Policy as an incentive and barrier to data sharing

• Funder policy motivates researchers to share data (88%\(^1\))

But

• Complying with funder policies is a challenge for more than half of researchers (54%\(^2\))

---

Supporting compliance: Research funders and data access

• More than 40 research funders globally have policies or mandates that require data archiving as a condition of grants\(^{(1)}\) e.g.
  • National Science Foundation (NSF)
  • Wellcome Trust
  • Bill and Melinda Gates Foundation
  • EU H2020 pilot
  • Research Councils UK (as part of open access policy)

• Some of these require data to be linked to publications including:
  • Engineering and Physical Sciences Research Council (EPSRC)

Journals’ research data policies are confusing

Data source: Linda Naughton, JISC Journal Research Data Policy Bank project presentation (n = 250)

“The evidence shows that the current research data policy ecosystem is in critical need of standardization and harmonization”

Policy Types

Type 1
Data sharing and data citation is encouraged but not required

Type 2
Data sharing and evidence of data sharing encouraged

Type 3
Data sharing encouraged and statements of data availability required

Type 4
Data sharing, evidence of data sharing and peer review of data required

Process

1. Identify and agree the most relevant policy type for individual journal

2. Implement standardised text and processes into relevant journal guides and publishing workflows

3. Provide a consistent and easy-to-follow journal data policy for authors, researchers and peer reviewers

http://www.springernature.com/gp/group/data-policy
All policy types:
- Preference **sharing of data via repositories** (rather than ESM/SI)
- **Allow citation of public datasets** in reference lists/bibliographies
- **Encourage use of publisher helpdesk** to ensure compliance with funder mandates
# Springer Nature Research Data Policy framework

<table>
<thead>
<tr>
<th>Feature</th>
<th>Explanation</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sharing via repositories supported</td>
<td>Details of sharing via repositories is referred to in journal guide to authors</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Data citation permitted</td>
<td>Journal style guide permits authors to cite publicly available datasets in reference lists</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Publisher helpdesk</td>
<td>Helpdesk contact details included in journal information for authors</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Public data deposition and dataset identifier checks for specific types of data</td>
<td>Data deposition checked as part of the publishing process where there is an established research community mandate</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Data availability statements</td>
<td>Statement in published articles explaining how supporting data can be accessed</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Public data deposition and dataset identifier required and verified</td>
<td>Data made publicly available and data identifiers provided for all published articles (with exceptions for sensitive/personal data)</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Data citations</td>
<td>Relevant dataset citations in reference lists provided and verified</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Peer review of data</td>
<td>Peer reviewer guidelines and process give guidance on accessing and reviewing data files</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Integrated data repository</td>
<td>Submission system/review process integrated with a journal-specific or general repository, such as figshare or dryad</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

● Mandatory  ○ Optional  ○ Not Required
Policy Types

**Type 1**
Data sharing and data citation is encouraged but not required

**E.g.**
Numerous Springer engineering, computer science and maths journals

**Type 2**
Data sharing and evidence of data sharing encouraged

**E.g.**
Plant and Soil & many Springer life sciences, medicine and physics journals

**Type 3**
Data sharing encouraged and statements of data availability required

**E.g.**
Palgrave Communications
BMC Genomics & most BMC journals
Heredity
Nature & Nature journals

**Type 4**
Data sharing, evidence of data sharing and peer review of data required

**E.g.**
Scientific Data
Genome Biology
Research Data Support helpdesk @Springer Nature

Support for editors:

• Identifying and implementing a data policy
• Identifying data repositories for their audience(s)
• Dealing with peer review of sensitive/human data
• Good practice for data-literature integration

Support for authors:

• Information on the data policy of their target journal(s)
• Identifying and using data repositories
• Compliance with funders’ and institutions’ data sharing policies
• Data reporting standards

http://www.springernature.com/gp/group/data-policy/helpdesk
Other resources for authors and editors include

