

This manuscript is a preprint. A preprint is an early version of an academic article that has been made available by the author for others to read for free online before it has been peer reviewed or published by an academic journal.

The manuscript was peer-reviewed and published in October 2020 in Data Science Journal: <https://doi.org/10.5334/dsj-2020-037>

Earth science and biodiversity journals can improve support for data publication

Andreas Hübner
Library and Information Services

GFZ German Research Centre for Geosciences
Potsdam, Germany
huebner@gfz-potsdam.de

 ORCID ID: <https://orcid.org/0000-0001-7342-9789>

Contents

Abstract	1
Introduction.....	2
Results	3
Information on data publishing.....	3
Content of information on data publishing.....	3
Discussion	5
Quality of journals' information on data publishing	5
FAIR and open data publication	6
Recommendations	6
Progress.....	7
References.....	7

Abstract

This study reviews research data policies and author instructions of 31 journals from the Earth sciences and from biodiversity that are published by German learned societies or research institutions. The statements on data publishing of the journal's data policies / author guidelines were matched to 14 defined features of journal research data policies. A brief discussion on quality of data policies is presented to raise awareness of German learned societies / research institutions and to guide them towards improved data policies of their journals.

Introduction

Statements on data publishing in journals' data policies and/or author guidelines recommend or mandate to make data related to an article available. Alongside to data publishing technical infrastructure and non-technical aspects like scientists' positive personal attitudes towards data publishing (Stieglitz et al., 2020), these statements are important drivers to increase data publications for the purpose of scientific transparency and data reuse. Learned societies and research institutions frequently publish scientific journals, and being the "owners" of the journals, they have control (or, being co-publishers, at least some influence) on the existence and the contents of statements on data publishing of the journals.

Traditionally, journal articles pass on information on the underlying data in chapters on data collection procedures and analysis techniques as well as description of results. Further, results are often reported in diagrams and data tables embedded in the text document. Most journals allow for data tables too large to be included in the text document and other research products, such as movies, audiofiles, or detailed information related to the research described in the article, to be published as "Supplemental electronic material" along with the journal article on the publishers' website.

However, modern data publication standards have evolved much beyond the practices described above: they call for data (including software, models etc) to be routinely shared in ways that allow easy discovery, recombination, and reuse, and information about samples, methods, and tools should be standardized, available, and linked across publications. Current best practice for data publication has been described for several research domains, including the Earth, space, and environmental sciences (Enabling FAIR Data Community, 2018a, 2018b) and biodiversity research (Penev et al. 2017). A common request in these recommendations is that all core research outputs should be directed to FAIR-aligned repositories. This stipulates that supplements on journal publishers' websites will no longer be used as the primary archive for data.

To adopt best practices in data publishing and guidance for authors, journals should have comprehensive, clear and up-to-date information in place that describe the general approach to data publishing as well as the details of what is expected from authors in terms of provision of research data. This should be stated in research data policies that are related to (or part of) the author instructions.

Research data policies and author instructions of 31 journals from the Earth sciences and from biodiversity that are published by German learned societies or research institutions are reviewed in this study. The focus is on journals that are publishing mainly original research articles. Since data publication is also relevant to Phd theses and "Habilitationsschriften", the "SDGG Schriftenreihe der Deutschen Gesellschaft für Geowissenschaften" is also included. Improved information on data publishing will increase the attractiveness of the journals to authors which are increasingly willing to, as well as mandated to, publish research data that is related to journal articles.

This survey identified whether journals had an explicit data policy in place and/or addressed the topic of data publishing in the author instructions. The statements on data publishing of the journals' data policies / author guidelines (status May 2020) were then matched to the 14 features of journal research data policies as defined by Hrynaszkiewicz et al. (2020) (Data table: Hübner, 2020).

Results

Information on data publishing

Information on the journals of this study as well as the availability of information on data publishing is presented in Tab. A [located at the end of this document].

12 out of 31 surveyed journals don't address data publishing at all: they don't have a data policy and neither address data publishing in the author instructions nor on the society/institution (journal owners) journal website.

