134	27.8%

Table 1. Accessibility of Pangaea cited articles in 2010 and 2015
* Many articles were available from multiple sources.

Gold

There were 744 peer-reviewed articles linked to Pangaea data sets that were published in 2010, of which 9.7% (72) were published in fully open access journals. An additional 6.7% (50) were published as gold open access articles in hybrid journals, bringing the total available as gold open access upon publication to 16.4% (112). The 2015 data set included 482 peer-reviewed articles, the supplemental data sets of which were stored in Pangaea. Of those 482 articles, 30.9% (149) were published in fully open access journals. An additional 10.4% (50) were published as gold open access articles in hybrid journals. Together they represented 41.3% (199) of articles from 2015. From 2010 to 2015, the proportion of articles linked to data sets in Pangaea that were published as either full or hybrid gold open access increased about 25%.

Multiple sources

Many articles were available through multiple sources. For example, a single article might be available as gold open access on a publisher website and also in a repository and/or Researchgate. As a result of this overlap, there is some duplication in access. Table 2 includes a count of unique access articles that were available only through one source, such as green but not gold, or through Researchgate but not green or gold.

The impact of this overlapping access can be seen when considering green open access articles, Researchgate availability, and delayed access. Green open access articles, those deposited in institutional or subject repositories, represented 45.8% (341) of articles in 2010 and 33.8% (163) in 2015. Of those articles available as green open access, 6.9%

(51) were available uniquely through a repository or website in 2010 and 5.4% (26) in 2015.

	2010		2015	
Gold/Full OA	5	0.7%	36	7.5%
Gold/Hybrid	8	1.1%	9	1.9%
Green	51	6.9%	26	5.4%
Researchgate	188	25.3%	84	17.4%
Total available from one source only	**252**	**33.9%**	**155**	**32.2%**
TOTAL ARTICLES	**744**		**482**	

Table 2. Unique availability of Pangaea cited articles in 2010 and 2015

While searching for the availability of articles online, it is hard to miss the impact of Researchgate. Sixty-five percent (478) of 2010 articles and 49% (238) of 2015 articles were available in full on Researchgate. Articles posted in PDF format on Researchgate and not available freely elsewhere increased access to another 25% (188) of 2010 articles, and another 17% (84) of 2015 articles. The percentage of articles in Researchgate decreased 16% from 2010 to 2015. Without knowing if and when articles will be added to Researchgate post-publication, it is possible that the 2015 number will grow over time.

Finally, some publishers have implemented a rolling embargo or delayed open access after which articles are open. For example, the American Geophysical Union (AGU/Wiley) has a 24 month embargo. The number of open access articles in 2010 increased by 113 when embargoed articles were included, yet only 18 of those were not also available through some other means. It should be noted that from the vantage point of 2017, it is impossible to know when 2010 articles were added to one or multiple repositories, immediately, within 12 months, or sometime between publication and today. There is likely some lag due to rolling embargoes that are still in effect for 2015. We can predict that of the 2015 data set an additional 51 articles will be opened in 2017, yet only 20 of those are not available already through a repository or elsewhere.

Taken together, 75% of articles from 2010 and 72% of articles from 2015 are freely available, yet only 34% (2010) and 32% (2015) are available through a unique source. Over the course of the study period, articles were increasingly available from multiple sources, and the mix of those sources has changed as well. There was a stronger preference for gold and a decreased, or possibly delayed, participation in green open access and social media sharing through Researchgate.

DISCUSSION

To what extent are Pangaea data depositors likely to also open their related articles? Two notable findings emerged to shed light on this question. First, Pangaea data depositors expressed an increased preference for gold open access throughout the study years (2010-15). Second, a majority of articles sampled are available without subscription access in multiple venues for the case study years: 75% in 2010 and 72% in 2015. The implications of both results will be considered below.

Article availability: Strong uptake of Gold Open Access

From 2010 to 2015, researchers who deposited their data in Pangaea increasingly published their articles in gold open access venues. Their preference for gold open access in full OA and hybrid journals increased from 16.4% of articles to 41.3% of articles in that time period. This finding is unexpected as prior sources suggested a much lower uptake of gold open access in the Earth Sciences and leads to the question: how does this rise in gold open access articles among Pangaea depositors compare to Earth Science researchers in general? Bjork et al. (2010) studied the proportion of peer-reviewed scholarly journal articles that were available online in 2008, and found that 7% of Earth Science articles were gold. Gargouri et al. (2012) found an average of 2% of articles in the Earth Sciences were gold open access for the years 2005-2010. Figure 3 attempts to bring those numbers forward in time and compares the proportion of gold open access articles associated with Pangaea data sets to gold open access articles marked in Web of Science (within the subject designations of Geology, Geochemistry & Geophysics, Meteorology & Atmospheric Science). Web of Science indexed articles show a slow rise from 8.6% open access in 2010 to 11.8% in 2015. While not a perfect comparison, as Pangaea covers marine science and other subtopics that may not be represented equally in the Web of Science coverage, the comparison, along with the data from prior studies, does highlight the fact that the community of Pangaea users have adopted gold open access at a greater rate than Earth Sciences researchers in general.

