

Implementing Data Sharing Policies at De Gruyter

by *Lyndsey Dixon, Agnieszka Bednarczyk-Drag, and Katharina Appelt*

Research as a whole is moving towards a greater openness. This no longer only refers to the articles published; the building blocks on which articles are based, the research data, are increasingly becoming the focus of attention.

When researchers make their research data public they demonstrate the robustness and validity of the research presented in an article through enabling others to reproduce and interrogate their findings and to reuse data for teaching and further research. It builds integrity and openness as part of the fabric of the research. For the researcher there are multiple benefits, including increasing exposure to their research. Indeed, promoting future collaborations and sharing can increase the impact of research, eventually by boosting the number of citations by as much as 50 % [1-5].

By direct citation of primary or secondary data as support for findings, just as literature is cited, the value of the data is recognized, ensuring credit to those who generated the data. Moreover, by collecting information on accessibility of the research data upon submission, publishers provide scientists with a strong handle to claim originality of their work. As of 2019, 22 % of funders are mandating or encouraging data sharing to maximize the impact of the research they fund by encouraging or requiring data sharing [6]. Improved reliability, reduced duplications and costs: sharing data may pave the way for “more open, ethical, and sustainable science” [7-9].

In February 2022 the National Institute of Health (NIH) announced that NIH funded researchers would be mandated to share their data in future [10]. What exactly this looks like in practice is not yet entirely clear.

De Gruyter has been collaborating with researchers, institutes, funders, repositories and other stakeholders in the research data infrastructure to make data sharing the new normal. As a member of the STM Research Data Group “2020 Research Data Year” initiative we are participating in the drive to achieve the harmonization and standardization of the research data ecosystem that is crucial for user-friendly processes and good data management [11-12].

To achieve this, we are following the guidelines of the 6-policies framework developed by the Research Data Alliance (RDA) that provides maximum flexibility in adapting the level of commitment to the requirements of the publication [13-14].

In this way, De Gruyter intends to help authors and journals to comply easily with funder mandates, to



increase the visibility and connectivity of their articles and data, and to improve reader and author service with more consistent links to data.

In brief, the six types of research data policy consist of research data policy features (14 in total), with the type 1 policy being the most relaxed, recommending provision of article data and type 6 the most stringent, mandating data sharing which will be peer reviewed as part of the submission process.

The De Gruyter journal portfolio is comprised of approximately 430 titles of which about 25% are Gold Open Access. Our portfolio ranges across 29 subject areas which is significant to mention as it is very important when talking about data to recognise that data means vastly different things to researchers and Editors depending on the subject. Across the portfolio “data” can range from primary and secondary sources, from pieces of art through to 3D models, from interview transcripts to gene sequences: when we talk about data, it is diverse and disparate and a lot of the time, very disorganised. Chemists have a long history of curating and working with data, but for many other disciplines, this is a relatively new undertaking for many involved, so ensuring that there is shared understanding across all stakeholders is paramount.

In the adoption of policies De Gruyter has been closely collaborating with the Editors of our journals to understand the needs of the research communities in which we work, collecting their feedback and questions, and determining an appropriate policy together, what would be the norm, what would be exceptions, what policy and therefore processes could the journal realistically take on.

When considering the policy to adopt it is

14 journal research data policy features arranged as 6 policy types (tiers)

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	○	○	○	○	○	○
Exceptions to policy	○	○	○	○	●	●
Embargoes	○	○	○	●	●	●
Supplementary materials	○	○	○	●	●	●
Data repositories	○	○	○	●	●	●
Data citation	○	○	○	○	●	●
Data licensing	○	○	○	○	○	○
Researcher/ author support	●	●	●	●	●	●
Data availability statements		○	●	●	●	●
Data formats and standards				○	○	●
Mandatory data sharing (specific data types)				●	●	●
Mandatory data sharing (all papers)				○	●	●
Peer review of data				○	○	●
Data Management Plans (DMPs)				○	○	○

○ Provide information
The text for the policy feature will be included in the policy template but it is clear that the feature will not be enforced and checked as part of the publishing or peer review process

● Provide information and action
The text of the policy feature is included and makes clear where applicable that the feature will be checked and enforced in the publishing or peer-review process

Figure 1: 14 journal research data policy features arranged as 6 policy types (tiers) (reproduced from [14]).

imperative to consider all the necessary resources to fully implement the policy, including the people, the communications, and the technology. Adopting higher tier policies may provide more benefits, enhance reader and user experience, increase the reusability of the data and visibility of the articles, and thus lead to more transparent research, but it also involves more technological needs, operational complexity and costs, and more involvement from the authors, editors, and reviewers.

