RESEARCH PAPER

The Open Data Challenge: An Analysis of 124,000 Data Availability Statements and an Ironic Lesson about Data Management Plans

Chris Graf†, Dave Flanagan, Lisa Wylie & Deirdre Silver

Wiley, 9600 Garsington Road, Oxford OX4 2DQ, UK

Keywords: Data availability statement (DAS); FAIR data; Machine learning; Trends; Journal; Policy

Citation: C. Graf, D. Flanagan, L. Wylie & D. Silver. The open data challenge: An analysis of 124,000 data availability statements and an ironic lesson about data management plans. Data Intelligence 2(2020), 554–568. doi: 10.1162/dint_a_00061
Received: November 26, 2019; Reviewed: July 13, 2020; Accepted: July 22, 2020

ABSTRACT

Data availability statements can provide useful information about how researchers actually share research data. We used unsupervised machine learning to analyze 124,000 data availability statements submitted by research authors to 176 Wiley journals between 2013 and 2019. We categorized the data availability statements, and looked at trends over time. We found expected increases in the number of data availability statements submitted over time, and marked increases that correlate with policy changes made by journals. Our open data challenge becomes to use what we have learned to present researchers with relevant and easy options that help them to share and make an impact with new research data.

1. INTRODUCTION

This study looks at data availability statements submitted to Wiley journals for useful information about how researchers actually share research data. This is important because researchers, particularly those with grant funding, are increasingly required to share the data they create [1, 2]. Our challenge becomes to present researchers with relevant and easy options that help them to share and make an impact with new research data.

† Corresponding author: Chris Graf (E-mail: cgraf@wiley.com; ORCID: 0000-0002-4699-4333).

2. RELATED WORK

More than a decade ago the US National Institutes of Health (NIH) said "data sharing is essential for expedited translation of research results into knowledge, products, and procedures" and began requiring data sharing plans in all grants greater than US$500,000 [3]. Since 2018, the General Office of the State Council in China has required that "all scientific data derived from the science and technology plans … shall be archived in the relevant science data centers" [4]. State Council governs the funding made available by the Chinese Academy of Sciences and also the Ministry of Science and Technology, under which sits the National Natural Science Foundation of China. The EUR100 billion Horizon Europe funding program from the European Commission will challenge funded researchers to deliver "open access to publications, data, and to research data management plans" starting in 2021[5]. Private not-for-profit funders, like the Wellcome Trust and the Bill & Melinda Gates Foundation, are among those with the most progressive requirements. Funders argue that sharing data creates more value and impact from every grant they award and enhances trustworthiness and potential reproducibility. This is the open data challenge: Sharing data first, and then realizing the value from doing so.

While the arguments for data sharing are compelling and the rhetoric often exciting ("researchers are creating, gathering and using data in hitherto unimagined-volumes" [6]), reservations have been expressed that reflect researchers' concerns, including the absence of infrastructure and incentives, and the presence of disincentives such as the fear of getting scooped and concerns about misinterpretation or misuse of shared data [7,8]. Whatever your position in those arguments, it is fair to say that the open data challenge is a big challenge, and it is important to recognize that communities of researchers are ready to meet it in different ways and to different degrees [9]. For example, life scientists have been reported as the most ready to share the data they create [10,11], and early career researchers may be particularly well-prepared to do so [12, 13]. In fact, researchers in most data-oriented disciplines have embraced the challenge to an extent, and even where research data are associated with complex ethical issues like consent and privacy, the obligation to share data is recognized [14]. It helps that researchers can look forward to greater impact [15, 16]. It also helps that understanding of what "open" data really mean has become quite sophisticated, aided by the FAIR Data Principles [17] and promoted with the much-used soundbite that research data should be "as open as possible, as closed as necessary" [18].

Looking at this through a publishing-focused lens, when researchers are ready to share data, publishers and journals can play a useful role in enabling and realizing the benefits. They help communicate and explain standards and expectations [19, 20]. They help researchers meet the data sharing requirements including those set by their funders [21]. They increase the discoverability of shared data, perhaps 1000-fold (although there may be more correlation in that number than simply causation) [22]. They prompt researchers to "plan for the longevity, reusability, and stability of the data" [23].

Opinions about data sharing among researchers continue to be widely surveyed [24, 25]. Actual data sharing practices have been investigated by looking at data availability statements published in journal articles [26, 27, 28]. Data availability statements describe whether and how researchers' newly analyzed

research data have been made available, and the conditions under which they can be accessed. When research authors have shared data in appropriate research data repositories, their data availability statements can include permanent identifiers to link between the journal article and the data. Studies of data availability statements have been used to assess the impact and increase effectiveness of data sharing policies at journals [26, 27, 28], and study how data sharing practices are changing. Some conclude that practices fall short of the study authors' expectations [26, 29, 30]. This reminds us how important it is for publishers and journals to set reasonable expectations, and to support those expectations with robust policy and process [31]. It also speaks to the pace of change in different communities as they become familiar with, interested in, able to, and required to share research data. Publishers and journals need to match that pace, and they can also lead change. Measuring, interpreting, and acting on data sharing trends ensures that publishers and journals continue to serve researchers well.

