SPRINGER NATURE

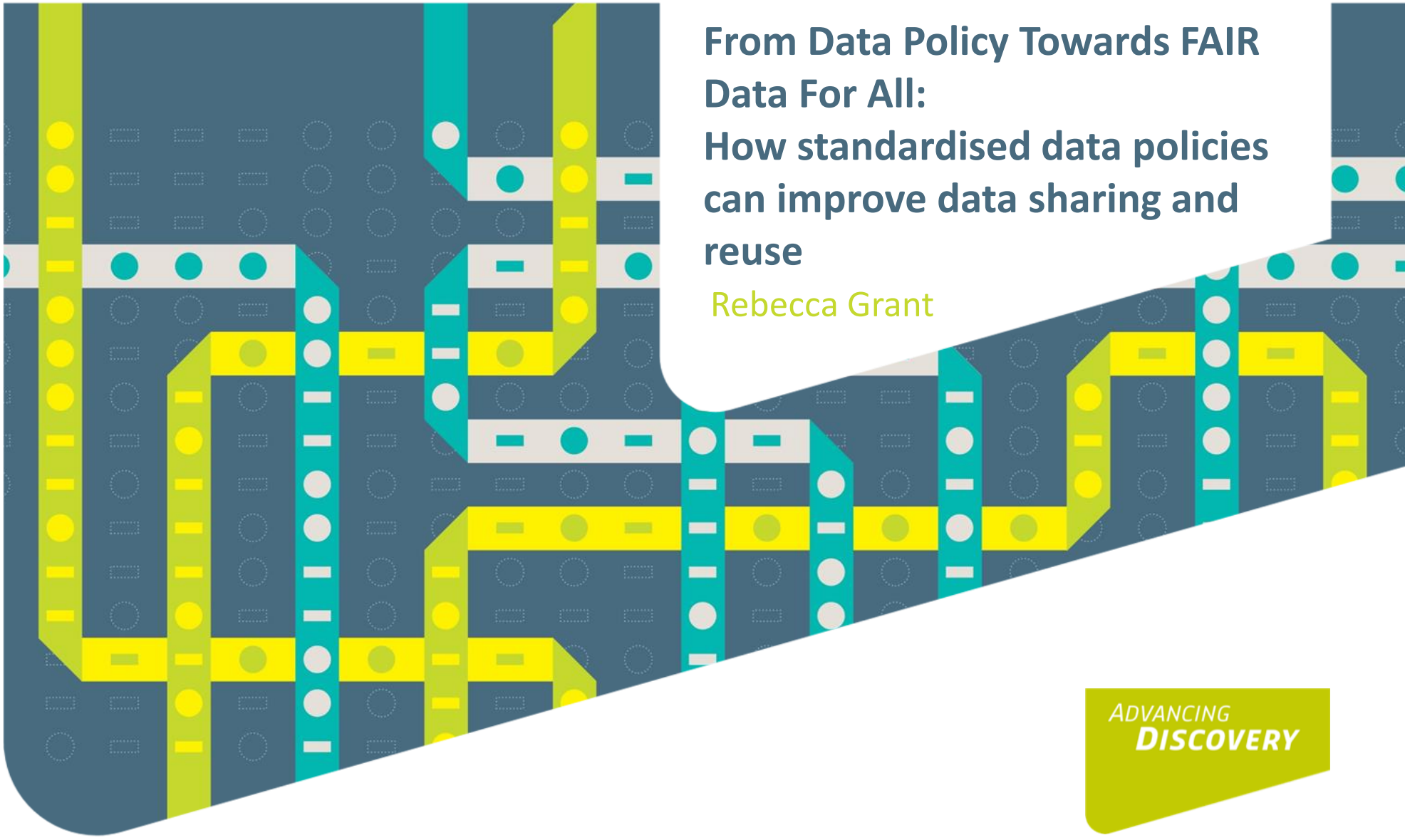


Illustration inspired by the work of Alan Turing

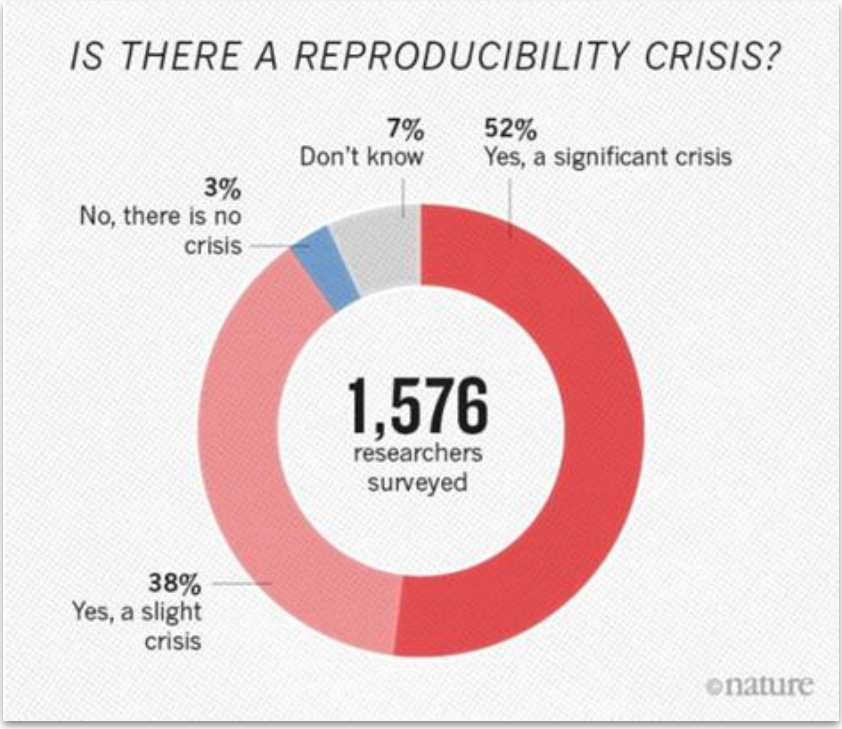
From Data Policy Towards FAIR Data For All:
How standardised data policies can improve data sharing and reuse

Rebecca Grant

ADVANCING DISCOVERY

The context for data sharing: 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)

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 context for data sharing: Benefits to researchers

