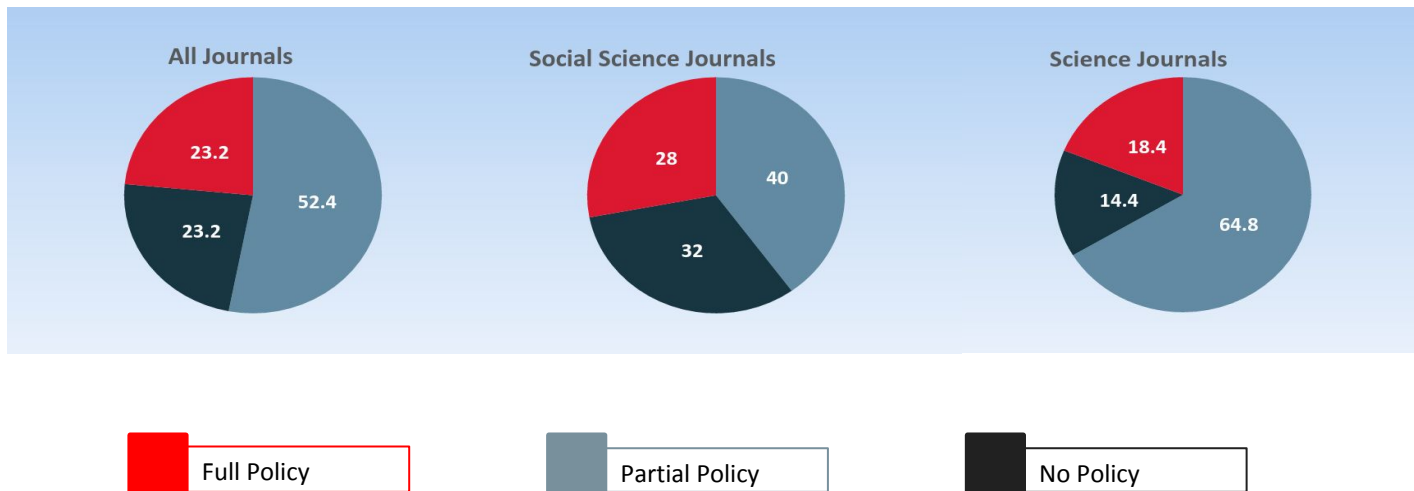# A pragmatic and practical approach to journal research data policies

Figshare webinar
Iain Hrynaszkiewicz, Head of Data Publishing, Springer Nature

Journals' research data policies can be confusing

All Journals
- 23.2
- 23.2
- 52.4

Social Science Journals
- 28
- 32
- 40

Science Journals
- 18.4
- 14.4
- 64.8

| Full Policy | Partial Policy | No Policy |

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*"

-- Naughton, L. & Kernohan, D., (2016). Making sense of journal research data policies. Insights. 29(1), pp.84–89. DOI: http://doi.org/10.1629/uksg.284