3. METHODS

We used topic modeling, an unsupervised machine learning technique, to identify topics from 124,000 data availability statements submitted by research authors to 176 Wiley journals between 2013 and 2019. The complete workflow is available at GitHub[①]. The workflow is managed with Snakemake [32].

Wiley's electronic editorial office systems allow for the inclusion of custom questions on a journal-by-journal basis. We first extracted all of the records that contained the term "data" in either the question or the answer, but then limited the selection to questions that mentioned "data availability" or "data accessibility".

We then used spaCy [33] to tokenize the answers, limiting the tokens to nouns, proper nouns, and adjectives. We also added some custom stop words to ignore like "Wiley", "url", "et", and "al".

Then, we used scikit-learn [34] to create a term frequency-inverse document frequency (TF-IDF) matrix [35] of the tokenized answers, followed by Latent Dirichlet Allocation (LDA) [36]. We initially used 20 topics to cluster the documents. We used pyLDAvis [37] to visualize the topics estimated by the model.

Finally, we labeled the topics where possible for further analysis and discussion, using Wiley's Data Sharing Policy Author Templates [38] as a starting guide.

4. RESULTS AND ANALYSIS

Simply counting the number of answers to the custom questions that contain the term "data availability" or "data accessibility" shows a dramatic uptick in volume starting in early 2019 (Figure 1). This coincides with the rollout of Wiley's Expects Data policy, which added data availability statement requirements to more than 100 journals starting in December 2018 [19].

① https://github.com/DWFlanagan/data-availability-statements

Figure 1. Cumulative data availability statements (left) and approximate cumulative number of submissions (right) from August 2012 to October 2019. The number of data availability statements increased dramatically in the first half of 2019.

Visualization of the topics estimated by the LDA (Figure 2) shows that the initial choice of 20 topics is reasonably spread out. As described in the original paper on LDAvis [37]: "In this view, we plot the topics as circles in the two-dimensional plane whose centers are determined by computing the distance between topics, and then by using multidimensional scaling to project the intertopic distances onto two dimensions, as is done in [38]. We encode each topic's overall prevalence using the areas of the circles, where we sort the topics in decreasing order of prevalence". Lambda is a relevance metric that can be adjusted to alter the rankings of terms in order to aid topic interpretation. The keyword frequencies are ranked in the right panel for the complete topic model, and hovering over an individual topic shows how that topic compares to the complete model.

Selecting the number of topics in an LDA analysis is an iterative process, and there is no formula for predicting what will be the "best" number of topics. Too few topics and they will be too general; too many, and they will be too specific. In other projects, we have used the pyLDAvis tool shown in Figure 2 to evaluate how well a topic model probably covers the topic space, looking at how much overlap there is between topics. As shown in Figure 2, there are some clusters, but only a couple of topics with significant overlap (5 and 15, which we labeled as "Third-party restrictions" and "Genetics databases"). Based on our experience with other projects, this is a good starting point for a topic model.

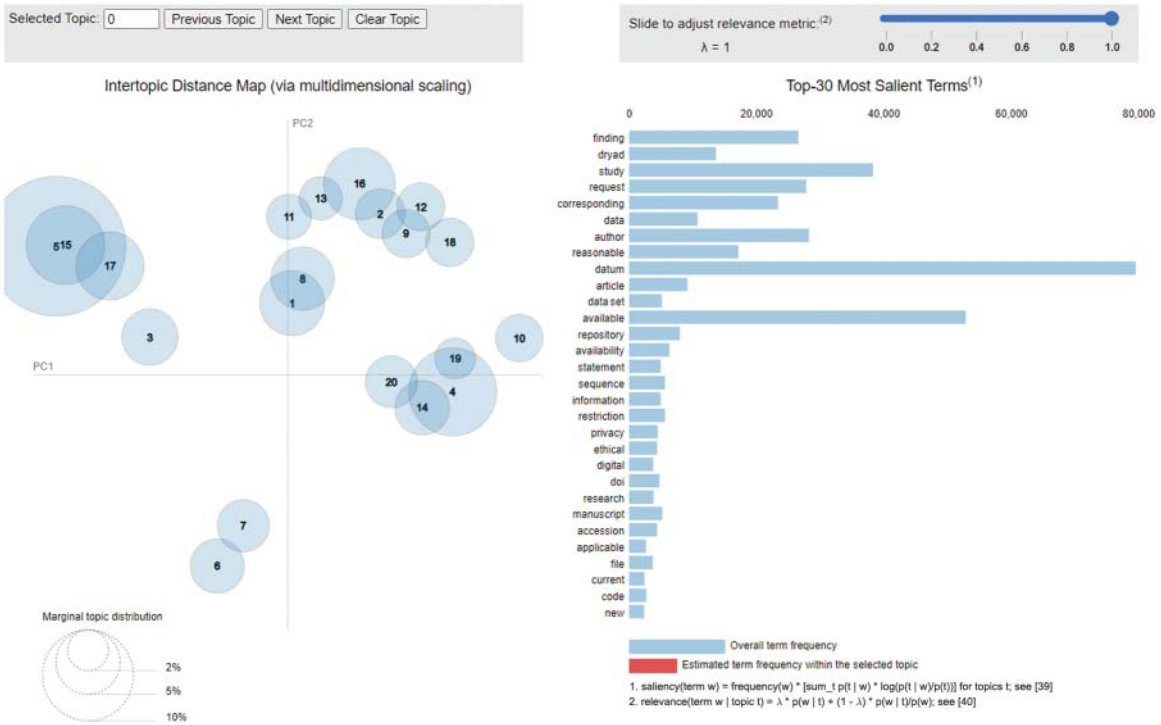

Figure 2. Visualization of the topics estimated by the LDA using pyLDAvis. Interactive plots available at DOI: 10.22541/au.157253515.58528497.