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

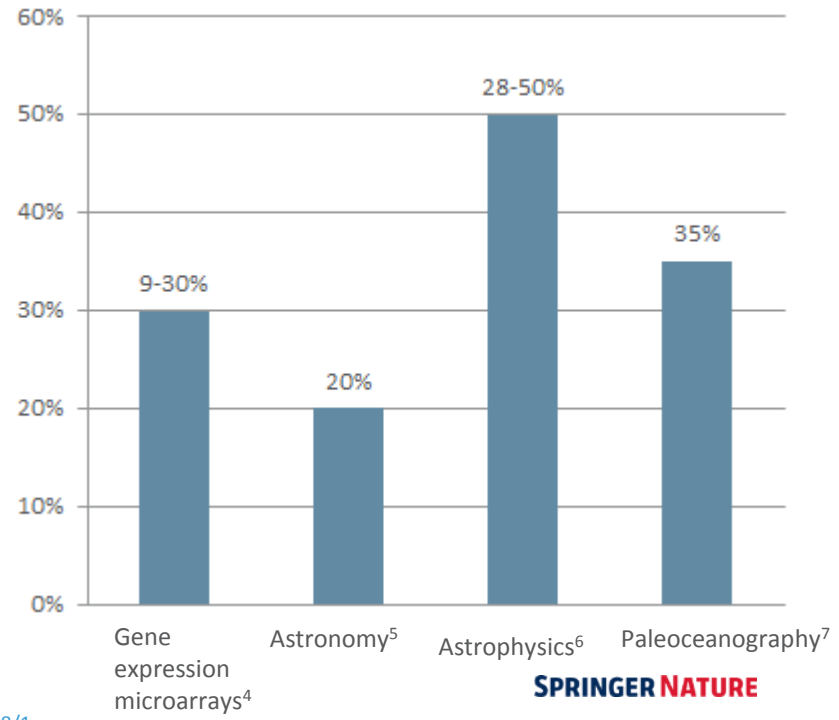
A study[3] of 7000 NSF and NIH research projects in social sciences found that:

- **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

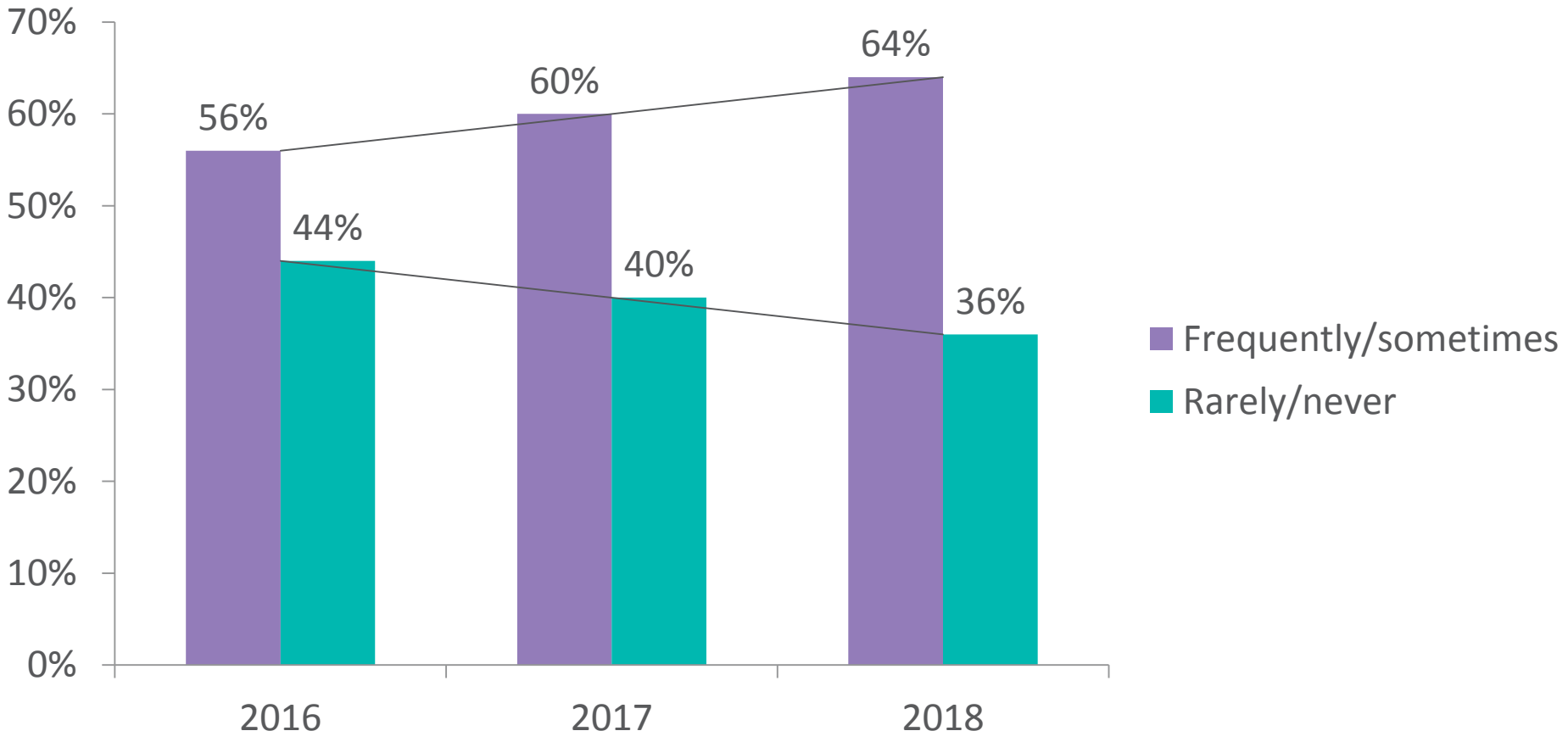


SPRINGER NATURE

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

Data are being shared

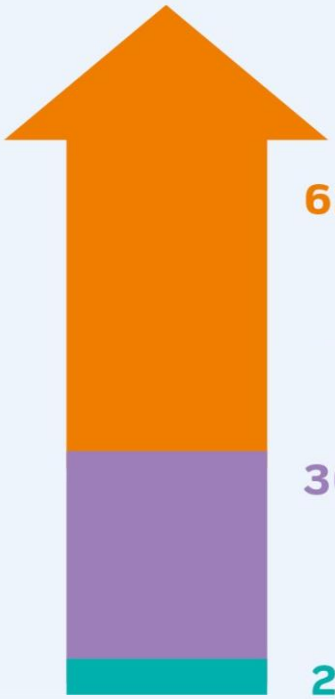
The S*tate of Open Data* report shows steady growth in the number of researchers sharing their data, up consistently year on year to 64% in 2018.