19 journals address data publishing, although in very variable ways. Some journals provide very little information on data publication, the only offer the possibility to add "supplemental material" to the journal article, to be published along with the article on the publisher's website. Other journals offer more specific information on the conditions of supplemental material publication. Of the 10 journals that have an explicit data policy, the policy may be an individual document/website or the policy is integrated into the author instructions. Only two societies, Deutsche Geophysikalische Gesellschaft DGG and Deutsche Mineralogische Gesellschaft DMG, do address data publishing on their respective websites (Tab. A [located at the end of this document]).

Content of information on data publishing

The statements on data publishing of the journal's data policies / author guidelines were matched to the 14 features of journal research data policies as defined by Hrynaszkiewicz et al. (2020). 19 out of 31 journals address data publishing either in their author guidelines or in a data policy. The only one feature that is addressed by all 19 journals is feature #4 "Supplementary materials" (Tab. 1). For five journals, this is the only feature that is addressed (Tab. 2). Please note that merely just mentioning the possibility of submission of supplemental material, without any further information, already qualified a journal to be included in the group of journals that address data publishing. Two journals address two features: feature #4 plus feature #10 "data formats and standards", and two journals combine three features: feature #4 plus features # 5 "Data repositories" and feature #9 "Data availability statements". Other journals combine more/other features in various ways. Three features are not addressed in any of the journals' texts about data publishing: #3 "Embargoes", #13 "Peer review of data", and #14 "Data Management Plans".

Table 1: Features of data policies as defined by Hrynaszkiewicz et al. (2020) and the number of journals that address this feature in data policies / author instructions.

Feature (#)	Number of journals that address this feature
Supplementary materials (4)	19
Data repositories (5)	12
Data citation (6)	10
Data availability statements (9)	9
Data formats and standards (10)	8
Definition of research data (1)	7
Data licensing (7)	5
Researcher/author support (8)	3
Definition of exceptions (2)	2
Mandatory data sharing (all papers) (12)	2
Mandatory data sharing (specific papers) (11)	1
Embargoes (3)	0
Peer review of data (13)	0
Data Management Plans (14)	0

Addressing data publishing: formats

10 journals have an explicit “data policy”. This policy may be a separate text in form of an individual document or individual website, but in most cases, it is a sub-chapter of the author guidelines. In the latter case information on data publishing is present in different parts of the author instructions, not exclusively in the chapter “data policy”, making it difficult for authors to get the full picture on data publishing policy of the respective journal.

Table 2: Number of features (Hrynaszkiewicz et al., 2020) that are addressed in data policies / author instructions of the surveyed journals. Journals ordered according to score.

Journal name	Number of features addressed
Journal of Photogrammetry, Remote Sensing and Geoinformation Science PFG	9
Marine Biodiversity	8
Palaeobiodiversity and Palaeoenvironments	8
Deutsche Entomologische Zeitschrift	7
Zoosystematics and Evolution	7
Geophysical Journal International GJI	5
European Journal of Mineralogy EJM	5
E&G Quaternary Science Journal	5
Fossil Record	5
Journal of Applied Entomology	4
International Journal of Earth Sciences (Geologische Rundschau) IJES	3
Paläontologische Zeitschrift PalZ	3
Standort	2
Arthropod Systematics & Phylogeny	2
Journal of Plant Nutrition and Soil Science	1
Contributions to Entomology – Beiträge zur Entomologie	1
Elements	1
Archiv für Molluskenkunde	1
Soil Organisms	1

Discussion

About half (9 out of 19) of the journals from earth sciences in this study don't address data publishing at all. This share is higher than the results of Malicki (2019), where only one third (5 out of 15) journals from the sub-group of "Earth and Planetary Sciences" journals don't address data sharing, including accepting data(sets) as supplementary materials in journals' instructions to authors. For the whole dataset of Malicki (2019) (= 835 scientific journals from all disciplines), 60% of journals don't mention data sharing in journals' instructions to authors. The differences for earth science journals in this and Malickis (2019) study indicates that especially earth science journals owned by German societies / institutions are lagging behind in addressing data publishing.

Quality of journals' information on data publishing

The number of addressed features in information on data publishing in data policies / author instructions indicates quality in a way that it reflects on completeness or comprehensiveness of the information, with journals in the upper ranges of Table 3 as positive examples. The specific effects of individual features in statements on data publication is unkown for many of the features. Clearly, mandating data publication is a strong stance, but it could have the adverse effect that authors will refrain from publishing in that journal because they don't want, for whatever reasons, to publish the underlaying data. The role of data availability statements in citation advantages have been investigated, showing that articles that include data availability statements that link to data in a repository have an up to about 25% higher citation impact on average (Colavizza et al., 2020). More in-depth studies of this kind are necessary to evalutate the (positive) effects of individual subjects in journal statements on data publishing.

