# The role of journals in (and publishers) in data sharing

Open in practice, Reading
30th March 2017

Iain Hrynaszkiewicz

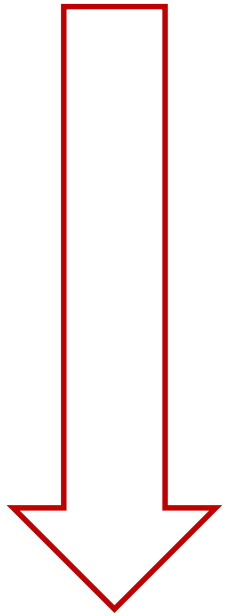


SPRINGER NATURE

# Not all research data are open data

Different levels of openness in research data publishing:

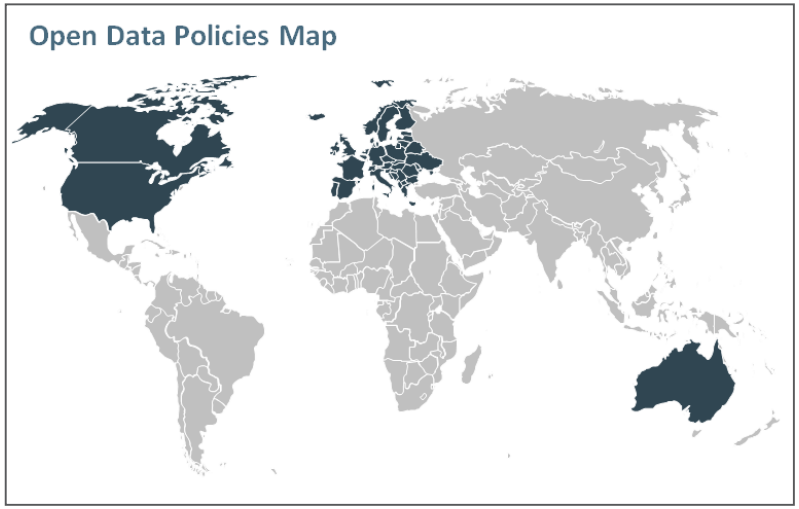
1. Accessible only to an individual researcher/group
2. Accessible to others on (reasonable) request
3. Published as electronic supplementary material
4. Deposited in a general or institutional data repository (e.g. figshare)
5. Deposited in a subject/community specific data repository



**More open**

# Research funders' policies

- More than 40 research funders globally have policies or mandates that require data archiving as a condition of grants (1), including:
  - National Science Foundation (NSF)
  - National Institutes of Health (NIH)
  - Wellcome Trust
  - Bill and Melinda Gates Foundation
  - Research Councils UK
    - Medical Research Council
    - BBSRC
    - ESRC

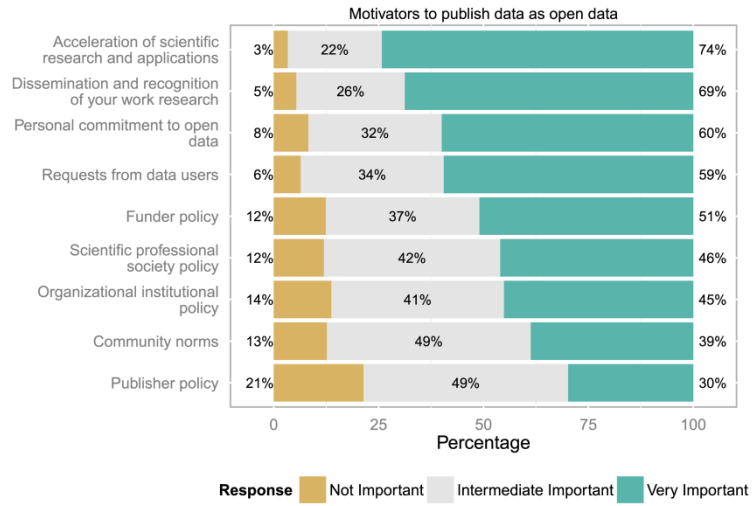



- **Some of these require data to be linked to publications including:**
  - Engineering and Physical Sciences Research Council (EPSRC)

(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)