- **Recommended repositories list**
  
  http://www.springernature.com/gp/group/data-policy/repositories

- **Guidance on and published examples of data availability statements**
  
  http://www.springernature.com/gp/group/data-policy/data-availability-statements

<table>
<thead>
<tr>
<th>Statement type description</th>
<th>Template/example text</th>
<th>Published example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data generated during the study are subject to a data sharing mandate and available in a public repository that does not issue datasets with DOIs</td>
<td>[Data type e.g. “Sequence”] data that support the findings of this study have been deposited in [repository name e.g. “GenBank”] with the [primary] accession codes [list accession codes with links e.g. “KP253039”]</td>
<td>BMC Biology, Nature Communications</td>
</tr>
<tr>
<td>Data available in a public (institutional, publicly accessible) data repository</td>
<td>The [data type] data that support the findings of this study are available in</td>
<td>Nature Communications</td>
</tr>
</tbody>
</table>
Policy implementation progress – 21st February 2017

- **~800 (>30%)** Springer Nature journals have adopted a standard policy – so far
- Includes all Nature and BioMed Central journals; Springer Research Group journals being added weekly
- **~4000** data policy website visits per month since launch in July
- List of data repositories and data availability statement guidance most visited
- Policies released CC BY Dec 2016 to enable wider policy adoption and development

Community engagement through RDA & being open

Data policy standardisation and implementation

6th April 2017 - RDA 9th Plenary Meeting
IG Meeting 2-3.30pm

Background and motivations

Increasing the availability of research data for reuse is in part being driven by research data policies and the number of funders and journals and institutions with some form of research data policy is growing. The research data policy landscape of funders, institutions and publishers is however too complex (Ref: http://insights.uksg.org/articles/10.1629/uksg.284) and the implementation and implications of policies for researchers can be unclear. While around half of researchers share data, their primary motivations are often to carry out and publish good research, and to receive renewed funding, rather than making data available. Data policies that support publication of research need to be practical and seen in this context to be effective beyond specialist data communities and publications.

The prevalence of research data policies from institutions and research funders (such as the UK research councils and European Commission) is increasing, so publishers and editors are paying more attention to, standardisation and the wider adoption of data sharing policies. The international
Policy implementation & helpdesk progress

Total no. of policy types implemented

- Policy Type 1: 343
- Policy Type 2: 217
- Policy Type 3: 177
- Policy Type 4: 8

Enquires to Helpdesk by month

- July: 7 (7 total)
- August: 8 (8 total)
- September: 15 (9 total)
- October: 24 (19 total)
- November: 74 (31 total)
- December: 92 (18 total)
Measuring costs and benefits

Time needed to add data availability statement (minutes)
Statement types by journal

- **1** = Data available on request
- **2** = Data with the article and/or SI
- **3** = Data/some data publicly available
- **4** = Figure source data

<table>
<thead>
<tr>
<th>Journal Type</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life science A</td>
<td>1</td>
<td>16</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Physical sciences A</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life sciences B</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Physical sciences B</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

n = 82
Average time by statement type (minutes; 16 journals)

- 1 = Data available on request
- 2 = Data with the article and/or SI
- 3 = Data/some data publicly available

n= 185
Lessons learned (so far)

- Start with a simple and achievable position
- Make the process easy for editors
- Consult with other stakeholders (e.g. librarians, funders)
- Don’t try and solve everything at once
- Acknowledge community and operational differences – one size doesn’t fit all
- Identify examples relevant to all research disciplines
- Work with early adopters/pilot titles
- Accommodate different platforms/websites
- Be prepared to your hands dirty – journal-by-journal
- (Any kind of) Data policy can help a journal begin its data sharing journey
Questions?

Please contact: iain.hrynaszkiewicz@nature.com on behalf of the Springer Nature Research Data Policy Group:

Amye Kenall (BMC and SpringerOpen), Sowmya Swaminathan (Nature Research), Ralf Gerstner (Springer Research Journals), Iain Hrynaszkiewicz (Project lead), Mathias Astell (Open Research Group), Grace Baynes (Marketing and Development Director)

Further acknowledgements:
Varsha Khodiyar & Andrew Hufton (Scientific Data)
Andreas Vogel & Astrid Huizer (Springer Research Group)