After checking that the number of topics seemed reasonable, we labeled the individual topics by hand by reading the top 20 statements for each topic and making our best guess what the cluster was about.

Some of the topics correspond well with one of the standard templates we encourage authors to use on our Data Sharing Policy page [38] and were easy to label, such as "#5: Third-party restrictions", which matched with "Data subject to third-party restrictions".

Other labels were more problematic. "6: Uncharacterizable" was a cluster that included experimental sections and actual data that the authors had copied and pasted into the Data Availability statement, perhaps highlighting the need for better author instructions. "7: Mixed" had many different kinds of statements that the LDA algorithm with the given parameters had combined. Tuning the text preprocessing parameters or the LDA parameters (number of topics and other hyperparameters) might resolve this mixed topic.

Some labels are also repeated. Topics 8 and 9 are both examples of "Available on reasonable request", although the LDA algorithm has resolved them into two separate topics based on the words contained in the statements themselves.

The percentages of each topic are shown in Figure 3.

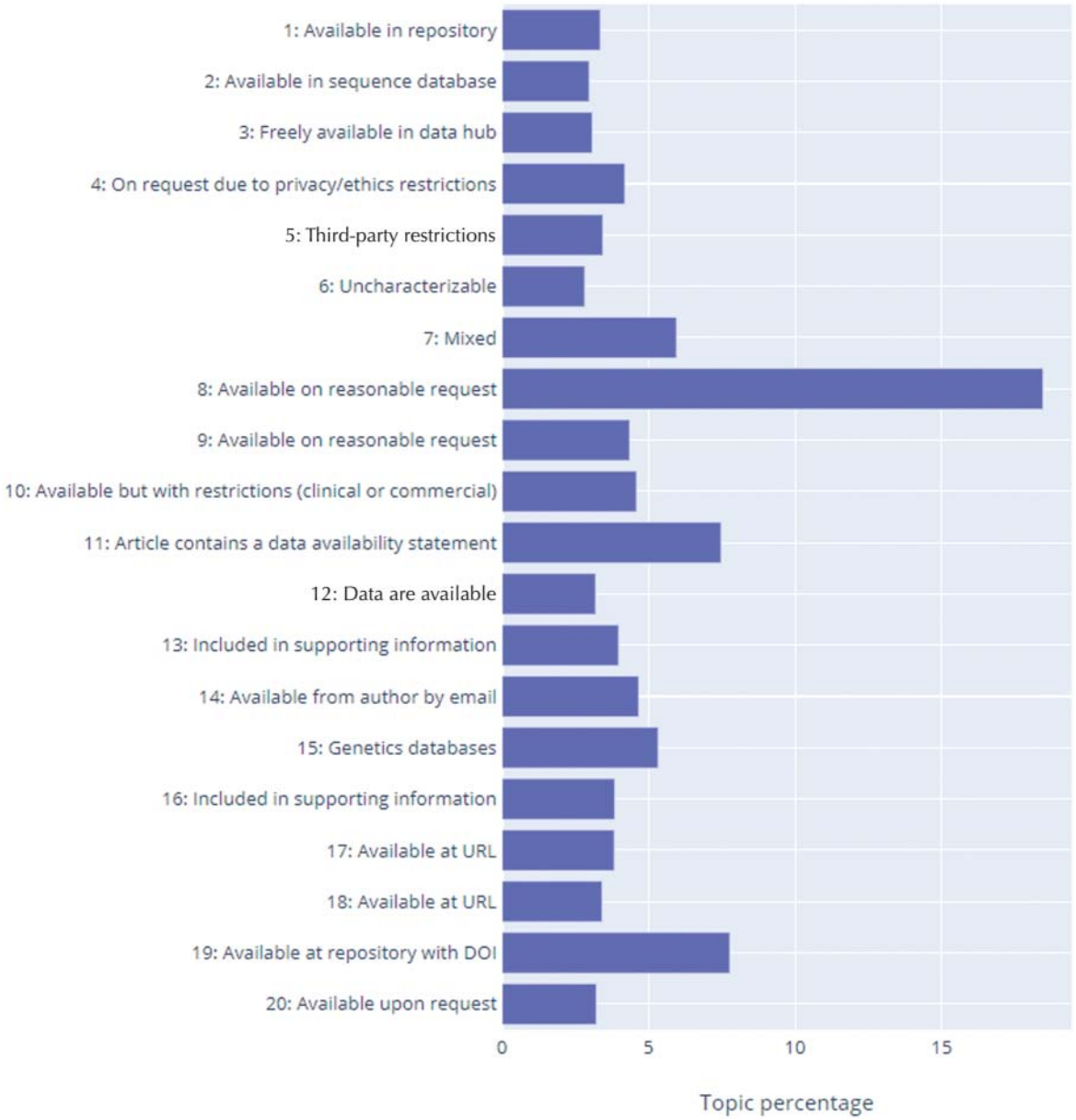

Figure 3. Topic distribution by percentage. Interactive plots available at DOI: 10.22541/au.157253515.58528497.

