Policy and publishing developments for sharing data and code

CERN, 31st October 2017

Iain Hrynaszkiewicz

SPRINGER NATURE

More funders and institutions are introducing data policies

- More than 50 research funders globally have policies or mandates that require data archiving as a condition of grants, including:

  - National Science Foundation (NSF)

  - National Institutes of Health (NIH)

  - Wellcome Trust

  - Bill and Melinda Gates Foundation

Open Data Policies Map
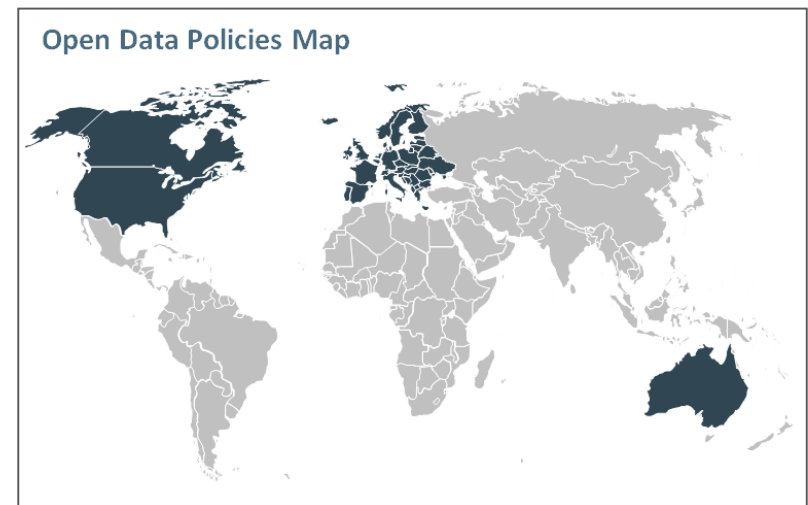
- **Some of these require data to be linked to publications including:**

  - Research Councils UK (as part of open access policy)

  - Engineering and Physical Sciences Research Council (EPSRC)

1. Hahnel, M: Global funders who require data archiving as a condition of grants. *figshare*.
https://dx.doi.org/10.6084/m9.figshare.1281141.v1 (2015)

Springer Nature research data policy initiative (July 2016)

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

SPRINGER NATURE

# What is a data availability statement (DAS)?

A statement about where data supporting the results reported in the article can be found

- The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

- All data generated or analysed during this study are included in this published article (and its supplementary information files).

- The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].

Required by many journals/publishers e.g. PLOS, Royal Society, Nature, BioMed Central, BMJ, Hindawi

SPRINGER NATURE

Data availability statements (DASs) at Nature Physics

Frequency of DAS types in Nature Physics 2016-17 (n=146)
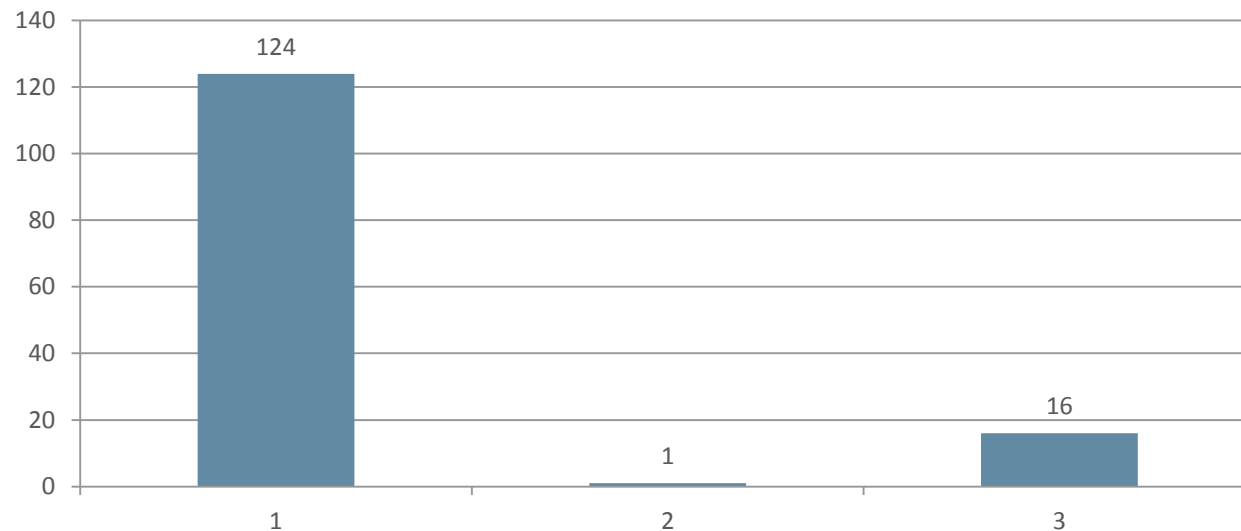
DAS coding key

