Guidelines for journals that wish to establish a "data policy" related to their publications

March 2021 (French original version), June 2021 (English version)

French Committee for Open Science (COSO)
Research Data College

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
Liberté
Égalité
Fraternité

OUVRIR
LA SCIENCE !

Guidelines for journals that wish to establish a "data policy" related to their publications
Recommandations aux revues souhaitant définir une « politique de données » liées aux publications

This document is designed for journals and editorial boards that wish to establish a data policy. A data policy defines what the journal expects from its authors in terms of managing and sharing the data related to its publications.

This document is intended in particular for editors of journals in the humanities and social sciences, as they have been relatively less active in this area than their counterparts in science, technology and medicine. However, it can be useful to all editors, regardless of the disciplinary scope of their journal.

Data policies differ depending on the nature of the incentives and requirements they provide, in particular:
- Do they encourage or require that all or part of the data underlying the publications be made available?
- Are there specific conditions concerning the availability of the data: deadline, format, licenses...?
- Are the data submitted to a peer review process as are the publications?

In order to progressively set up their data policy, journals can refer to existing typologies (e.g. RDA offers 6 types of data policies, Springer defines 4).

Research data include all "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results" [1]

This document is organised into 7 sections and 4 columns:
- The 1st column contains the name of the section.
- The 2nd column describes the section being presented.
- The 3rd column specifies the issues of the section and what questions the journals should address.
- The 4th column provides examples of wordings that are given **as guidance**.

This document was produced by the Research Data College of the French Committee for Open Science. It is distributed under a Creative Commons CC-BY license. It is based on the following references (amongst others):
- Iain Hrynaszkiewicz, Natasha Simons, Azhar Hussain, Rebecca Grant, Simon Goudie. "Developing a Research Data Policy Framework for All Journals and Publishers." *Data Science Journal*, 19 (1). 2020. DOI: https://doi.org/10.5334/dsj-2020-005;
- and its French adaptation by the University of Toulouse-Jean Jaurès: Chloée Fabre, Françoise Gouzi. *Proposition de modèle de politique pour les revues et éditeurs quant aux données de la recherche*. 2020. ⟨hal-03026731⟩.

---

[1] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information http://data.europa.eu/eli/dir/2019/1024/oj

[2] Normandie Université, Université de Lille

[3] Université Toulouse – Jean Jaurès

[4] Open Edition Center (UAR 2004), CNRS, Université d'Aix-Marseille, EHESS, Université d'Avignon

[5] Huma-Num (UAR 3598), CNRS, Université d'Aix-Marseille, Campus Condorcet

[6] Huma-Num (UAR 3598), CNRS, Université d'Aix-Marseille, Campus Condorcet

[7] Maison des sciences de l'homme de Dijon (UAR 3516 CNRS, uB, UBFC)

Section	Description	Issues and questions to be addressed	Examples of wordings
1. Definition of Research Data and exceptions	Describes which data the policy applies to. Specifies any exceptions to this policy.	**Issues** ● Allow authors to select the data affected by the journal policy. ● Allow authors to curate the data in order to make them available in connection with the publication. **Questions to consider** ● Which data are affected by the policy? ● Should the affected data be the raw data or the processed data that underpinned the results presented in the publication? ● Do the data contain sensitive content that falls within the scope of the General Data Protection Regulation (GDPR)? ● If so, what processing should the data undergo to comply with the GDPR? ● What data are exceptions to the policy?	*This policy applies to research data that would be necessary to check the results presented in the publications of the journal. Research data include data produced by the authors as well as data from other sources that are analysed by the authors in their study. These data can be presented in various forms: images, videos, statistical tables…* *Research data that are not necessary to check the results reported in publications are not covered by this policy.* *This policy will be limited by the legitimate exceptions regulated by law, for example with regard to professional confidentiality, industrial and commercial secrets, personal data or content protected by copyright.*

| 2. Data (and metadata) standards and formats | Lists the main standards (and/or resources to find them) used for data and associated metadata.

Necessarily includes the dissemination protocols mainly associated with the metadata. | **Issues**
● To be able to durably find, read and interpret data associated with publications.

● Recommend the use of open and standardized file formats

● Alert authors on the importance and utility of using standards to structure data and metadata.

