# Accelerate progress with a journal research data policy framework

STM webinar, February 2020

Iain Hrynaszkiewicz, Publisher, Open Research, PLOS

# Data availability makes research more reliable

- Studies have found that unavailability of data and insufficient data curation reduce the repeatability / reproducibility (reliability) of research[1,2]

- Journal policies that promote transparency support publication of more reliable research, and are the first, simple, logical thing a publisher can do to raise awareness of issues[3]

- Journal policies support STM year of research data objective 1: **SHARE**
  - Increase the number of journals with data policies and articles with Data Availability Statements (DAS)
    - DAS report if and where data supporting the results reported in a published article are available – including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study.

1. Ioannidis JPA et al (2009). Nat Genet 41:149–155
2. Hardwicke TE et al (2018) R Soc Open Sci 5:180448
3. Hrynaszkiewicz I. (2019) In: Handbook of Experimental Pharmacology. https://doi.org/10.1007/164_2019_290

# The research data policy landscape is evolving
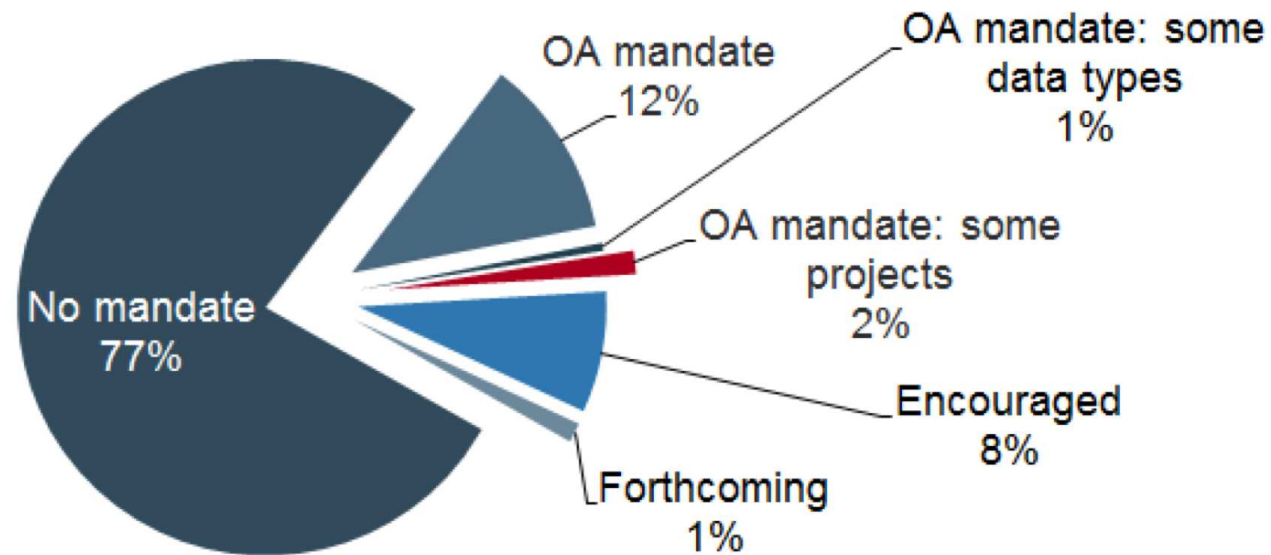


**85 (~22%) funders mandate or encourage data sharing (up from ~50 funders in 2016).**
Over 300 have no stated policy yet.

Source: Springer Nature (2019)