1 = Data available on request

2 = Included with supplementary information files

3 = In a public repository

- 4 papers from CERN authors in study period
  - 3 used HEPdata; 1 stated data are available on request
- 10 different repositories used by authors including figshare (2), Zenodo (1) and some institutional repositories

Policy implementation and helpdesk progress – Oct 2017

- **More than 1,100** (~50%) Springer Nature journals have adopted a standard policy

- Includes all Nature and BioMed Central journals and many from Springer

- Around 10 enquiries per week received by Research Data Support Helpdesk

  - Advises authors on data policy compliance, finding repositories, writing data availability statements

  - Supports editors in identifying and implementing a data policy

- Policies and recommended repository list released under CC BY (open access) in Dec 2016 to enable wider policy adoption and development across other publishers and stakeholders

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive

Archive > Volume 537 > Issue 7619 > Editorial > Article

NATURE | EDITORIAL

Announcement: Where are the data?

07 September 2016

Community and stakeholder engagement via RDA

Co-chairs:

Natasha Simons (ANDS), Simon Goudie (Wiley), TBC (Jisc), Iain Hrynaszkiewicz (Springer Nature)

Proposed group activities can build on and be informed by research carried by Jisc, ongoing activities of ANDS and work of Springer Nature on data policy

Data policy standardisation and implementation

Home » Working And Interest Groups » ...

ℹ Group details

Status: Under community review
Secretariat Liaison: Lynn Yarmey

— History

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

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

SPRINGER NATURE

Other recent publisher initiatives on data policy

- **FORCE11 Data Citation Implementation: Publisher early adopters group** (July 2016)[1]
  - Defining and implementing data citation consistently

- **Chemistry Data Publishers Policies - supported by the RDA Chemistry Research Data Interest Group** (April 2017)[2]
  - List journal requirements for chemistry data and identify opportunities to improve practice

- **Information Providers in Astronomy, Astrophysics and High Energy Physics** (AAHEP9, May 2017)[3]
  - Publishers to implement data availability statements

- **Elsevier Research Data Guidelines** (September 2017)[4]
  - Introduced framework and guidelines for all journals' research data policies

- **Wiley Data Sharing and Citation policies** (September 2017)[5]

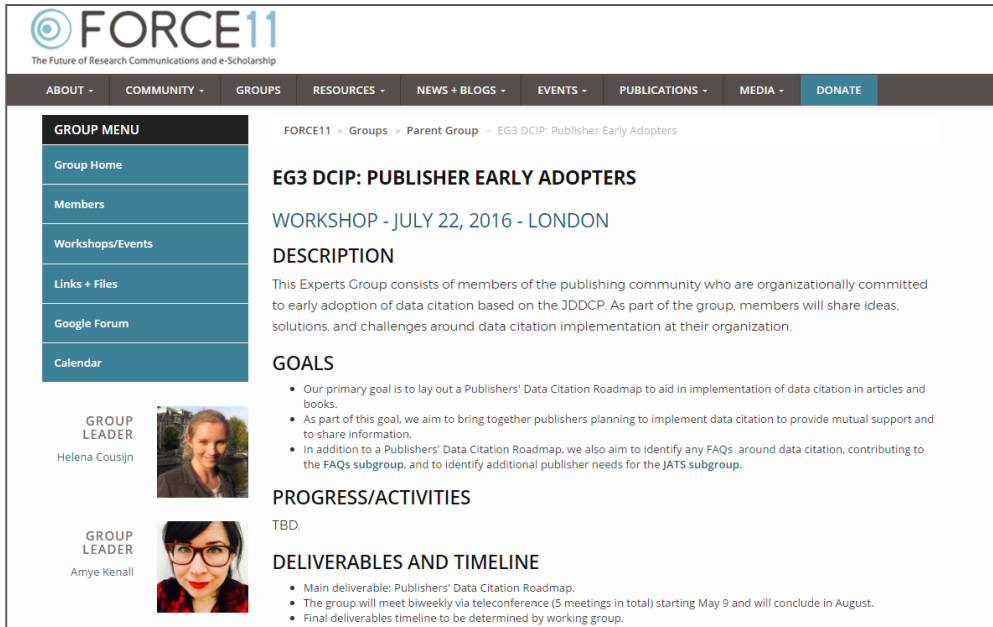1. *A Data Citation Roadmap for Scientific Publishers. bioRxiv 100784; doi:* https://doi.org/10.1101/100784
2. https://www.rd-alliance.org/groups/chemistry-research-data-interest-group.html
3. https://docs.google.com/document/d/1Ok3F3qRdz1_HIz9y9wC6titTyiRgp2HdwcdtEJWQ0vA/edit?usp=sharing
4. https://www.elsevier.com/authors/author-services/research-data/data-guidelines
5. https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-open-access/open-access/data-sharing.html

SPRINGER NATURE

Publishers collaborating to implement data citation

https://www.force11.org/group/dcip/eg3publisherearlyadopters

- Primary goal is to lay out a Publishers' Data Citation Roadmap to aid in implementation of data citation in articles and books.
- Key part of roadmap is data policy and how these are presented to researchers/readers at publisher and/or journal level
- Data citation is a key part of all journal/publisher data policies
- Also defines content production (XML) standards for identifying data citations

Reference:

A Data Citation Roadmap for Scientific Publishers.

