

Illustration inspired by the work of John Maynard Keynes

Sharing research data: what publishers want authors to know

Dr. Rebecca Grant

Research Data Manager

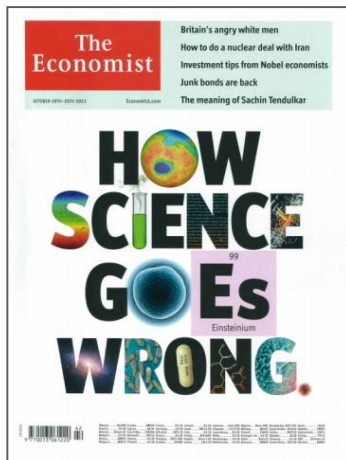
December 2020

ADVANCING
DISCOVERY

- Why do journals have research data policies?
- What are the key features of journal data policies?
- What are the most common methods of data sharing at journals?
- How do you choose the right repository for your data?

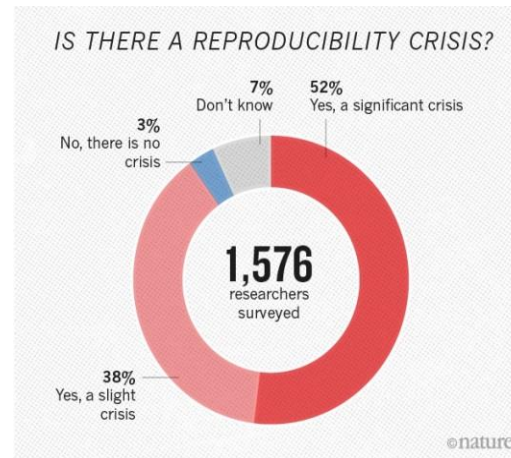
The “replication crisis”

Evidence is mounting on costs & scale of the issue



- Irreproducible biology research costs US \$28 billion per year¹
- Pharma companies report 75%+ failure rates replicating conclusions of peer-reviewed papers^{2,3}

A recent *Nature* survey⁴ highlights concern in the research community



>50% of researchers couldn't reproduce their own experiments
>70% couldn't reproduce the work of others

There is evidence that data availability increases reproducibility

A study⁵ of eighteen *Nature Genetics* papers found :

- Two could be reproduced fully
- Six were reproduced partially
- Ten could not be reproduced

“The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis.”
— *Nature Genetics* **41**, 149–155 (2009)