Additional to completeness, *clarity* and *up-to-dateness* are further and most important parameters that are essential to overall quality of author information on data publishing.

Whatever the policy of a journal is, it should phrase expectations to authors with clarity, in an understandable, precise and concrete manner. However, in some of the surveyed journals' texts, ambiguous and inconsistent statements were encountered, making it hard for authors to identify the expectations of the journal on data publishing. For example, the statement "Where a widely established research community expectation for data archiving in public repositories exists, submission to a community-endorsed, public repository is mandatory." unfortunately leaves it to the authors to judge if they may be part of a research community where these expectations are "widely established" and thus, if this mandate applies to them or not.

Because publication of supplementary information is most widely offered, clarity is needed about the demarcation of supplementary information (published along with the article on publishers' websites) and data that is to be published elsewhere (preferably in a FAIR-aligned repository). Authors need information about what is still deemed as supplementary information and what (and why) data should be directed to repositories. This also underpins the necessity of defining what is understood by research data. The advantages of repository deposition above supplementals on publishers' websites as well as the current cultural change towards repository publication is illustrated by Kwon (2019).

Information on data publishing should be up-to-date and reflect modern best practice and current guidelines, especially with the FAIR principles and its evolving specifications (eg. see Hrynaszkiewicz et al. (2020), Enabling FAIR Data Community (2018a, 2018b), Davidson et al. (2019)). Like with clarity, a wide spectrum of up-to-dateness was encountered in data policies / author instructions in this

survey. For example, recommendations for data table file formats as standard MS office or PDF format could be updated to meet today's requirements towards non-proprietary formats and machine readability of data files.

Examples of high-quality-information on data publishing with respect to completeness, clarity, and up-to-dateness in this survey are the data policies / author instructions of journals published by Copernicus Publications and Pensoft. Both refer to recently published community guidelines: Copernicus Publications supports (and is a signatory of) the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) commitment statement and the Enabling FAIR Data Commitment Statement in the Earth, Space, and Environmental Sciences (Enabling FAIR Data Community, 2018a), and the two journals of this survey that are published by Pensoft promote recent, community-endorsed guidelines for scholarly publishing of biodiversity data (Penev et al. 2017).

FAIR and open data publication

The features defined by Hrynaszkiewicz et al. (2020) do support FAIR data publishing, however, other initiatives provide further features to be considered. The Author Guidelines for scientific data (Enabling FAIR Data Community, 2018b) describe four features that are intended to align the journal author instructions for the submission of data in the Earth, space, and environmental sciences, for all affiliated publishers. Three of the features match with features defined by Hrynaszkiewicz et al. (2020). The fourth one, the call for provision of "unrestricted access to all data and materials underlying reported findings for which ethical or legal constraints do not apply" has no clear-cut equivalent to the features of Hrynaszkiewicz et al. (2020). It can be interpreted to call for full open access of the data. None of the evaluated journals in this study have such a strong statement on unrestricted or open access in their data policy or author guidelines.

The FAIR Data Policy Landscape Analysis (Davidson et al., 2019) identified elements that support or hinder FAIR data practice in various national, funder, publisher, and institutional data policies. The analysis includes, to a small extent, also journal policies. Additional to the features identified by Hrynaszkiewicz et al. (2020) and the resources of the Enabling FAIR data project, the analysis questioned data policies about statements on the monitoring of policy compliance. Additionally, the importance that policies can be interpreted unambiguously by not only humans but also machines is emphasized. To be machine-readable and actionable, a structured data markup schema may be used. Furthermore, policies should be versioned, indexed and semantically annotated in a policy registry, such as the FAIRsharing registry¹, a curated service of interlinked standards, repositories and data policies. FAIRsharing registry supports machine-actionability by asking contributors to make clear the status of the resource being described in terms of effective machine actionability.