Helena Cousijn, Amye Kenall, Emma Ganley, Melissa Harrison, David Kernohan, Fiona Murphy, Patrick Polischuk, Maryann Martone, Timothy Clark
bioRxiv 100784; doi: https://doi.org/10.1101/100784

SPRINGER NATURE

Publishers collaborating to improve data-article links – Scholarly Link Exchange (Scholix)

Scholix enables connections between articles and datasets across publishers and repositories.

Article-data links should be exchanged in a dynamic and standardized way, enhancing discoverability of all contributors' content.

Information about these links is stored in open hubs (e.g. Crossref, DataCite) so other systems can extract relevant connections.

http://www.scholix.org/guidelines
https://www.rd-alliance.org/groups/rdawds-publishing-data-services-wg.html

Policies and pilots on code sharing and peer review

Code share

Papers in Nature journals should make computer code accessible where possible.

A theme in *Nature*'s ongoing campaign for the replicability and reproducibility of our research papers is that key components of publications should be available to peers who wish to validate the techniques and results.

A core element of many papers is the computer code used by authors in models, simulations and data analysis. In an ideal world, this code would always be transportable and easily used by others. In such a world, our editorial policy would be to insist on sharing to allow free use, as we already do (as far as is practicable) with data and research materials. Unfortunately, such an ideal is not easy to attain owing to the amount of extra funding and effort it would require to render some major pieces of code shareable. Nevertheless, we at *Nature* and the Nature research journals want to encourage as much sharing as possible.

Climate modellers have made some strides in this regard. The journal *Geoscientific Model Development* has a good example of such a policy (see go.nature.com/iv8g1w), and an article in *Nature Geoscience* dis-cu

the obstacles (S. M. Easterbrook *Nature Geosci.* 7, 779–781; 2014).

As a leading example of transparency policies in other disciplines, the data journal *GigaScience* requires code used in its papers to be available, and hosts it in a way that allows others to analyse the data in publications. One point made by Easterbrook is that even if the code is shared, others might often make little or no use of it, but on some occasions the take

Nature and the N
of practices in the d
puter code in all cas
at least indicating v
mandates that whe
require a statement
out any restriction
where they conside
sharing will be eval
a paper if importa
dedicated section
code can be placed
put together best-p

⊃ NATURE.COM
To comment online,
click on Editorials at:

5 3

Sharing Software

We expect that all researchers submitting to PLOS submissions in which
relevant software available without restrictions upon publication of the wor
time regardless of versions or upgrades. If the original software is not able
facsimile.

Software manuscripts

Software submitted to PLOS must meet the following requirements:

> Based on open source standards
> Conform to the Open Source Definition
> Deposited in an open software archive (see "Depositing software," below)
> Included in the submission as supporting information
> Linked directly from the manuscript file

If the software or algorithm is not central to the manuscript, we also encourage authors to make all available.

Software and code

Any previously unreported software application or custom code described in the manuscript should be available for testing by reviewers in a way that preserves their anonymity. The manuscript should include a description in the Availability of Data and Materials section of how the reviewers can access the unreported software application or custom code. This section should include a link to the most recent version of your software or code (e.g. GitHub or Sourceforge) as well as a link to the archived version referenced in the manuscript. T
a DOI or other un
published, the sof
to use it for non-c
material transfer
manuscript shoul
discuss the tool in

nature neuroscience

Home | Current issue | Comment | Research | Archive ▾ | Authors & referees ▾ | About the journal ▾

home ▸ archive ▸ issue ▸ editorial ▸ full text

NATURE NEUROSCIENCE | EDITORIAL

Extending transparency to code

Nature Neuroscience **20**, 761 (2017) | doi:10.1038/nn.4579
Published online 25 May 2017

📄 PDF | ⬇ Citation | 🔑 Rights & permissions | 📊 Article metrics

Reproducibility initiatives seek to promote greater transparency and sharing of scientific reagents, procedures and data. Less recognized is the need to share data analysis routines. *Nature Neuroscience* is launching a pilot project to evaluate the efficacy of sharing code.

New publishing options in response to these policy shifts

Thank you

iain.hrynaszkiewicz@nature.com

researchdata@springernature.com

@iainh_z

SPRINGER NATURE