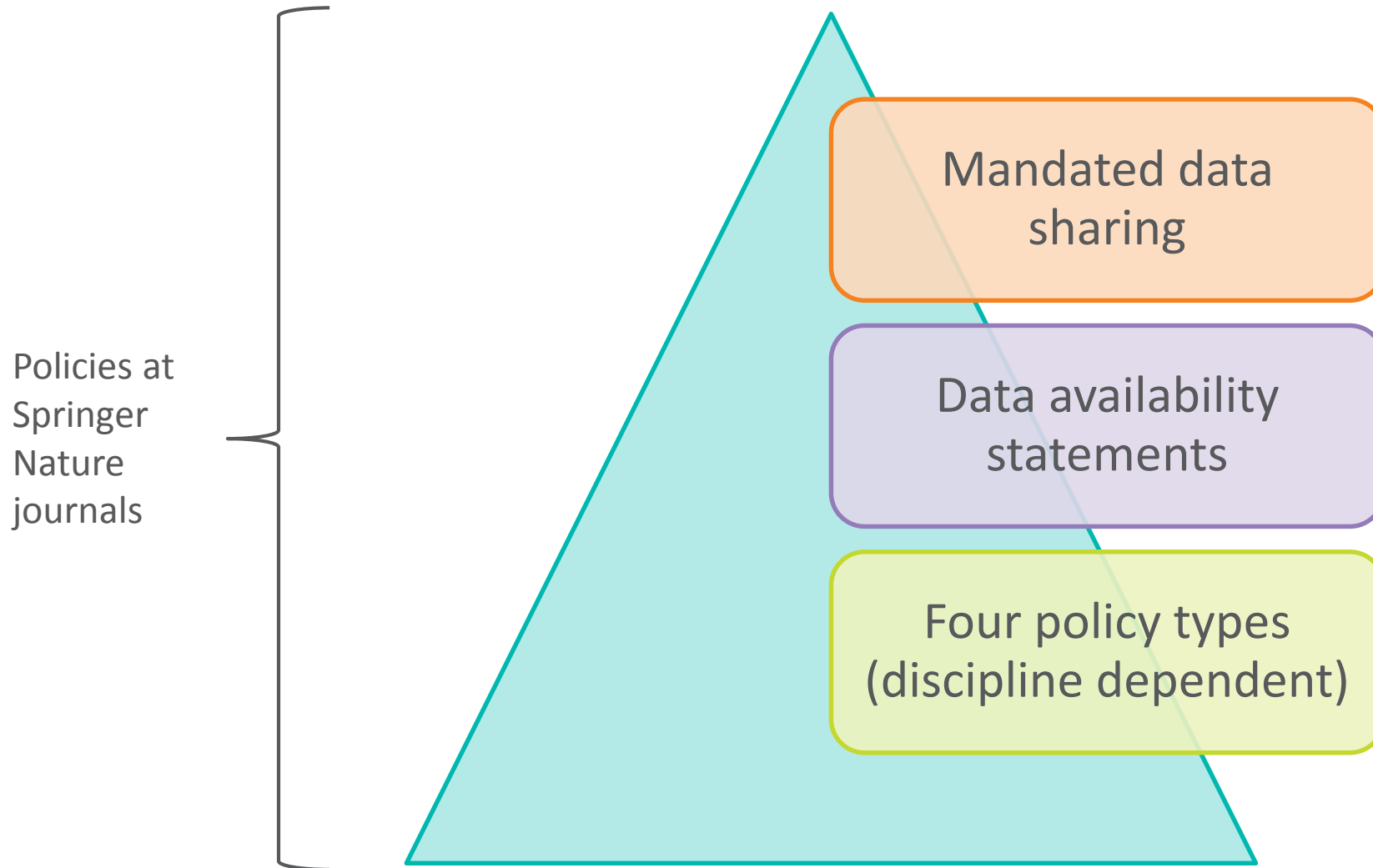
1. Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. *PLoS Biol.* **13**, e1002165 (2015) <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165>

2. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012), 3. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10**, 712 (2011)