SPRINGER NATURE

The State of Open Data Report 2018, https://doi.org/10.6084/m9.figshare.7195058.v2

How researchers commonly share their data



62% have shared data both privately and publicly

36% have only shared data privately

2% of respondents have only shared their data publicly

The three most common methods of private sharing were:

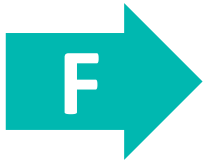
- email (65%)
- USB or flash drives (41%)
- file sharing services (39%)

The three most common ways of public sharing were: (n=569)

- supplementary information to journal articles (51%)
- lab or personal website (27%)
- subject specific repository or data archive (25%)

Source: Springer Nature researcher survey "Practical Challenges for Researchers in Data Sharing" 2018. Total respondents: 7719

Standards for sharing: the FAIR Data Principles

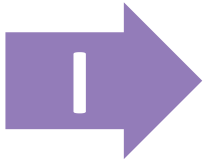


Findable

by leveraging metadata and persistent identifiers

Accessible

through free and open communications protocols



Interoperable

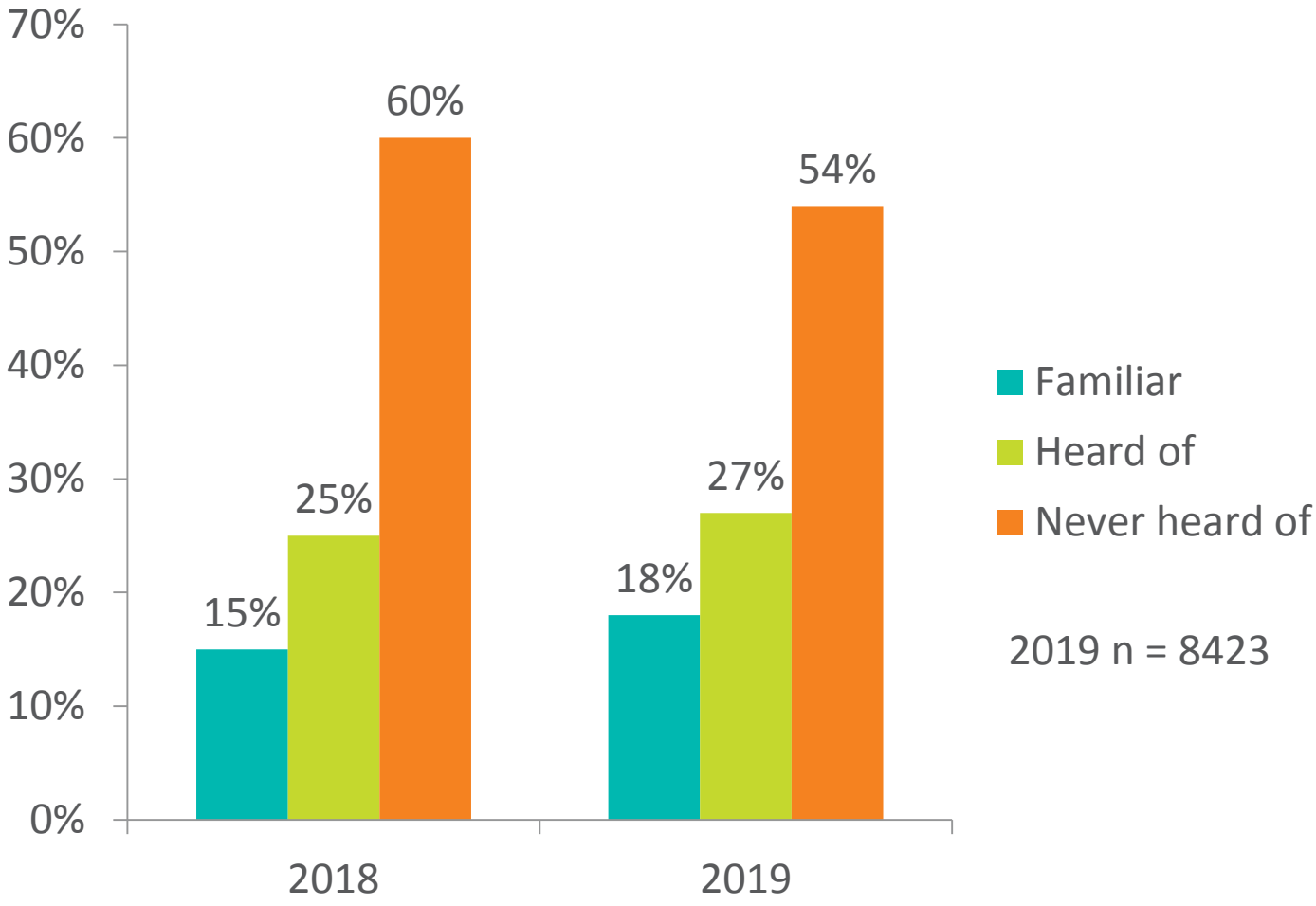
by using controlled vocabularies, implementing machine-readability and including references where appropriate

Reusable

by highlighting clear licence statements that enable the greatest possible reusability

119 organisations endorse the FAIR Principles, including…

How familiar are you with the FAIR Principles?