We can visualize which topics correspond to which document, as shown in Figure 4. In Figure 4, each row corresponds to a single document, and each column to a topic. The intensity of the color corresponds to the topic weighting for that document as calculated by the model. We would expect that a document would correspond strongly to one or two topics. For example, Document 1 corresponds most strongly to Topic 11, while Document 2 corresponds most strongly to Topic 15, but also to Topic 16.

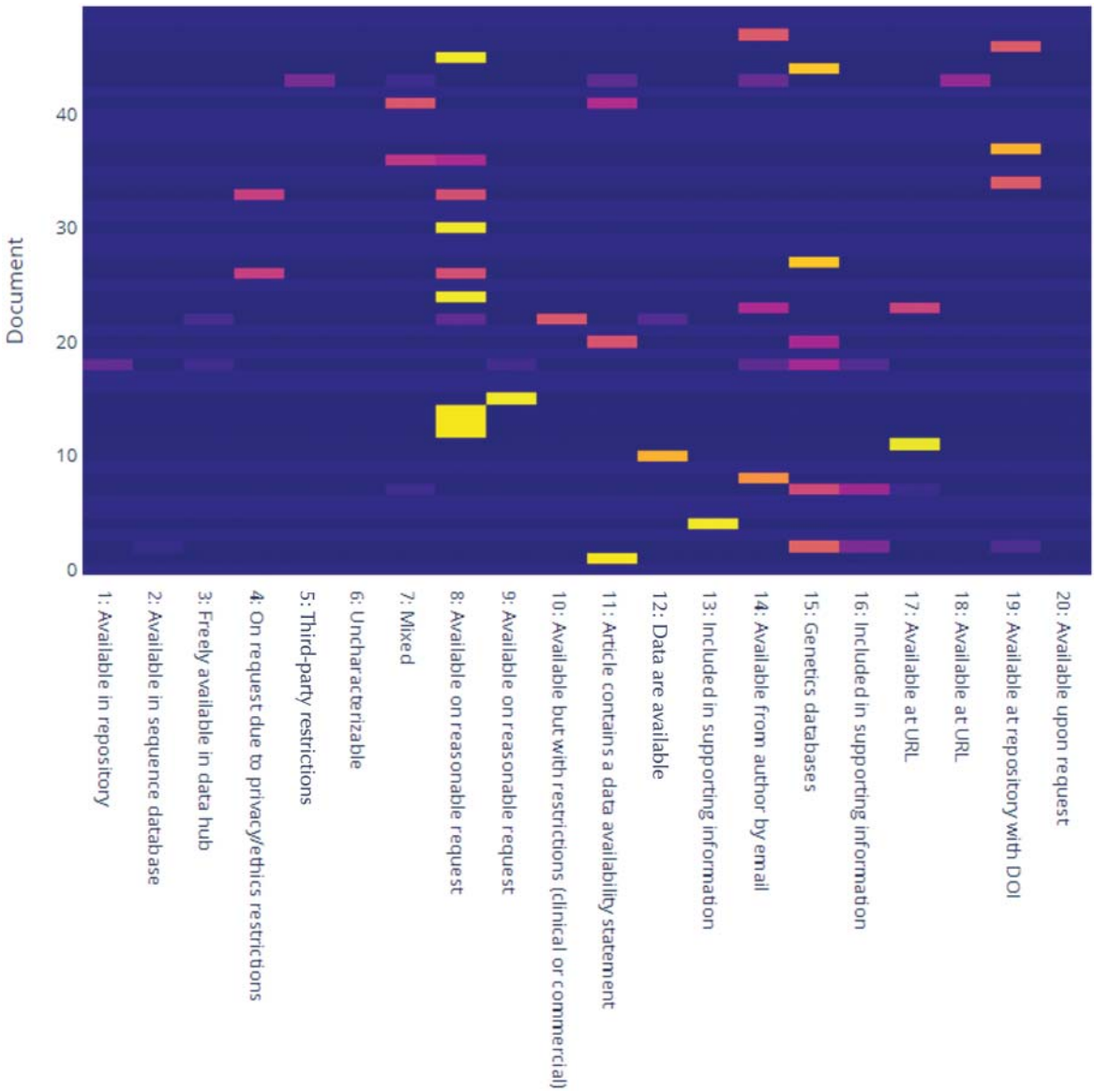

Figure 4. Document/topic matrix for the first 50 documents in the data set. Topic weighting ranges from low (blue) to high (red). Interactive plots available at DOI: 10.22541/au.157253515.58528497.

We can also visualize which words correspond most strongly to each topic, as shown in Figure 5. As expected, words like "datum" (the tokenized version of "data"), "available", "reasonable", and "request" correspond strongly to Topic 8, "Available on reasonable request". Topic 19, "Available at repository with DOI", is strongly associated with terms like "dryad" and "repository".

