SPRINGER NATURE

Illustration inspired by the work of Alan Turing

11,000 researchers gave their views on data sharing: here's what we learned

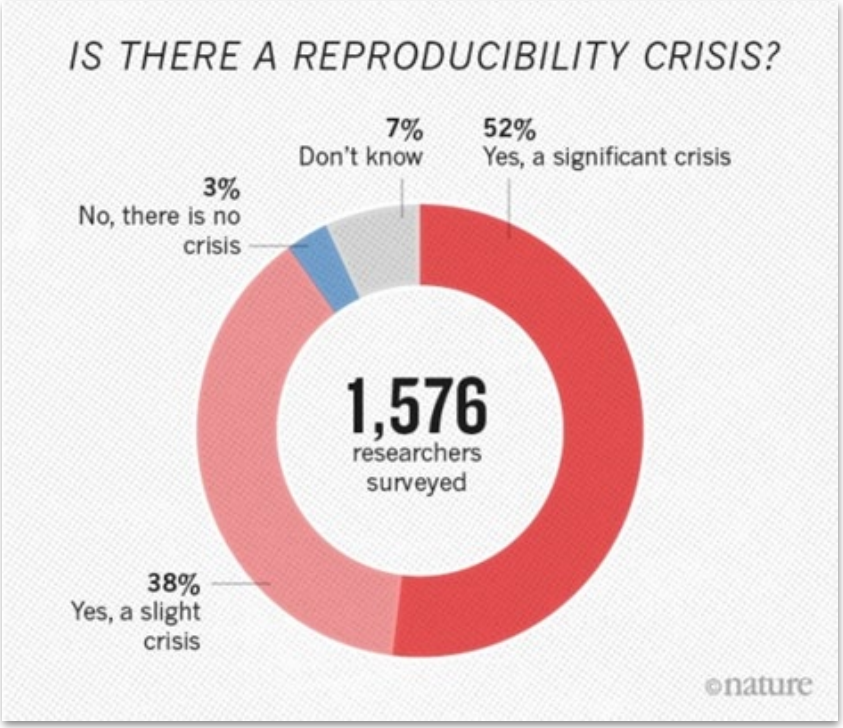
Samuel Winthrop

14th Munin Conference on Scholarly Publishing 2019

ADVANCING DISCOVERY

The case for data: Reproducibility

A *Nature* survey from 2015[1] highlights concern in the research community



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

There is evidence that data availability increases reproducibility

A study[2] 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)

11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

1. Baker (2015) http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970
2. Ioannidis et al (2009) https://www.nature.com/ng/journal/v41/n2/full/ng.295.html

The case for data: Benefits to researchers and science

Data archiving can **double** the publication output of studies

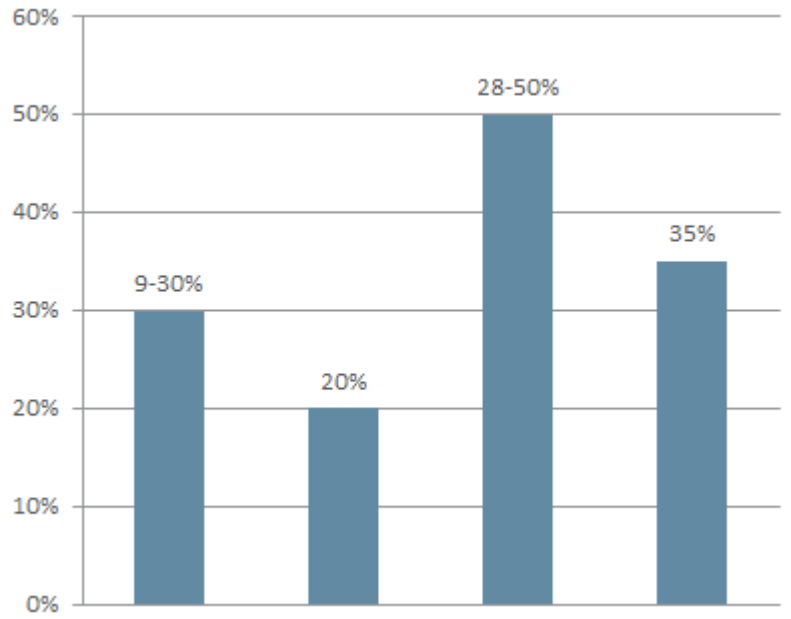
A study[3] of 7000 NSF and NIH research projects in social sciences found that:

- **Those with archived data resulted in ten (median) publications**
- **Those without archived data resulted in five publications**

Principal investigators who archived their data were more likely to publish more articles per project, and to see others build on their work

Research articles with open data are associated up to **50% more citations**

Analysis shows that articles with data available are cited 9-50% more, depending on the field



3. Pienta et al (2010) https://deepblue.lib.umich.edu/handle/2027.42/78307
4. Piwowar & Vision (2013) https://doi.org/10.7717/peerj.175
5. Henneken & Accomazzi (2011) https://arxiv.org/abs/1111.3618
6. Dorch et al (2015) https://arxiv.org/abs/1511.02512
7. Sears et al (2011) https://figshare.com/articles/Data_Sharing_Effect_on_Article_Citation_Rate_in_Paleoceanography/1222998/1

The case for data: Societal benefits

CASE STUDY: Human Genome Project

$1 trillion: Estimated contribution to the US economy, as reported by the Battelle Memorial Institute[1]

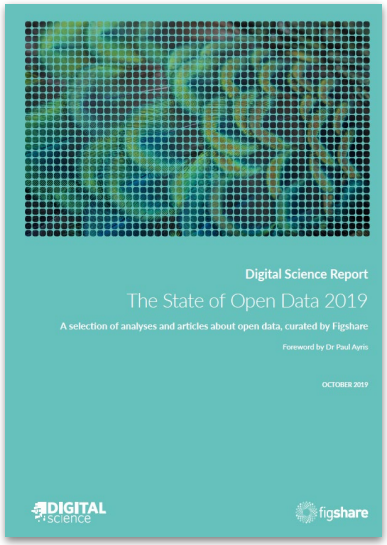
CASE STUDY: European Bioinformatics Institute

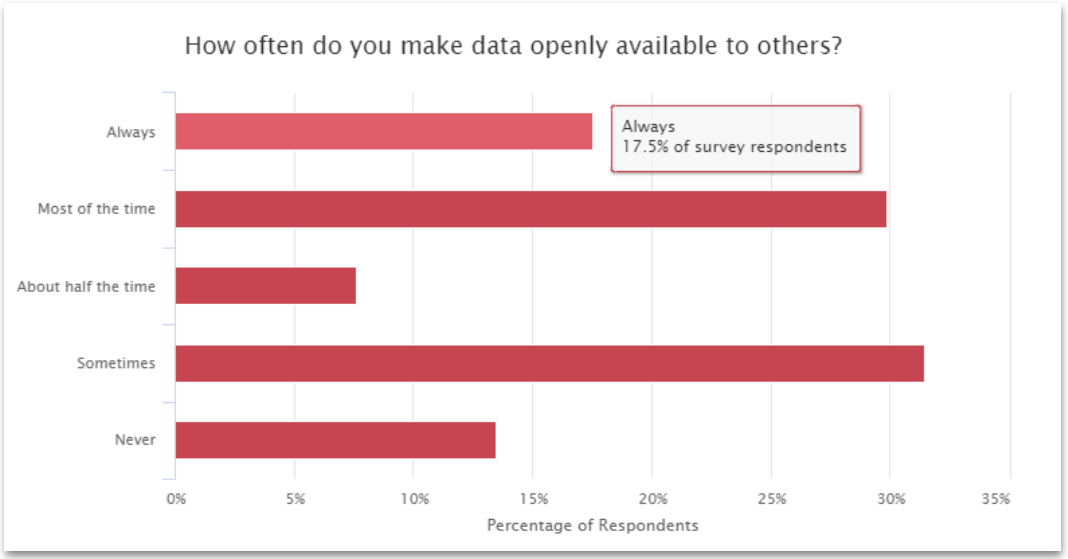
£1 billion: Annual efficiency savings to researchers worldwide, according to an independent report[2]

