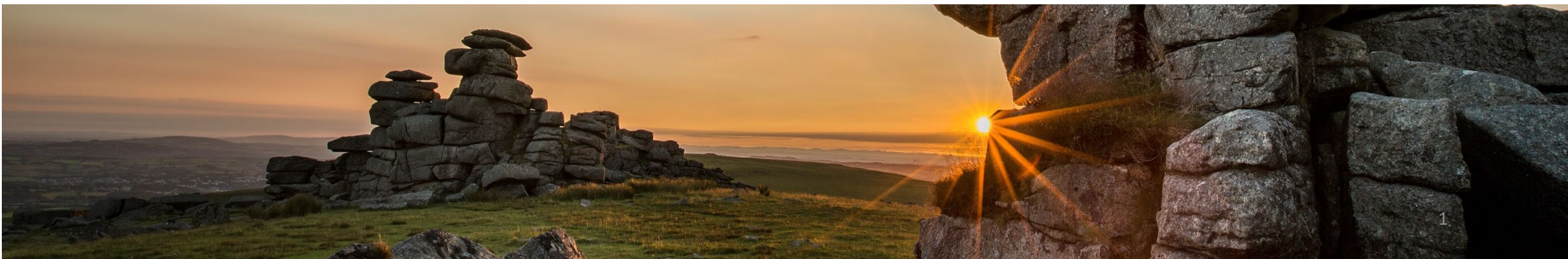# Developing a Data Policy:
A Society Perspective, 2015–2019

Sarah Gibbs, Head of Production, Geological Society of London
sarah.gibbs@geolsoc.org.uk

About the Geological Society of London

- UK's professional body for Earth science
  - founded in 1807, oldest geological society in the world
    - Burlington House, London
    - Publishing House, Bath
  - worldwide membership of >12,500
  - world-leading communicator of Earth science via
    - scholarly publishing
    - library and information services
    - scientific conferences
    - education activities
    - outreach to the general public
    - support policy-making and public debate

GSL Data Policy 2015

- Responding to UKRI and NSF
  o created GSL Data Policy
  o co-signatory on the Statement of Commitment for the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS) in 2015
  www.copdess.org/statement-of-commitment
- GSL Data Policy
  o reflected COPDESS and Force11 data citation guidance
  o available on our website at https://www.geolsoc.org.uk/Publications/Data-Policy
  o directed authors to it
- Also set up a figshare portal
  o data that was 'homeless'
  o authors' supplementary material
- promoted new policy... **but** ...

GSL Data Policy 2015–2018

- … not much happened in terms of author behaviour and data citation habits

- Became apparent we needed to
  - encourage and support authors and editorial boards
  - engage with our authors

GSL Data Policy 2019

- Time to revisit and update Data Policy
  - in 2016 "FAIR Guiding Principles for scientific data management and stewardship" published (https://www.nature.com/articles/sdata201618)
    - **f**indable, **a**ccessible, **i**nteroperable and **r**eusable
  - in 2018 COPDESS set up own '**Enabling FAIR Data Project'**
    - http://www.copdess.org/enabling-fair-data-project/enabling-fair-project-overview/

Enabling FAIR Data Project

- GSL are currently seeking Council agreement to becoming a signatory

- Becoming a signatory to updated COPDESS means we will:
  - review the commitment elements
  - develop a plan to address gaps
  - complete the plan within one year of signing (or soonest)

- New COPDESS principles impact on **all areas** of Society activity:
  - actions for the Publishing House
  - actions for the Society

Actions for the Publishing House

- Insert Data Accessibility Statements

- Actively seek out and highlight references to data & software sources

- Mandate ORCID for corresponding authors

- Encourage use International Geo Sample Numbers (IGSN)

- Insert and send downstream author contributions (CRediT)

- Ensure datasets & software are correctly tagged and sent downstream

- Update our instructions for authors, editors and reviewers highlighting data sharing and citation, the FAIR DATA project and data management practices

Actions for the Geological Society

- Develop a FAIR data landing page linking to:
  - information for students and researchers
  - winners of research grants

- Make GSL's London premises, Burlington House, available to Specialist Groups or other scientists that wish to hold meetings or workshops on data best practice

- Encourage nominators for awards to consider whether their nominee has engaged in FAIR data practices, if relevant

Watch this space

We are a work in progress and will be happy to report back next year ….