4. Baker (2015) <http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

5. Ioannidis et al (2009) <https://www.nature.com/ng/journal/v41/n2/full/ng.295.html>

Research data policies at Springer Nature

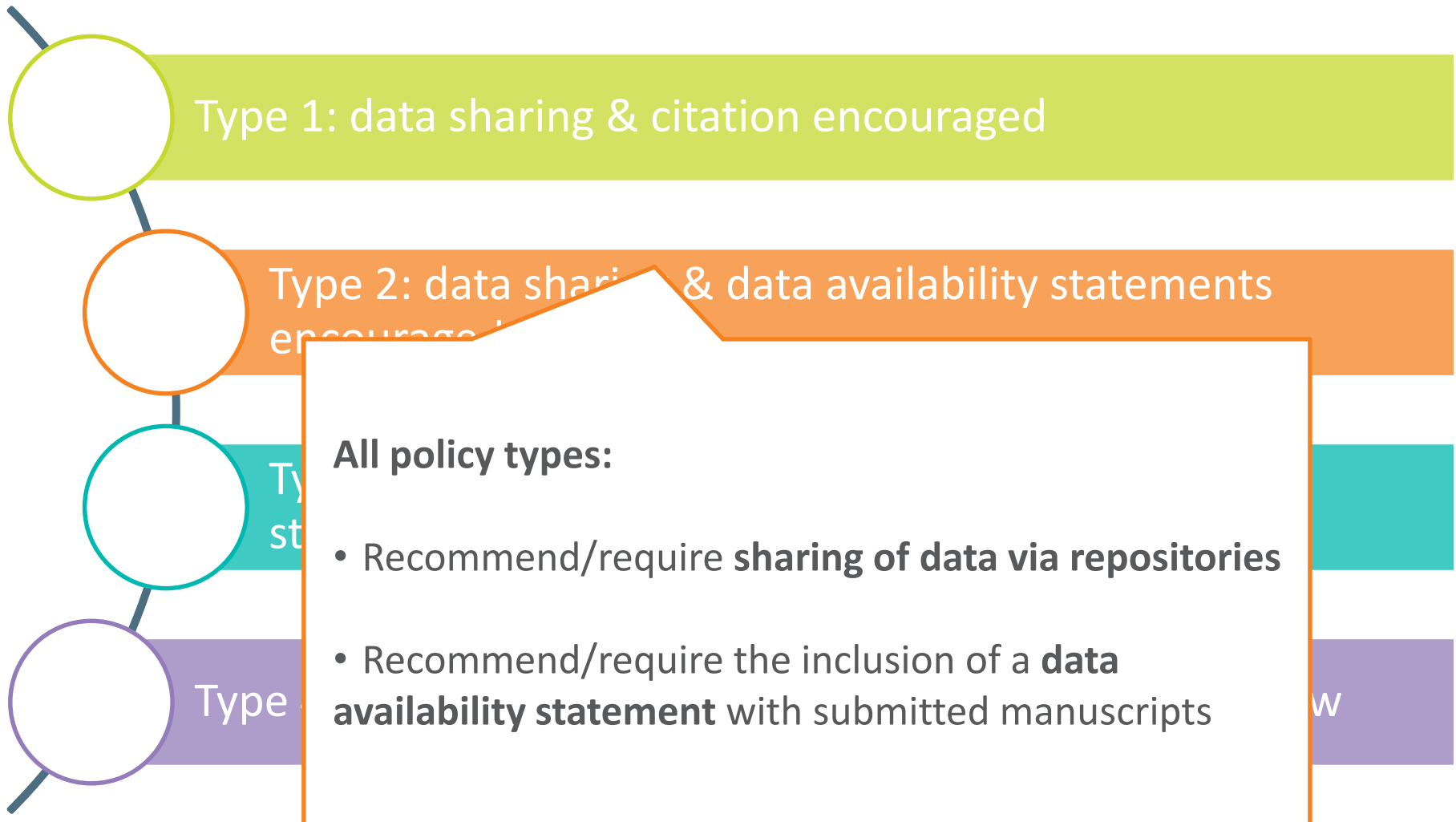


Data sharing policies

Publisher data policies

- **All major publishers** now have research data policies for their authors and are working together to explore consistent approaches.
- Policies relate to the **data underpinning the articles** submitted to journals.
- **Policies can differ** depending on the discipline and journal.
- **Before manuscript submission**, check your chosen journal's policy and what you may be asked to do.