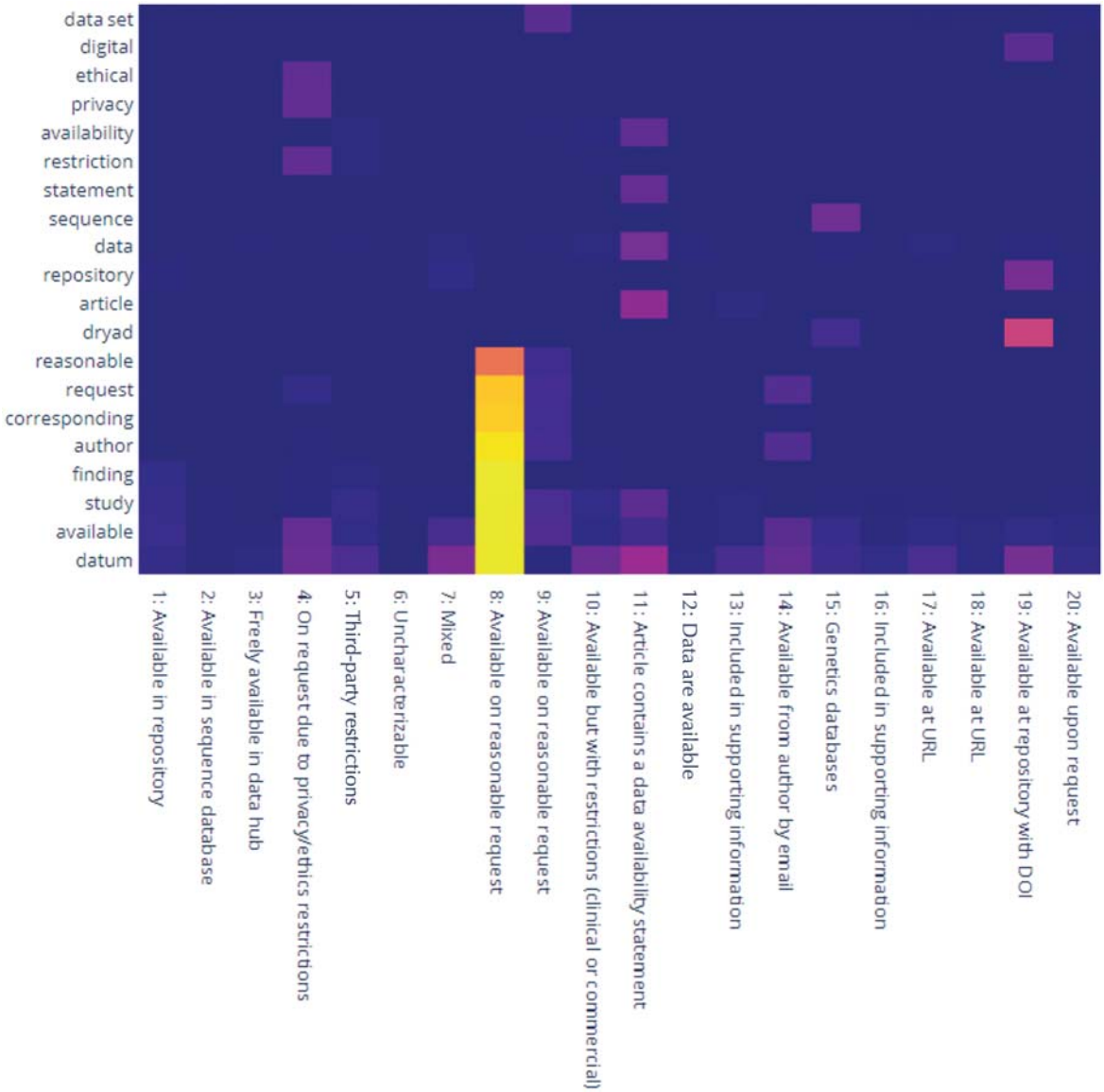

Figure 5. Word/topic matrix for the 20 most popular overall words in the data set. Topic weighting ranges from low (blue) to high (red). Interactive plots available at DOI: 10.22541/au.157253515.58528497.

Finally, we can visualize the trends in topic growth over time (Figure 6) by assigning each document to the highest-weighted topic. A significant number of statements (approximately 46,000) were not predicted to belong to any topic category, so we filtered these out. This could be investigated in future work – it might be noise spread out over the entire data set that could be reduced by tuning the model. We can see that after December 2018 the number of authors making data available on reasonable request (Topic 8) made a sharp increase.