Implementing data citation

Kirsty McCormack

Publishing Services Manager

The Company of Biologists

Data sharing policy

2014: Force 11 - Joint Declaration of Data Citation Principles (2014)
Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.)
San Diego CA: FORCE11; 2014 https://doi.org/10.25490/a97f-egyk

2015: Transparency Openness Promotion (TOP) guidelines (https://cos.lo/top)

In early 2016, we added a Data Availability section to the footnotes all research papers, and prompt authors at proof stage

Data availability
Any datasets supporting your work should be made publicly available at the time of publication. Please provide details of repository name, accession numbers or other identifiers such as a doi and, where possible, include a hyperlink to the web address of the dataset. For more specific information regarding the types of data that must be disclosed, please see our journal policies page.

Data availability statement (DAS)

gene.
using
rimer

Acknowledgements
We thank all patients who donated tissues for this research. This work would not have been possible without their generosity.

Competing interests
The authors declare no competing or financial interests.

probe
with
l was
and
kage.
r 3.1,
tional
ction.
iance
lar to
of the
1224
taset.
rmed
finity
in R
luster
arson
o be
luded
ed by
each
ative,

Author contributions
Conceptualization: F.S., L.C.L., M.M.P.; Methodology: F.S., D.P., A.W., M.M.-Z.; Validation: F.S., D.P., O.F., A.W.; Formal analysis: F.S., M.K., C.T., L.C.L.; Investigation: F.S., D.P., A.W.; Resources: F.S., D.P., K.A.N.R., K.K.N., C.-W.C., M.M.-Z., D.R.N.; Writing - original draft: F.S.; Writing - review & editing: M.K., A.W., C.-W.C., L.C.L., M.M.P.; Visualization: F.S., M.M.P.; Supervision: D.R.N., L.C.L., M.M.P.; Funding acquisition: M.M.P.

Funding
This work was supported by the National Institute of Child Health and Human Development (R01-NIH HD07110 to M.M.P.). M.M.P. and L.C.L. were also supported by a grant from the California Institute for Regenerative Medicine (RN3-06396 to M.M.P.). Deposited in PMC for release after 12 months.

Data availability
Microarray data have been deposited in the Gene Expression Omnibus (GEO) database under accession numbers GSE100053, GSE100052, GSE100051, GSE127824 and GSE100279.

Supplementary information
Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.156273.supplemental

Data sharing with Dryad

- In Feb 2016, we partnered with Dryad

- Integrated with BenchPress to allow authors to submit their data at submission

- Data can be made securely available to referees and automatically made public when the paper is published