To successfully implement a policy it needs a joined-up approach across the functions of journal publishing from editorial, to production to digital to communications. Technical steps are needed to be able to ensure that the data sharing policies achieve true sharing and aren't just data dumping.

For example, to ensure that data citations are correctly captured, and the author receives credit for the data that they have generated and shared, the data set needs to have a Digital Object Identifier (DOI) or a Compact Identifier associated with data references. The data citations should include a Permanent ID (such as a DOI) and should ideally include the minimum information recommended by DataCite and the FORCE11 data citation principles. The production department and its vendor(s) must ensure all data citations provided by the author in the reference list are processed appropriately using the correct XML tags. All references need to be delivered to and registered with Crossref. By sending these data citations to Crossref, they become available in a Scholix compliant

Implementing Data Sharing Policies at De Gruyter

way (<http://www.scholix.org/>) to ensure that the researcher is getting recognition for where their data is cited through the ScholixExplorer.

Similarly, the policy must specify what license is applied to research data published in the journal itself. It must also specify that copyright in research data is not transferred to the publisher. The preference is for Open Data conformant licenses (such as Creative Commons Attribution License, CC BY, Creative Commons Public Domain Waiver, CC0) to try to make the data accessible and reusable.

The Data Availability Statements (DAS) are a core component of making sure that the data is findable, so we've endeavoured to make them as useful as possible by providing templated text for use. One of the issues is that the DAS are sometimes included as part of the main body of the article, so these statements must be made findable - within the production process they need to be identified, processed, tagged, and converted to a separate document. This information also needs to appear alongside the metadata of the article in front of any paywall should there be one.

What have we achieved?

Our first learning is that this takes time to implement. In fact, getting up to speed took the best part of eight months to get the policies rolled out and the operational side of things fully- functioning.

So far policies have mainly been introduced for approximate 70 journals, and policy texts have been put on all individual websites. The majority of policies adopted are tiers 1 or 2. We elected not to adopt tiers higher than tier 4 given the requirements on technological and operational developments.

Aside from the process changes needed, the most important element to get right was the communication of the whys, the hows, and the whats. This communication and education is two-way with us—learning from the Editors as well as explaining the benefits of undertaking this effort. Responses were mixed, with some Editors nervous that this is complicating the process for authors, whereas others were adamant that for the good of research we should be adopting the most stringent policies. On the whole, we've erred on the side of caution to start slowly and work up, being mindful of the subject norms and what researchers will be used to across other publishing outlets in their field. The more we can automate the better, but at the moment there is additional workload for the higher policies, particularly on the part of the peer reviewers. The more this practice becomes prevalent, the more

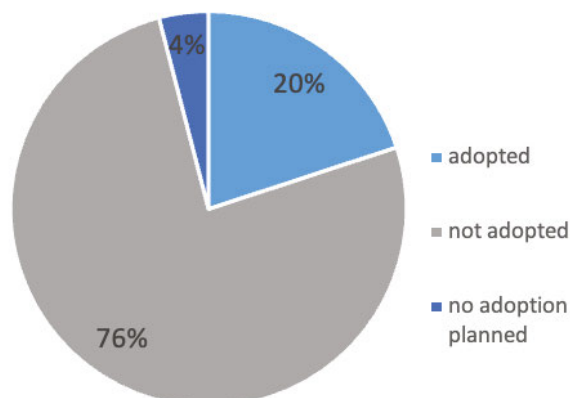


Figure 2: De Gruyter journals which have adopted a data sharing policy

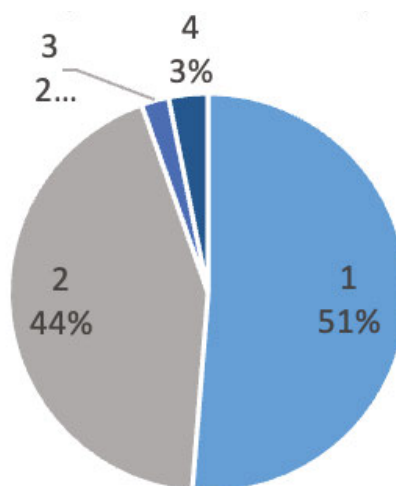



Figure 3: The policy tiers adopted by De Gruyter Journals

normal this will become, but currently it requires education and explanation.

There are a few barriers to a rapid expansion. First, internally, we are creating a central repository of information to provide easy access to the policies to facilitate quicker adoption. Second, there is a lack of suitable automated workflows for capturing Data Availability Statements (DAS) or links to data sets in submission systems. To this end, the STM Association is in talks with a number of system providers, including Editorial Manager and ScholarOne, to provide a potential solution to the problem. Third, the ethics standards, best practices, guidelines, or recommendations for data publications are currently not well established. The Committee on Publication Ethics (COPE) is collaborating with the FORCE11 Research Data Publishing Ethics Working Group to develop

Implementing Data Sharing Policies at De Gruyter

relevant documentation, resources and workflows, and we are monitoring this process and will introduce their recommendations.