Recommendations

Funding bodies, infrastructure providers and publishers are key stakeholders in providing a consistent and easy-to-use environment for FAIR data publishing. Learned societies as well as research institutions that own and publish journals (either self-publishing or with support of a professional publisher) are an important part of that publisher stakeholder group. For the learned societies as well as research institutions it is recommended to review their attitude towards data publishing. For the journals that have no data policy in place, the owners of the journal should

¹ <https://fairsharing.org/>

strongly consider to adopt a data policy. All other owners should revisit existing data policies in terms of completeness, clarity and up-to-dateness, considering the features described and issues raised in for example Hrynaszkiewicz et al. (2020), Author Guidelines for scientific data and the Commitment statement (Enabling FAIR Data Community, 2018a, 2018b) as well as in the FAIR Data Policy Landscape Analysis (Davidson et al., 2019).

Also, geoscience learned societies should consider signing the “Commitment statement in the Earth, space, and environmental sciences” to complement their data policy efforts with other actions to foster FAIR and open data publishing and to show public support for this community-driven initiative. FID GEO², the specialized information service for geosciences, is ready to guide and support geoscience learned societies and research institutions in developing data policies for their journals.

Progress

This survey was conducted in May 2020 and was sent thereafter to the journal owners for consideration. Several owners responded, emphasizing the importance of the topic, announcing future changes to current policies / author guidelines (planned or nearly finalized) and/or asking for support for drafting new or amending existing policies / author guidelines. As to the knowledge of the author of this study, at the end of August 2020, the editorial office and the editing committee of the journal Hydrologie und Wasserbewirtschaftung, HyWa have formulated and published a data policy, and the editors of the journal Archiv für Molluskenkunde finalised an update of the Instructions for Authors.

References

- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., and McGillivray, B. (2020) The citation advantage of linking publications to research data. PLoS ONE 15(4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>
- Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., and Whyte, A. (2019) D3.1 FAIR Policy Landscape Analysis (Version v1.0_draft). FAIRsFAIR. <https://doi.org/10.5281/zenodo.3558173>
- Enabling FAIR Data Community (2018a) Commitment statement to enabling FAIR data in the earth, space, and environmental sciences. Zenodo 8 October 2018.
<https://doi.org/10.5281/zenodo.1451971>
- Enabling FAIR Data Community, de Waard, A., Cousijn, H., Heber, J., Hanson, B., Bradford, M., ... Yeston, J. (2018b) Author Guidelines for Enabling FAIR Data in the Earth, Space, and Environmental Science (Version 2). Zenodo. <http://doi.org/10.5281/zenodo.1447108>
- Hübner, A. (2020) Features of research data policies of Earth science and biodiversity academic society journals [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4009195>
- Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R. and Goudie, S. (2020) Developing a Research Data Policy Framework for All Journals and Publishers. Data Science Journal, 19: 5, pp. 1–15.
<https://doi.org/10.5334/dsj-2020-005>

² <https://www.fidgeo.de/en/research-data/>

Kwon, D. (2019) The Push to Replace Journal Supplements with Repositories. *The Scientist*, <https://www.the-scientist.com/news-opinion/the-push-to-replace-journal-supplements-with-repositories--66296>

Malički, M. (2019) “Project: Fostering Transparent and Responsible Conduct of Research: What can Journals do?”, Mendeley Data, v3 <http://dx.doi.org/10.17632/53cskwpdn.3>

Malički, M., Aalbersberg, I.J., Bouter, L., and ter Riet, G. (2019) Journals’ instructions to authors: A cross-sectional study across scientific disciplines. *PLoS ONE* 14(9): e0222157. <https://doi.org/10.1371/journal.pone.0222157>

Penev, L., Mietchen, D., Chavan, V., Hagedorn, G., Smith, V., Shotton, D., Ó Tuama, É., Senderov, V., Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017) Strategies and guidelines for scholarly publishing of biodiversity data. *Research Ideas and Outcomes* 3: e12431. <https://doi.org/10.3897/rio.3.e12431>

Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., Lopez, A., et al. (2020) When are researchers willing to share their data? – Impacts of values and uncertainty on open data in academia. *PLoS ONE* 15(7): e0234172. <https://doi.org/10.1371/journal.pone.0234172>

Table A: Details of the journals in this survey.