Journal policies do influence researcher behaviour
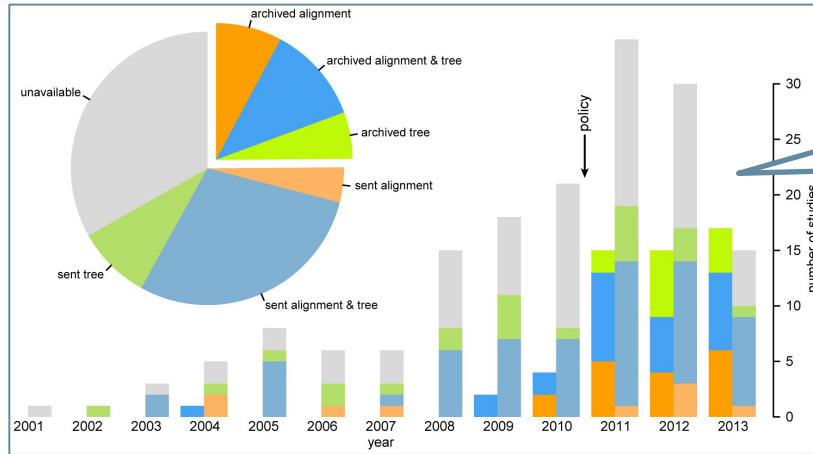
Magee *et al* (2014) show increase in sharing and deposition of data from evolution research after 2010, after this research community and its journals adopted stronger research data policies
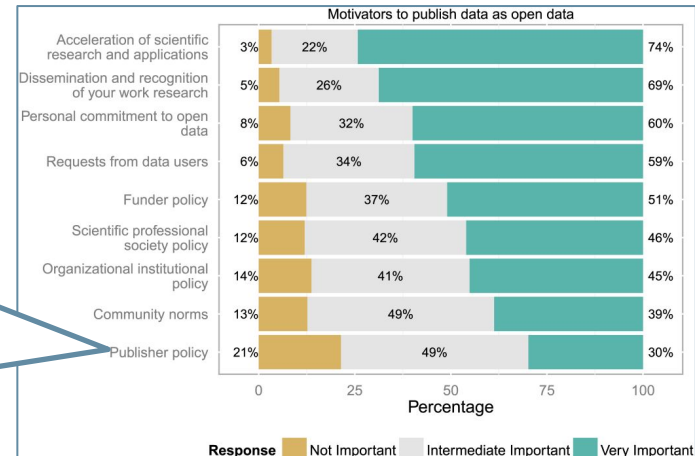
Citation and image credit (CC BY)
https://doi.org/10.1371/journal.pone.0110268

Survey data from Schmidt *et al* (2016) (n=843) identifies publisher policy as a motivator to publish data as open data

Citation and image credit (CC BY)
https://doi.org/10.1371/journal.pone.0146695

There are benefits to a stronger data policy

Data archiving can **double** the publication output of studies

A study of 7,000 NSF and NIH research projects in social sciences found that:

- Those with archived data resulted in 10 (median) publications;
- Those without archived data resulted in 5 publications[1]

Principal investigators who archived their data were more likely to publish more articles per project, and to see others build on their work

1. Pienta et al (2010) https://deepblue.lib.umich.edu/handle/2027.42/78307
2. Piwowar & Vision (2013) https://doi.org/10.7717/peerj.175
3. Henneken & Accomazzi (2011) https://arxiv.org/abs/1111.3618
4. Dorch et al (2015) https://arxiv.org/abs/1511.02512
5. Sears et al (2011) https://figshare.com/articles/Data_Sharing_Effect_on_Article_Citation_Rate_in_Paleoceanography/1222998/1

