

PRACTICE PAPER

Developing a Research Data Policy Framework for All Journals and Publishers

An output of the Data policy standardisation and implementation Interest Group (IG) of the Research Data Alliance (RDA)

Iain Hrynaszkiewicz¹, Natasha Simons², Azhar Hussain³, Rebecca Grant⁴ and Simon Goudie⁵

¹ Public Library of Science (PLOS), Cambridge, UK

² Australian Research Data Commons, Brisbane, AU

³ Jisc, London, UK

⁴ Springer Nature, London, UK

⁵ Wiley, Melbourne, AU

Corresponding author: Iain Hrynaszkiewicz (ihrynaszkiewicz@plos.org)

More journals and publishers – and funding agencies and institutions – are introducing research data policies. But as the prevalence of policies increases, there is potential to confuse researchers and support staff with numerous or conflicting policy requirements. We define and describe 14 features of journal research data policies and arrange these into a set of six standard policy types or tiers, which can be adopted by journals and publishers to promote data sharing in a way that encourages good practice and is appropriate for their audience's perceived needs. Policy features include coverage of topics such as data citation, data repositories, data availability statements, data standards and formats, and peer review of research data. These policy features and types have been created by reviewing the policies of multiple scholarly publishers, which collectively publish more than 10,000 journals, and through discussions and consensus building with multiple stakeholders in research data policy via the Data Policy Standardisation and Implementation Interest Group of the Research Data Alliance. Implementation guidelines for the standard research data policies for journals and publishers are also provided, along with template policy texts which can be implemented by journals in their Information for Authors and publishing workflows. We conclude with a call for collaboration across the scholarly publishing and wider research community to drive further implementation and adoption of consistent research data policies.

Keywords: Data policy; journals; publishing; Research Data Alliance; Data citation; peer review

Introduction

An increasing number of publishers and journals are implementing policies that require or recommend that published articles be accompanied by the underlying research data (Jones, Grant & Hrynaszkiewicz, 2019). These policies are an important part of the shift toward reproducible research and contribute to the availability of research data for reuse (Vines, Andrew, Bock, *et al.*, 2013).

While uptake of journal data policies is on the rise, there is wide variation between policies on aspects such as their content, their discoverability, their ease of interpretation, infrastructure integration and support for compliance. This makes it challenging for journal editors to develop and support a data policy, difficult for researchers in understanding and complying with data policies, and complex for infrastructure providers and research support staff to assist with data policy compliance. There is clear benefit in a more standardised approach, as evidenced in the findings of the Jisc UK Journal Data Registry Project and the pioneering work of publishers, such as Springer Nature, to develop and support standard policy types for their journals (Naughton & Kernohan, 2016; Hrynaszkiewicz, Birukou, Astell, *et al.*, 2017).

This research data policy framework is intended to help journal editors and publishers to navigate the creation or enhancement of a research data policy. It reflects international efforts by the Research Data Alliance (RDA) Data Policy Standardisation and Implementation Interest Group (Hrynaszkiewicz, Simons, Goudie, *et al.*, n.d.) to identify the key elements of a good data availability policy and to standardise data policies.

Methods

The initial list of research data policy features included in this policy framework was developed by reviewing, combining and harmonising requirements from existing scholarly publishers' research data policies – Springer Nature, Elsevier, Wiley, PLOS (Anon, n.d., Anon, n.d., Anon, n.d., Anon, n.d.). A number of publisher and funding agency policies – notably the European Commission – refer to the Findable, Accessible, Interoperable and Reusable (FAIR) data principles [10], and some policies aim to enable compliance with the FAIR principles as an outcome of the policies [1]. While the FAIR principles are well established amongst the research data policy making, curation and infrastructure communities, for simplicity the policy framework and guidance in this paper avoids using the FAIR acronym, as fewer than half of researchers [11, 12] are familiar with FAIR. A number of journal policy features defined are relevant to the FAIR principles however, such as data licensing to enable reuse. The CODATA best practice guidelines for research data policy (Hodson & Molloy, 2015) and the TOP guidelines (Anon, 2014) were also included in the review of existing policy frameworks. The first version of the framework also incorporated feedback on, and requirements for, research data policy gathered during RDA plenary meetings and community conference calls/web meetings that were conducted during 2017. The first draft version of the framework (v1.2) (Hrynaszkiewicz, Simons, Goudie, *et al.*, n.d.) was made available for public comment for a period of three weeks, shortly before the March 2018 11th RDA Plenary meeting in Berlin. More than 30 comments were received from nearly 20 reviewers. The draft framework and a synthesis of the comments received were presented at the Berlin meeting, with further feedback received from attendees. The present version of the framework aims to address important feedback received from the community of reviewers on issues of scope, presentation, and clarity. It also aims to serve as a tool for editors and publishers to understand and implement standardised research data policies at their journals.

Results and Discussion

Table 1 defines what each of the 14 research data policy features are and the reason for their inclusion as part of the policy framework. **Figure 1** summarises which features are included in which policy type and provides a visual representation of how the feature is implemented.

The features are arranged into six types of research data policy, with increasing numbers of features and policy stringency as one progresses from the first type of policy through to the sixth.