# What motivates researchers to share data?

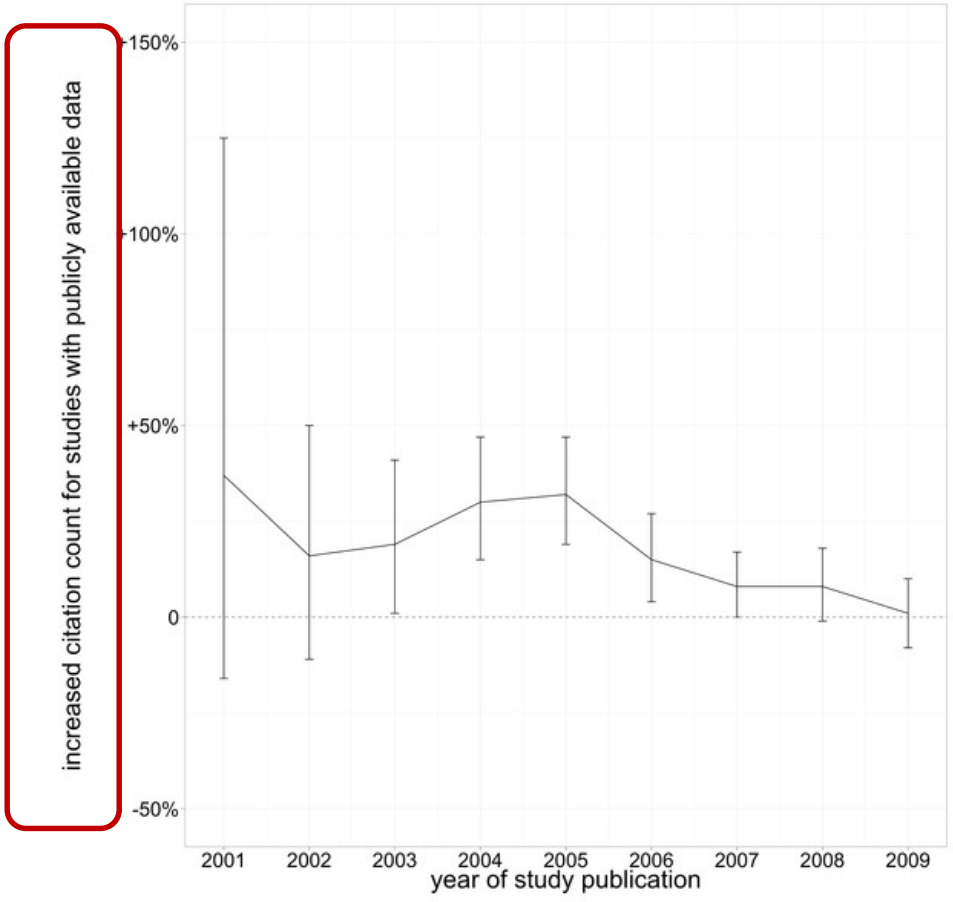
- **97%** - to accelerate research and its applications[1]
- **96%** - increased visibility and discovery  of their research data[1,2]
- **95%** - increased usability of their research data[2]
- **>90%** - credit mechanism for deposit of data[1,2]
- **88%** - to comply with funder policy[1]



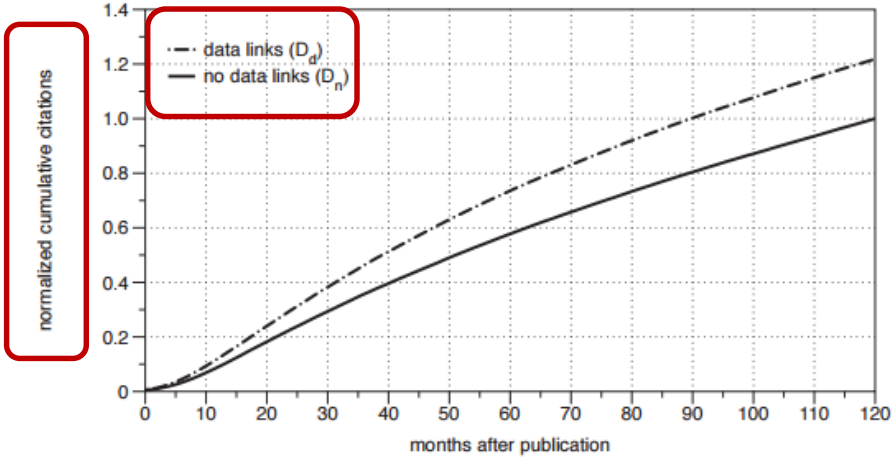

1. Schmidt et al. (2016). PLoS ONE 11(1): e0146695. doi:10.1371/journal.pone.0146695 (**n=1248**) (& image credit, CC BY)
2. Nature Publishing Group (2014): Data publication survey - raw data. figshare. http://dx.doi.org/10.6084/ m9.figshare.1234052 (**n=387**)

SPRINGER NATURE

# Publicly available data/links to data may = more citations



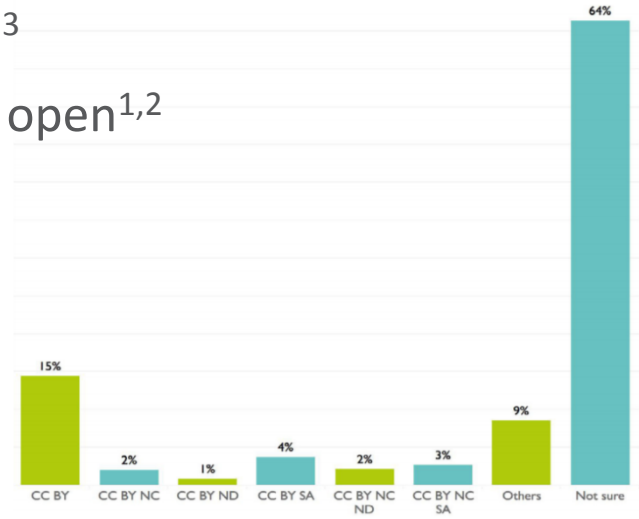

Source: https://doi.org/10.7717/peerj.175/fig-2




Source: http://arxiv.org/pdf/1111.3618v1.pdf

# What challenges do researchers face?

- **64%** unsure about open licensing of research data[1]
- **56%** do not use a metadata standard[2]
- **54%** would like more guidance complying with funder policies[1]
- **54%** do not have enough time to make data available[1]
- **45%** unaware of a repository for some of their data[3]
- **39%** uncertain about meeting costs of making data open[1,2]




