The State of Open Data

2021

Mark Hahnel

Founder & CEO, Figshare

Greg Goodey

Research Analyst, Springer Nature

History of the State of Open Data

Been running for 6 years

Had over 21,000 respondents from 192 countries over that time

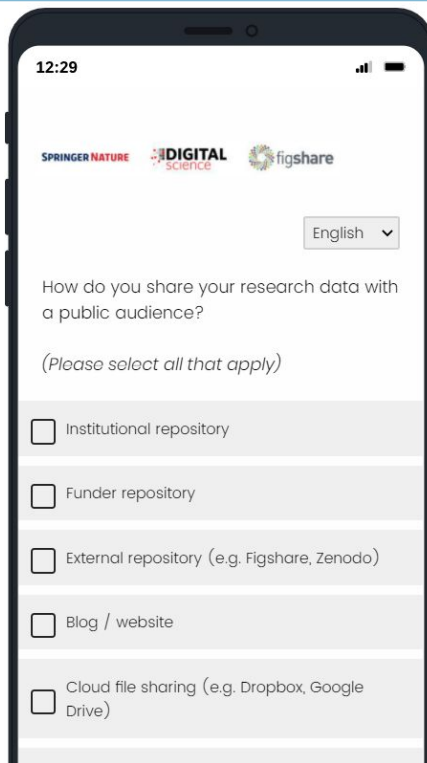Provided us with a sustained look at the state of open data over time

Why is Springer Nature involved?

To move to a truly open research environment, we as a community need to get better at taking concrete steps to drive forward appropriate measures to better support our research community. Springer Nature is firmly committed to this and we continue to work closely with our partners, such as Figshare and Digital Science, to better develop solutions and understanding around data sharing together.

Eugenie Regan, Vice President, Research Solutions, Springer Nature

SPRINGER NATURE

Research methodology

- **Survey hosted on Qualtrics**

- **Translated into Chinese, Japanese and German**

- **Ran from May 25th until August 2nd 2021**

- **Distributed via a number of channels including:**

  - Springer Nature, Digital Science and Figshare marketing email lists
  - Social media (including Facebook, Twitter, WeChat)
  - Blog post

  - **Total of <u>4,491 responses</u>**

SPRINGER NATURE

Demographics

2021 survey population

Region

North America: 17%
US: 15%

Europe: 36%
Germany: 6%; UK 5%

Asia: 30%
India: 11%; China: 3%

Africa: 7%
Nigeria: 2%

South America: 7%
Brazil: 3%

Australasia: 3%
Australia: 2%
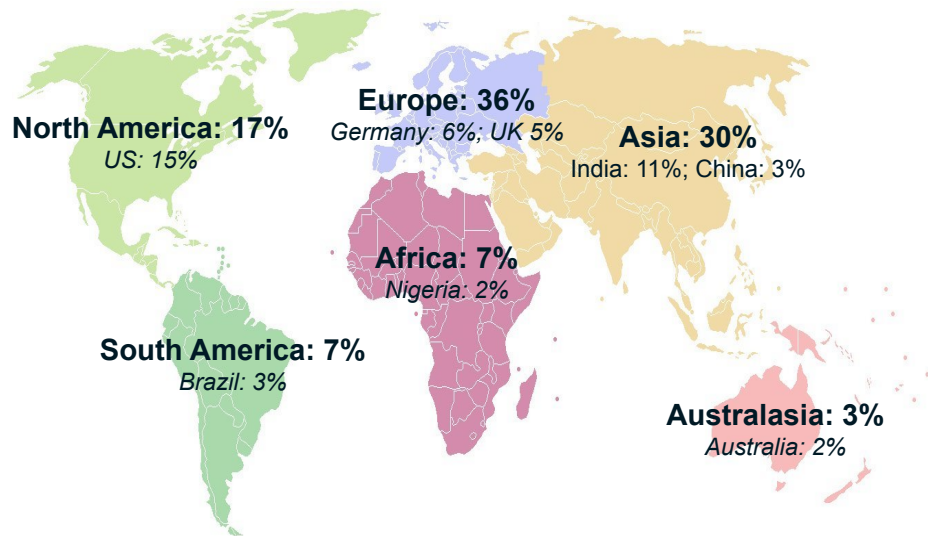
Field of interest