Standard data policies at Springer Nature



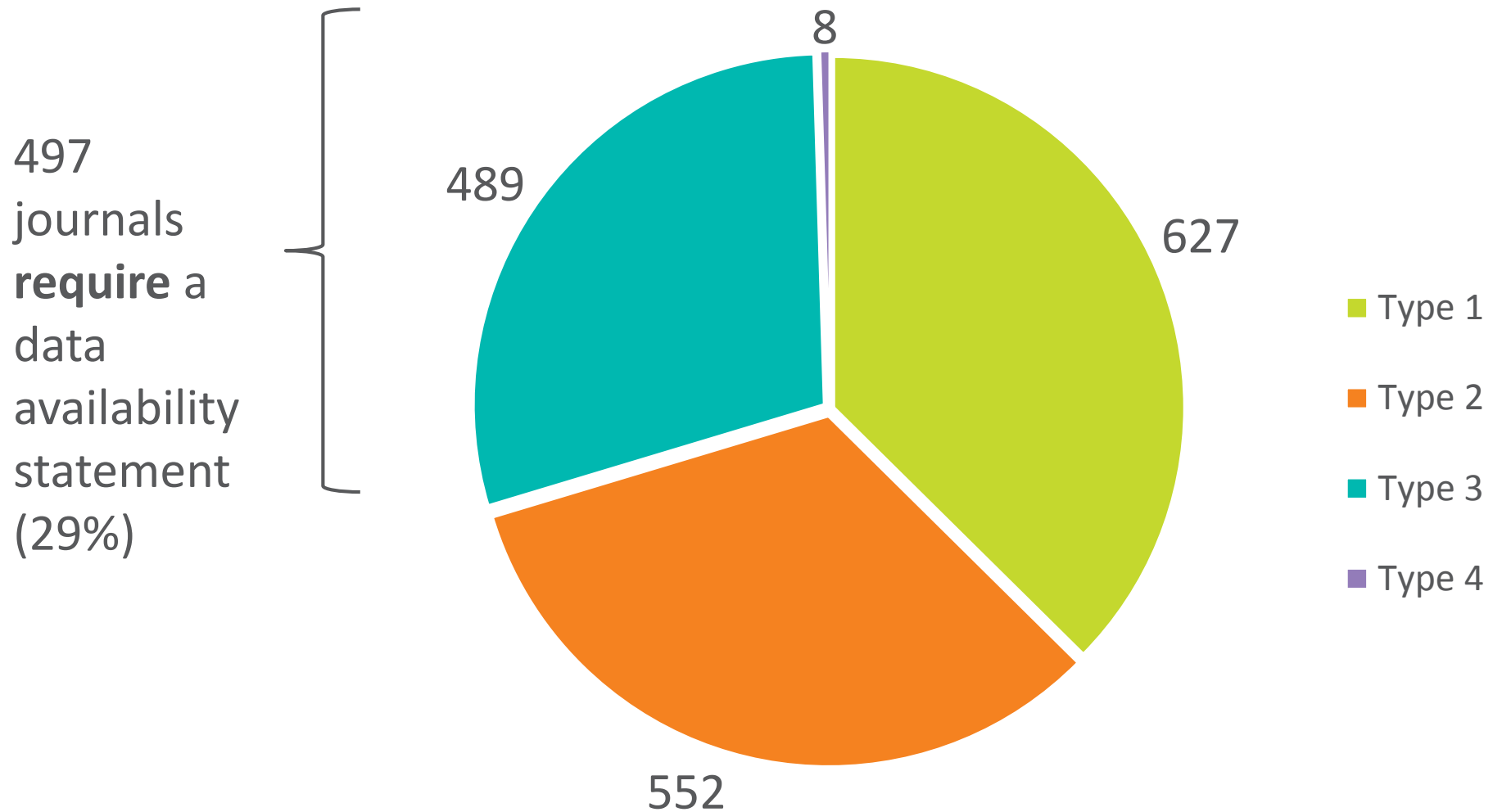
What is a data availability statement (DAS)?

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

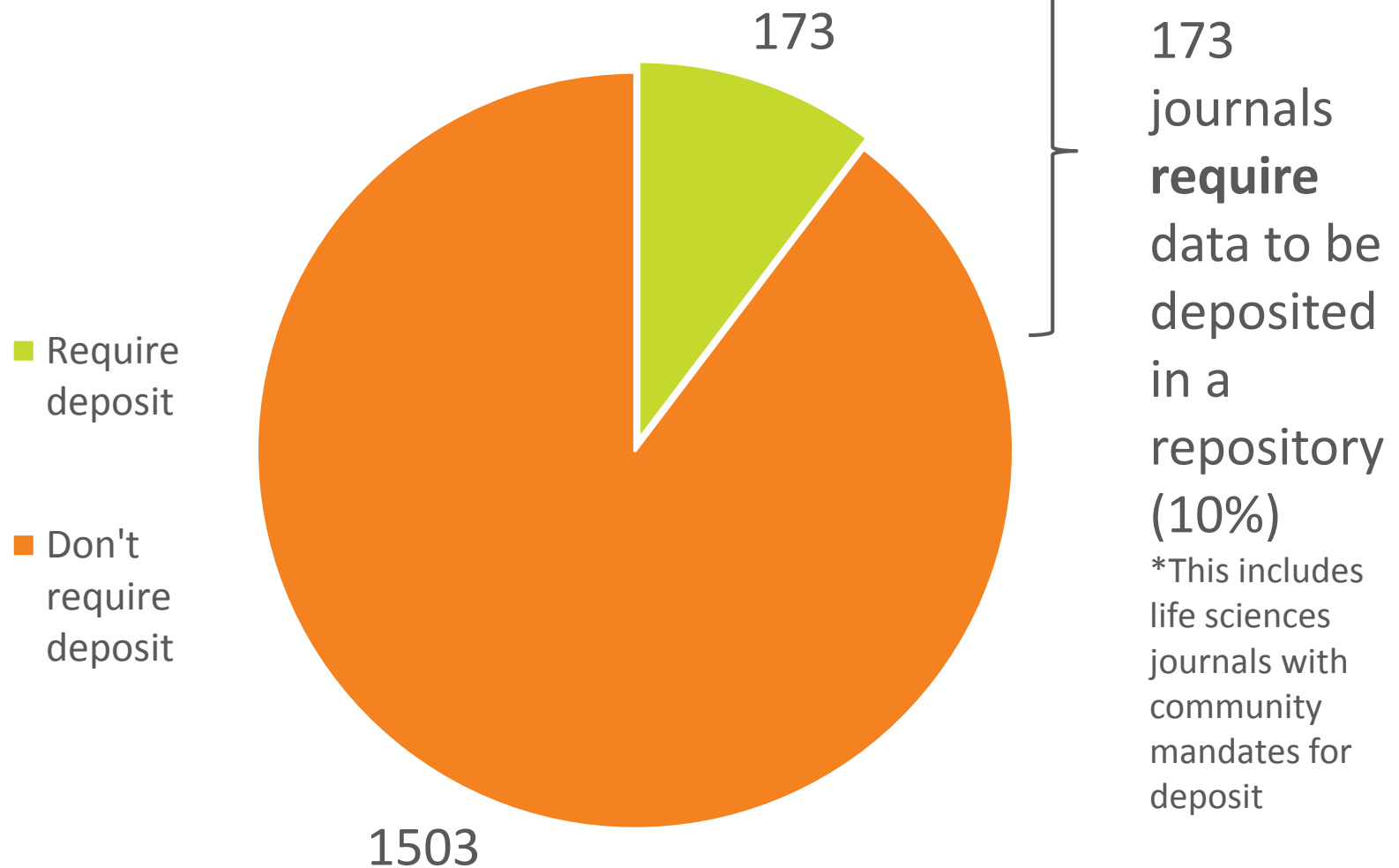
“The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.”