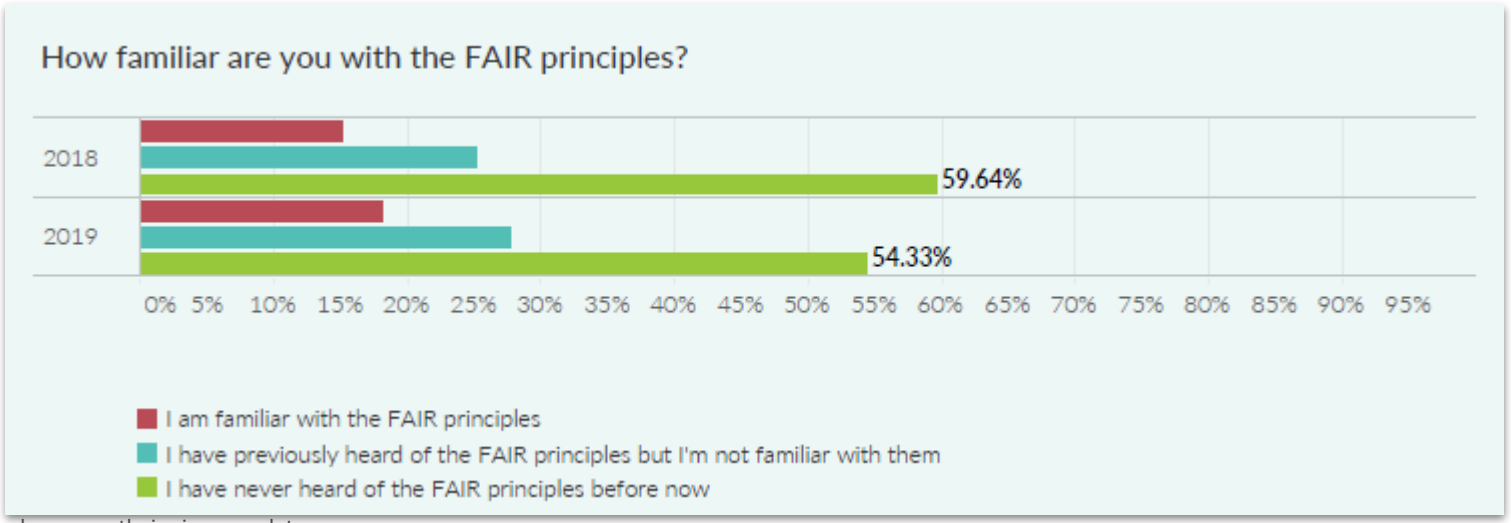
11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

8. http://www.unitedformedicalresearch.com/advocacy_reports/the-impact-of-genomics-on-the-u-s-economy
9. http://www.ebi.ac.uk/about/news/press-releases/value-and-impact-of-the-european-bioinformatics-institute

There is still work to be done







11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

10. The State of Open Data Report 2019. figshare. Report. https://doi.org/10.6084/m9.figshare.9980783.v2

SPRINGER NATURE

Working to understand researchers needs and challenges

11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

11. Practical Challenges for researchers in data sharing. figshare. https://doi.org/10.6084/m9.figshare.5975011.v1
12. The State of Open Data Report 2018. figshare. https://doi.org/10.6084/m9.figshare.7195058.v2
13. Challenges and Opportunities for Data Sharing in China. (2019). https://doi.org/10.6084/m9.figshare.7718441.v1
14. Challenges and Opportunities for Data Sharing in Japan https://doi.org/10.6084/m9.figshare.7999451.v1

Looking at the big picture

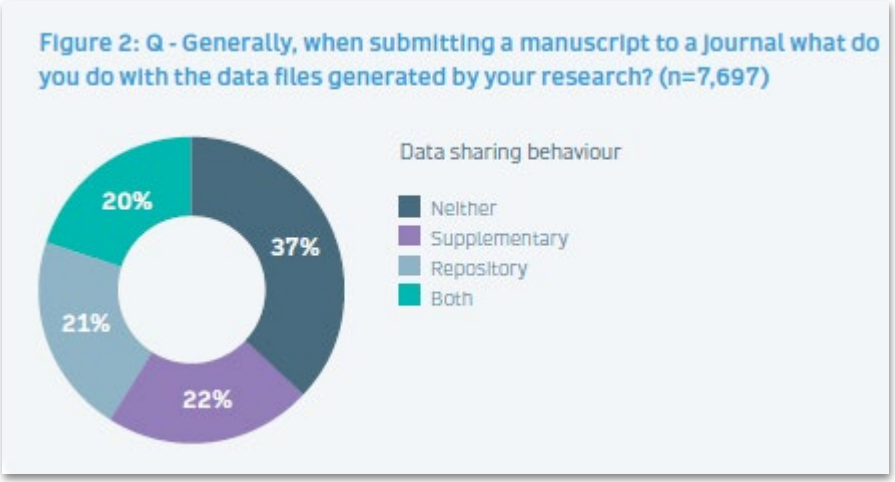
In 2019, Springer Nature published a whitepaper, *Five Essential Factors for Data Sharing*[15], which looked at the key challenges in data management and data sharing, based on responses collected in 2017-19 of over 11,000 researchers across the globe.