**Key**: "OA" means "open access data", "data sharing" or "data management"

*More funding agencies (and institutions) are introducing research data policies, which publishers and journals are obliged to support*
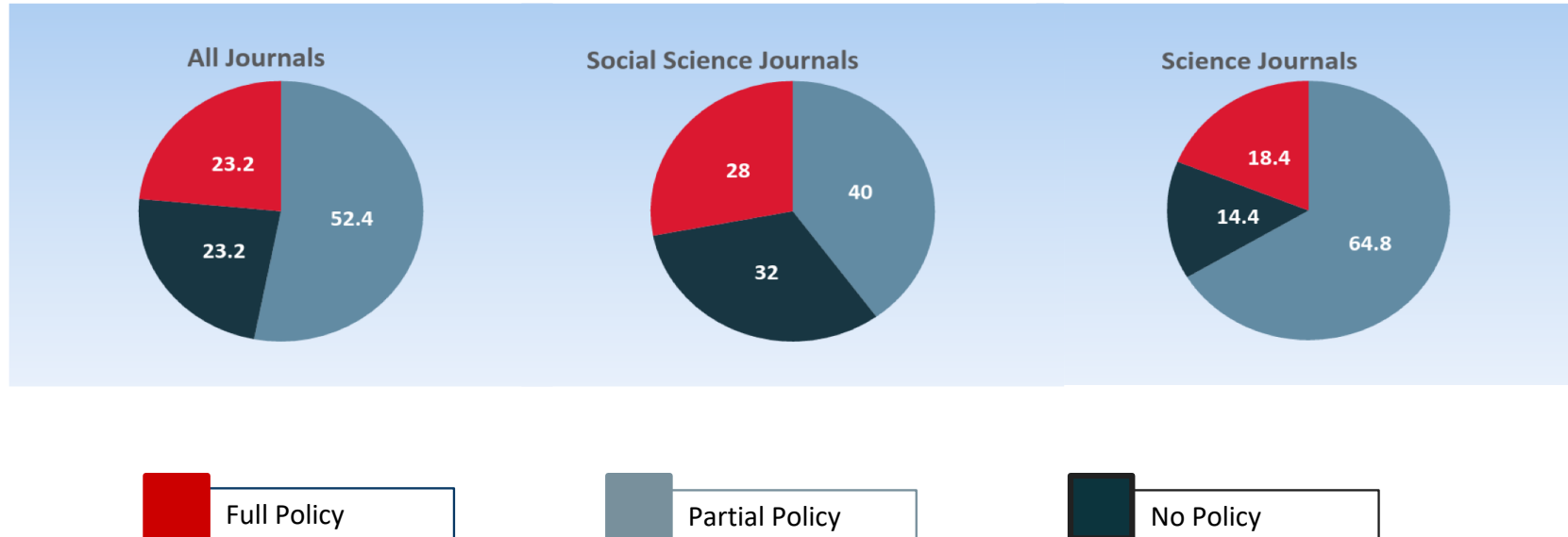
# Some research data policy early adopters

- Requirements in some communities for more than 25 years (genetic sequences, protein structure)
- Since 2014 PLOS required authors to make all data underlying the findings described in their manuscript available without restriction at the time of publication, with rare exceptions, and provide a "Data Availability Statement" (DAS)

- Since 2011, some BMC (BioMed Central) journals either required a DAS *or* encouraged authors to provide a DAS

- In 2015, all BMC journals (250+ journals) mandated a DAS for all articles

# TOP guidelines and data transparency (2015)

| | Not implemented | Level I | Level II | Level III |
|---|---|---|---|---|
| Data transparency | Journal encourages data sharing or says nothing | Article states whether data are available and, if so, where to access them | Data must be posted to a trusted repository. Exceptions must be identified at article submission | Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication |

Nosek B et al (2014) Transparency and openness promotion (TOP) guidelines. https://osf.io/vj54c/

# Journal research data policies can be confusing



All Journals
- 23.2 (Full Policy)
- 23.2 (No Policy)
- 52.4 (Partial Policy)

Social Science Journals
- 28 (Full Policy)
- 32 (No Policy)
- 40 (Partial Policy)

Science Journals
- 18.4 (Full Policy)
- 14.4 (No Policy)
- 64.8 (Partial Policy)

Legend:
- 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

# Since 2016, more stakeholders introduce policies

- 2016: Springer Nature: 4 policy types

- 2017: Elsevier: 5 policy options

- 2017: Wiley: 3, latterly 4 policy types

- 2018: Taylor & Francis: 4 policy types; BMJ Group: 3 policy tiers

- Meanwhile, more community/society initiatives such as:

  - Findable, Accessible, Interoperable and Reusable (FAIR) Data principles

  - Coalition for Publishing Data in the Earth and Space Sciences (COPDESS), American Geophysical Union Enabling FAIR data

# Challenges and opportunities

- Multiple similar but non-identical and non-agreed standard policies
- Mandatory and enforced policies are much more effective in ensuring data accessibility[1]
- Mandatory and enforced policies are more costly to implement[2]
- Different needs and expectations for data sharing in different research communities
- Increasing support for data sharing from large publishers
- Publisher and journal policies motivate researchers to share data[3]

1. Vines et al (2013) https://doi.org/10.1096/fj.12-218164
2. Grant & Hrynaszkiewicz, IJDC 2018 https://doi.org/10.2218/ijdc.v13i1.614
3. The State of Open Data Report 2019. figshare. Report. https://doi.org/10.6084/m9.figshare.9980783.v2