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

What our journal policies require



What our journal policies require



Options for research data sharing

How can you comply with journal data policies?

Type 1 & 2 policy journals

Add DAS if
desired

Any data
sharing option
allowed

Type 3 journals

DAS is
required

Any data
sharing option
allowed*

Type 4 journals

DAS is
required

Data must be
in a repository

Some data are subject to community mandates for sharing

Mandatory deposition	Suitable repositories
Protein sequences	Uniprot
DNA and RNA sequences	Genbank
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP
	dbVar
	European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP
	The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Gene expression data (must be MIAME compliant)	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database
Proteomics data	PRIDE
*Earth, space & environmental sciences	Recommended Repositories

Thinking outside of data policy requirements..

- Community expectations to share certain data in a repository even at Type 2 and 3 journals - this will be clear from the journal's instructions for authors.
- Peer reviewers or editors may require you to share other data types in repositories, at their discretion.
- You will NOT be forced to share your data if it would be inappropriate to do so, e.g. human data without consent to share, data you do not own.



Cite

Download (44.62 kB)

Share

Embed

+ Collect (you need to log in first) ...

Capture-mark-recapture data modelling survival rates of *Microcebus murinus* in relation to glucocorticoid level, parasite infection and body condition

Dataset posted on 01.09.2017, 09:26 by Josué H Rakotonjaina, Peter M Kappeler, Eva Kaesler, Anni M Hämäläinen, Clemens Kirschbaum, Cornelia Kraus

This dataset consists of an Excel spreadsheet containing capture-mark-recapture data, which were used to model survival of *Microcebus murinus* in different contexts.

These were:

- a Multistate modelling approach to model semi-annual survival relative to hair cortisol concentration (HCC) and scaled mass index (SMI). Median or third quartile were used as categorization cut-off.

- a Cormack-Jolly-Seber (CJS) modelling approach to model survival over the productive season relative to hair cortisol concentration, scaled mass index, and pattern of parasitism which was measured as the parasite species richness (number of distinct parasite morphotypes found per individual), the multiple infection (presence of more than one parasite morphotype), and the overall parasite prevalence (presence of at least one parasite morphotype).