1. Treadway et al. (2016). figshare. https://dx.doi.org/10.6084/m9.figshare.4036398.v1 (**n= 2061**) (& image credit CC BY)
2. Tenopir et al. (2011). PLoS ONE 6(6): e21101. doi:10.1371/journal.pone.0021101 (**n=1315**)
3. Nature Publishing Group (2014). figshare. http://dx.doi.org/10.6084/ m9.figshare.1234052 (**n=387**)

# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships

# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships

SCIENTIFIC DATA

A peer-reviewed open-access journal
Biodiversity Data Journal
Making your data count! ISSN 1314-2828 (online)

DATA IN BRIEF
MAKE YOUR DATA COUNT

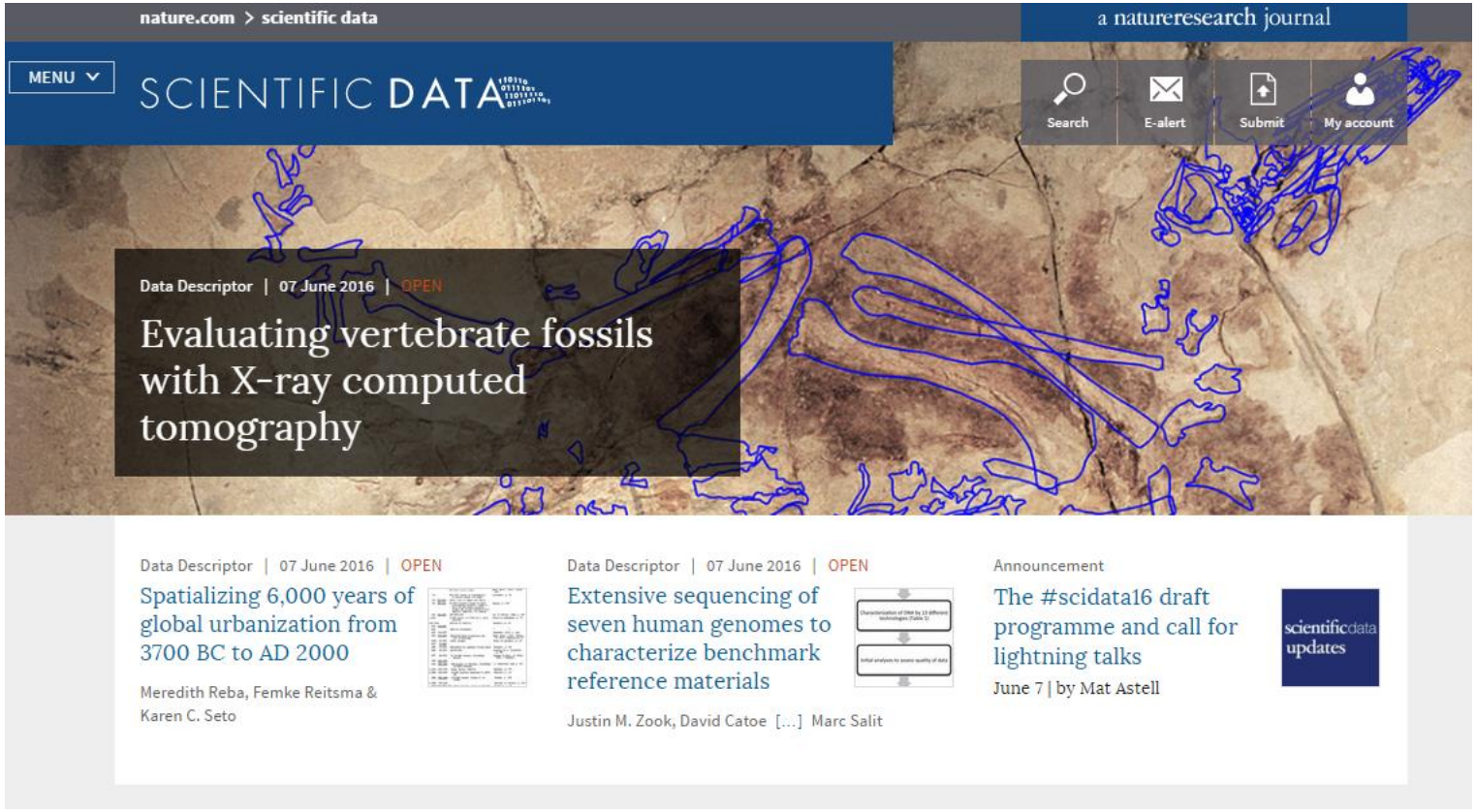
(GIGA)n SCIENCE

BMC Research Notes


F1000Research

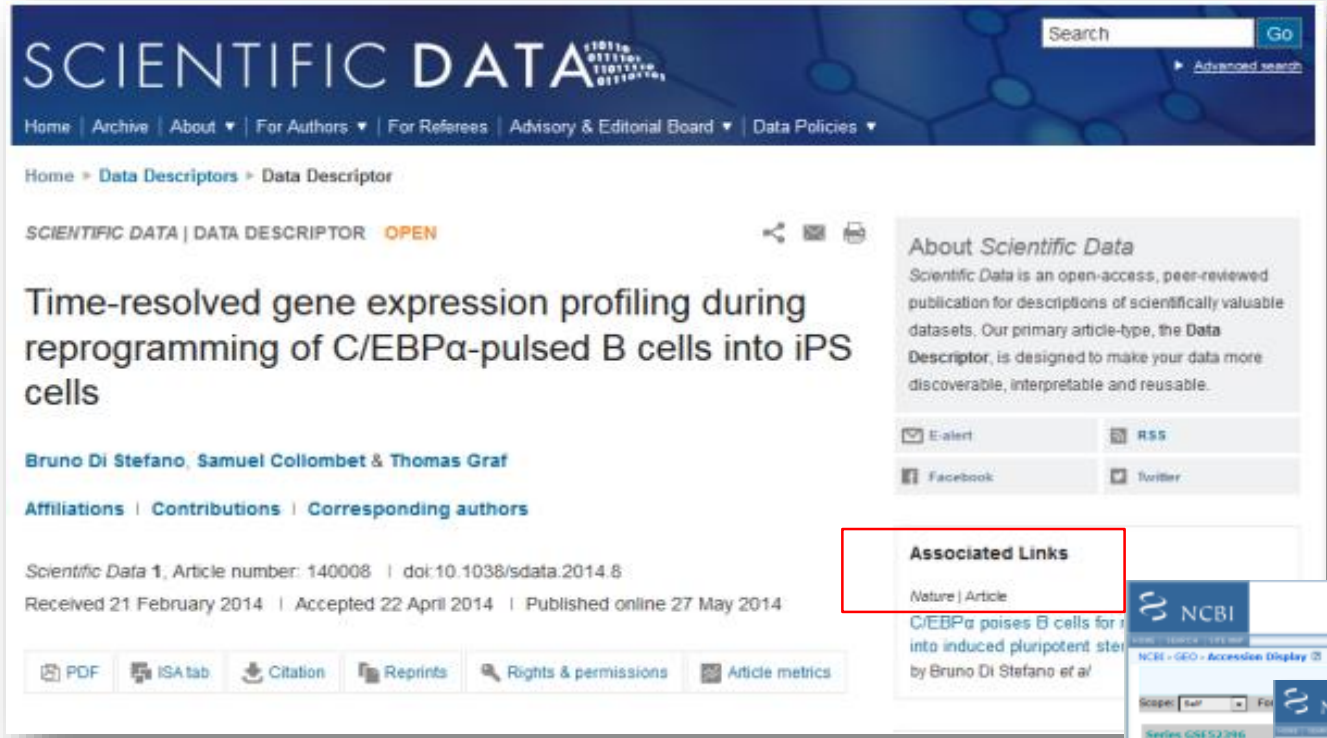
SPRINGER NATURE

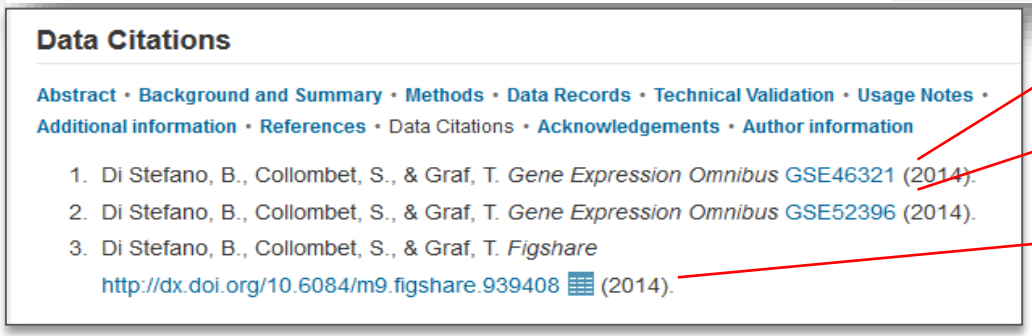
# *Scientific Data*

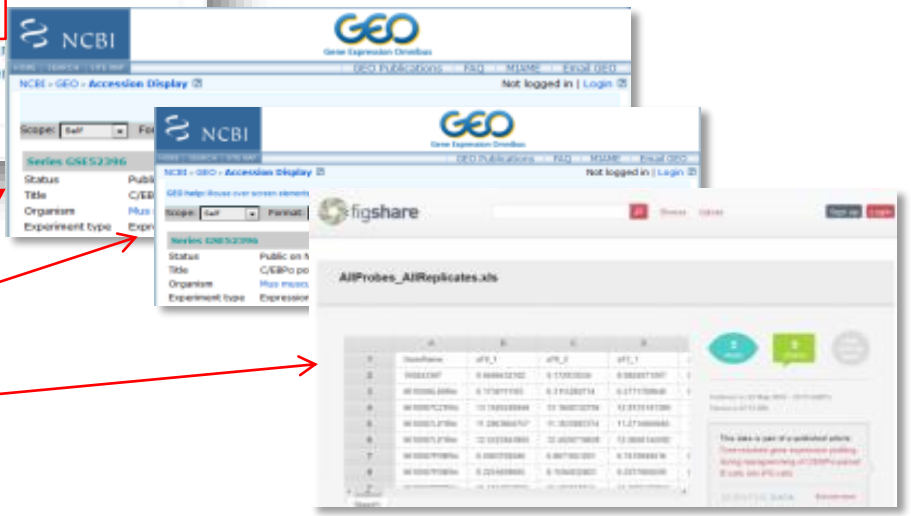


# The data descriptor article

- **Associated *Nature* Article**
- Data at **fig**share & NCBI GEO
- Integrated **fig**share data viewer



# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships

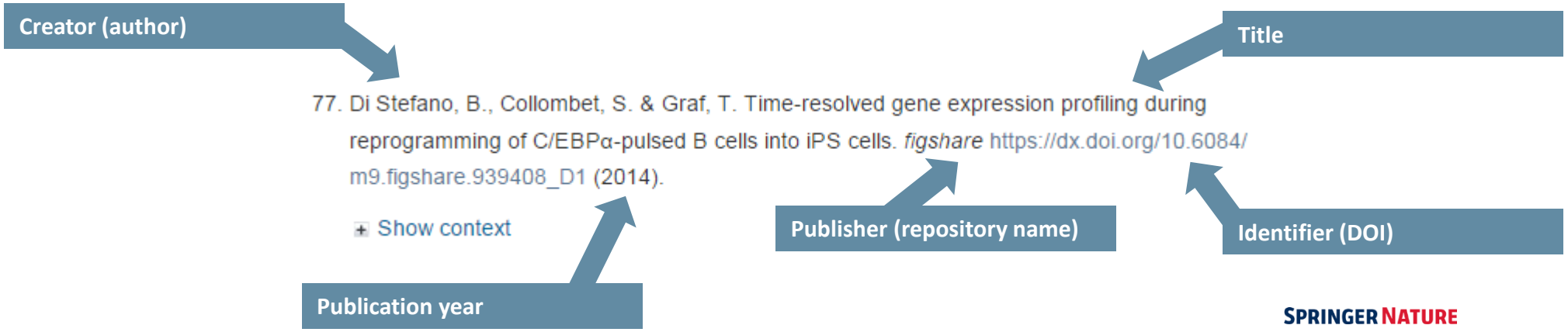
BioMed Central
The Open Access Publisher

DC1
Data Citation Principles

nature

# What is a data citation?

- *A reference made to data in the same way as researchers routinely provide a bibliographic reference to journal articles and books*
- When public datasets have Digital Object Identifiers (DOIs), **very similar to citing a journal article**
- Include the minimum information recommended by DataCite and follow Nature style i.e.
  - authors, title, publisher (repository name), identifier, year



# Data are increasingly recognised as a 1st class object

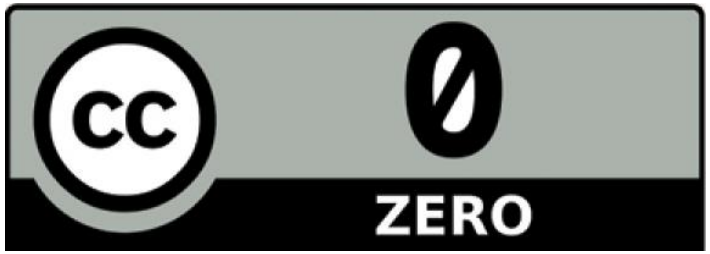
**HEFCE and the REF**: http://blog.hefce.ac.uk/2015/09/01/opening-up-research-data/ *"Via the REF,* ***we fully recognise data as an equally valid form of research output****, and, through our open access policy for the next REF, we plan to reward research environments that deliver open access to a wider set of outputs than just journal articles and conference papers."*

**National Science Foundation** (US) *"For all new grant applications from 14 January, the US National Science Foundation (NSF) asks a principal investigator to* ***list his or her research "products" rather than "publications" in the biographical sketch section.*** *This means that, according to the NSF, a scientist's worth is not dependent solely on publications.* ***Data sets, software and other non-traditional research products will count too****."*
http://www.nature.com/nature/journal/v493/n7431/full/493159a.html

# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships

BioMed Central
The Open Access Publisher



# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships

SCIENTIFIC DATA

# Making data peer review a reality

**Peer review at *Scientific Data* focuses on:**

- Completeness (can others reproduce?)
- Consistency (were community standards followed?)
- Integrity (are data in the best repository?)
- Experimental rigour and technical quality (were the methods sound?)

**Does not focus on:**

- Perceived impact/importance
- Size/complexity of data

# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships


# Helping researchers find the right repository

- Recommended repositories list (>80 repositories)
- http://www.springernature.com/gp/group/data-policy/repositories

***What makes a good data repository?***

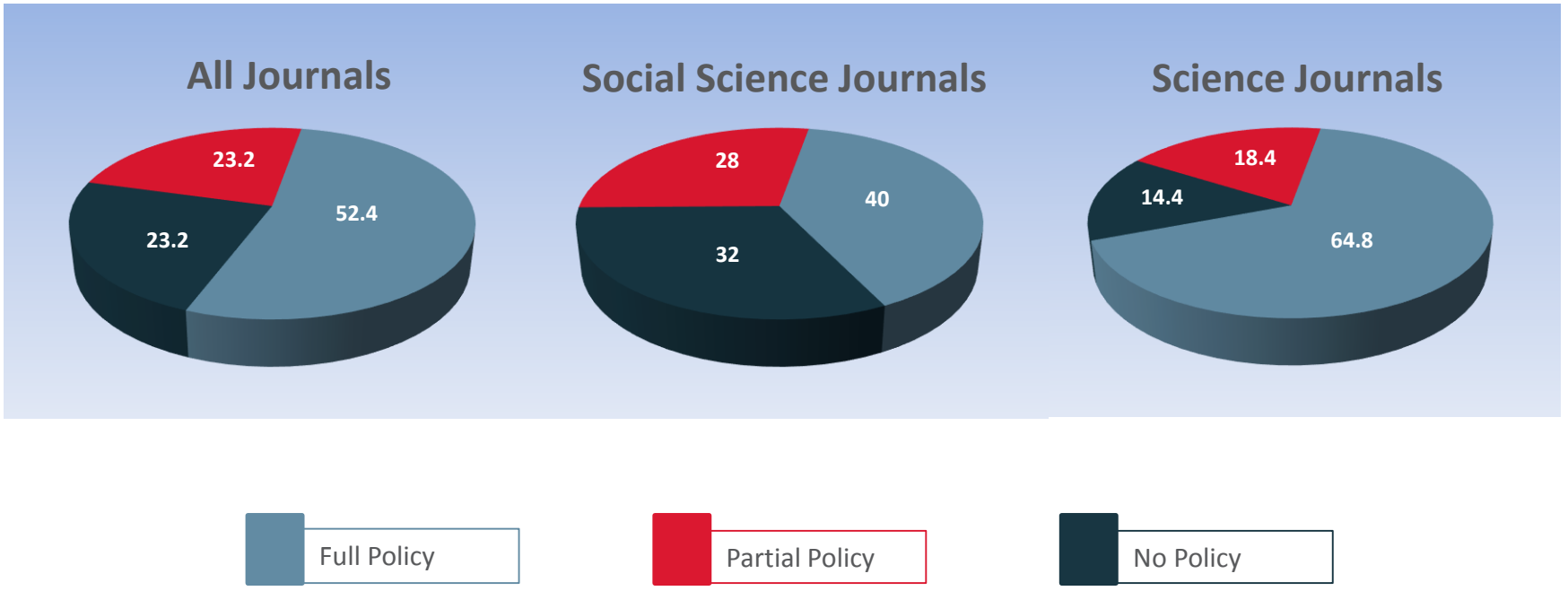
Data repositories for data supporting peer-reviewed publications generally should:

I. Ensure long-term persistence and preservation of datasets
II. Be recognized by a research community or research institution
III. Provide deposited datasets with stable and persistent identifiers, such as Digital Object Identifiers (DOIs)
IV. Allow access to data without unnecessary restrictions
V. Provide a clear license or terms of use for deposited datasets

# What are journals & publishers doing about it?

- **Content types** e.g. data articles and journals
- **Credit and incentives** e.g. data citation and data articles
- **Encouraging reuse** e.g. open licenses
- **Data quality** e.g. data peer review
- **Data discoverability** e.g. linking data to publications; supporting repositories
- **Raising awareness** e.g. editorials, outreach
- **Guidance and policy** e.g. information for authors, policy harmonization
- **Technology** e.g. platform developments, repository integration and other partnerships

# How many journals have a research data policy?

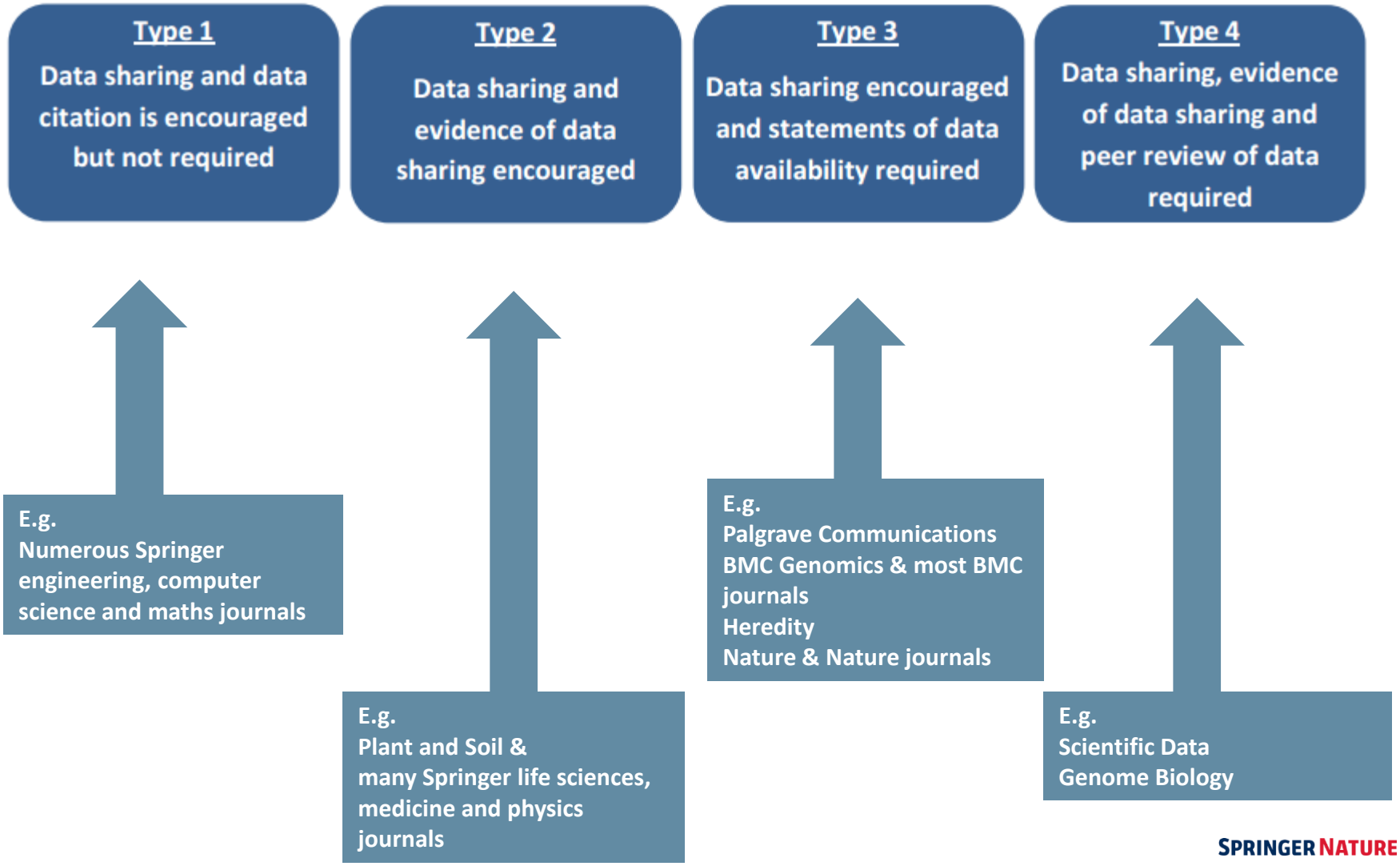



*Data source: Linda Naughton, JISC Journal Research Data Policy Bank project presentation (n = 250)*

***"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

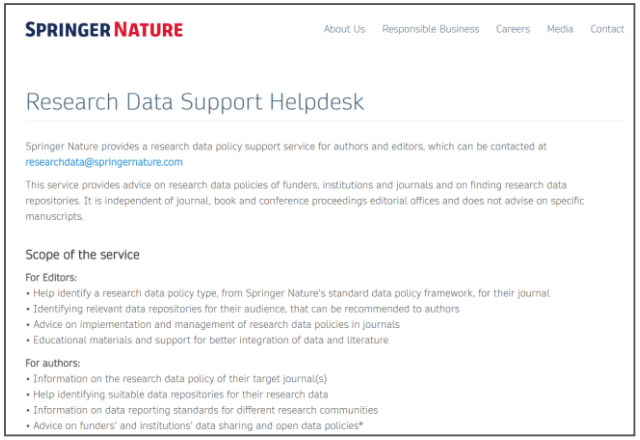
# Policy Types



# 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


SPRINGER NATURE

About Us Responsible Business Careers Media Contact

Research Data Support Helpdesk

Springer Nature provides a research data policy support service for authors and editors, which can be contacted at researchdata@springernature.com

This service provides advice on research data policies of funders, institutions and journals and on finding research data repositories. It is independent of journal, book and conference proceedings editorial offices and does not advise on specific manuscripts.

Scope of the service

For Editors:

- Help identify a research data policy type, from Springer Nature's standard data policy framework, for their journal
- Identifying relevant data repositories for their audience, that can be recommended to authors
- Advice on implementation and management of research data policies in journals
- Educational materials and support for better integration of data and literature

For authors:

- Information on the research data policy of their target journal(s)
- Help identifying suitable data repositories for their research data
- Information on data reporting standards for different research communities
- Advice on funders' and institutions' data sharing and open data policies*

**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

# 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 in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
- 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).

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

http://www.reading.ac.uk/reas-DataAccessStatements.aspx

# Data availability statements

- Guidance on and published examples of data availability statements

  http://www.springernature.com/gp/group/data-policy/data-availability-statements

| Statement type/description | Template/example text | Published example |
|---|---|---|
| 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 | [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 "] | BMC Biology<br>Nature Communications |
| Data available in a public (institutional, | The [data type] data that support the findings of this study are available in | Nature Communications |

# Future directions

- A research data policy for every journal
  - Expect to provide a data availability statement to publish in a reputable journal
- More and better links between datasets/repositories and articles
  - Impact/reuse information for research datasets
- More services from publishers to support data sharing
- Integration of larger datasets into online article pages
- Information about data sharing required during submission
  - More peer review visibility for data
- More evidence on costs and benefits of sharing data supporting publications