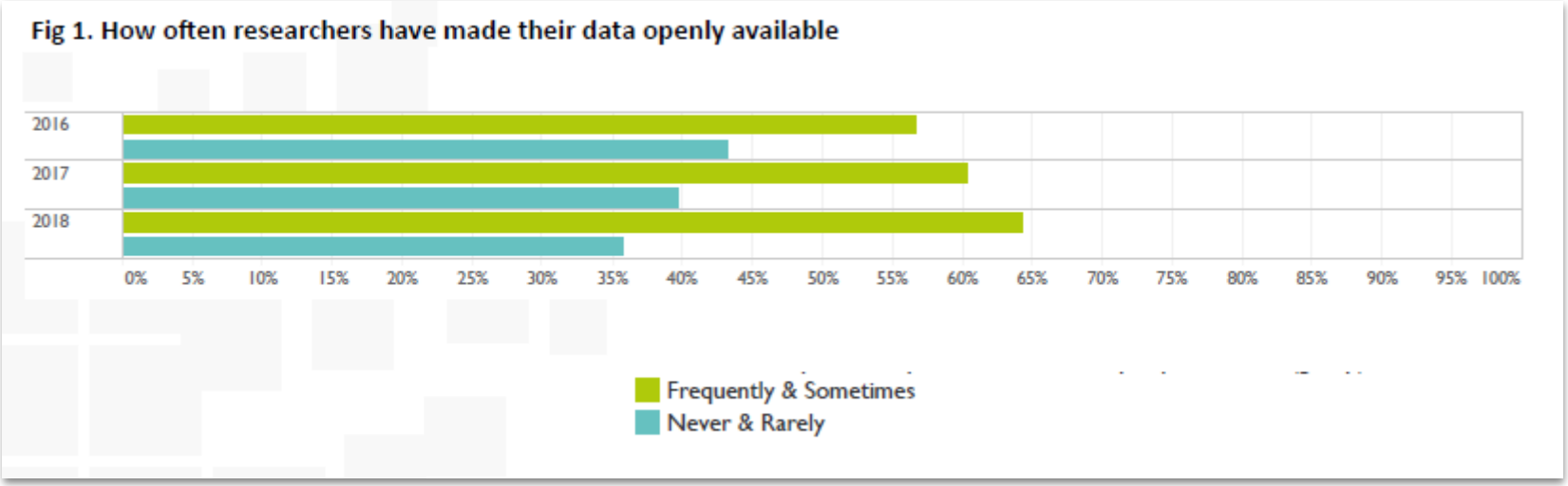
11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

15. Five Essential Factors for Data Sharing (2019): https://doi.org/10.6084/m9.figshare.7807949.v2

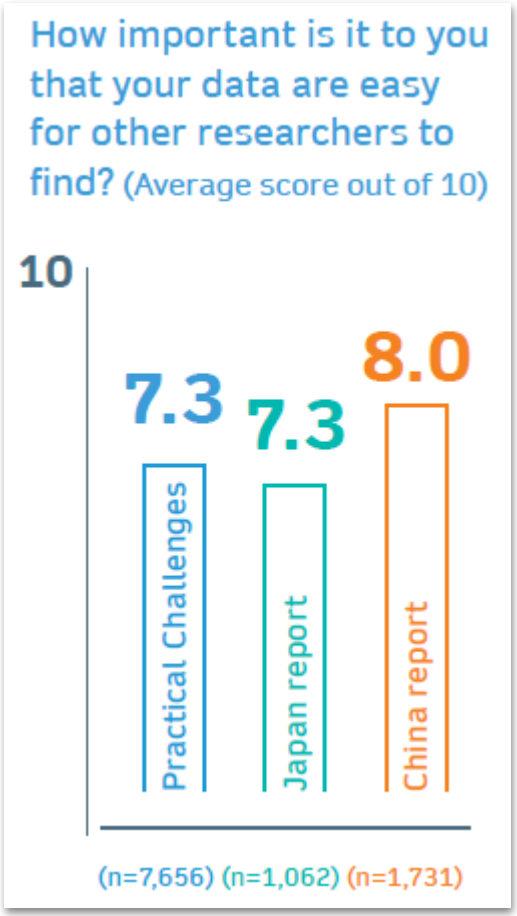
Data is being shared more

The *State of Open Data* report shows steady growth in the number of researchers sharing their data, up consistently year on year to 64% in 2018. Our *Practical Challenges* report shows similar evidence of data sharing, with 63% generally submitting research data files at the point of publishing a research article.