The list of policy features and whether they are enforced through action is prescriptive. However, exactly how each policy feature and its requirements are implemented is not prescriptive, as the operations and resources available to different journals varies greatly. We, however, provide some implementation guidance and templated policy text for editors and publishers, which journals are encouraged to reuse. We acknowledge that wording and implementation methods will vary between journals, publishers and research disciplines. The scope of this document does not extend to supporting guidance and resources that are linked to from several policy features. For example, lists of recommended data repositories and criteria for assessing data repositories are not in scope. The scope of this document also does not include detailed guidance on preparing data availability statements, or detailed guidance on implementing data citation at scholarly publishers. Where appropriate, this document links to other initiatives that have or are defining more detailed guidance in these areas (Cousijn, Kenall, Ganley, *et al.*, 2018; Murphy & Samors, 2018).

Features arranged by policy type and implementation method

The 14 features are arranged into six types or tiers of policy, with more features and requirements as one moves from policy one through to six (**Figure 1**). The six tiers allow for more nuanced, step-wise and robust implementation of policies by different journals. This tiered approach to policy guidelines and frameworks is already in place at numerous large publishers, which from 2018 also includes Taylor & Francis (Anon, n.d.), and BMJ (Anon, n.d.). This tiering also acknowledges that the later features require the most effort to implement. At a low tier, Policy 2 enables a journal to provide full information on data sharing standards and good practice, but without the need to enforce any aspect. An option is also available for journals that wish to mandate data availability statements, but which do not have the means to check the contents of those statements in detail or enforce any data sharing mandates (policy 3). Journals that can commit to enforcing

Table 1: Policy feature definitions.

Feature	What	Why
Definition of research data	Define which research data the policy applies to, and the types of research data covered by the policy.	This enables the policy to define its scope and, where appropriate, provide general or discipline specific information on research data and file and format types (Hodson & Molloy, 2015). Specifying non-numeric data types (images, video, text etc) helps ensure relevance and applicability across research disciplines.
Definition of exceptions	Define what data do not need to be, or should not be made publicly available, under the policy and the alternative options for describing the availability of these data.	Ensures data policy is applicable to all research publications, but acknowledges legitimate exceptions and makes clear the policy does not create new legal or ethical precedents.
Embargoes	Define if and what embargoes on data release are permitted.	Researchers' reasonable right of first use of data generated during their research is a widely accepted principle of data sharing (Anon, 2016b), but reasonable lengths of embargo may vary by discipline, data type and study.
Supplementary materials	Define the journal/publisher's position on data sharing via supplementary materials, and if and when sharing data as supplementary materials is permitted under the policy.	While many policies preference sharing data in repositories (McCarthy, 2009; Santos, Blake & States, 2005; Evangelou, Trikalinos & Ioannidis, 2005), sharing data as supplementary materials remains very common. Some journals have strong data sharing policies and specify supplementary materials as the mechanism for data sharing. Supplementary materials are often a solution for researchers without discipline specific repositories and the definitions of supplementary material and research data often overlap.
Data repositories	State position on the use of data repositories. Data repositories are the preferred mechanism for sharing data with community/discipline specific repositories preferred to general repositories, where they are available.	Lack of an appropriate repository or lack of awareness of repositories are common reasons reported by researchers for not sharing data (Stuart, Baynes, Hrynaszkiewicz, et al., 2018). Journal and publisher information for authors is an important way of raising awareness of the availability of repositories for the majority of research data (Schmidt, Gemeinholzer & Treloar, 2016).
Data citation	Statement on the journal/publisher's support for the provision of persistent identifiers for research data that support publications, and statement of support for data persistent identifiers to be included in the reference list as formal citations. Includes whether data citation is encouraged or required.	Citing and linking to data increases visibility of research, increases academic credit and has been correlated with published articles receiving more citations (Piwowar & Vision, 2013; Piwowar, Day & Fridsma, 2007; Anon, n.d.; Henneken & Accomazzi, 2011; Dorch, Drachen & Ellegaard, 2015). This benefits researchers, journals, publishers, and society. Data citation in reference lists occurs in a fraction of published literature but is steadily increasing (Anon, n.d.). To ensure data citation happens consistently in published articles requires additional effort from authors and editors, and therefore operational costs, but enhances reader and use experience by consistently linking important research outputs (Cousijn, Kenall, Ganley, et al., 2018).

(Contd.)

Feature	What	Why
Data licensing	Define position on licensing and copyright for research data.	Lack of understanding of copyright and licensing of research data is a common reason why researchers don't share data (Stuart, Baynes, Hrynaskiewicz, <i>et al.</i> , 2018). Journal/publisher policy can help increase awareness and prevalence of explicit, and ideally open, licenses for research data. However, many established repositories do not have open data-conformant licenses and this is unlikely to change in the foreseeable future (Anon, 2017). Publishers are frequently asked whether the journal/publisher requires copyright transfer for datasets. Publishers issued a joint statement in 2006 declaring they would not take copyright in research data (STM & ALPSP, 2006).
Researcher/author support	Information on who authors should contact at the journal or publisher for more information on complying with the policy.	Research data sharing remains a new concept for some journals and disciplines and common questions can be answered by journal and publishing staff, such as writing data availability statements, finding repositories and on exceptions to the policy (Astell, Hrynaskiewicz, Grant, <i>et al.</i> , n.d.).
Data availability statements (DASs)	Define position on provision of data availability statements.	Data availability statements are a simple, consistent, human and, increasingly, machine-readable way of expressing data availability and policy compliance. They are already encouraged, expected or required by many journals and publishers and some funding agencies (Anon, 2016a; Murphy & Samors, 2018).
Data formats and standards	State position on the use of community/discipline-specific data standards – whether encouraged, required in some cases, or required in all cases. Also state whether certain file formats, such as open formats, are preferred or required.	Data prepared according to community standards are more interoperable and reusable, and data available in open formats are more accessible (Sansone, McQuilton, Rocca-Serra, <i>et al.</i> , 2019). Data standards are distinct from reporting standards, which are not within the scope of a research data policy (e.g. MIAME as a reporting standard for papers describing for microarray experiments).