The data used to assess the link between semi-annual survival rates and HCC includes the results of capture sessions held in October 2012, 2013, 2014, April 2013, and March 2014, during which a total of 171 individuals (74 females, 97 males) were captured. The same dataset, excluding the October 2014 session, was used to assess the effect of SMI on semi-annual survival probabilities, for a total of 149 individuals (63 females, 86 males). The dataset used for the CJS models includes data collected during monthly trapping sessions between September 2012 and April 2013, for a total of 48 individuals (16 females, 32 males).

All research activities conducted in Madagascar received official approval from the Ministère de l'Environnement, de l'Ecologie, and de la Mer et des Forêts, and comply with national animal care legislation of Madagascar.

FUNDING

Deutscher Akademischer Austauschdienst (A/12/90426) and Deutsche Forschungsgemeinschaft (KR 3834/4-1)

RESEARCH DATA SUPPORT

Research data support provided by Springer Nature.

1000
views36
downloads1
citations

17

READ THE PEER-REVIEWED PUBLICATION

Hair cortisol concentrations correlate negatively with survival in a wild primate population

SPRINGER NATURE

CATEGORIES

- Parasitology
- Behavioural Ecology
- Ecological Physiology
- Physiology
- Population Ecology

KEYWORD(S)

- Microcebus murinus
- Cormack-Jolly-Seber models
- Survival Analysis
- Glucocorticoid Function
- Capture mark-recapture
- Mouse Lemur Primates
- Hair cortisol levels
- parasitisms
- ecological
- Madagascar
- Cort-fitness hypothesis
- Multistate Survival Models
- Multistate models

LICENCE



CC BY-NC-SA

Dataset title

Author names

Altmetrics provides information on downloads and citations

Link to the peer reviewed article connected to this dataset

Relevant categories and keywords allow other researchers to find the dataset

Licence added to make reuse conditions clear

Comprehensive description including the data context of the study and data gathering method

Source: <https://doi.org/10.6084/m9.figshare.5259415>

Common features of data repositories

- Provide a place to **share your data and metadata**.
- Assign a **persistent and unique identifier and URL** for your data.
- Provide a **landing page** for your dataset.
- Most guarantee **data preservation** for a defined period of time (most).
- Some provide guidance on the landing page as to **how your data should be cited**.
- Some manage **data access requests** on your behalf.



How do I pick a data repository?

1. Start by using an appropriate **repository list or indexing service** for your specific needs.
2. If there is a **discipline or data-specific repository** for your data type, upload your data there.
3. If no discipline or data-specific repository exists for your data type, select a **generalist repository** for your data.

- Be aware that your funder, institution or journal may mandate deposit in a particular data repository.
- Be aware that you may need to upload your data to several repositories, depending on what you aim to deposit.