- Dryad charges a small fee, waivers are available

- Data packages are given a permanent, citable DOI

- Bidirectional linking between the data and the paper
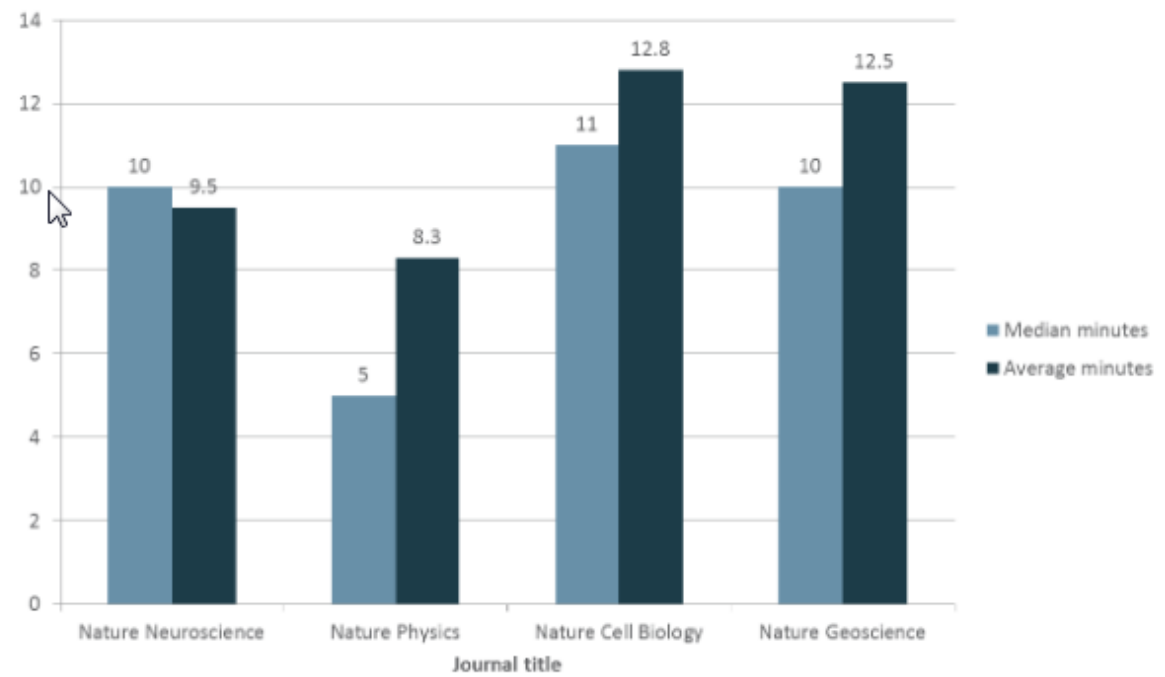
Impact of DAS on publication workflow

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

The Impact on Authors and Editors of Introducing Data Availability Statements at Nature Journals, Rebecca Grant, Iain Hrynaszkiewicz, Springer Nature https://doi.org/10.2218/ijdc.v13i1.614

Data citation policy

Force 11 - JOINT DECLARATION OF DATA CITATION PRINCIPLES (2014)

In late 2016, we updated our policy to *recommend* that authors include data citations in the reference list, in addition to the Data Availability section:

Development endorses the Force 11 Data Citation Principles and recommends that all publicly available datasets be fully referenced in the reference list with an accession number or unique identifier such as a DOI.

Only a requirement for Dryad data deposits.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P. and Beerli, P. (2001). Data from: The strength of phenotypic selection in natural populations. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.166.

Open citation policy

April 2017, Initiative for Open Citations (I4OC) https://i4oc.org/

Established by OpenCitations, the Wikimedia Foundation, PLOS, eLife, DataCite, and the Centre for Culture and Technology at Curtin University

"a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data"

July 2017, we added references to our metadata deposits:

- Crossref turn on reference distribution for all DOI prefixes
- HighWire add references to article metadata deposited with Crossref

Add data citations to reference metadata

What's up with Data Citations, Scholarly Kitchen (2018), Tim Vines
https://scholarlykitchen.sspnet.org/2018/05/28/whats-up-with-data-citations/

Crossref: In the metadata deposit, publishers can deposit links to data for their content in two places:

1. Bibliographic references
2. Relations type: Publishers include the data link in the relationship section of the metadata

Crossref recommend: if the dataset is generated as part of the article's research results, use the relation type "is-supplemented-by"

if the dataset was produced by a different set of researchers than the article authors, use the relation type "references"

Once deposited, data citations across journals (and publishers) are then aggregated and made freely available

Specify relationship type in xml

For data citations in JATS xml:

- Ask HighWire to enable data citation deposits in Crossref

- To allow accession numbers to link correctly, the file provider should tag data deposits using appropriate ext-link-type database (HighWire can provide list of supported databases, e.g. ncbi:gene, ncbi:geo, genpept, swissprot)