(Contd.)

Feature	What	Why
Mandatory data sharing (specific papers)	Statement on whether data sharing is mandatory for specific types of research data, such as where there is a community or journal-specific mandate, and the mechanism(s) by which these types of data must be shared. Examples include DNA and RNA sequence data, and macromolecular structure data.	Where there are established community mandates for data sharing, journals and publishers have an obligation to support editors and researchers in upholding community standards as part of their service to the research communities they serve (Anon, n.d.).
Mandatory data sharing (all papers)	Sharing of research data via an external mechanism (repositories or supplementary information) is a condition of submission or publication for all articles published.	Mandatory data sharing policies that are enforced during the peer-review and publishing process, and supported with suitable data repositories, are the most effective policies (Vines, Andrew, Bock, <i>et al.</i> , 2013). These policies can also be more costly (time consuming) to implement and have the greatest impact on editors and authors (Grant & Hrynaszkiewicz, 2018). They could however have the most benefits in terms of increasing citations and visibility of papers.
Peer review of data	Statement on whether peer review of data is expected or required, and if so what the expectations of peer reviewers are in their assessment of data files. Reviewers can also or alternatively be asked to assess compliance with research data policy.	Where data are made available with research articles they are accessible to peer reviewers, but for journals with a strong focus on data, such as data journals, consistent review of the data and the description of those data can be required. Peer review traditionally focuses on manuscripts rather than data, but more consistent availability of data for validation and reuse can improve the reproducibility – and quality – of published research (Anon, 2016c).
Data Management Plans (DMPs)	State position on sharing of DMPs.	This is currently uncommon in journal and publisher policy although encouraging their provision is analogous to the number of medical journals which encourage or require sharing or publication of study protocols. Furthermore, they are increasingly required by funding agencies. Some journals, such as RIO journal, publish them as articles.

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	<input type="radio"/>					
Exceptions to policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Embargoes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Supplementary materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data repositories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data licensing	<input type="radio"/>					
Researcher/ author support	<input checked="" type="radio"/>					
Data availability statements		<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data formats and standards				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (specific data types)				<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (all papers)				<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Peer review of data				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Data Management Plans (DMPs)				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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: Fourteen journal research data policy features arranged as six policy types (tiers).

all relevant mandates for their communities adopt policy 4. The six-policy approach also provides a specific policy for journals that mandate data sharing but do not carry out data peer review routinely (policy 5) and culminates by incorporating data peer review as a feature of the highest tier, policy 6.

Explanation and implementation requirements of the research data policy framework features

Policy feature: Definition of research data

Every policy must define research data as being the data that support the findings or claims made in the published article. This definition helps to focus researchers on identifying and sharing research data that enable replication or validation of claims made in the paper, and avoids potential for confusion of researchers (authors) about the journal’s expectations. This definition also manages the expectations of researchers who may often produce much greater volumes of research outputs from a study than are necessary to reproduce or evidence claims made in a particular paper.

Policies should also specify what kinds of data are included in the policy, such as tabular data, code, images, audio, video, maps, raw and/or processed data. Data can be digital and non-digital. This aims to increase the relevance of the policy to all researchers who, across multiple disciplines, will have different interpretations of the meaning of “research data”. Qualitative research may produce interview videos and transcripts as its research data, rather than numerical tables, for example.

Policies should further include and define their coverage of:

- Data produced by the authors for the study being reported (“primary data”)
- Data reused or analysed by the authors for the study being reported (“secondary data”)
- Additional data from experiments or observations, such as raw, unprocessed data or data from time points not reported in the article
- Any other materials that might be required to reproduce or replicate the results

Implementation notes

Journal or publisher’s editorial policy text or information for authors must include the definition of research data. See template policy text for an example.

Policy feature: Definition of exceptions

The policy must define the types of data that it does not expect to be shared publicly. It should also define, if applicable, research data that are not covered by the policy – if these are not already explicit in preceding feature, Definition of research data.

Data that a journal or publisher does not expect to be shared publicly may include personal or sensitive data, such as quantitative or qualitative data that could identify an individual, data which participants did not consent to be shared,, locations of endangered species, and data subject to other legitimate restrictions on public availability.

Other types of sensitive data must also be defined if they are applicable to the journal/publisher’s content. Alternative options for public sharing of these data and describing their availability should be given, such as:

- Controlled access repositories
- Anonymisation and de-identification of data
- Sharing metadata only
- Working with data access committees and implementing Data Use Agreements

Aspects of this policy feature and text may be superseded, substituted or modified by the feature “Mandatory data sharing (all papers)”.

Implementation notes

For policies 1–4: Journal or publisher’s editorial policy text or information for authors must include examples of the types of data that they do not expect to be shared publicly. See template policy text for an example.

The journal or publisher, via its editorial or peer-review process, must be able to identify if authors are sharing sensitive or personal data without appropriate consent. Where this is identified, it must advise authors of appropriate action, referring to resources on alternatives to public data sharing where appropriate.

For policies 5 and 6: These journals require mandatory data sharing for every publication, evidenced by datasets cited in reference lists. In such cases, raw data, such as individual participant data from clinical studies that are not anonymised, might not be publicly available but the data must be archived in a secure repository that provides a persistent identifier and landing page for the data so that the data can be cited.