Society/Institution ("owner")	Name of the journal	Publisher ("service provider")	Data Policy available	Information on data publishing in Instructions for authors or equivalent	Information on data publishing on website of society/institution ("owner")
Bundesanstalt für Gewässerkunde on behalf of the German federal and state administrations	Hydrologie und Wasserbewirtschaftung, HyWa	Self-published	no	no	no
Deutsche Bodenkundliche Gesellschaft DBG	Journal of Plant Nutrition and Soil Science	John Wiley & Sons	no	yes	no
Deutsche Geologische Gesellschaft – Geologische Vereinigung DGGV	ZDGG Zeitschrift der Deutschen Gesellschaft für Geowissenschaften / Journal of Applied and Regional Geology	Schweizerbart Science Publishers	no	no	no
Deutsche Geologische Gesellschaft – Geologische Vereinigung DGGV	IJES International Journal of Earth Sciences (Geologische Rundschau)	Springer	no	yes	no
Deutsche Geologische Gesellschaft – Geologische Vereinigung DGGV	SDGG Schriftenreihe der Deutschen Gesellschaft für Geowissenschaften	Self-published	no	no	no
Deutsche Geophysikalische Gesellschaft DGG	Geophysical Journal International	Oxford Academic	yes	yes	yes
Deutsche Gesellschaft für allgemeine und angewandte Entomologie DGaaE	Contributions to Entomology – Beiträge zur Entomologie	Self-published (OJS, Goethe University Frankfurt)	no	yes	no
Deutsche Gesellschaft für allgemeine und angewandte Entomologie DGaaE	Journal of Applied Entomology	Wiley Blackwell	yes	yes	no
Deutsche Gesellschaft für Moor- und Torfkunde DGMT	TELMA	Self-published	no	no	no

Table A (continued): Details of the journals in this survey.

Deutsche Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation	PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science	Springer	yes	yes	no
Deutsche Meteorologische Gesellschaft DMG	Meteorologische Zeitschrift (Contributions to Atmospheric Sciences)	Schweizerbart Science Publishers	no	no	no
Deutsche Mineralogische Gesellschaft DMG	European Journal of Mineralogy (EJM)	Copernicus Publications	yes	yes	yes
Deutsche Mineralogische Gesellschaft DMG	Elements	Mineralogical Society of America, jointly with 17 societies	no	yes	no
Deutsche Quartärvereinigung DEUQUA	E&G Quaternary Science Journal	Copernicus Publications	yes	yes	no
Deutschen Malakozoologischen Gesellschaft DMG	Archiv für Molluskenkunde	Schweizerbart Science Publishers	no	yes	no
Deutscher Verband für Angewandte Geographie DVAG	Standort	Springer	no	yes	no
Fachsektion Hydrogeologie in der DGGV FH-DGGV	Grundwasser	Springer	no	no	no
Gesellschaft der Metallurgen und Bergleute GDMB	World of Metallurgy – ERZMETALL	GDMB Verlag	no	no	no
Gesellschaft für Erdkunde zu Berlin	DIE ERDE	Self-published (OJS, FU Berlin)	no	no	no
Gesellschaft für Geodäsie, Geoinformation und Landmanagement DVW	zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement	Wißner-Verlag	no	no	no
Museum für Naturkunde Berlin	Fossil Record	Copernicus Publications	yes	yes	no

Table A (continued): Details of the journals in this survey.

Museum für Naturkunde Berlin	Deutsche Entomologische Zeitschrift	Pensoft	yes	yes	no
Museum für Naturkunde Berlin	Zoosystematics and Evolution	Pensoft	yes	yes	no
Paläontologische Gesellschaft PalGes	PalZ Paläontologische Zeitschrift	Springer	no	yes	no
Senckenberg Gesellschaft für Naturforschung	Arthropod Systematics & Phylogeny	Self-published	no	yes	no
Senckenberg Gesellschaft für Naturforschung	Marine Biodiversity	Springer	yes	yes	no
Senckenberg Gesellschaft für Naturforschung	Palaeobiodiversity and Palaeoenvironments	Springer	yes	yes	no
Senckenberg Gesellschaft für Naturforschung	Soil Organisms	Self-published	no	yes	no
Senckenberg Gesellschaft für Naturforschung	Studia dipterologica	Self-published	no	no	no
Senckenberg Gesellschaft für Naturforschung	Vertebrate Zoology	Self-published	no	no	no
Senckenberg Museum für Naturkunde Görlitz	PECKIANA	Self-published	no	no	no