At De Gruyter, we continue to gather feedback from this community-driven initiative, working within the frameworks established, with our authors and editors on the relevance, usefulness, and uptake of these policies and once the issues above have been resolved, we will move on to full-scale roll-out across the portfolio. 

References

1. Piwowar HA, Day RS, Fridsma DD. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2007; 2:e308; <https://doi.org/10.1371/journal.pone.0000308>
2. Dorch B, Drachen T, Ellegaard O. The data sharing advantage in astrophysics. *Proc Int Astron Union* 2015; 11:172-175; <https://doi.org/10.1017/S1743921316002696>.
3. Drachen TM, Ellegaard O, Larsen AV, Dorch SBF. Sharing data increases citations. *LIBER Quart* 2016;26:67-82; <https://doi.org/10.18352/lq.10149>
4. Christensen G, Allan D, Edward M, Moore DA, Rose AK. A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLoS One* 2019; 14:e0225883; <https://doi.org/10.1371/journal.pone.0225883>
5. Colavizza G, Hrynaskiewicz I, Staden I, Whitaker K, McGillivray B. The citation advantage of linking publications to research data. *Arxiv* 2019; <https://arxiv.org/pdf/1907.02565.pdf>
6. Lucraft M, Baynes G, Allin K, Hrynaskiewicz I, Khodiyar V. Five Essential Factors for Data Sharing. *figshare. Journal contribution*. 2019; <https://doi.org/10.6084/m9.figshare.7807949.v1>
7. Ioannidis J, Allison D, Ball C, Coulibaly I, Cui X, Culhane AC et al. Repeatability of published microarray gene expression analyses. *Nat Genet* 2009; 41:149-155; <https://doi.org/10.1038/ng.295>.
8. Hardwicke TE, Mathur M, MacDonald K, Nilsson G, Banks GC, Kidwell MC et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. *R Soc Open Sci* 2018; 5:180448; <https://doi.org/10.1098/rsos.180448>
9. Figueiredo AS. Convert Challenges into Opportunities. *Front Public Health* 2017; 5; <https://doi.org/10.3389/fpubh.2017.00327>
10. Kozlov M, NIH issues a seismic mandate: share data publicly. *Nature* 602, 558-559 (2022) <https://doi.org/10.1038/d41586-022-00402-1>
11. Naughton L, Kernohan D. Making sense of journal research data policies. *Insights* 2016; 29:84-89; <http://doi.org/10.1629/uksg.284>
12. <https://www.stm-researchdata.org/>

Data-Sharing Seminar Series for Societies.

Societies have a unique role in bringing awareness of developing practices and supporting the necessary discussions within disciplines to bring their voice to the larger community. In 2021, IUPAC co-hosted monthly seminars that brought speakers discuss topics specific to society engagement to help with data sharing, credit, transparency and more.

Recordings are resources are available at <https://wesharedata.org/>

13. <https://www.rd-alliance.org/>

14. Hrynaskiewicz I, Simons N, Hussain A, Goudie S. Developing a research data policy framework for all journals and publishers. *Data Sci J* 2020; 19:5; <http://doi.org/10.5334/dsj-2020-005>

Further background reading:

The Coalition for Publishing Data in the Earth and Space Sciences commitment statement <http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/>

STM Research Data program:

<https://www.stm-researchdata.org/>

Data Citation : <https://www.force11.org/datacitationprinciples>

Crossref : https://www.crossref.org/community/linking_data/

Scholix: www.scholix.org

Wilkinson M, Dumontier M, Aalbersberg I, Appleton G,

Axton M, Baak A et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016; 3:160018; <https://doi.org/10.1038/sdata.2016.18>

Facts and Figures for open research data [https://](https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en)

ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor/facts-and-figures-open-research-data_en

For more information, contact at De Gruyter: Agnieszka Bednarczyk-Drag (Agnieszka.Bednarczyk-Drag@degruyter.com) or Katharina Appelt (Katharina.Appelt@degruyter.com). Lyndsey Dixon (Lyndsey.dixon@degruyter.com) (ORCID <https://orcid.org/0000-0002-4747-5295>) Headquartered in Berlin, Germany, De Gruyter is an international, independent academic publisher. Operating for over 270 years, the company publishes more than 1,300 new book titles each year and over 430 journals in the humanities, the social sciences, medicine, mathematics, engineering, computer sciences, natural sciences, and law. The company also offers a wide range of digital media, including open access journals and books. www.degruyter.com