Policy feature: Embargoes

The policy must include a statement about the journal or publisher’s position on embargoes. This may need to consider community norms, funding agency policies (where applicable) and enabling researchers a reasonable right of first use. The policy should provide information on any relevant community-specific embargoes.

Implementation notes

Journal or publisher’s editorial policy text or information for authors must include information on embargoes. The journal or publisher must be prepared to respond to and resolve unreasonable embargo periods on data included in its policy’s definition of research data, if it is made aware of them. See template policy text for an example.

Note that for policies 5 and 6, these tiers do not permit embargoes on data access and data must be accessible to readers at the publication date, and at minimum have been accessible to editors and peer reviewers before publication.

Policy feature: Supplementary materials

Data repositories are the preferred method for sharing data supporting publications and this must be stated in the policy. Research data published as supplementary materials files are less persistent and findable than research data deposited in data repositories (Anderson, Tarczy-Hornoch & Bumgarner, 2006; Evangelou, Trikalinos & Ioannidis, 2005). The policy must also specify whether sharing research data via supplementary materials, or an equivalent method by which data objects are archived by the publisher as part of the published article, is permitted.

Implementation notes

Journal or publisher's editorial policy text or information for authors must include information on whether data sharing via supplementary materials is permitted. This statement may need to reference and be consistent with the publisher's existing policy on supplementary materials. Journals and publishers that deposit supplementary materials files in third-party repositories, such as figshare, that assign persistent identifiers to each file should provide further information on these services. See template policy text for an example.

For policies 5 and 6: Data sharing via supplementary materials is not permitted and the journal must support this requirement with checks in the editorial or peer review process to ensure that datasets supporting the claims in the paper are deposited in appropriate repositories and cited in the reference list.

Policy feature: Data repositories

Data policies must be supported with data preservation, which will require a list of recommended, trusted or supported data repositories. This could be the journal/publisher's own list, a community/discipline-specific list, or a curated and trusted third-party list, such as those available from FAIRsharing.org or a repository finder tool or service such as <https://repositoryfinder.datacite.org/>. Data policies must also preference the use of community/discipline-specific data repositories over general data repositories, where community/discipline-specific repositories exist. Community/discipline specific repositories are preferred because they typically require deposition of data and metadata in standard, common formats, enabling more efficient discovery and reuse of data. They also typically provide professional data curation (Anon, n.d.). However, general repositories fulfill a vital function for much research data, which do not have a relevant community/discipline specific repository. The provided list must include general repositories, if community/discipline-specific repositories cannot support all research data included in the definition of the policy.

If the publisher's own list of repositories is provided, it must include criteria for adding repositories to the list and a position statement on its support for institutional data repositories. Different standards for assessing trusted data repositories exist – such as the CoreTrust Seal and Springer Nature/*Scientific Data's* criteria for recommended repositories – but it is beyond the scope of this document to define standard criteria for trusted data repositories.

Implementation notes

For policies 1–3: Journal or publisher's editorial policy text or information for authors must include information on data repositories. See template policy text for an example. Information on data repositories, and a reference to the research data policy in general, should also be communicated to authors at an appropriate point during submission of manuscripts. This could be communicated via the journal or publisher's manuscript submission system and/or in standard email correspondence sent to authors during the editorial and peer-review process. Journals and publishers must be prepared to respond to requests from authors for advice on finding appropriate data repositories.

For policies 4–6: The use of data repositories, for specific (policy 4) or all (policies 5 and 6) datasets supporting publications is mandatory. This requirement must be enforced by checks in the editorial or peer review process to ensure datasets are deposited.

Policy feature: Data citation

The policy must enable, and for policies 5 and 6 require, authors to cite datasets in the reference lists (bibliographies) of their articles. It must also include the journal/publisher's style(s) for referencing datasets. One or more examples of data citation should be included. The policy should also specify if the journal

or publisher has any restrictions on which datasets can be cited in reference lists, such as those that have particular types of persistent identifier (e.g. Digital Object Identifiers [DOIs], accession codes, etc).

The policy should include links to more examples of data citation in published articles, and further information on the benefits of citing and linking data are desirable.

Implementation notes

For policies 1–4: Journal or publisher’s editorial policy text or information for authors must include information on data citation. See template policy text for an example. Journals and publishers must also ensure that authors who cite data in their references do not receive conflicting information during the publishing process, nor should data citations be arbitrarily removed from reference lists.

For policies 5 and 6: Accurate and consistent provision of data citations must be enforced through editorial, peer-review and/or article production procedures. This requirement must be included in the policy and supported by checks on manuscripts that ensure publicly available, persistently-identified datasets are cited in reference lists. Automation to identify dataset identifiers, by publishers, can aid the identification of datasets that must appear in reference lists.

Implementation of data citation by publishers also has implications for content structure and XML production workflows, which is beyond the scope of this document. Journals and publishers that are implementing data citation should consult the data citation roadmap for scholarly publishers (Cousijn, Kenall, Ganley, *et al.*, 2018).

Policy feature: Data licensing

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.
- If the journal or publisher has expectations for the licenses authors make their research data available under, and what those expectations are.

The policy must express a preference for Open Data conformant licenses (such as Creative Commons Attribution License, CC BY, Creative Commons Public Domain Waiver, CC0). Licensing can also be addressed in part with recommended repositories, as criteria for trusted repositories often include requirements for licensing.