Most researchers think sharing data is important

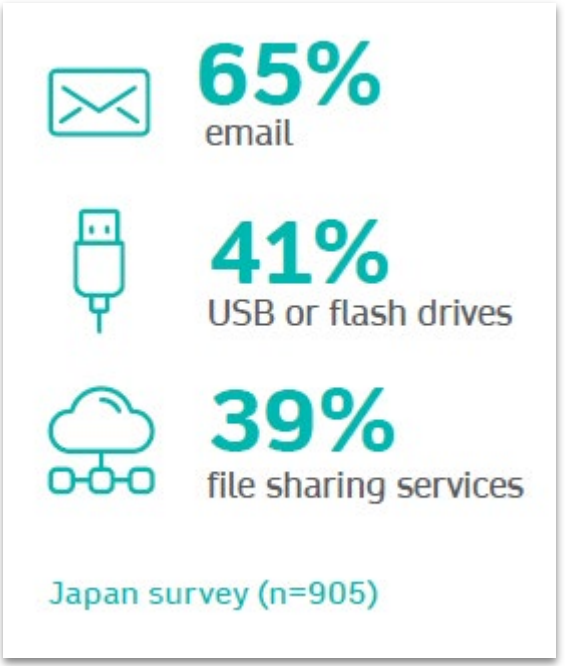


In *Practical Challenges*, when asked about the importance of making their data discoverable, researchers gave an average rating of 7.3 out of 10, with the most popular rating being the maximum rating, 10 out of 10 (25%).

In our follow up regional surveys with researchers in China and in Japan, we saw similar levels of agreement, with an average score of 8.0 for researchers in China and 7.2 in Japan.

Data sharing is suboptimal

When asked in the State of Open Data Report where researchers publish their data, 35% of respondents had published their data as an appendix to a research article, with little change from 2017 (34%). As shown by the data from China and Japan, most researchers share data (person to person) via email or flash drives.



11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

The Five Essential Factors

Clear Policy

State of Open Data Report: What circumstances would motivate you to share your data? (n=1,359) (multiple select)		
Answer	%	Count
Increased impact and visibility of my research	62%	841
Public benefit	59%	802
Transparency and re-use	48%	652
Getting proper credit for sharing data	46%	621
Journal/publisher requirement	44%	599
Trust the person requesting my data	41%	561
Institution/organisation requirement	38%	522
It was made easy and simple to do so	36%	485
Funder requirement	33%	453
Freedom of information request	26%	352
Other (please specify)	5%	63
I would never share my data	1%	17
Total	100%	1,359

1. Clear policy

More communication is needed on funder requirements to increase awareness and uptake

Is there a relationship between data sharing mandates and data-sharing behaviour?