# Research Data Alliance (RDA) helps engage different stakeholders on shared problems



Iain Hrynaszkiewicz (PLOS), Natasha Simons (ANDS), Simon Goudie (Wiley), Azhar Hussain (Jisc), Rebecca Grant (Springer Nature)

*Formed in 2017, Group activities build on research carried by Jisc, ongoing activities of Australian Research Data Commons and work of journal publishers on data policy*

https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation

# Interest Group objectives

- Define common frameworks for different stakeholders' research data policies allowing for different levels of commitment and requirements

- Identify priority areas/stakeholders where policy frameworks can be defined **beginning with journal/publisher policy**, then considering funder policy

- For these prioritised areas, stimulate activities to:
    - Produce guidance for on complying with and implementing research data policy
    - Facilitate greater understanding of the landscape of research data policies across disciplines, institutions and learned societies

- Increase adoption of (standardised) research data policies by all stakeholders

# A research data policy framework for all publishers?



| | Data availability statement is published[1] | Data has been shared[2] | Data has been peer reviewed[3] | Example Wiley journals |
|---|---|---|---|---|
| Encourages Data Sharing | Optional | Optional | Optional | |
| Expects Data Sharing | Required | Optional | Optional | British Journal of Social Psychology |
| Mandates Data Sharing | Required | Required | Optional | Ecology and Evolution |
| Mandates Data Sharing and Peer Reviews Data | Required | Required | Required | Geoscience Data Journal American Journal of Political Science |
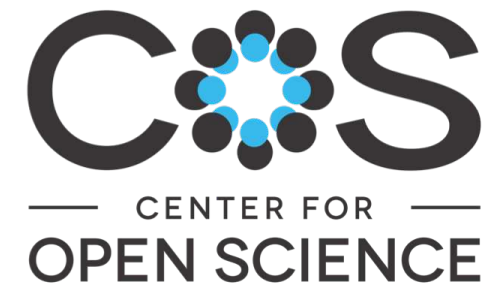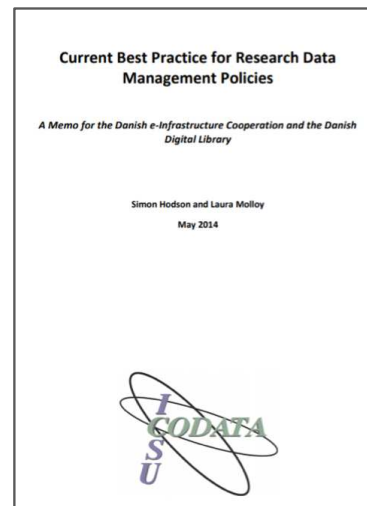
# Methodology: open development of framework

- 2017: Community calls to gather requirements from different stakeholders (researchers, publishers, funders, librarians, societies)
- 2018: First public draft (v1.2) made available for comment
- More than 30 comments received from more than 20 reviewers
- 2018: Discussion and review at RDA Plenary meetings
- Late 2018 - early 2019:
  - Revision of framework
  - Exploration of different presentation formats for tables
  - Draft Implementation requirements
  - Creation of policy templates
- Jun 2019: Publication of preprint on figshare
- Jan 2020: In press after peer review, *CODATA Data Science Journal*

# Results: 14 policy features, 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) | | | | ○ | ○ | ○ |

Hrynaszkiewicz, Iain; Simons, Natasha; Hussain, Azhar; Goudie, Simon (2019): Developing a research data policy framework for all journals and publishers. figshare. Preprint. https://doi.org/10.6084/m9.figshare.8223365.v1

**Key:**

○ = Information required

● = Information and action required

– = Not applicable

# Existing policies map to the framework

| | Policy 01 | Policy 02 | Policy 03 | Policy 04 | Policy 05 | Policy 06 |
|---|---|---|---|---|---|---|
| Definition of the research | ○ | ○ | ○ | ○ | ○ | ○ |
| | ○ | ○ | ○ | ○ | ● | ● |
| | | ○ | | ● | ● | |
| Supplementary materials | | ○ | | ● | ● | |
| Data repositories | ○ | ○ | | ● | | ● |
| | | ○ | | ○ | | ● |
| | | ○ | | | | |
| 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) | | | | ○ | ○ | ○ |

E.g. Springer Nature policy 4, Elsevier policy 5

E.g. Springer Nature policy 1

E.g. Elsevier policy 4

E.g. Wiley policy 1, Taylor & Francis Basic policy

E.g. Wiley policy 3, PLOS policy

E.g. Wiley policy 2, TOP level I

**Key:**

○ = Information required

● = Information and action required

- = Not applicable

# Example feature: Data repositories

| | | |
|---|---|---|
| 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). |

**Feature definition and evidence for its inclusion**

| | |
|---|---|
| **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. |

**Template text for journal information for authors**

# Policy summary: honest and clear requirements

Standard text for policy 3

**Policy summary for authors**

By publishing in the journal authors are required to provide a data availability statement in their articles. Authors are encouraged to share their data but not required to. The decision to publish will not be affected by whether or not authors share their research data.