Implementation notes

Journal or publisher’s editorial or publishing policy text or information for authors must include information on data licensing. See template policy text for an example. Journals and publishers cannot enforce specific licenses for research data that are deposited in third party repositories, as the licenses applied by repositories are generally outside of the publisher’s control. Journals and publishers must be prepared to respond to questions from authors about licensing and copyright of research data.

Policy feature: Researcher/author support

The policy must include information on who authors can contact with questions about compliance with the policy. This might include email addresses, phone numbers and/or web-based customer support tools. It may also include information on other services or organisations that researchers can approach for support for sharing research data.

Implementation notes

Journal or publisher’s editorial policy text or information for authors must include contact information for author support. See template policy text for an example.

Policy feature: Data availability statements

The policy must include a definition of a data availability statement (DAS) and where it should be placed in the manuscript. It must also specify if such statements are mandatory and must state if authors are permitted to make research data “available on reasonable request”. Numerous examples of DASs and template DASs exist. Defining standards for DASs is beyond the scope of this document but is the topic of other initiatives (Murphy & Samors, 2018).

Implementation notes

For policy 1: This feature does not apply to policy 1.

For policy 2: Journal or publisher's editorial policy text or information for authors must include information on DASs. Template DASs for the most common types of DAS should be provided in the policy and further guidance and examples linked to. Contextual examples, such as DASs from published articles including DASs, should be provided. See template policy text for example text.

Where DASs are mandatory

Where DASs are a mandatory part of published articles, this requirement must be communicated to authors at an appropriate point during submission of manuscripts as well as in the information for authors or editorial policy text. This can be communicated via the journal or publisher's manuscript submission system and/or in standard email correspondence sent to authors during the editorial and peer-review process. Journal staff or editors who are responsible for ensuring mandatory sections of articles are included in published articles must update their standard operating procedures and documentation, such as manuscript templates provided to authors, to support this requirement. Journals and publishers implementing mandatory DASs should determine the likely impact, in time and cost, of this change on their authors, journal staff, and editors, and modify the resources available to the journal to support this requirement. An analysis by the Nature Research journals found adding mandatory DASs increases the time it takes to process a manuscript by several minutes (Grant & Hrynaszkiewicz, 2018).

For policy 3: Accuracy of DASs is based on trust and there is no expectation for the journal or publisher to verify the accuracy of the statements for every publication.

For policies 4–6: Where data are not shared publicly, authors publishing in these journals must be willing to respond to reasonable requests from other researchers for copies of the data, where data are not publicly available, to verify or reproduce results reported in the paper. The journal and publisher must also facilitate readers' access to data supporting publications, for example if no response is received to requests for data or the response to a request for data is not consistent with the policy. Journals must take action where necessary if it transpires after publication that their data policy has not been adhered to. This might include contacting authors or their institutions directly, and in some cases publishing corrections, expressions of concern or retractions, in accordance with publication ethics guidelines.

Policy feature: Data standards and formats

The policy must express support for community-endorsed data/metadata standards and formats, if and where any may be applicable to the journal or publisher's publications. The policy should also provide one or more specific examples of data standards and formats. A data standard is a common and interoperable way of representing, labeling or structuring data and metadata. A data format refers to the way data are stored or archived, commonly the digital file type or extension. Depositing data in a community/discipline specific data repository can often achieve adherence to domain-specific data standards.

The policy must also define its position on open and proprietary formats, and encourage the most interoperable file formats where this is practical to achieve. For example, encouraging or requiring open file formats (e.g. CSV for tabular data). Resources such as FAIRsharing.org should be linked to.

Implementation notes

For policies 1–3: This feature does not apply to policies 1–3.

For policies 4–5: Journal or publisher's editorial policy text or information for authors must include information on data formats and standards. See template policy text for example text.

For policy 6: Datasets must be shared in the appropriate standard and format, and this must be enforced through the editorial or peer review process. Enforcing deposition in community (discipline) specific data repositories can often achieve this requirement as community specific repositories often require data submission in specific formats and according to specific standards.

Policy feature: Mandatory data sharing (specific papers)

The policy must specify the data sharing mandate(s) that must be followed as a condition of submission and/or publication and the mechanisms for demonstrating compliance, such as deposition in specific repositories. Data sharing mandates typically relate to specific types of data, for which data sharing is an

established norm and for which community/discipline specific data repositories exist for the data type(s) covered by the mandate. The policy must also specify if and how these mandates are enforced by the journal, such as by checks by editors, reviewers or journal staff. These mandates mostly apply to specific types of research data generated in life science disciplines. A list of these established community data sharing mandates is available from Nature Research (Anon, n.d.).

Implementation notes

For policies 1–3: This feature does not apply to policies 1–3.

For policies 4–6: These mandates are enforced and journals and publishers will need procedures in place to ensure they are enforced consistently. For journals and publishers that publish in multiple research disciplines, and where mandates may only apply to certain papers, implementation of its enforcement mechanism needs special attention and the impact of introducing enforcement measures on authors and editors determined. Enforcement can be enabled by editorial checklists (as used for example by the Nature Research journals (Anon, 2018)) and, potentially, supported by artificial intelligence tools, such as <https://www.penelope.ai/>.

Policy feature: Mandatory data sharing (all papers)

For these journals, “available on reasonable request” DASs are not acceptable. The mechanism(s) for complying with the policy, such as integrated data repositories available to the journal/publisher, must be specified. The data must be available in a data repository or with the article as supplementary material and this must be verified as part of the publishing or peer review process. For clinical or sensitive data published under this policy, public sharing of raw data may not be required but deposition in a repository that supports controlled access and has independent governance procedures (such as data access committees and data use agreements) is required. Journals that wish to permit other types of exception to the policy, such as commercial restrictions should not adopt policy 5 or 6.