SPRINGER NATURE

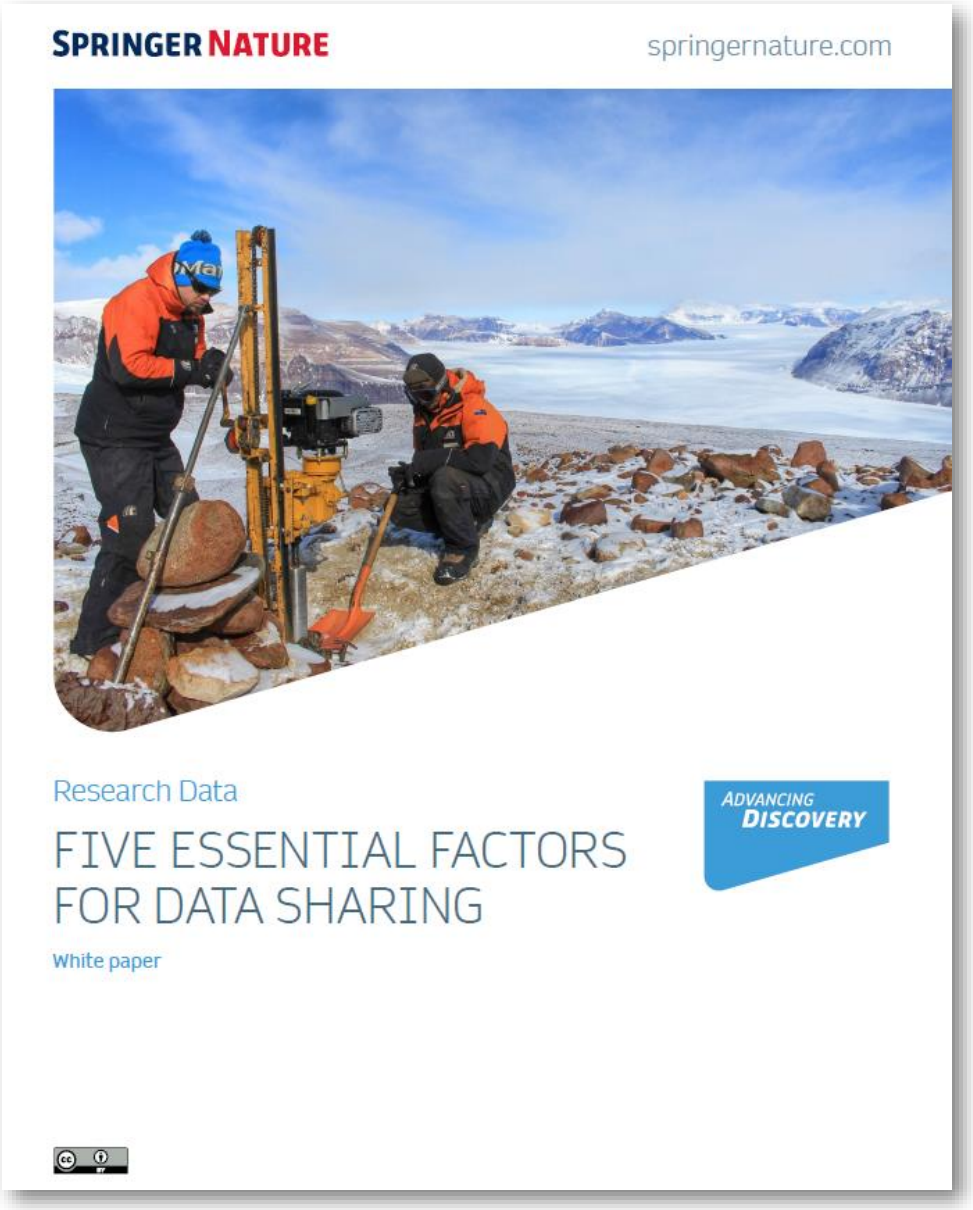
The State of Open Data Report 2019. figshare. Report. https://doi.org/10.6084/m9.figshare.9980783.v2

Working to understand researchers needs and challenges

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

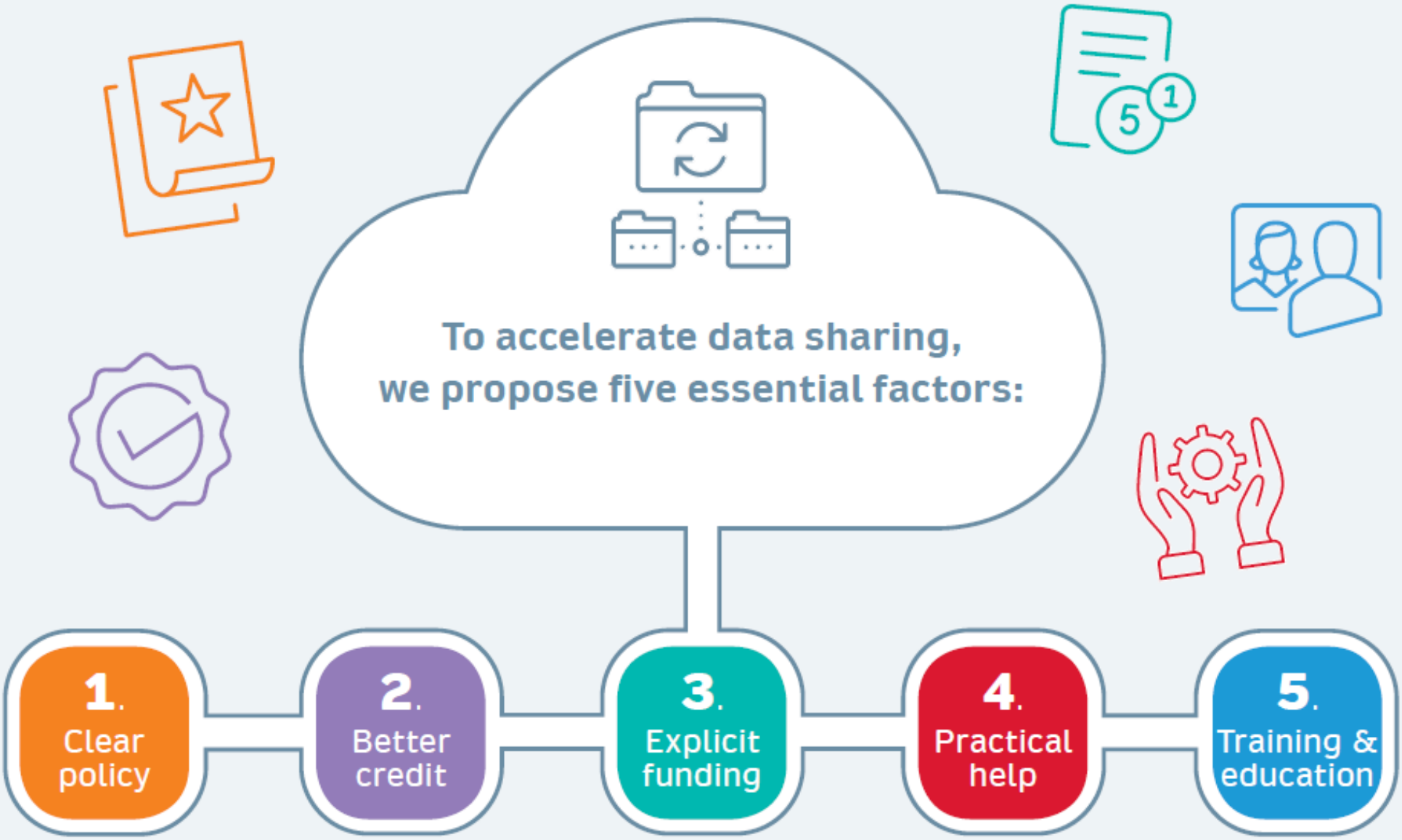
Looking at the big picture

In 2019, Springer Nature published a whitepaper, *Five Essential Factors for Data Sharing*, 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.