The rising influence of full OA journals is seen in the increasing preference for the suite of journals published by the European Geosciences Union (from 7% to 15%). Seven out of the top 10 open access journals in this study are also published by EGU, and three EGU journals, Cryosphere, Geoscientific Model Development, and Biogeosciences are now among the top 25 Earth Sciences journals by impact factor. The study period also covered a period of growth in the number of EGU publications. Five titles were launched during the study period: *Solid Earth*, *Earth System Dynamics*, *Geoscientific Instrumentation*, Earth *Surface Dynamics*, and *Natural Hazards*. The increased use of discussion journals, similar

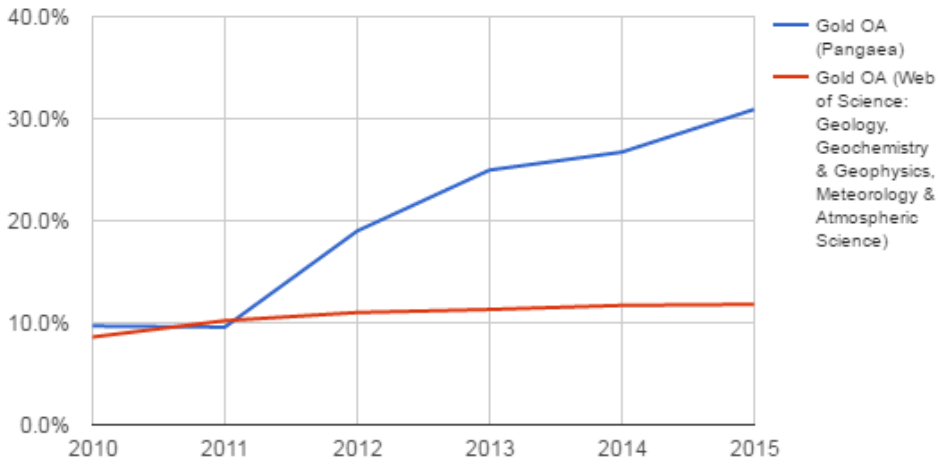


Figure 3. Percentage of Articles published in Gold Open Access Journals in the Earth Sciences

to preprint repositories like arxiv.org, was also noted.[1] While discussion journals were cited only 36 times throughout the study period, Earth System Science Data Discussions was the 7th most cited OA journal in the study. Pangaea records were not updated to show whether an article had been promoted from the discussion journal to the main journal, but the usage of discussion journals supports the idea that researchers are interested in engaging openly with reviewers and commenters and broadcasting their research early, ahead of official publication. Similar findings are found in the biological sciences in the recent support for open preprints (Desjardins-Proulx et al., 2013) and with the longstanding participation of physicists in ArXiv (Ginsparg, 2011).

Overlapping access

This study began by considering the extent to which Pangaea data depositors who shared their data also opened their articles, initially focusing on gold open access, but expanding scope to cover different types of access as it became apparent how much overlap in access there was. It was both unexpected and notable that in addition to the rise in gold open access availability, the majority of articles were available freely and through more than one

[1] A description of the EGU discussion journals purpose and scope can be found here: http://www.egu.eu/about/statements/position-statement-on-the-status-of-discussion-papers-published-in-egu-interactive-open-access-journals/

source. Seventy-five percent of 2010 articles were open, and 55% of those were available through more than one source; 72% of 2015 articles were open, and 61% of those were available through multiple venues. Figure 4 represents the overlap in free access of articles linked to Pangaea data sets. It reiterates both the growth in gold open access from 2010 to 2015 and the influence of Researchgate, demonstrated in particular by the overlap of Researchgate with green access in the 2010 dataset. As noted in the Results above, Researchgate saw a decline in the percentage of Pangaea articles deposited from 2010 to 2015. This might be attributed to a delay in posting articles to social media sites after publication, as well as the nature of Researchgate's methods for acquiring articles. The site has been criticized in the past for creating false profiles and populating them with PDFs scraped from the web and for sending out automated emails claiming to come from colleagues or co-authors (Van Noorden 2014). Although the practice seems to have been discontinued ("Inviting colleagues," 2016), it may have influenced deposits of articles in the first year of the study.