Implementation notes:

For policies 1–4: This feature does not apply to policies 1–4.

For policies 5 and 6: Journal or publisher’s editorial policy text or information for authors must include its mandatory data sharing policy. See template policy text for an example. Journals must carry out checks on every manuscript that is sent for peer review to ensure that any datasets on which the claims are based are available in accordance with the policy. How this is achieved will depend on the systems and operations of the journal but the procedures build logically upon those that implement Data Availability Statements, mandatory data sharing for specific papers, and Data citation. The availability of data, such as links to datasets in repositories, must be visible to peer reviewers. Some manuscript submission systems can offer integration with general data repositories that enable confidential access to data during peer review, such as figshare and Dryad, to enable authors and journals to comply efficiently with this policy.

Policy feature: Peer review of data

The policy must state:

- What the expectations are for peer reviewers to access data
- Whether peer reviewers should or must assess data policy compliance
- Whether peer reviewers should or must assess sufficiency of description of data files for understanding and reuse
- Whether peer reviewers should or must assess the data files themselves, such as for structure, completeness, reusability, etc
- Whether peer reviewers are expected to validate, replicate or reproduce claims/statistics reported in the paper

Criteria for peer reviewers’ assessment of data files are included in Springer Nature’s data policy framework (Anon, n.d.), and a guide to data peer review has been produced by PLOS (<https://plos-marketing.s3.amazonaws.com/Marketing/Peer+Reviewing+Datasets.pdf>).

Implementation notes

For policies 1–3: This feature is not applicable to Policies 1–3.

For policy 4: Peer reviewers are not expected to routinely access and assess supporting datasets, although are not discouraged from doing so. For policy 4, peer reviewers are expected to include in their assessment of papers recommendations on whether the authors have complied with the journal's data sharing policy, rather than on the data files themselves. This requirement must be communicated, such as in the journal or publisher's guide to peer reviewers or peer reviewer forms. See template policy text for example text.

For policy 5: All reviewers, under this policy, will have the opportunity to see supporting data files and guidelines for reviewers must provide information on what reviewers should consider when accessing and assessing datasets. See template policy text for example text.

For policy 6: Peer reviewers are required to access supporting datasets, enabled by the journal's editorial process and data policy. Reviewers must be aware of the journal's expectations for data peer review and utilise these in their assessment of manuscripts and supporting datasets. Including these requirements in peer reviewer forms and checklists is highly desirable.

Where applicable, implementation will need to address issues with double blind peer review, as datasets may contain information that can identify the authors.

Policy feature: Data Management Plans

The policy must define if and how it incorporates the preparation and sharing of Data Management Plans (DMPs). Options include:

- Always encouraging or requiring preparation of DMPs
- Requiring DMPs when funding agencies or institutions also require them
- Encouraging publication or sharing of DMPs

Implementation notes

For policies 1–3: Feature does not apply to policies 1–3.

For policies 4–6: Journal or publisher's editorial policy text or information for authors must include information on the preparation and/or sharing of DMPs. See template policy text for an example. Under no policy type is the use, or sharing with the journal, of DMPs mandatory or enforced, as this reflects current practice in scholarly publishing and in funding agencies' policies.

Conclusion and next steps

This paper provides a comprehensive journal research data policy framework that can be adopted by and aligns with the policy requirements of all scholarly journals and publishers. It is an output of the Data Policy Standardisation and Implementation Interest Group of the RDA and has been produced with open, research community and publishing industry consultation over a period of two years. The framework is practical and pragmatic, enabling any journal to implement a research data policy that is compatible with the editorial model and procedures of the journal, and the level of support for data sharing in the journal's author and reader community. While some of the policy types in this framework might be viewed as unambitious, "overreach" has been identified as a factor associated with policy failure (Neylon, 2017) and our goal is to provide a research data policy framework that is usable by the widest possible audience. Implementation, adoption and endorsement of this framework by journals and publishers is critical to its success and a partnership has been formed between this RDA group and the STM Association in 2019, to increase adoption in the publishing industry (<https://www.stm-researchdata.org/>). Success of this initiative can be measured, in the short term, by the number of journals and publishers who adopt this policy framework or align their existing data policy options with this framework. Longer term, success should be measured by increased levels of data sharing and reuse, which means enabling journals, editors and researchers to implement the policy types 3 and above. Policy types 3 and above require data availability statements in published articles and these are a recognised compliance monitoring and data-discovery tool. Policy implementation should be combined with ongoing evaluation of the impact (costs, as well as benefits) of data policies. It is also assumed that data reuse and reproducible research are enabled by data sharing, and further research is needed to test these assumptions in large cohorts and in multiple research disciplines.

Definitions

This section aims to provide definitions of common terms used in the document:

Primary data: Data that are collected directly from first-hand sources, using methods such as surveys, interviews, or experiments.

Secondary data: Data gathered from studies, surveys, or experiments that have been conducted by other people or for other research.

Paper, article, publication: In this document paper, article and publication are used to refer to outputs that are published in journals. Beginning as unpublished manuscripts of research not previously published, these typically undergo a peer-review process by one or more academic referees before being accepted or rejected for publication within a journal.