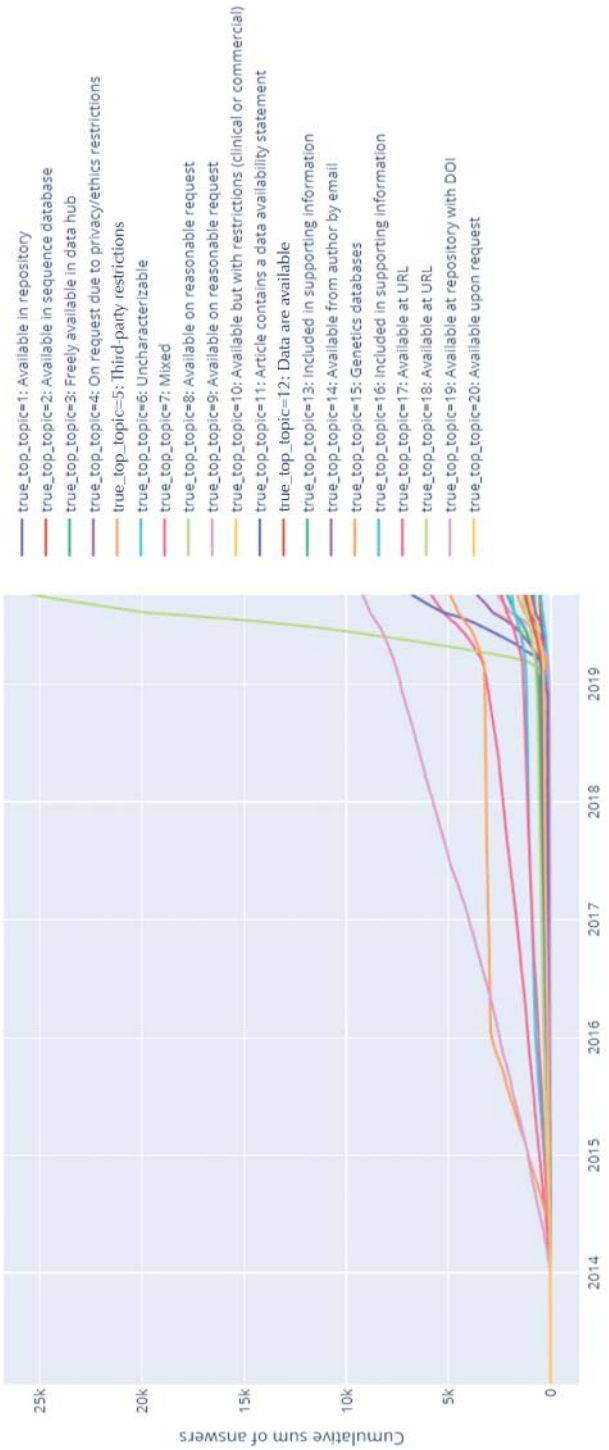

Figure 6. Growth of individual topics over time. Interactive plots available at DOI: 10.22541/au.157253515.58528497.

5. DISCUSSION AND CONCLUSIONS

We see growth across all the topics we used to categorize data availability statements submitted to 176 journals between 2013 and 2019 (Figure 1).

Figure 3 and Figure 4 illustrate how our methods delivered results more than offering analysis; respectively, they show overall percentage of topics identified, and the relationship of topics to individual documents analyzed. Figure 5 shows relationship of topics to words used by authors, and offers readers a simple visual validation for our methods and results; for example, the intense yellow area above Topic 8 "Available on reasonable request" indicates strong relationships with the words: reasonable, request, corresponding author, finding, study, available, and datum. Figure 6 offers trends in topics over time, and these are discussed in more detail below.

We see a particularly sharp increase in growth in early 2019 after launching the Expects Data policy which, for journals that adopt it, requires a data availability statement in every article [31]. Implementing this requirement for data availability statements correlates with many more submitted data availability statements (as is to be expected of a successful implementation). It also correlates with many more declarations made by researchers that data are available on request (for example, Topic 8 in Figure 6, and the related Topics 4, 9, 10, 12, 14, and 20). This is an improvement over the absence of any statement about data. It seems reasonable to anticipate that as researchers become familiar with, interested in, able to, and required to share research data the high proportion of data availability statements categorized as Topic 8 (and related Topics listed above) will gradually be replaced by data availability statements that describe shared data (like Topics 1, 2, 3, 15, 17, 18, and 19).

For data that have been shared, Topic 19 is a good standard to aspire to. It indicates that data are shared in a repository with a permanent digital object identifier (DOI). The number of data availability statements categorized as Topic 19 shows steady growth over six years. This is reassuring, but Topic 19 does not show the sharp increase in 2019 that we might expect to correlate with launch of our Expects Data policy. Several related topics that also describe data having been shared online (Topics 1, 2, 3, 15, 17, and 18) do show the expected sharp increase in early 2019. For Topic 19, consistent growth may be real and based on author behavior, or it may be an artefact of the analysis that we could investigate in future work.

Data that are available in genetics databases, per Topic 15, also show an interesting trend: steeper growth between 2014 and 2016; a distinct flat period between 2016 and 2018; and then steep growth in 2019, correlating with launch of our Expects Data policy. This could be an area for future analysis. It is also interesting to note continuing presence and moderate growth in Topics 13 and 16, which indicate that data have been shared in journal supporting information.

To conclude, if our goal is simply to enable research authors to describe in their journal articles whether or not they have shared the new data they have created then this can be achieved using a policy that requires data availability statements. If our goal is to increase data sharing, then launching a policy and studying the data collected from it may also be valuable: it creates insights into how to enable and support

better experiences for researchers, more data sharing, and higher-quality data sharing. For example, data from this study could help identify which kinds of articles without shared data are similar to those with shared data, and to which journals both are submitted. With that information we could design and launch supportive policies and services where they are more likely to be welcomed by researchers, and therefore where they are most likely to have a positive impact.