Biomedical sciences: 20%

Biological sciences: 18%
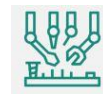
Applied sciences: 15%

HSS: 12%

Physical sciences: 8%

Earth & environmental sciences: 7%

Other: 19%

Inferred career stage*

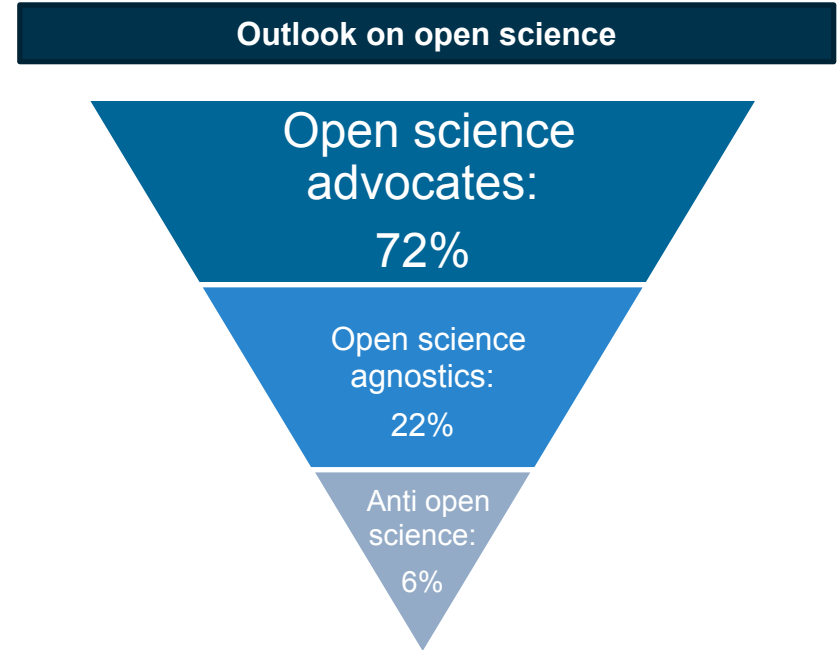Late-career researcher: 58%

Mid-career researcher: 13%

Early-career researcher: 29%

*Career staged inferred from year that the respondent first published a peer review article (LCR - >10 years ago; ECR - < 5 years ago) and tenure status where available

Receptiveness to Open Science

Respondents were asked to what extent they agree or disagree with the following statements:

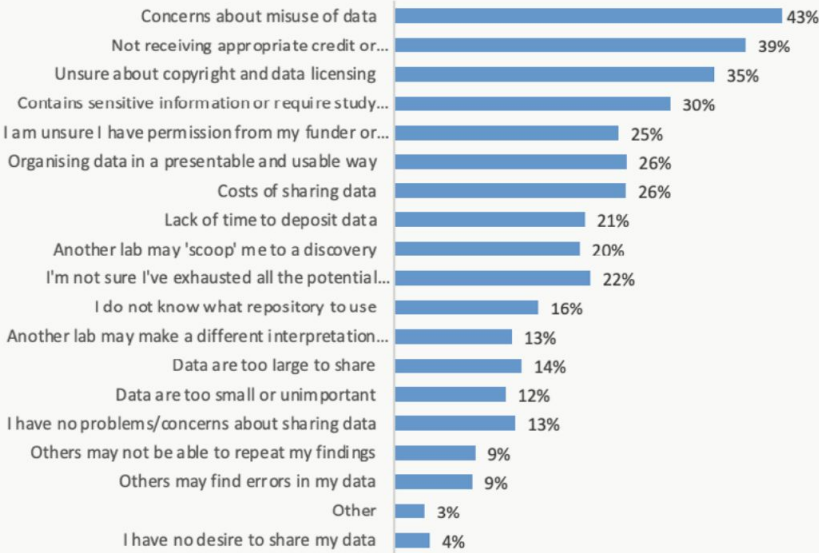- *"Making research articles open access should be common scholarly practice."*