Data availability/accessibility statements (DAS): A data availability statement (also referred to as a 'data accessibility statement') indicates where the data associated with a paper is available, under what conditions the data can be accessed, including links (where applicable) to the data set.

Community/discipline-specific repository: A public data repository designed for housing data for a given domain of research.

Data standard: A common and interoperable way of representing, labelling or structuring data.

Data format: The way data are stored or archived, commonly the digital file type or extension.

Data citation: A reference to a published or unpublished data source, for the purpose of acknowledging the relevance of the works of others concerning the topic being discussed.

Supplementary materials: Any material that adds detail, background, or context to an article by providing, for example, multimedia objects such as audio clips and applets; additional XML-tagged sections, tables, or figures; raw data in a spreadsheet, or a software application in a repository.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Policy feature standard texts for policy template construction. DOI: <https://doi.org/10.5334/dsj-2020-005.s1>

Acknowledgements

The authors thank all members of the Research Data Policy Standardisation and Implementation IG and any other individuals who provided comments on previous versions of this paper and/or contributions to our community calls, and plenary meetings. We also acknowledge previous co-chairs of the group, David Kernohan and Simone Taylor for their contributions in establishing this initiative.

This paper was published as a preprint in June 2019 (Hrynaszkiewicz, Simons, Hussain, *et al.*, 2019).

Competing Interests

At the time of writing the original manuscript Iain Hrynaszkiewicz and Rebecca Grant were employees of the publisher Springer Nature and Simon Goudie an employee of the publisher Wiley. At the time of submission to the journal Iain Hrynaszkiewicz is an employee of PLOS. None of these employers required approval or review of the text before publication.

References

- Anderson, NR, Tarczy-Hornoch, P and Bumgarner, RE.** 2006. On the persistence of supplementary resources in biomedical publications. *BMC Bioinformatics* [Online], 7260. DOI: <https://doi.org/10.1186/1471-2105-7-260>
- Anon.** 2014. *Transparency and Openness Promotion (TOP) Guidelines* [Online], 12 August 2014. Available from: https://osf.io/xd6gr/?_ga=2.251468229.297610246.1542300800-587952028.1539080384 [Accessed: 15 November 2018].
- Anon.** 2016a. Announcement: Where are the data? [Online], 537(7619): 138. DOI: <https://doi.org/10.1038/537138a>

- Anon.** 2016b. *Concordat on Open Research Data* [Online]. Available from: <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/> [Accessed: 19 January 2017].
- Anon.** 2016c. Let referees see the data [Online], 3160033. DOI: <https://doi.org/10.1038/sdata.2016.33>
- Anon.** 2017. Open for business, 4. DOI: <https://doi.org/10.1038/sdata.2017.58>
- Anon.** 2018. Checklists work to improve science [Online], 556(7701): 273–274. DOI: <https://doi.org/10.1038/d41586-018-04590-7>
- Anon.** n.d. *Availability of data, materials, code and protocols: authors & referees @ npg*. [Online]. Available from: <https://www.nature.com/authors/policies/availability.html> [Accessed: 1 May 2019a].
- Anon.** n.d. *Data Policy Types|Authors|Springer Nature* [Online]. Available from: <https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096> [Accessed: 1 May 2019b].
- Anon.** n.d. *Data sharing – BMJ Author Hub* [Online]. Available from: <https://authors.bmj.com/policies/data-sharing/> [Accessed: 2 May 2019d].
- Anon.** n.d. *Data Sharing & Citation|Wiley* [Online]. Available from: <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/index.html> [Accessed: 1 May 2019c].
- Anon.** n.d. *Data Sharing Effect on Article Citation Rate in Paleoceanography* [Online]. Available from: https://figshare.com/articles/Data_Sharing_Effect_on_Article_Citation_Rate_in_Paleoceanography/1222998/1 [Accessed: 28 March 2019e].
- Anon.** n.d. *Glad You Asked: A Snapshot of the Current State of Data Citation* [Online]. Available from: <https://blog.datacite.org/citation-analysis-scholix-rda/> [Accessed: 16 November 2018f].
- Anon.** n.d. *Making the Case for Disciplinary Data Repositories* [Online]. Available from: <https://deepblue.lib.umich.edu/handle/2027.42/135733> [Accessed: 20 January 2020g].
- Anon.** n.d. *PLOS Data availability policy* [Online]. Available from: <https://journals.plos.org/plosone/s/data-availability> [Accessed: 9 November 2012h].
- Anon.** n.d. *Research Data Guidelines* [Online]. Available from: <https://www.elsevier.com/authors/author-resources/research-data/data-guidelines> [Accessed: 1 May 2019i].
- Anon.** n.d. *Understanding our data sharing policies – Author Services*. [Online]. Available from: <https://authorservices.taylorandfrancis.com/understanding-our-data-sharing-policies/> [Accessed: 2 May 2019j].
- Astell, M, Hrynaszkiewicz, I, Grant, R, Smith, G,** et al. n.d. *Have questions about research data? Ask the Springer Nature Helpdesk* [Online]. Available from: https://figshare.com/articles/Providing_advice_and_guidance_on_research_data_a_look_at_the_Springer_Nature_Helpdesk/5890432 [Accessed: 30 April 2018].
- Cousijn, H, Kenall, A, Ganley, E, Harrison, M,** et al. 2018. A data citation roadmap for scientific publishers. *Scientific data* [Online], 5180259. DOI: <https://doi.org/10.1038/sdata.2018.259>
- Dorch, BF, Drachen, TM and Ellegaard, O.** 2015. The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union* [Online], 11(A29A): 172–175. DOI: <https://doi.org/10.1017/S1743921316002696>
- Evangelou, E, Trikalinos, TA and Ioannidis, JPA.** 2005. Unavailability of online supplementary scientific information from articles published in major journals. *The FASEB Journal* [Online], 19(14): 1943–1944. DOI: <https://doi.org/10.1096/fj.05-4784lsf>
- Grant, R and Hrynaszkiewicz, I.** 2018. The impact on authors and editors of introducing Data Availability Statements at Nature journals. *International Journal of Digital Curation* [Online], 13(1): 195–203. DOI: <https://doi.org/10.2218/ijdc.v13i1.614>
- Henneken, EA and Accomazzi, A.** 2011. *Linking to Data – Effect on Citation Rates in Astronomy* [Online], 4. Available from: <http://arxiv.org/abs/1111.3618> [Accessed: 14 February 2013].
- Hodson, S and Molloy, L.** 2015. *Current Best Practice for Research Data Management Policies* [Online]. Available from: https://zenodo.org/record/27872#.WlczsIVl_IV [Accessed: 11 January 2018].
- Hrynaszkiewicz, I, Birukou, A, Astell, M, Swaminathan, S,** et al. 2017. Standardising and harmonising research data policy in scholarly publishing. *International Journal of Digital Curation* [Online], 12(1): 65. DOI: <https://doi.org/10.2218/ijdc.v12i1.531>
- Hrynaszkiewicz, I, Simons, N, Goudie, S and Hussain, A.** n.d. *Journal and publisher research data policy master framework (RDA IG draft output) DRAFT v1.2 Feb 2018* [Online]. Available from: https://docs.google.com/document/d/1DTAfOKkE1a2n2f_1hGcrXlrw-5Tq_AL5tk-ju8B82_E/edit?usp=sharing [Accessed: 1 May 2019a].
- Hrynaszkiewicz, I, Simons, N, Goudie, S and Hussain, A.** n.d. *Research Data Alliance Interest Group: Data policy standardisation and implementation* [Online]. Available from: <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation> [Accessed: 30 April 2018b].