- File provider should use the <ext-link> element and add a special specific-use attribute
    @specific-use="dataset references"
    @specific-use="dataset is-supplemented-by".
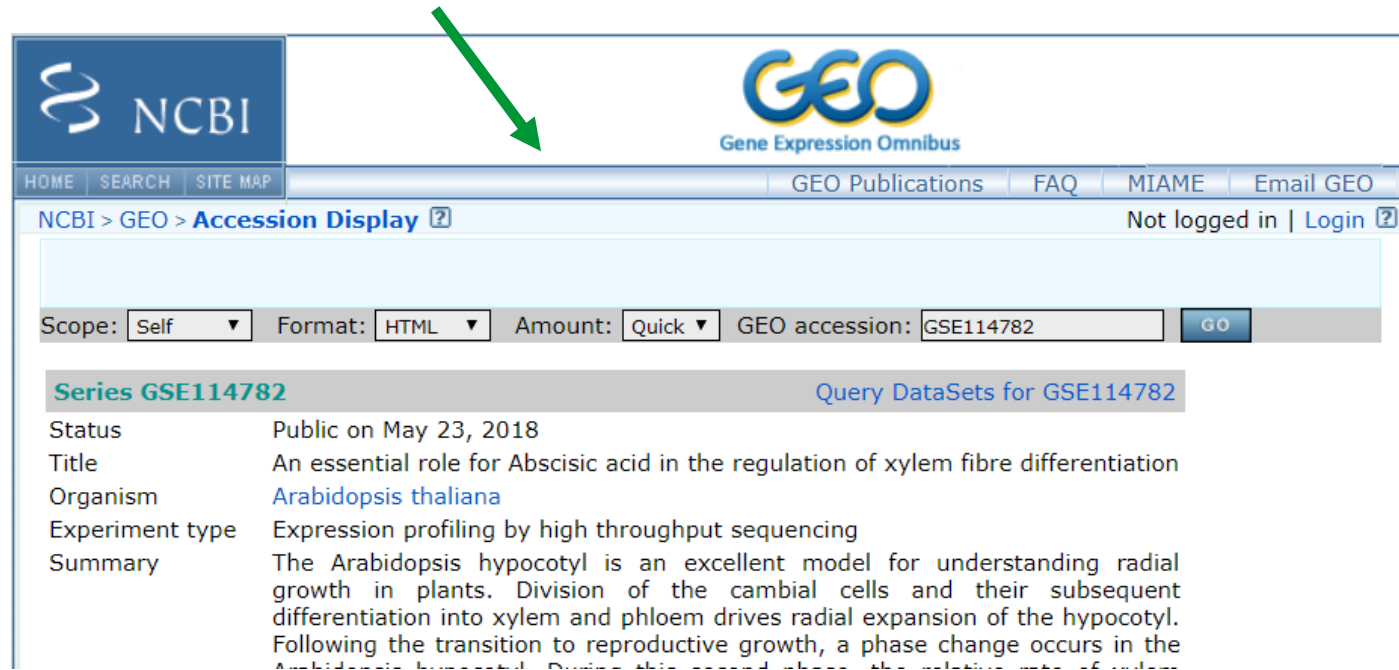
e.g. RNAseq data has been deposited in GEO under accession number GSE103403.

<ext-link ext-link-type="ncbi:geo" xlink:href="GSE103403" specific-use="dataset is-supplemented-by">GSE103403</ext-link>

Accession numbers link to corresponding database

What's next?

Research Data Alliance:

Developing a research data policy framework for all journals and publishers

Iain Hrynaszkiewicz, Natasha Simons, Azhar Hussain and Simon Goudie

Figshare Preprint: https://doi.org/10.6084/m9.figshare.8223365.v1

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 [34].

Thank you for listening

Any questions?

The Company of Biologists