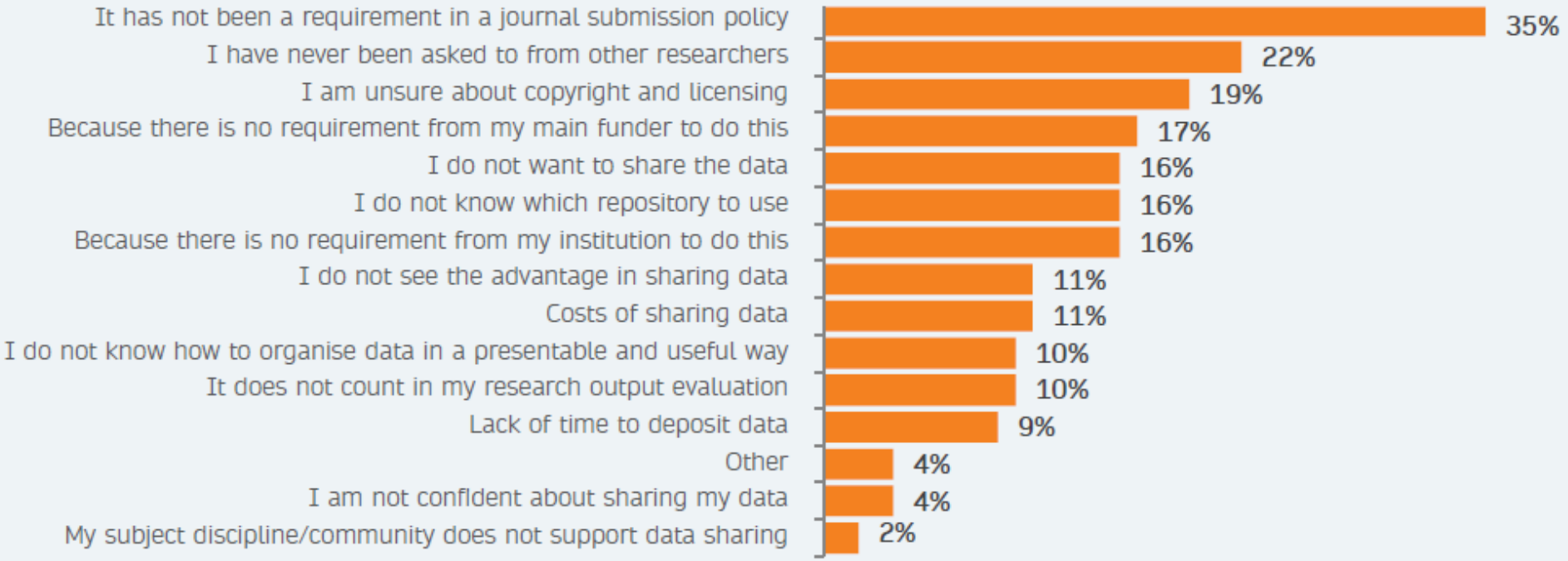
Clear Policy

1. Clear policy

The impact of journal policies on data sharing is higher than both funder and institutional requirements, based on reporting in both the State of Open Data Report and our surveys with researchers in Japan and China.

China report

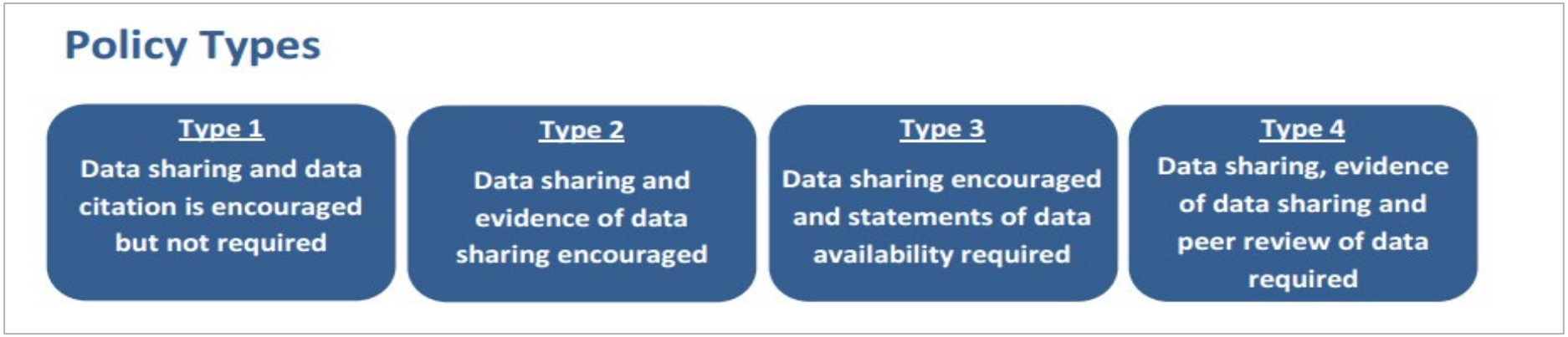
Why have you not shared data generated by your research? (n=108)



Clear Policy

1. Clear policy

Even where journal standards have been introduced, challenges remain around author awareness and support offered.



- **More than 1600 Springer Nature journals have adopted a data policy, including all of BMC (type 3 and 4), all of Nature and Nature Partner Journals (type 3).**
- **Similar initiatives in other publishers.**
- **Co-chairing the data policy standardisation working group of the Research Data Alliance.**

Credit

Most researchers do not think they get enough credit for sharing their data. Data citation scores highest as a means of credit.