- Hrynaszkiewicz, I, Simons, N, Hussain, A and Goudie, S.** 2019. Developing a research data policy framework for all journals and publishers. *Figshare* [Online]. DOI: <https://doi.org/10.6084/m9.figshare.8223365.v1>
- Jones, L, Grant, R and Hrynaszkiewicz, I.** 2019. Implementing publisher policies that inform, support and encourage authors to share data: two case studies. *Insights the UKSG journal* [Online], 32(1). DOI: <https://doi.org/10.1629/uksg.463>
- McCarthy, J.** 2009. Supplementary online material: potential and precautions. *Augmentative and alternative communication (Baltimore, Md.: 1985)* [Online], 25(1): 4–6. DOI: <https://doi.org/10.1080/07434610902744041>
- Murphy, F and Samors, RJ.** 2018. *Belmont Forum Data Accessibility Statement Policy and Template – Endorsed 18 October 2018.*
- Naughton, L and Kernohan, D.** 2016. Making sense of journal research data policies. *Insights the UKSG journal* [Online], 29(1): 84–89. DOI: <https://doi.org/10.1629/uksg.284>
- Neylon, C.** 2017 Building a culture of data sharing: policy design and implementation for research data management in development research. *Research Ideas and Outcomes* [Online], 3e21773. DOI: <https://doi.org/10.3897/rio.3.e21773>
- Piwovar, HA, Day, RS and Fridsma, DB.** 2007. Sharing detailed research data is associated with increased citation rate. *Plos One* [Online], 2(3): e308. DOI: <https://doi.org/10.1371/journal.pone.0000308>
- Piwovar, HA and Vision, TJ.** 2013. Data reuse and the open data citation advantage. *PeerJ* [Online], 1e175. DOI: <https://doi.org/10.7717/peerj.175>
- Sansone, S-A, McQuilton, P, Rocca-Serra, P, Gonzalez-Beltran, A, et al.** 2019. FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology* [Online], 37(4): 358–367. DOI: <https://doi.org/10.1038/s41587-019-0080-8>
- Santos, C, Blake, J and States, D.** 2005. Supplementary data need to be kept in public repositories. *Nature*, 438(7069): 738. DOI: <https://doi.org/10.1038/438738a>
- Schmidt, B, Gemeinholzer, B and Treloar, A.** 2016. Open data in global environmental research: the belmont forum's open data survey. *Plos One* [Online], 11(1): e0146695. DOI: <https://doi.org/10.1371/journal.pone.0146695>
- STM & ALPSP.** 2006. *Databases, data sets, and data accessibility – views and practices of scholarly publishers.*
- Stuart, D, Baynes, G, Hrynaszkiewicz, I, Allin, K, et al.** 2018. *Whitepaper: Practical challenges for researchers in data sharing* [Online]. Available from: https://figshare.com/articles/Whitepaper_Practical_challenges_for_researchers_in_data_sharing/5975011 [Accessed: 30 April 2018].
- Vines, TH, Andrew, RL, Bock, DG, Franklin, MT, et al.** 2013. Mandated data archiving greatly improves access to research data. *The FASEB Journal* [Online], 27(4): 1304–1308. DOI: <https://doi.org/10.1096/fj.12-218164>

How to cite this article: 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. DOI: <https://doi.org/10.5334/dsj-2020-005>

Submitted: 20 October 2019 **Accepted:** 29 January 2020 **Published:** 21 February 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.