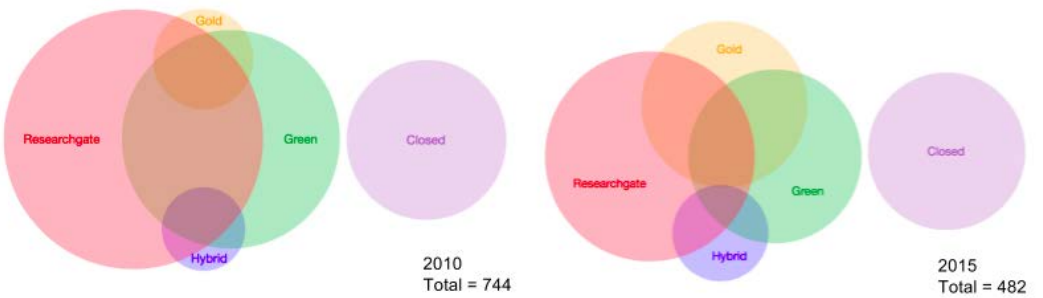


Figure 4. Overlap in free availability of articles linked to Pangaea data sets

Visualizing the overlap in access might also help elucidate the differences between researcher and librarian perspectives on open access publishing. For example, in this study, the articles of Pangaea researchers were well represented in green repositories, both institutional and subject related (though this study did not quantify the specific type). Green open access represents the largest proportion of OA papers, but also the most varied (Archambault et al. 2014). Librarians are well placed to understand the options for green open access and to educate our users on issues related to permanence, embargoes, and citation metrics. Those who are charged with promoting campus open access policies and the use of institutional repositories might find justification or explanation from this data for the slow uptake of institutional mandates (Swan, Gargouri, Hunt, and Harnad 2015) as these repositories represent just one venue of many possible options for researchers to share their work. Our own institutional open access policy has seen participation from

about 18% of faculty in the Earth and Planetary Sciences Department since our publication management system was launched in November of 2015. The use of these methods to analyze how publications are linked to their associated datasets might aid librarians in more fully assessing their patrons' adoption of open science principles.

Limitations and future work

This study attempted to shed light on the publishing habits of Earth Science researchers who share data. The choice of the Pangaea repository imposed some limitations. As mentioned above, Pangaea focuses on some subtopics within the Earth Sciences but is not necessarily representative of the field as a whole. Pangaea is also a European based system and EU researchers or co-authors may have favored Pangaea over other US based repositories. Although Pangaea is not an official repository for a particular publisher or journal, there does seem to be an unofficial preference for Pangaea among EGU journal authors as well as some Elsevier journals such as *Earth and Planetary Science Letters* and *Quaternary Science Reviews*. An examination of repositories such as EarthChem or Dryad, both endorsed by the American Geophysical Union (AGU) backed Coalition on Publishing Data in the Earth and Space Sciences (COPDESS), might yield more insights into US focused practices. Nonetheless, coupling the examination of data sharing and article sharing provided a fuller picture of researcher habits and demonstrated how a particular set of researchers, in this case those depositing in Pangaea, might differ from the general population of researchers. What this method does not achieve is an understanding of the motivations and influences of researchers who shared both their data and their articles in this time period.

Future work might consider the influence of funder and institutional mandates in the decision to publish open access. Funding sources or institutional affiliations of researchers might be examined to demonstrate whether Pangaea depositors were influenced by a particular open access mandate. An examination of researcher funding streams could determine whether European, British, and American policies influenced the steep rise in gold open access in Pangaea in 2012 and 2013. For example, Bjork (2016)'s recent observations note the influence of "major research funders, particularly in Europe, in the creation of earmarked budgets for the funding of gold OA articles." The Research Councils UK (RCUK) policy in 2012 requiring open access for all RCUK funded research expressed a preference for gold over green (Suber 2012), but in the US, the Federal Research Public Access Act in 2012, and its successor, the Fair Access to Science and Technology Research Act (FASTR) in 2015, would have required green open access to articles stemming from federally funded research within twelve months after publication. In between these efforts in February 2013, the White House released OSTP memo directing federal agencies with more than $100M in research and development expenditures to make articles and data

supported by federally funded research freely available within one year of publication. Many of the agencies covered by the OSTP memo implemented effective dates ranging from 2014 through 2016 ("Implementation of public access," 2016). Though these public access policies are becoming the default, they likely would have had limited impact on the articles covered here. Based on this study, it is not possible to say whether US and European mandates were effective carrots for Pangaea researchers, though their influence might be reconsidered in future work.

CONCLUSION

Open science has many aspects. This study is just one snapshot in time, and Pangaea is just one data repository in the Earth Sciences, but it illustrates that the drive toward openness is tangible and holistic. Librarians might consider whether the practices of researchers who use Pangaea are indicative of future trends in other disciplines and look for opportunities to link the promotion of research data management to outreach in open access and scholarly communications. Moving forward more work needs to be done to extend the study of openness in data and in publications beyond Earth Science to other disciplines and to consider the influence of funder and institutional mandates on the uptake of open access in all of its forms.