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

The Five Essential Factors



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

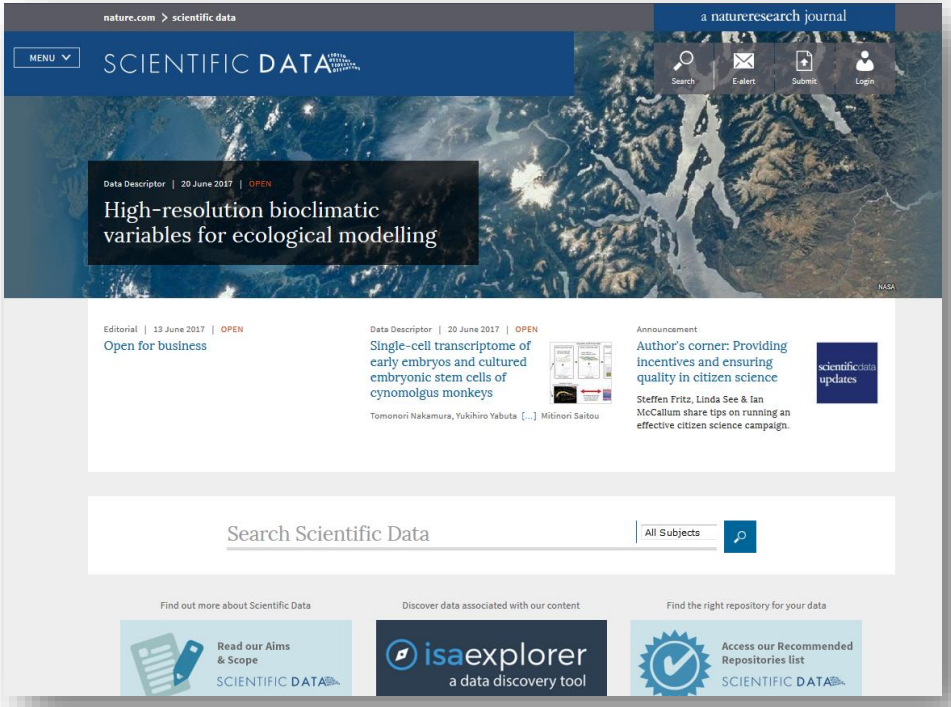
Credit

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



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

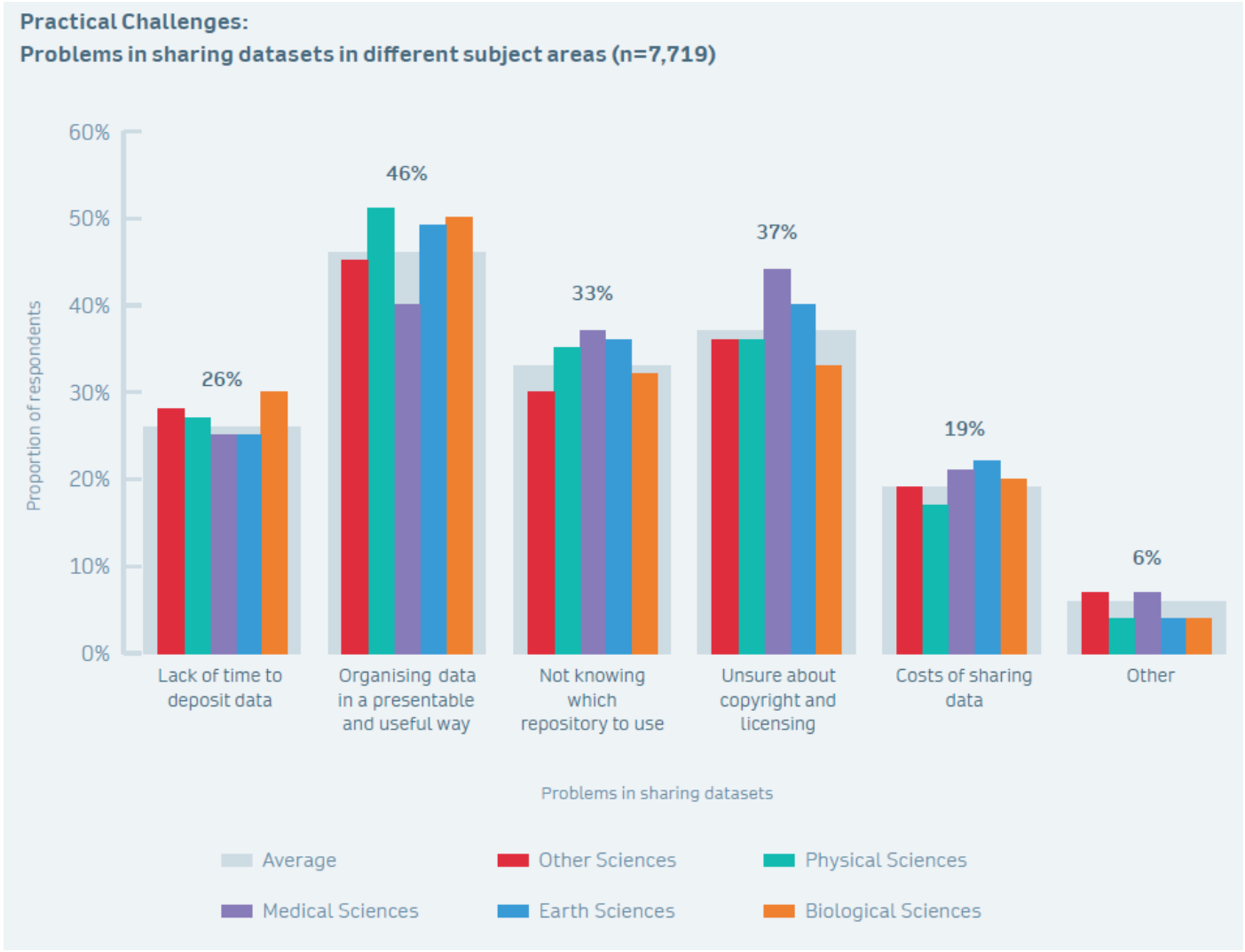
41%

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

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

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

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?