It is ironic to write an article about data availability while not sharing the data set. The data availability statements we analyzed were submitted by researchers to Wiley as part of journal articles, some of which we published. We analyzed this information to improve our understanding of researchers' practices, and to improve our products and services. That kind of use is covered by both our privacy policy and the license researchers give us to publish their work. We did not ask those researchers when they submitted their articles whether we could share data about their data availability statements, and for this reason we chose not to share the data set. With that in mind, the final and perhaps most important lesson for us from this study is an appreciation for the value of careful study designs and data management plans [41], created before starting a study.

AUTHOR CONTRIBUTIONS

All authors, C. Graf (cgraf@wiley.com), D. Flanagan (dflanaga@wiley.com), L. Wylie (lwylie@wiley.com) and D. Silver (dsilver@wiley.com) made substantial contributions to the design of this paper. D. Flanagan and L. Wylie contributed to the methodology part of this paper and all authors participated in the investigation and data collection. C. Graf and D. Flanagan wrote the first draft. All authors approved the version to be published and are accountable for the paper.

ACKNOWLEDGEMENTS

Thanks to Elisha Morris at Wiley for the literature search and analysis we used to write our introduction. Thanks to Yan Wu at Wiley for insights into data sharing requirements in China. Thanks to Gary Spencer at Wiley for useful discussions about author behavior and manuscript submission processes. Thanks to Alex Moscrop at Wiley for providing our data. Written collaboratively and preprinted using Authorea; thanks to Alberto Pepe and the Authorea team.

REFERENCES

[1] M. Hahnel. Global funders who require data archiving as a condition of grants. Available at: https://figshare.com/articles/Global_funders_who_require_data_archiving_as_a_condition_of_grants/1281141/1.

[2] G. Popkin. Data sharing and how it can benefit your scientific career. Nature 569(2019), 445–447. doi: 10.1038/d41586-019-01506-x.

[3] NIH Guide: Final NIH statement on sharing research data. Available at: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.

[4] China open science and open data mandate released. Available at: https://www.enago.com/academy/china-open-science-open-data-manadate-released/.
[5] European Union. EU budget for the future: Horizon Europe. EU funding for research and innovation 2021–2027. doi: 10.2777/101500.
[6] Realizing the potential – Final report of the Open Research Data Task Force. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf.
[7] L. Bezuidenhout & E. Chakauya. Hidden concerns of sharing research data by low/middle-income country scientists. Global Bioethics 29(1)(2018), 39–54. doi: 10.1080/11287462.2018.1441780.
[8] A. Meadows. To share or not to share? That is the (research data) question…. Available at: https://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/.
[9] Two competing visions for research data sharing. Available at: https://scholarlykitchen.sspnet.org/2019/10/14/competing-visions-research-data/.
[10] L. Jones, R. Grant, & I. Hrynaszkiewicz. Implementing publisher policies that inform, support and encourage authors to share data: Two case studies. Insights 32(1)(2019), 11.
[11] R. Grant & I. Hrynaszkiewicz. The impact on authors and editors of introducing Data Availability Statements at Nature journals. International Journal of Digital Curation 13(1)(2018), 195–203. doi: 10.2218/ijdc.v13i1.614.
[12] D. Sholler, K. Ram, C. Boettiger, & D.S. Katz. Enforcing public data archiving policies in academic publishing: A study of ecology journals. Big Data & Society 6(1)(2019): 1–18. doi: 10.1177/2053951719836258.
[13] H.A. Campbell, M.A. Micheli-Campbell, & V. Udyawer. Early career researchers embrace data sharing. Trends in Ecology & Evolution 34(2)(2019), 95–98. doi: 10.1016/j.tree.2018.11.010.
[14] D.B. Taichman, P. Sahni, A. Pinborg, L. Peiperl, C. Laine, A. James, S.-T. Hong, … & J. Backus. Data sharing statements for clinical trials. JAMA 317(24)(2017), 2491–2492. doi:10.1001/jama.2017.6514.
[15] G. Colavizza, I. Hrynaszkiewicz, I. Staden, K. Whitaker, & B. McGillivray. The citation advantage of linking publications to research data. arXiv preprint. arXiv:1907.02565v2, 2019.
[16] N.A. Vasilevsky, J. Minnier, M.A. Haendel, & R.E. Champieux. Reproducible and reusable research: Are journal data sharing policies meeting the mark? PeerJ 5(2017), e3208. doi: 10.7717/peerj.3208.
[17] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, … & B. Mons. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3(2016), Article No. 160018. doi: 10.1038/sdata.2016.18.
[18] H2020 Program, Guidelines on FAIR Data Management in Horizon 2020. Available at: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
[19] Y. Wu, E. Moylan, H. Inman, & C. Graf. Paving the way to open data. Data Intelligence 1(4)(2019), 60–72. doi: 10.1162/dint_a_00021.
[20] I. Hrynaszkiewicz, N. Simons, A. Hussain, & S. Goudie. Developing a research data policy framework for all journals and publishers. Available at: https://figshare.com/articles/Developing_a_research_data_policy_framework_for_all_journals_and_publishers/8223365/1.
[21] C. Graf. Why share and cite my research data? A guide to making open research easier. Available at: https://www.wiley.com/network/researchers/writing-and-conducting-research/why-share-and-cite-my-research-data-a-guide-to-making-open-research-easier.
[22] T.H. Vines, R.L. Andrew, D.G. Bock, M.T. Franklin, K.J. Gilbert, N.C. Kane, J.-S. Moore, … & S.Yeaman. Mandated data archiving greatly improves access to research data. The FASEB Journal 27(2013), 1304–1308. doi: 10.1096/fj.12-218164.
[23] F. Murphy. Belmont Forum data accessibility statement policy and template - Endorsed 18 October 2018. Available at: https://zenodo.org/record/1476871#.XxqA94P_ysA.