**Required**
- Data availability statements

**Optional**
- Data citation
- Data sharing via repositories for all research data

**Policy summary for Editors**

The journal will encourage data sharing and ensure that every published article includes a data availability statement. The decision to publish will not be affected by whether or not authors share their research data.

**Required**
- Respond to questions about this policy.
- Data availability statements

**Optional**
- Data citation
- Data sharing via repositories for all research data

**Makes clear when sharing is optional and when it is not, and when the journal and authors must perform certain actions**

Hrynaszkiewicz, Iain; Simons, Natasha; Hussain, Azhar; Goudie, Simon (2019): Developing a research data policy framework for all journals and publishers. figshare. Preprint. https://doi.org/10.6084/m9.figshare.8223365.v1
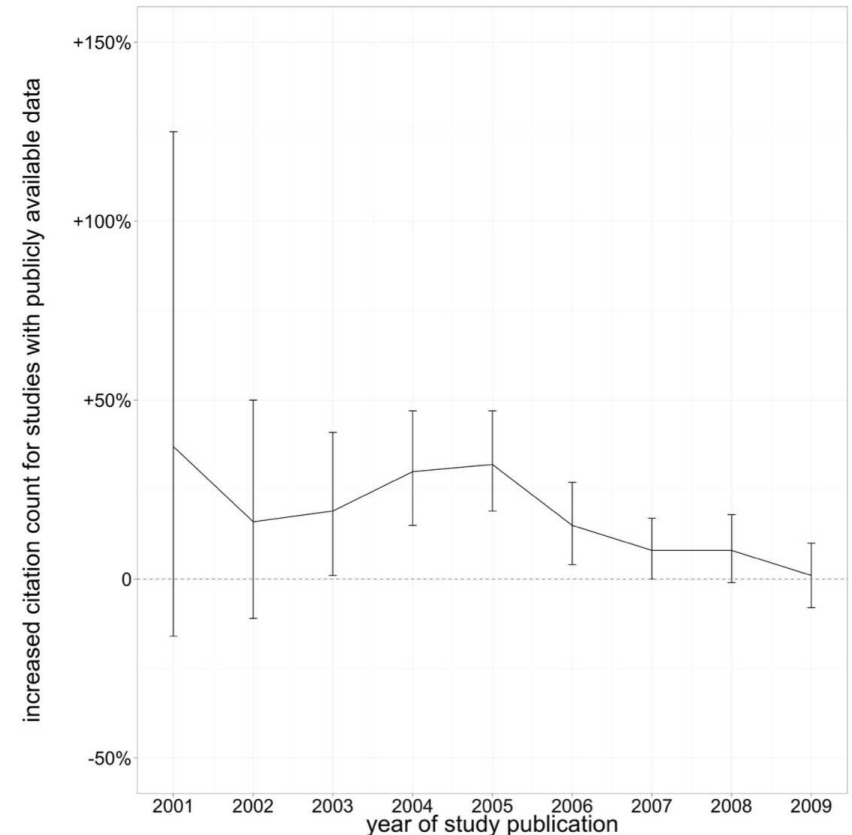
# Considerations when choosing a policy

- What are your objectives in implementing a policy?
  - Raising awareness and signalling importance of an issue (policies 1-2)
  - Increasing transparency about data availability and compliance (policy 3)
  - Increasing availability of data (policies 4, 5)
  - Increasing and verifying reusability of data (policy 6)
- What are the needs of the research discipline(s) you serve?
- Higher policy tiers require greater effort (cost) to implement
  - Implementing a data availability statement for every article will likely increase manuscript processing time
  - Simple, unverified "Available on request" statements will be faster to implement than statements that link to data in a repository (Grant & Hrynaszkiewicz, IJDC 2018 https://doi.org/10.2218/ijdc.v13i1.614)

# There are benefits to stronger policies

- Several studies of specific disciplines have found that data sharing is associated with increased citations to papers

- The largest study, of papers published in PLOS and BMC, found that linking to research data in a repository via the data availability statement was correlated with **a 25% increase in citations**