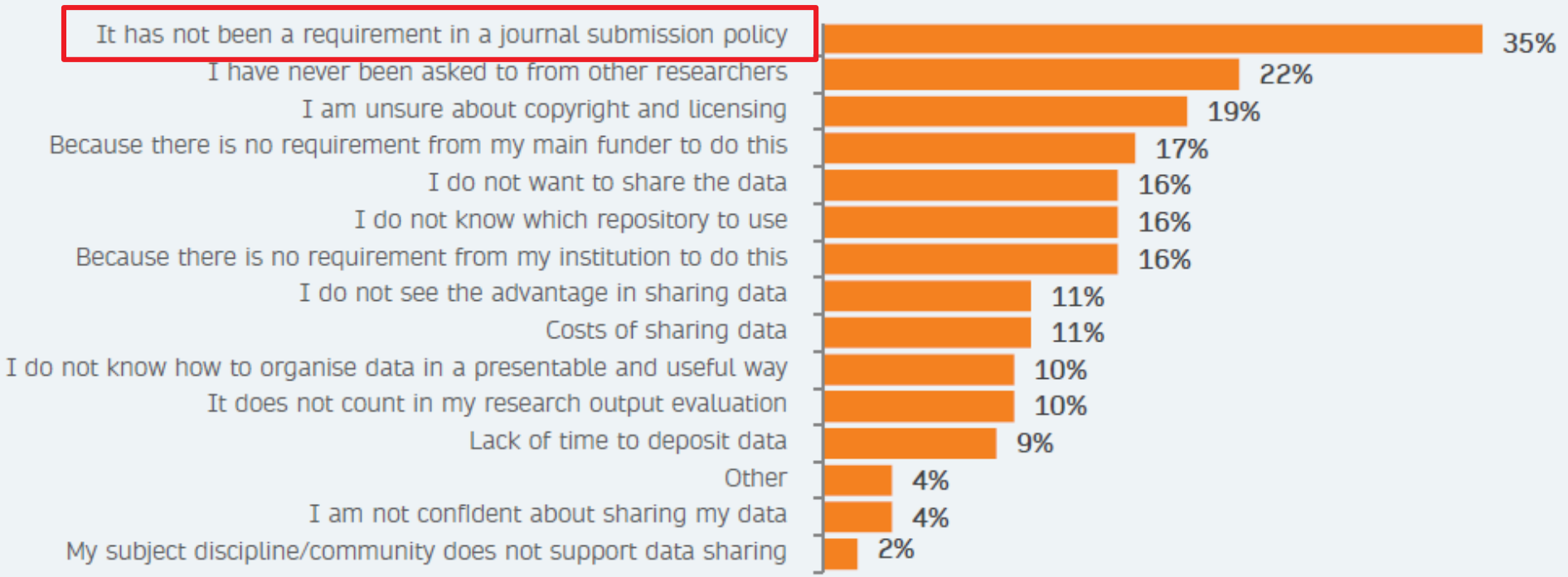
Clear Policy

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.

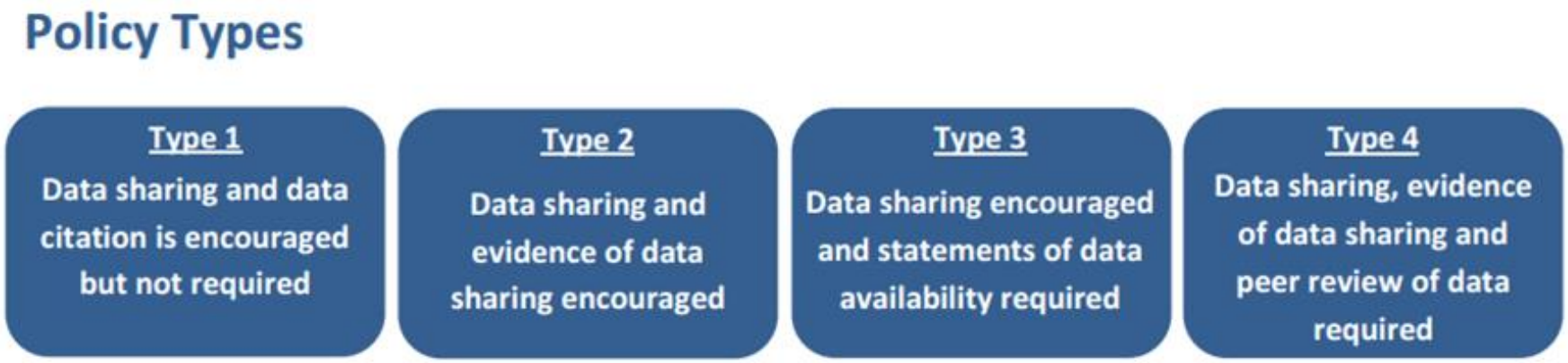
China report

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



Standard research data policies at Springer Nature

- Rolling out standard research data policies since 2016.
- More than 1,600 (~65%) Springer Nature journals have one of these policies.
- Approach is practical and pragmatic, enabling all journals to adopt a policy even if they are new to data sharing.