● Build on existing national and international initiatives, such as the RDA (Research Data Alliance) working groups: https://rd-alliance.org/groups/.

Questions to consider
● Check the existence of standards (e.g. structuring of metadata, vocabularies, file formats…) used in the disciplines covered by the journal. The metadata used to describe the dataset during its dissemination are dependent on the choice of the data repository (see section 3).

● Published datasets must at the very least be described by the mandatory metadata from the schema published by DataCite namely:
   - Creator of the Dataset,
   - Title of the dataset,
   - Publisher or host of the data,
   - Year of publication,
   - Identifier and its type (DOI, handle...),
   - Type of resource. | *The journal encourages authors to use open and standard formats. For example, the compliance of data file formats with CINES recommendations for long term preservation can be checked at: https://facile.cines.fr (in French)*

Descriptive metadata must be structured using recognized standards, at the very least Dublin Core. The standards can be either disciplinary or more generic. (https://en.wikipedia.org/wiki/Dublin_Core).

The use of "controlled" (or reference) vocabularies, either disciplinary or more generic ones,, to express these metadata is strongly recommended (e.g., to reference an author https://orcid.org; to reference a place https://www.geonames.org). |

| 3. Data access and hosting | Indicates how the data should be hosted to ensure that access is secured and guaranteed for the longest possible time.

Specifies whether a specific repository is recommended and, if so, its characteristics (e.g. certification, degree of compliance with FAIR principles, relevance to the discipline...) | **Issues**
● Ensure the preservation, visibility and access to data by depositing them in a repository

● Facilitate their sharing and reuse, and provide the elements to build scientific evidence.

● Recommend the use of a repository that will guarantee the security of the data and their long term accessibility.

● Recommend the use of a disciplinary repository adapted to the journal.

● Specify the criteria for the choice of a repository, for example, avoid using a private repository.

● Specify that FAIR principles must be implemented when data are made available

Questions to consider
● Does the journal wish to suggest a specific repository or leave the choice of repository open?

● Does the journal cover a disciplinary area for which a specific repository can be recommended?

● Do authors need to use a CoreTrustSeal certified repository (see certification criteria and list of certified repositories on the CoreTrustSeal website).

● If the repository is not certified, what other selection criteria does the journal recommend?
  - Community recognition
  - Assignment of permanent identifiers (DOI, handle...).
  - Distribution license.
  - Hosting location.
  - Long-term preservation.
  - Public status.
  - Accepted file formats.
  - Link between data and publication. | *The data that contributed to the writing of the publication must be deposited in a data repository that will guarantee secure storage and access to the data, in particular through the attribution of a permanent identifier.*
We advise authors to avoid the use of private repositories whose roadmap is not transparent in terms of economic model, governance, sustainability ... (e.g. Figshare).

If the journal wishes to recommend a specific repository
The journal recommends that data be deposited in the disciplinary repository [Name of the repository] (e.g. Nakala for Social Sciences and Humanities).
In this case, describe the repository and the link between the journal and the repository: support offered to authors, presence of a specific collection for the journal on the repository…

If the journal wishes to make general recommendations
The journal recommends data be deposited in a repository, whether it is generalist (e.g. Zenodo), institutional (e.g. Data INRAE) or disciplinary (e.g. beQuali for qualitative survey data).
In all cases, authors should check that the chosen repository meets the following main quality criteria: See https://doranum.fr/depot-entrepots/criteres-choix-entrepot/ *(in French)* |

| 4. Data availability procedures | Explains how the data will be made available and in what timeframe

Specifies whether and how data are peer-reviewed | **Issues**
Authors
● Prior to submission, allow authors to know when to make their data available to the editorial board members or the reviewers.

● Inform authors that the license chosen affects the ability to reuse the data. Journals should advocate the use of open licenses (e.g. Creative Commons licenses).

● The following resources (*in French*) help to know how authors can disseminate their data:
   ○ DoraNum document (Inist-CNRS) ;
   ○ Logigramme (INRAE).
Reviewers
● State whether peer review of the data is planned or required, and what criteria apply. If applicable, these should be made publicly available on the journal's website. Example of data assessment criteria: https://zenodo.org/record/1065991#.X7T0MxJCcdU

● Specify whether reviewers are also or alternatively asked to assess whether the supporting data are consistent with the journal's policy.