Repository resources, indexes and lists

Indexing services



Curated lists



Certified repositories

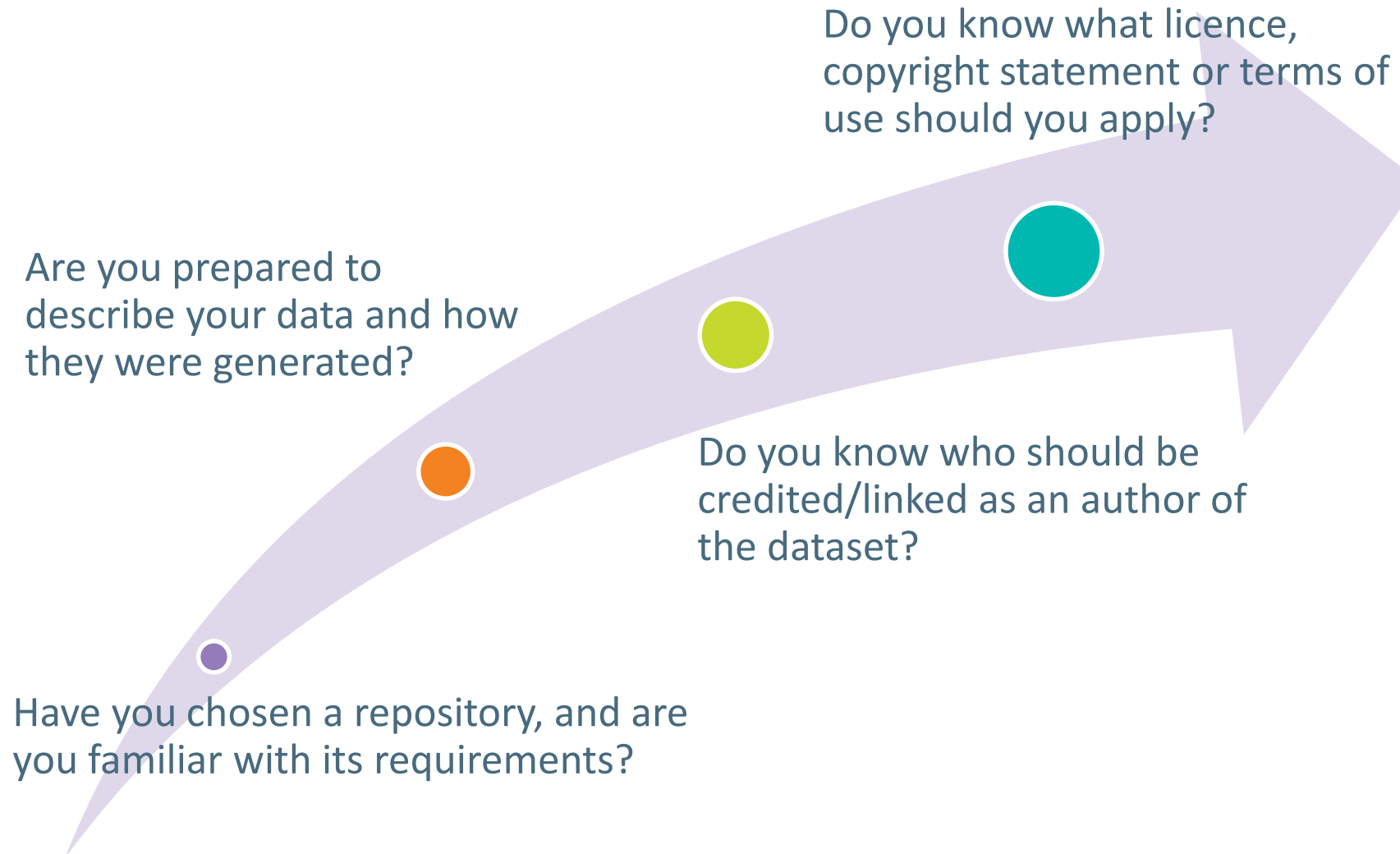


Tools to help select repositories



<https://repositoryfinder.datacite.org/>

Steps to data deposit



Benefits of data sharing in repositories



- Linking papers to their supporting data in a repository was associated with an average **25% increase in citations**
- Study based on **over half a million** open access articles

<https://doi.org/10.1371/journal.pone.0230416>

In summary: Journal policies and data sharing

- ✓ Check the policy requirements that apply to you.
- ✓ Don't assume that all journals have the same policy.
- ✓ At most journals, you should expect to include a data availability statement with your manuscript.
- ✓ Policies give you high-level guidance but editors and peer reviewers can use their discretion and request data sharing.
- ✓ Editors will not force you to share something that you shouldn't.
- ✓ There are many repositories available for your data, and tangible benefits to using them.



Useful resources

- Writing a data availability statement: <https://tinyurl.com/DASTips>
- FAQs and Author Support: <https://tinyurl.com/SNdataFAQs>
- Recommended repositories list: <https://tinyurl.com/SNrepos>
- Repositories index: www.re3data.org
- Repositories, policies and data standards: www.fairsharing.org

Research data training from Springer Nature

- Training from Nature Research Academies:
<https://partnerships.nature.com/product/nature-research-academies-research-data-workshops/>
- Nature Masterclasses Online “Managing Research Data to Unlock its Full Potential” (free taster):
<https://masterclasses.nature.com/online-course-in-managing-data/18365540>



Thank you for your attention.

Dr. Rebecca Grant – Research Data Manager
Rebecca.Grant@springernature.com

Join our Research Data community:
www.researchdata.springernature.com

Image credits:

Icon by dDara on Nounproject

Icon by Iconixar on Nounproject

Photo by Dana Marin on Unsplash

Photo by Markus Winkler on Unsplash

Photo by Samuel Zeller on Unsplash

Photo by Hush Naidoo on Unsplash

Photo by Ashley Edwards on Unsplash