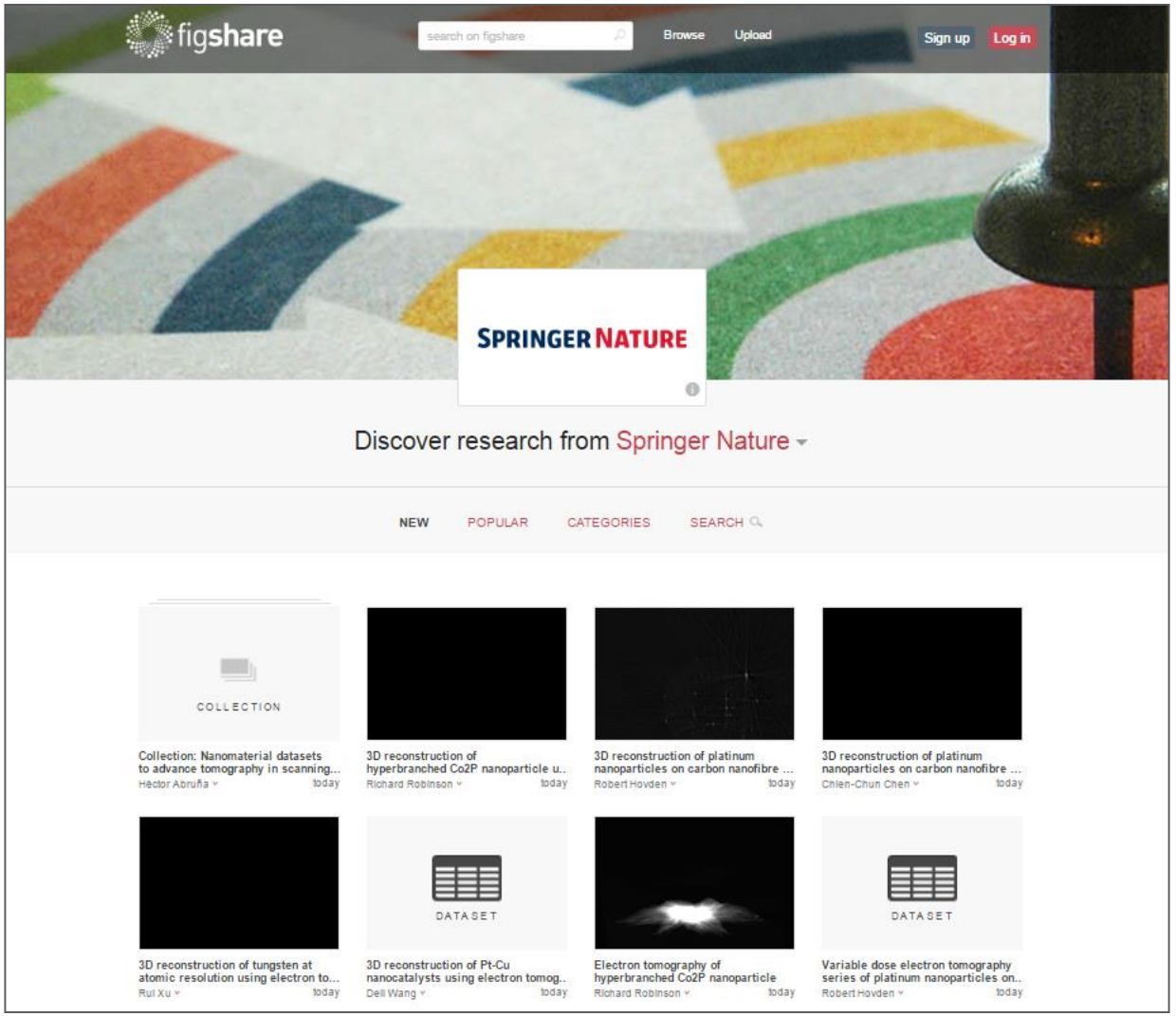
# Making supplementary data more discoverable

- Discoverability/visibility of articles
- Improve author service
  - Storage of larger datasets and files
  - Better metrics for data files
- Improve reader service
  - Better display/integration of a variety of file types e.g. video, large images
- Improve editor and peer reviewer service
  - Easier access to supporting data files

https://blogs.biomedcentral.com/bmcblog/2016/12/15/making-research-more-accessible-with-figshare/

figshare

# Figshare & Springer Nature



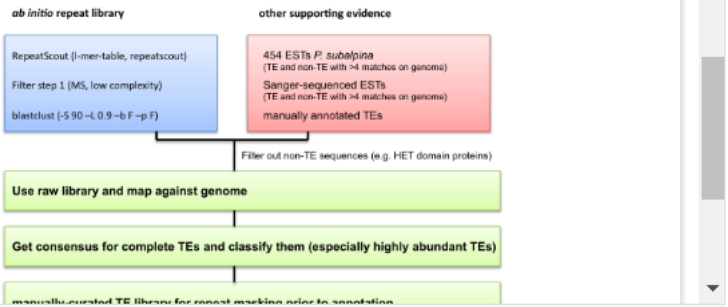

https://springernature.figshare.com/

# Figshare “widget” on all BMC & SpringerOpen journals

Globally distributed root endophyte Phialocephala subalpina links pathogenic and saprophytic lifestyles

Showing 1/9: 12864_2016_3369_MOESM8_ESM.pdf

36 views | 0 shares | 0 downloads




figshare 1 / 9 Share Download

Additional file 1: Mapping statics for the assembled 454 ESTs. (PDF 197 kb)

Additional file 2: Validation of low identity gene models. (PDF 190 kb)

Additional file 3: Putative secondary metabolite clusters identified in the *P. subalpina* genome (XLSX 12 kb)

Additional file 4: General statistics on the presence of InterPro accessions in the analyzed genomes. (XLSX 10 kb)

Additional file 5: CAZyme modules related to PCWD used for principle component analysis. (XLSX 12 kb)

Additional file 6: General genome statistics for the species included in the present study. (XLSX 15 kb)

Additional file 7: PCA analysis based on InterPro accessions. Placement of the 13 ascomycete species and *P. subalpina* in PCA based on InterPro accessions described in Soanes et al. 2008 to be enriched in pathogens. (PDF 359 kb)

Additional file 8: Strategy to construct the manually-curated repeat library. (PDF 114 kb)

Additional file 9: Overview of the annotation strategy applied. (PDF 120 kb)

http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3369-8

References

1. Sieber TN. Fungal root endophytes. In: Waisel Y, Eshel A, Kafkafi U, editors. Plant Roots: The hidden half. 3rd ed.

SPRINGER NATURE

# The role of journals & publishers: in summary

- Create incentives and increase motivations
- Remove barriers to data sharing
- Support community specific solutions
- Acknowledge community differences
- Implement/support funder and community policy requirements
- Help improve quality and value
- Increase transparency (to increase reproducibility)
- Provide clear, consistent guidance and support
- Make it easier to do the right thing
- Supporting transition to open access

# Better Science through Better Data #scidata17


SCIENTIFIC DATA

wellcome

A full day conference of talks and demos exploring issues in research data, **focusing on the needs of early career researchers**.

25th October 2017, London

# Thank you

iain.hrynaszkiewicz@nature.com

researchdata@springernature.com

@iainh_z



SPRINGER NATURE