    ○ Colavizza et al (2019) https://arxiv.org/pdf/1907.02565.pdf



Piwowar & Vision (2013)
https://doi.org/10.7717/peerj.175

# Analysis of PLOS & BMC shows mandates work
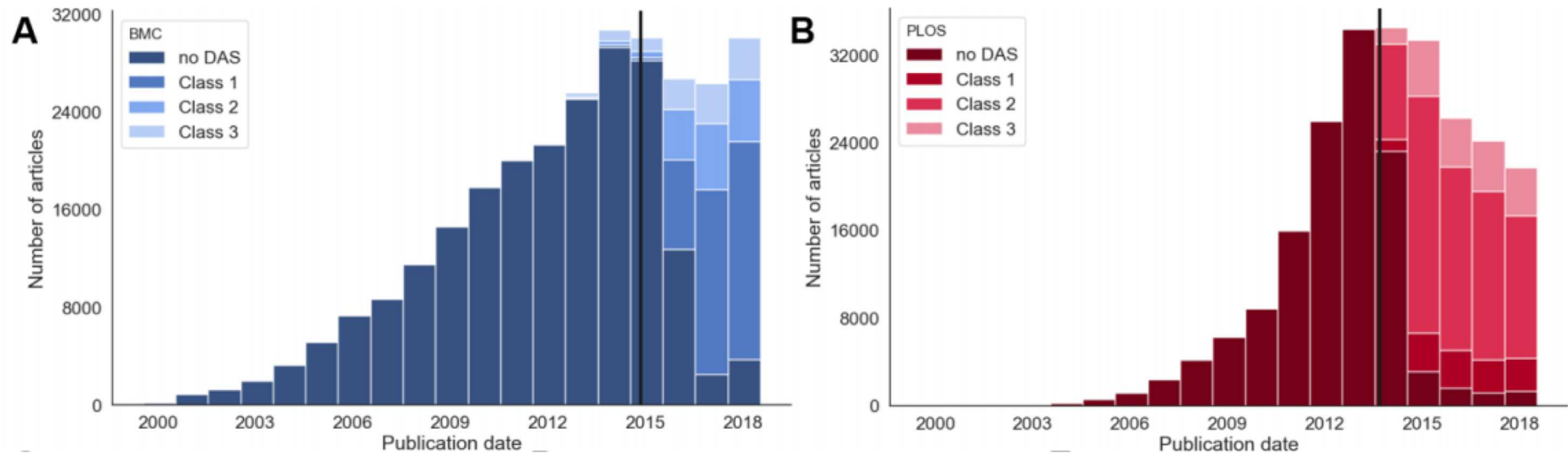


Mandatory data availability statements (DAS) introduced at PLOS in 2014, and at BMC in 2015;
DAS mostly optional at BMC 2012-2015

Colavizza et al (2019) https://arxiv.org/pdf/1907.02565.pdf

# Opportunities for review/ audit of established policies

- *Scientific Data* journal (Springer Nature; policy 6 in this framework) added a new feature to support Data Management Plans (feature 14 in this framework) in 2019
- PLOS also added DMP language to its policy in 2019
- Several large publishers have indicated intention to drive up standards, particularly for data availability statements (DAS; policy 3 and above)

### Data management plans

The journal encourages authors to prepare data management plans (DMPs) before conducting their research and encourages authors to make those plans available to editors, reviewers and readers who wish to assess them.

Authors who wish to share their DMPs in support of their submissions to the journal may provide them among their submitted files, or by email to the editors to supplement their manuscript submission. Note that some funding agencies have policies on the preparation and sharing of DMPs, and authors who receive funding from some agencies may be required to prepare DMPs as a condition of grants. More information on DMPs is available online from the Digital Curation Centre.

https://www.nature.com/sdata/policies/data-policies

### Data Management Plans

Some funding agencies have policies on the preparation and sharing of Data Management Plans (DMPs), and authors who receive funding from some agencies may be required to prepare DMPs as a condition of grants.

PLOS encourages authors to prepare DMPs before conducting their research and encourages authors to make those plans available to editors, reviewers and readers who wish to assess them.

The following resources may also be consulted for guidance on DMPs:

> Funders and institutions
> Digital Curation Centre
> DMPTool
> Data Stewardship Wizard

https://journals.plos.org/plosone/s/data-availability#loc-data-management-plans

# Thanks for listening

ihrynaszkiewicz@plos.org
rda-data-policy-standardisation-ig@rda-groups.org

For more information on the STM Research Data Year https://www.stm-researchdata.org/ please contact Joris van Rossum, STM Research Data Director, at rossum@stm-assoc.org