Standard research data policies at Springer Nature

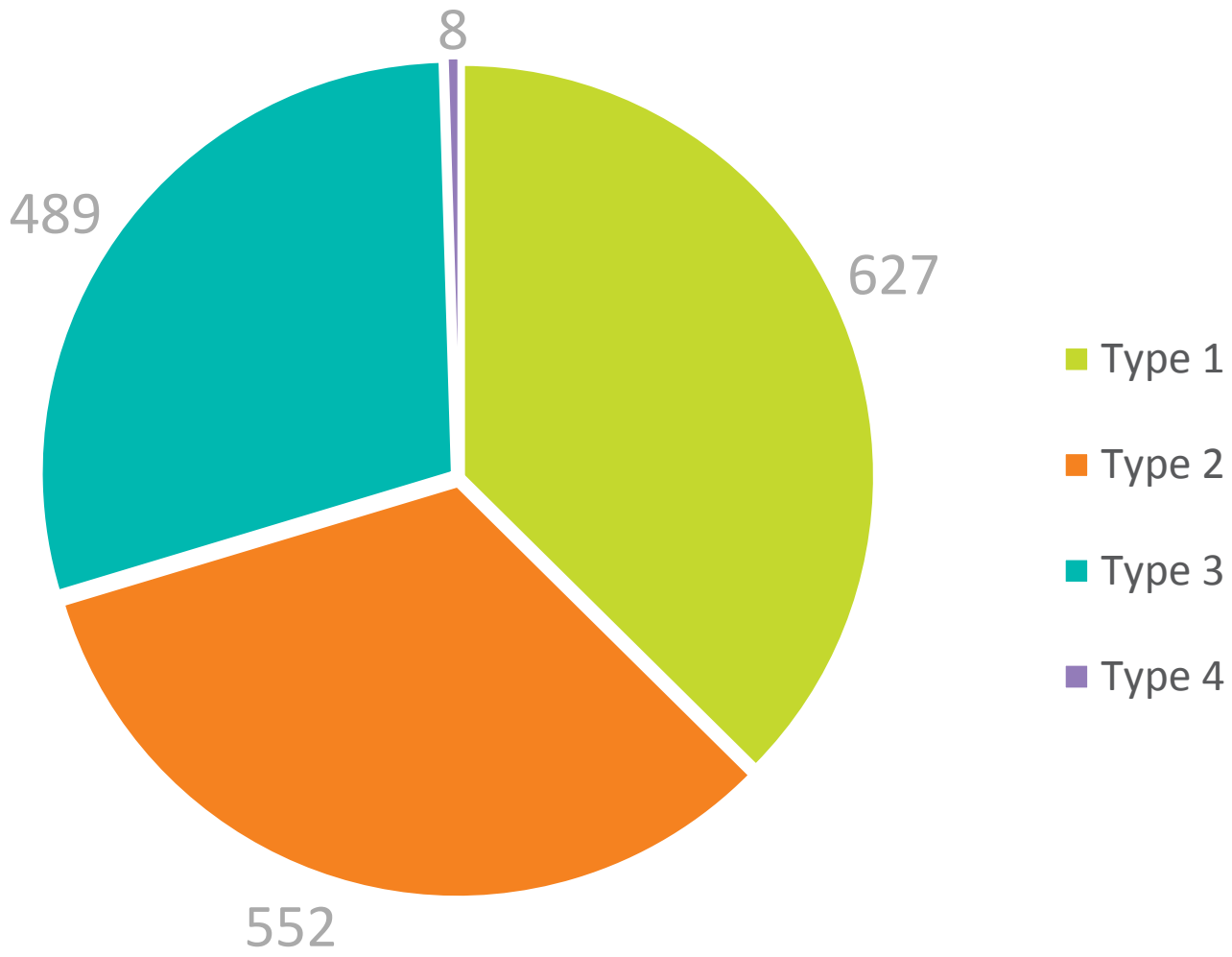
Policy Types

All policy types:

- Recommend **sharing of data via repositories**
- **Allow citation of public datasets** in reference lists/bibliographies
- **Encourage use of publisher helpdesk**

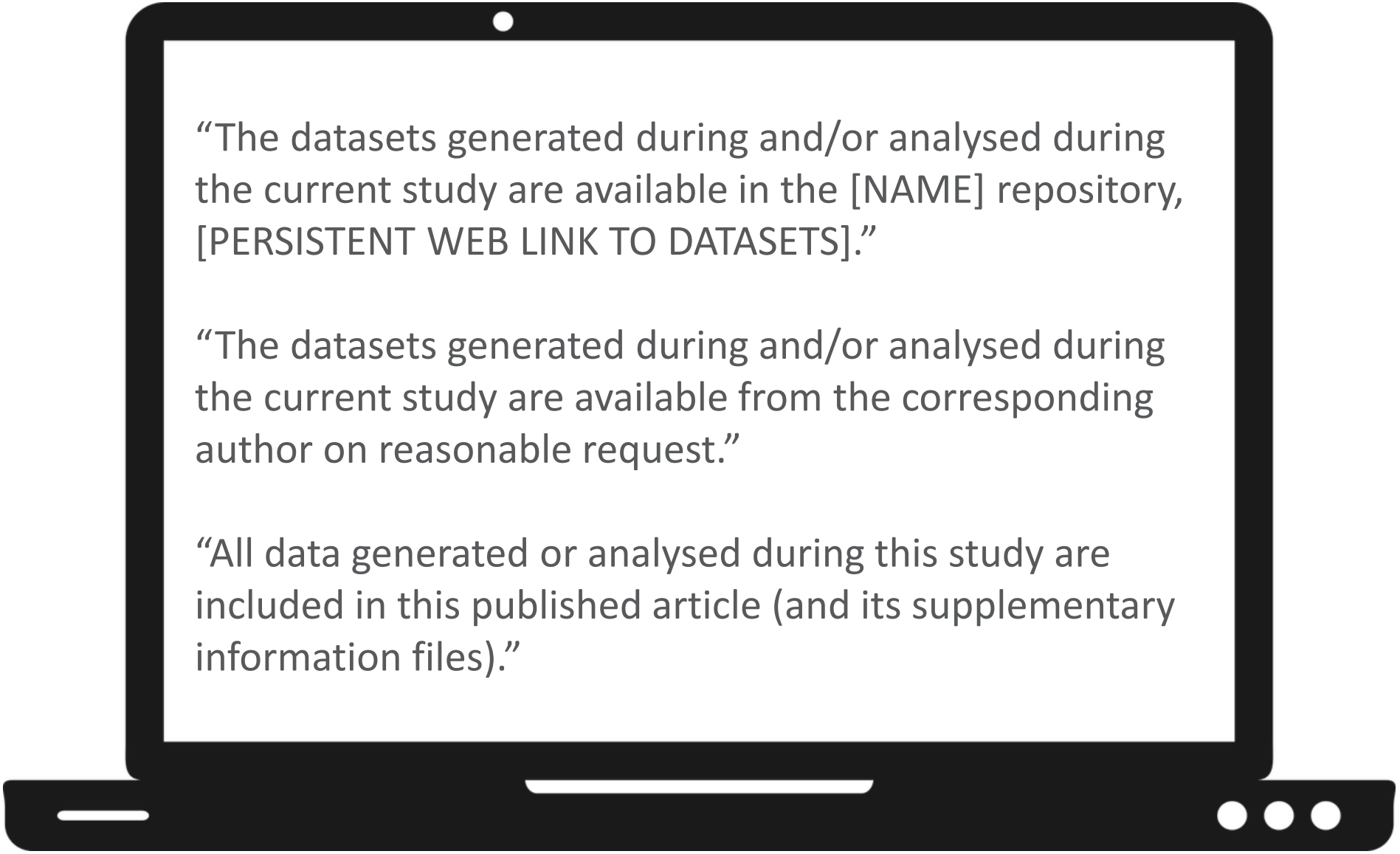
What our journal policies require

497 journals **require** a data availability statement (29%)