Questions to consider
● Does the journal want to access the data upon submission of the article, during its review, or only in the copy-editing phase?

● When deposition of the data with the paper is mandatory (apart from justified exceptions), is it possible to apply an embargo?

● How long are authors expected to provide access to their data?

● On what analysis grid, or what criteria, is the data assessment based? Is this grid public and available for the authors? | *Submission phase*
Authors are not required to transmit the data when submitting their contributions.

Peer reviewing phase
If the reviewers deem it necessary, the authors should make the data that support the results reported in their contribution available for reviewers.

Acceptance phase
Data should be available without embargo, or with the shortest embargo period possible; sharing terms must allow reuse, with an explicit link between the data and the publication they support (see sections 4 and 5).

The journal encourages authors to share data under open licenses that allow for their free reuse. Authors must use the licenses recommended by the repository where the datasets were deposited.

By publishing in this journal, authors commit to make the data and metadata publicly available for at least 5 years after their contribution has been published, either through a platform, or by individual provision if the data cannot be freely shared.

Alternatives to open access sharing of personal or sensitive data are:
● *Anonymization or pseudonymization of the data before open access release*
● *Data available on request for research purposes only*
● *Availability of the metadata of data only* |

| 5. Support for authors and reviewers | Describes the support that the journal offers both to its authors and to its reviewers. | **Issues**
• Provide information about services authors can contact for support in sharing research data.

• When the journal includes data evaluation, provide reviewers with recommendations and procedures for accessing, reading, and understanding the data that accompany publications.

Questions to consider
• Does one or more of the journal members have sufficient expertise in the topic to answer questions from authors? from reviewers?

• What support arrangements does the journal offer: online information? dedicated email? specific resources? | *For any questions about our journal's data policy, authors can write to: revue.donnees[at]*

We invite authors to contact their institution's support services as regards good practices of data management and sharing.

If a data management plan exists for these data, authors are strongly encouraged to consult it for answers to their questions. |

| 6. Publications and data linking | Describes association mechanisms between data and publications and the prerequisites to achieve them (e.g. systematic attribution of persistent identifiers) | **Issues**
● Provide visible access to the reciprocal links between data and publications in order to access the data from the publication and vice versa.

● Facilitate the reuse of data.

● Foster trust in the integrity of the scientific process and methods behind the research work being presented.

Questions to consider
● How can the citation of data and publications be facilitated and thus their cross-linking promoted?

● How can a standard for dataset citation be adopted? See for example: https://datacite.org/cite-your-data.html | *Authors are encouraged to cite the datasets underlying their publications in a specific "Research Data" section. This section must describe the available data, how to access them, and provide a permanent link to the data.*

The section may include one, or a combination, of the following:
● *The datasets generated during and/or analysed during the current study are available in the [NAME] repository; [DOI].*
● *The datasets generated during and/or analysed during the current study are not available in open access due to [specify reason] but are available from the author on reasonable request.*
● *Data sharing does not apply to this article because no datasets were generated or analyzed during the current study.*
● *The datasets on which the current study is based were not generated by the authors. They are available online: Creator (Year of publication). Title. Version. [Repository Name]. [DOI].* |

| 7. Non-compliance with the policy | Describes the risks that authors run in not following the policy to avoid litigation.

Specifies internal procedures for dealing with such cases. | **Issues**
● Provide the means to ensure the implementation of the journal's policy and avoid any litigation with authors.

● Ensure that the procedures for handling such cases are proportionate and realistic and that decisions are endorsed by the journal as a whole and not by particular individuals.

Questions to consider
● Does the journal have the means to identify cases of non-compliance or does it only deal with cases reported by readers or reviewers?

● What types of remediation, or even sanctions, are considered: banner on the article? retraction?

● What procedures are in place to deal with such cases? | **Case of an incentive policy**
This policy is only an incentive. There are no penalties for not complying with this policy.

Case of a coercive policy
If a contribution does not comply with the rules for making data available, the authors will be informed by the editorial board and will have one month to comply with the journal's policy.

If at the end of this period the contribution still does not comply with the journal's policy, a banner will be added at the top of the article to specify in what way the journal's policy is not respected.

If the failure to provide data calls into question the credibility of the contribution, or if it is demonstrated that some of the data on which the contribution is built are erroneous, the journal retains the right to withdraw it. |