Research articles with open data are cited up to **50% more**

Analysis shows that articles with data available are cited 9-50% more, depending on the field
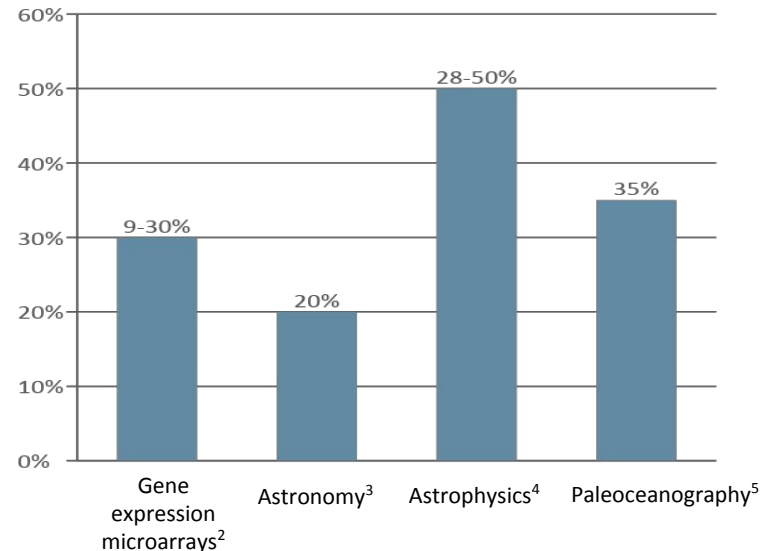
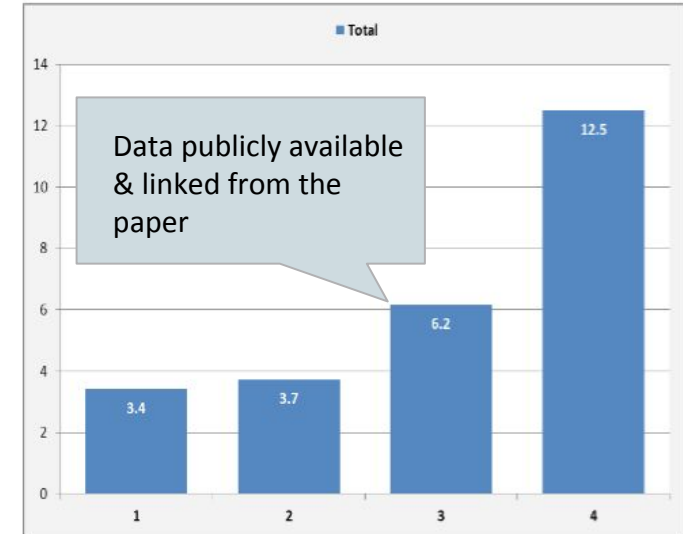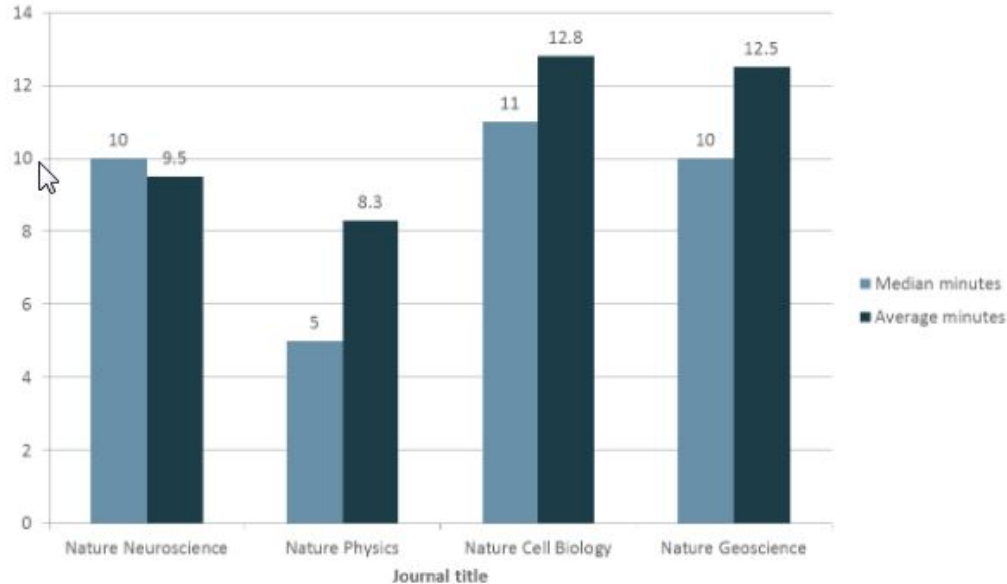# But there are also costs for publishers and editors

Figure 1. Median and average editor time to add DAS by journal (minutes).

Data publicly available & linked from the paper

Figure 3. Average time by statement type in minutes (n=329).

SPRINGER NATURE

Many developments in journal/publisher policy

Policy Types

Type 1	Type 2	Type 3	Type 4
Data sharing and data citation is encouraged but not required	Data sharing and evidence of data sharing encouraged	Data sharing encouraged and statements of data availability required	Data sharing, evidence of data sharing and peer review of data required

- Springer Nature launched a data policy standardisation initiative in 2016[1]
- Approach is practical and pragmatic, enabling all journals to adopt a policy even if they are new to data sharing
- All policies support data citation and deposition in repositories
- Similar initiatives since introduced at Elsevier, Wiley, Taylor & Francis, BMJ and others 2017-2018[2,3]