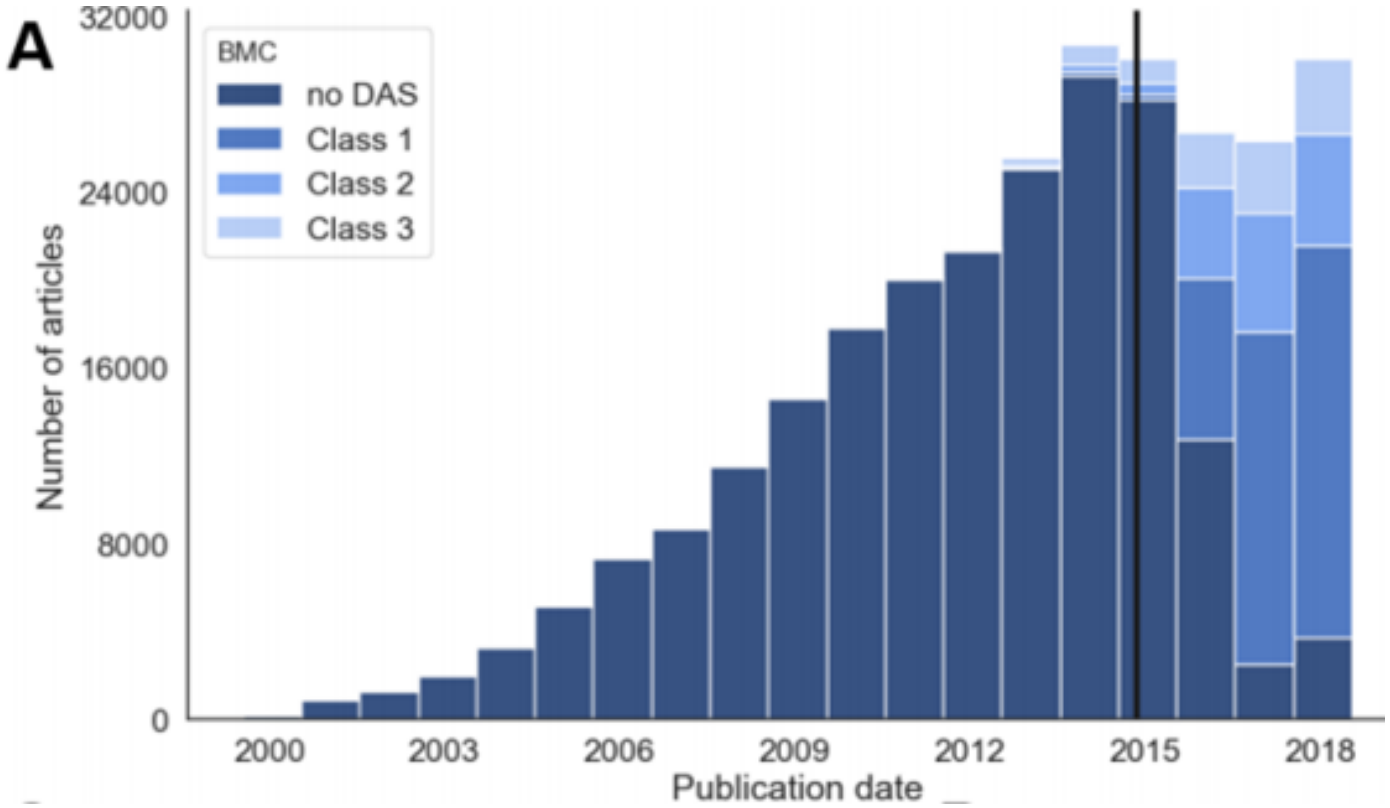
n = 1676

What is a data availability statement (DAS)?



Assessing the impact: prevalence of DAS at BMC journals

- No DAS
- Class 1 (data available on request)
- Class 2 (data contained within the article and supplementary materials)
- Class 3 (a link to archived data in a public repository)



"The citation advantage of linking publications to research data," Giovanni Colavizza, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker & Barbara McGillivray: https://arxiv.org/abs/1907.02565

Aligning policies across publishers

The Data policy standardisation and implementation IG at the Research Data Alliance worked across publishers to align journal data policies

- Springer Nature
- Wiley
- JISC
- Australian National Data Service
- PLOS

Pre-print of the framework "Developing a research data policy framework for all journals and publishers" available at https://doi.org/10.6084/m9.figshare.8223365.v1

14 journal research data policy features arranged as 6 policy types (tiers)

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	○	○	○	○	○	○
Exceptions to policy	○	○	○	○	●	●
Embargoes	○	○	○	●	●	●
Supplementary materials	○	○	○	●	●	●
Data repositories	○	○	○	●	●	●
Data citation	○	○	○	○	●	●
Data licensing	○	○	○	○	○	○
Researcher/ author support	●	●	●	●	●	●
Data availability statements		○	●	●	●	●
Data formats and standards				○	○	●
Mandatory data sharing (specific data types)				●	●	●
Mandatory data sharing (all papers)				○	●	●
Peer review of data				○	○	●
Data Management Plans (DMPs)				○	○	○

Key:

○ = Information required

● = Information and action required

- = Not applicable

Hrynaszkiewicz, Iain; Simons, Natasha; Hussain, Azhar; Goudie, Simon (2019): Developing a research data policy framework for all journals and publishers. figshare. Preprint. https://doi.org/10.6084/m9.figshare.8223365.v1

Next steps for data sharing and policy implementation

- Move approx. 900 journals to Type 3 policy (requiring a data availability statement) by end of 2020
- Align our journal policies with the Research Data Alliance framework
- Explore how we can support journals to encourage or require authors to share their data in repositories.
- Participate in the STM Research Data Year

SHARE-LINK-CITE
STM 2020 Research Data Year

STM launched a **dedicated action plan** to

- Increase the number of journals with data policies (**SHARE**)
- Expand the number of journals depositing data links (**LINK**)
- Grow the volume of citations to datasets (**CITE**)

- Workshops, webinars & eLearning modules, hands-on support, and on-site training, by experienced consultants
- Resources, best-practices and get-started toolkit on **stm-researchdata.org**
- Partnering and collaborating with many other organisations (e.g. RDA, CHORUS, Crossref, FAIRsFAIR, Center for Open Science)

For more information please contact:
Joris van Rossum, PhD, Director of Research Data, STM
rossum@stm-assoc.org

Thank you

Rebecca Grant
Research Data Manager, Springer Nature
Rebecca.grant@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

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