State of Open Data Report: What credit mechanisms do you think would encourage more researchers to share their data? – Coded (Base n = 623; total n = 1,874; 1,251 missing)	
Answer	%
Citation	30%
Co-authorship	18%
Acknowledgement	13%
Financial/discounts	7%
Counts towards tenure/grants	7%
Cultural	5%
Mandates	5%
Visibility/transparency on use	5%
Limit misuse/security	3%
Data index/dedicated system	3%
Making it easier/education	2%
Other	8%
Unsure	15%
None	2%

2. Better credit

58% of researchers don't think they get sufficient credit for sharing research data

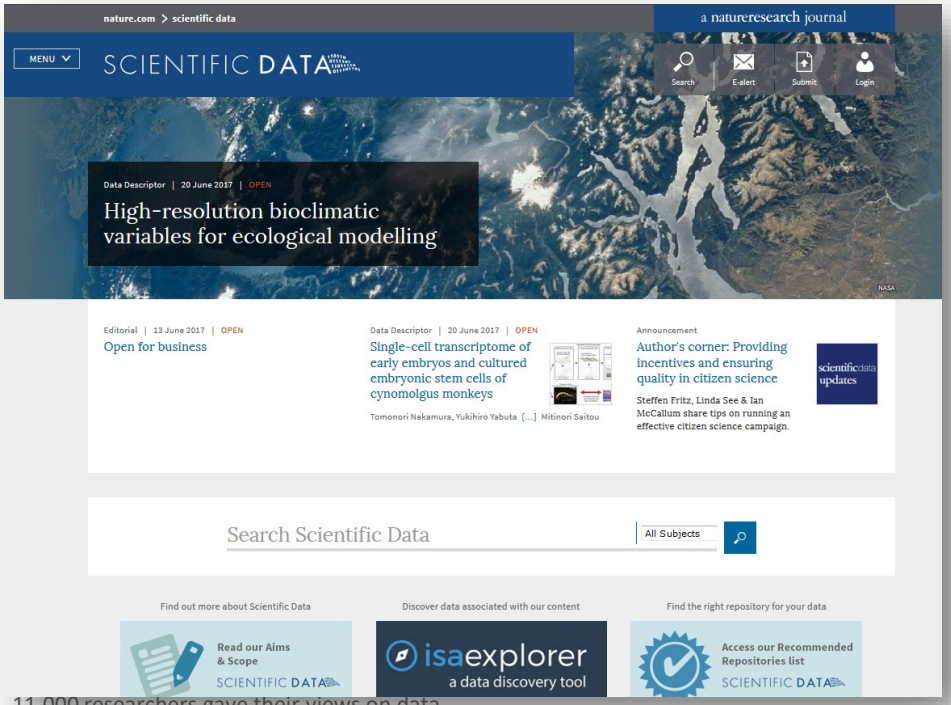
Credit

Data article publishing is increasing, but still at low levels. It is also over-reported in our 2018 surveys, suggesting further understanding is needed.

2. Better credit

SCIENTIFIC DATA

BMC



11,000 researchers gave their views on data sharing, here's what we learned / 27 November 2019

SPRINGER NATURE

Funding

3. Explicit funding

Few funders explicitly make funding available for data management

27%

of researchers do not know how they would meet the costs of making their research data openly available