1. **Standardising and harmonising research data policy in scholarly publishing**. Iain Hrynaszkiewicz, Aliaksandr Birukou, Mathias Astell, Sowmya Swaminathan, Amye Kenall, Varsha Khodiyar. *International Journal of Digital Curation* 2017; doi: https://doi.org/10.2218/ijdc.v12i1.531
2. https://cos.io/blog/landscape-open-data-policies/
3. **Implementing Publisher Policies That Inform, Support and Encourage Authors to Share Data: Two Case Studies**. Jones, Leila, Rebecca Grant, and Iain Hrynaszkiewicz. 2019. Insights 32 (1): 11. DOI: http://doi.org/10.1629/uksg.463

Developing a research data policy framework for all journals and publishers

Iain Hrynaszkiewicz, Natasha Simons, Azhar Hussain, Simon Goudie

Problems we want to solve with this approach

- Authors and editors just want to know if they have to share data or not
- Some journals lack capacity to enforce mandates or data sharing consistently but could introduce Data Availability statements relatively easily
- Enforcing some policy features, such as data citation consistently and ensuring peer reviewer access to research data, are resource-intensive
- Aspirational policies will not be implementable quickly
- Transparency in the requirements for authors and editors needed as well being clear about what an author's experience at the journal will be, are critical

Methodology

- Review of existing publisher policy frameworks (Springer Nature, Elsevier, Wiley), and well-established policy of PLOS in 2017
- Review of other policy frameworks that include data (Transparency and Openness Promotion; TOP Guidelines) and good practice guidelines for data policy - for funders and institutions - from CODATA
- Community calls with RDA group members and others in 2017 to gather requirements for data policy from different stakeholders
- First public draft (v1.2) made available for comment Feb 2018
- More than 30 comments received from more than 20 reviewers
- Revised draft presented at RDA Plenary 13 (April 2019)

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

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

○ **Provide information**

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

● **Provide information and action**

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

14 standardised research data policy features arranged as 6 policy types (tiers/levels)

Standard text for policy 4

Policy summary for authors

By publishing in the journal authors are required to make research data available to editors and reviewers, and to readers upon request. For some research data, deposition in repositories is required and this is encouraged for all research data. For some papers, the decision to publish will be affected by whether or not authors share their research data.

Required
- Data sharing via repositori
- Data availability statement
- Data sharing on request

Optional
- Data citation
- Data sharing via repositori
- Prepare and share Data M

	require authors to deposit these in an approved repository as a condition of publication.
Data repositories	The preferred mechanism for sharing research data is via data repositories. Please see <recommended repository list> or https://repositoryfinder.datacite.org/ for help finding research data repositories. Research data of the types listed in "Mandatory data sharing (specific papers)" must be uploaded to an appropriate repository. The journal will require authors to deposit these in an approved repository as a condition of publication.
Data citation	The journal encourages authors to cite any publicly available research data in their reference list. References to datasets (data citations) must include a persistent identifier (such as a DOI). Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style.
Data licensing	The journal encourages research data to be made available under open licences

Further work

- Pilots with early adopters with success measures defined
- Scientific Data journal agreed to pilot implementation of the "type 6" policy
- Journal and publisher outreach and engagement - collaboration with STM Association
- Creation of a simple web tool to help journals select an appropriate policy type
- Collaboration with research funders groups to enable a funder policy standardisation/good practice guideline activity
- Further research/analysis on costs and benefits of data policies….

New evidence of benefits

The citation advantage of linking publications to research data

(Submitted on 4 Jul 2019)

Giovanni Colavizza, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, Barbara McGillivray

We consider 531,889 journal articles published by PLOS and BMC which are part of the PubMed Open Access collection, categorize their data availability statements according to their content and analyze the citation advantage of different statement categories via regression. We find that, following mandated publisher policies, data availability statements have become common by now, yet statements containing a link to a repository are still just a fraction of the total. We also find that articles with these statements, in particular, can have up to 25.36% higher citation impact on average.

https://arxiv.org/abs/1907.02565