[24] Wiley. Wiley Open Science Researcher Survey 2016. Available at: https://figshare.com/articles/dataset/Wiley_Open_Science_Researcher_Survey_2016/4748332.
[25] B. Fecher, S. Friesike, & M. Hebing. What drives academic data sharing? PLOS ONE 10(2015), e0118053. doi: 10.1371/journal.pone.0118053.
[26] L.M. Federer, C.W. Belter, D.J. Joubert, A. Livinski, Y.-L. Lu, L.N. Snyders, & H. Thompson. Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLOS ONE 13(2018), e0194768. doi: 10.1371/journal.pone.0194768.
[27] S. Stall, L. Yarmey, J. Cutcher-Gershenfeld, B. Hanson, K. Lehnert, B. Nosek, M. Parsons, & L. Wyborn. Make scientific data FAIR. Nature 570(2019), 27–29. doi: 10.1038/d41586-019-01720-7.
[28] T.E. Hardwicke, M.B. Mathur, K. MacDonald, G. Nilsonne, G.C. Banks, M.C. Kidwell, A.H. Mohr … & M.C. Frank. Data availability reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. Royal Society Open Science 5(2018), 180448. doi: 10.1098/rsos.180448.
[29] J.D. Wallach, K.W. Boyack, & J.P.A. Ioannidis. Reproducible research practices transparency, and open access data in the biomedical literature, 2015–2017. PLOS Biology 16(2018), e2006930. doi: 10.1371/journal.pbio.2006930.
[30] A. Rowhani-Farid & A.G. Barnett. Has open data arrived at the British Medical Journal (BMJ)? An observational study. BMJ Open 6(10)(2016), e011784. doi: 10.1136/bmjopen-2016-011784.
[31] C. Graf. How and why we're making research data more open. Available at: https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-we-re-making-research-data-more-open.
[32] J. Koster & S. Rahmann. Snakemake—A scalable bioinformatics workflow engine. Bioinformatics 28(2012), 2520–2522. doi:10.1093/bioinformatics/bts480.
[33] spaCy. Industrial-strength natural language processing in Python. Available at: https://spacy.io/.
[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel,… & E. Duchesnay. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12(2011), 2825–2830. Available at: http://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.
[35] K.S. Jones. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 28(1972), 11–21. doi: 10.1108/eb026526.
[36] D.M. Blei, A.Y. Ng, & M.I. Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research 3(2003), 993–1022.
[37] C. Sievert & K. Shirley. LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 2014, pp. 63–70. doi: 10.3115/v1/W14-3110.
[38] Wiley. Data sharing policy. Available at: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html.
[39] J. Chuang, C.D. Manning, & J. Heer. Termite: Visualization techniques for assessing textual topic models. In: Advanced Visual Interfaces International Working Conference (AVI '12), 2012, pp. 1–4.
[40] C. Sievert & K.E. Shirley. LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 2014, pp. 63–70.
[41] DCC. Data management plans. Available at: https://www.dcc.ac.uk/resources/data-management-plans.

AUTHOR BIOGRAPHY

Chris Graf is Director, Research Integrity, in Wiley's Open Research team. His responsibilities include the implementation of policies and tools at Wiley that enable researchers and journals to adopt more open, transparent practices. Chris is past co-chair of the Committee on Publication Ethics (COPE), and a program committee member for the 7th World Conference on Research Integrity.
ORCID: 0000-0002-4699-4333

David Flanagan is Director of Data Science in Wiley's Research division, where his group develops applications of data science and machine learning to address scholarly publishing questions. Previously he was Editor-in-Chief of *Advanced Functional Materials* and general manager of ChemPlanner, Wiley's award-winning organic synthesis prediction tool. He received his PhD in Polymer Science and Engineering from the University of Massachusetts Amherst.
ORCID: 0000-0002-7364-4961

Lisa Wylie is Senior Data Product Manager at Wiley. With a background in chemistry and geology, her role incorporates 15 years of editorial experience with project management and data science skills. She is particularly interested in the application of topic modelling and natural language processing for visualizing and predicting trends in published research.
ORCID: 0000-0002-0148-6087

Deirdre Silver is Executive Vice President (EVP) and General Counsel at Wiley. Deirdre joined Wiley in 2002 as Legal Director of its Higher Education business and subsequently counseled Wiley's Professional/Trade, Talent Solutions and Research businesses prior to being appointed EVP and General Counsel. She is responsible for all aspects of legal support. She is a board member of the Copyright Clearance Center and a member of scientific technical and medical (STM)'s Copyright and Legal Affairs Committee. Deirdre is a graduate of Cornell University (B.A., Government, with Distinction in All Subjects) and New York University Law School (J.D.) and a member of the New York State Bar Association.
ORCID: 0000-0002-8648-8857