ACKNOWLEDGEMENTS

Code for the overlapping venn diagrams was taken from: https://github.com/benfred/venn.js/

REFERENCES

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013. Retrieved from http://science-metrix.com/en/publications/reports/proportion-of-open-access-papers-published-in-peer-reviewed-journals-at-the

Björk, B.C. (2016). The open access movement at a crossroad: Are the big publishers and academic social media taking over? *Learned Publishing*, *29*(2), 131–134. https://doi.org/10.1002/leap.1021

Björk, B.C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Guðnason, G. (2010). Open access to the scientific journal literature: Situation 2009. *PLOS ONE*, *5*(6), e11273. https://doi.org/10.1371/journal.pone.0011273

CENDI. (2016) Implementation of Public Access Programs in Federal Agencies. Retrieved from https://cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023–4038. https://doi.org/10.1098/rsta.2010.0165

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. *BMJ*, *337*, a568. https://doi.org/10.1136/bmj.a568

Desjardins-Proulx, P., White, E. P., Adamson, J. J., Ram, K., Poisot, T., & Gravel, D. (2013). The case for open preprints in biology. *PLoS Biol*, *11*(5), e1001563. https://doi.org/10.1371/journal.pbio.1001563

Diepenbroek, M., Grobe, H., Reinke, M., Schindler, U., Schlitzer, R., Sieger, R., & Wefer, G. (2002). PANGAEA—an information system for environmental sciences. *Computers & Geosciences*, *28*(10), 1201–1210. https://doi.org/10.1016/S0098-3004(02)00039-0

Elsevier. (2016) Supported data repositories: Earth, environmental and oceanographic data. Retrieved from https://www.elsevier.com/books-and-journals/enrichments/data-base-linking/supported-data-repositories#Earth

Elsevier. (2016) About Pangaea linked data. Retrieved from https://www.elsevier.com/books-and-journals/enrichments/pangaea-linked-data

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, ppen access increases citation impact for higher quality research. *PLOS ONE*,*5*(10), e13636. https://doi.org/10.1371/journal.pone.0013636

Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). Green and gold open access percentages and growth, by discipline. *arXiv:1206.3664 [cs]*. Retrieved from http://arxiv.org/abs/1206.3664

Ginsparg, P. (2011). ArXiv at 20. *Nature*, *476*(7359), 145-147. https://doi.org/10.1038/476145a

Grobe, H., Diepenbroek, M., Dittert, N., Reinke, M., & Sieger, R. (2006). Archiving and distributing earth-science data with the PANGAEA information system. In P. D. D. K. Fütterer, D. D. Damaske, P. D. https://doi.org/10.1007/3-540-32934-x_51

G. Kleinschmidt, P. D. D. h c H. Miller, & D. F. Tessensohn (Eds.), *Antarctica* (pp. 403–406). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/3-540-32934-X_51

Herold, P. (2015). Data sharing among ecology, evolution, and natural resources scientists: An analysis of selected publications. *Journal of Librarianship and Scholarly Communication. 3*(2), p.eP1244. https://doi.org/10.7710/2162-3309.1244

Hsu, L., Martin, R. L., McElroy, B., Litwin-Miller, K., & Kim, W. (2015). Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology*, *244*, 180-189. https://doi.org/10.1016/j.geomorph.2015.03.039

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

Poynder, R. (2012). Open access mandates: Ensuring compliance. Retrieved from: http://poynder.blogspot.fi/2012/05/open-access-mandates-ensuring.html

Researchgate. (2016). Inviting colleagues to Researchgate. Retrieved from https://explore.researchgate.net/display/support/Inviting+colleagues+to+ResearchGate

Schmidt B., Gemeinholzer B., Treloar A. (2016) Open data in global environmental research: The Belmont Forum's Open Data Survey. *PLoS ONE 11*(1): e0146695. https://doi.org/10.1371/journal.pone.0146695

Suber, P. (2012). Ensuring open access for publicly funded research. *BMJ*, *345*, e5184. https://doi.org/10.1136/bmj.e5184

Swan A., Gargouri G., Hunt M., & Harnad S. (2015) Working together to promote open access policy alignment in Europe – Work package 3 report: Open access policies. Retrieved from: http://eprints.soton.ac.uk/375854/1/PASTEUR4OA3.pdf

Van Noorden, R. (2013). NIH sees surge in open-access manuscripts. *Nature.* Retrieved from http://blogs.nature.com/news/2013/07/nih-sees-surge-in-open-access-manuscripts.html

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature.* https://doi.org/10.1038/512126a