How researchers would meet the costs of making data open

39%

would use money specifically for this purpose from a funder

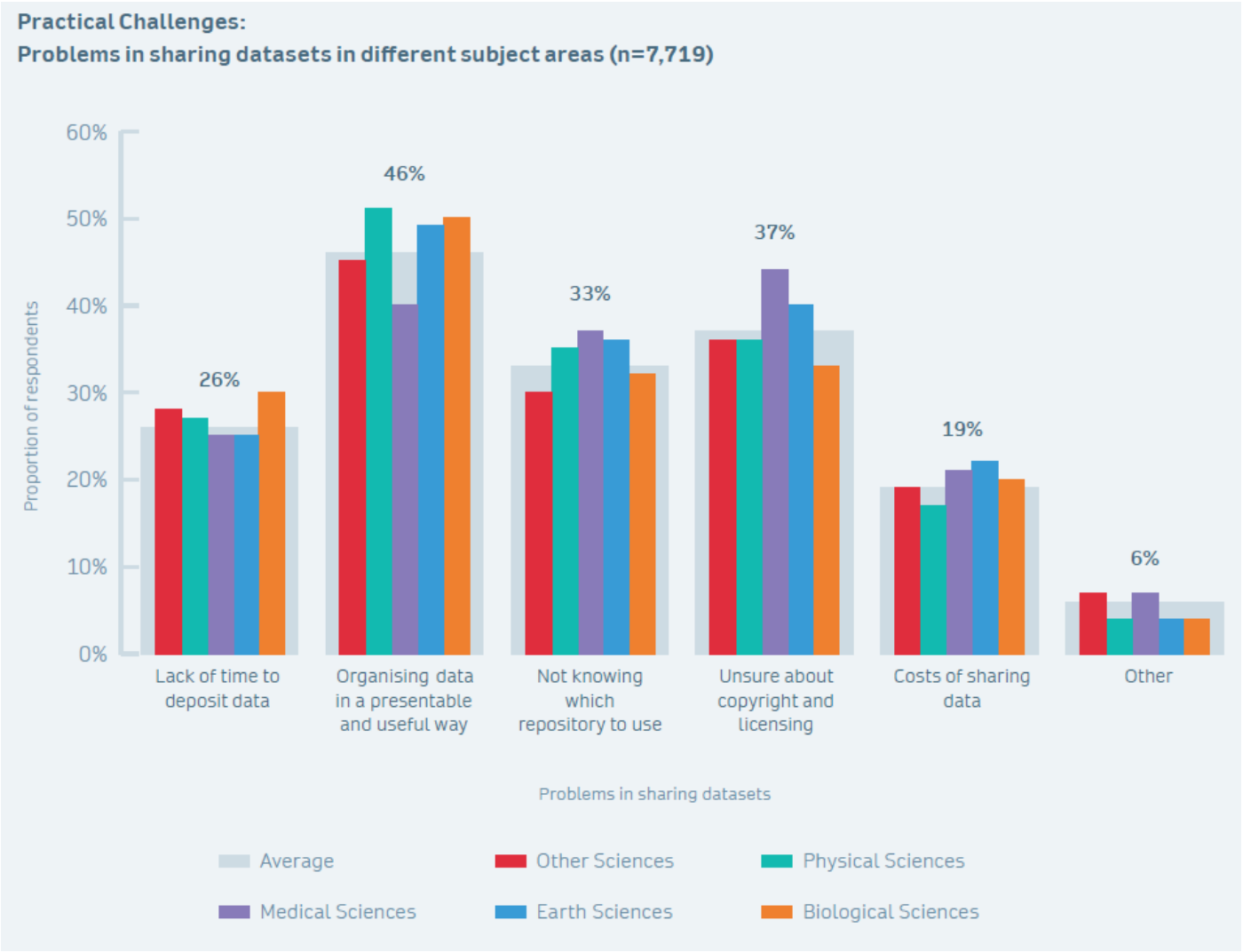
41%

would use funds identified in their grant

37%

likely to use their own funds for data sharing

Practical help



4. Practical help

The solutions outlined here require collaboration between researchers, institutions, funders, publishers, repositories, and other research data infrastructure providers

26%

We offer free guidance via our Research Data Helpdesk. The majority of enquiries to our Helpdesk service in 2018 related to appropriate repositories and depositing data

SPRINGER NATURE

Training & Education

What are the most commonly raised areas as to where education and training are needed?

- Copyright
- Repositories
- Misuse of data
- Sensitive data
- Cultural attitudes to sharing
- Size of data
- Data Management Plans

5. Training & education

65%

of researchers feel there is not sufficient training, support and advice in regard to data management

Where do we go from here?

The key messages:

- **More practical support and assistance is needed**
- **We need to collaborate (funders, institutions, publishers etc.) in order to change behaviour**
- **We need to continue to learn to understand researcher attitudes to data sharing, in order to effectively 'normalise' good data practices.**

Thank you

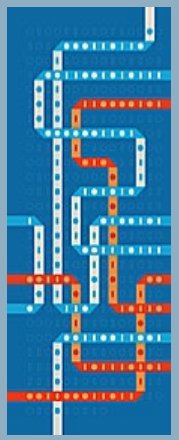
Samuel Winthrop
Journal Development Manager, Research Data
Springer Nature
samuel.winthrop@springernature.com

For more information on Research Data Support and other data-related activities at Springer Nature:

Email: researchdata@springernature.com
Website: http://go.nature.com/ResearchDataServices

Slide acknowledgements:
Grace Baynes, Rebecca Grant, Mithu Lucraft

The story behind the image



Alan Turing (1867–1934)

The scope of the achievements of Alan Turing, computer pioneer, wartime code-breaker and polymath, cannot be overstated. Renowned as the man who broke the Enigma code, Turing is also considered the father of computer science and artificial intelligence. His legacy is represented here with a visualisation of a "Turing Machine", a hypothetical device he devised to represent the logic of a computer. The binary code depicted translates to one of Turing's memorable quotes: Science is a differential equation. Religion is a boundary condition.

SPRINGER NATURE