- *"Making research data openly available should be common scholarly practice."*

- *"Making all research outputs openly available should be common scholarly practice."*

Outlook on open science

Open science advocates:

72%

Open science agnostics:

22%

Anti open science:

6%

*Developed using a latent class analysis

SPRINGER NATURE

3 **key takeaways**

1 **There is more concern about sharing data than ever before**

What problems/concerns, if any, do you have with sharing datasets?

|	
Concerns about misuse of data	43%
Not receiving appropriate credit or...	39%
Unsure about copyright and data licensing	35%
Contains sensitive information or require study...	30%
I am unsure I have permission from my funder or...	25%
Organising data in a presentable and usable way	26%
Costs of sharing data	26%
Lack of time to deposit data	21%
Another lab may 'scoop' me to a discovery	20%
I'm not sure I've exhausted all the potential...	22%
I do not know what repository to use	16%
Another lab may make a different interpretation...	13%
Data are too large to share	14%
Data are too small or unimportant	12%
I have no problems/concerns about sharing data	13%
Others may not be able to repeat my findings	9%
Others may find errors in my data	9%
Other	3%
I have no desire to share my data	4%

Problems/concerns with sharing data

over the last 4 years

" *The urgent need to better understand and treat the virus in 2020 brought unprecedented collective and collaborative action.* "

	2018	2019	2020	2021
Concerns about misuse of data	518 / 36%	2082 / 37%	1881 / 38%	1920 / 43%
Not receiving appropriate credit or acknowledgement	478 / 33%	1834 / 32%	1531 / 31%	1739 / 39%
Unsure about copyright and data licensing	508 / 35%	1858 / 33%	1479 / 30%	1585 / 35%
Contains sensitive information or requires consent	416 / 29%	1274 / 22%	1535 / 31%	1366 / 30%
I am unsure I have permission from my funder or institute	458 / 32%	1454 / 26%	1291 / 26%	1110 / 25%
Organising data in a presentable and usable way	456 / 32%	1434 / 25%	1137 / 23%	1150 / 26%
Costs of sharing data	274 / 19%	1497 / 26%	1192 / 24%	1147 / 26%
Lack of time to deposit data	300 / 21%	1389 / 24%	984 / 20%	942 / 21%
Another lab may 'scoop' me to a discovery	0	0	0	918 / 20%
I'm not sure I've exhausted all the potential findings yet	0	0	881 / 18%	971 / 22%
I do not know what repository to use	333 / 23%	1124 / 20%	861 / 17%	713 / 16%
Another lab may make a different interpretation of my data	0	678 / 12%	694 / 14%	581 / 13%
Data are too large to share	191 / 13%	674 / 12%	571 / 12%	630 / 14%
Data are too small or unimportant	245 / 17%	626 / 11%	513 / 10%	550 / 12%
I have no problems/concerns about sharing data	28 / 2%	661 / 12%	589 / 12%	600 / 13%
Others may not be able to repeat my findings	0	409 / 7%	379 / 8%	402 / 9%
Others may find errors in my data	0	416 / 7%	386 / 8%	387 / 9%
Other	100 / 7%	224 / 4%	131 / 3%	151 / 3%
I have no desire to share my data	0	145 / 3%	97 / 2%	177 / 4%

Sharing and reusing Covid data

About a **third** of respondents indicated that they have **reused** their own or someone else's **openly accessible data** more during the pandemic than before
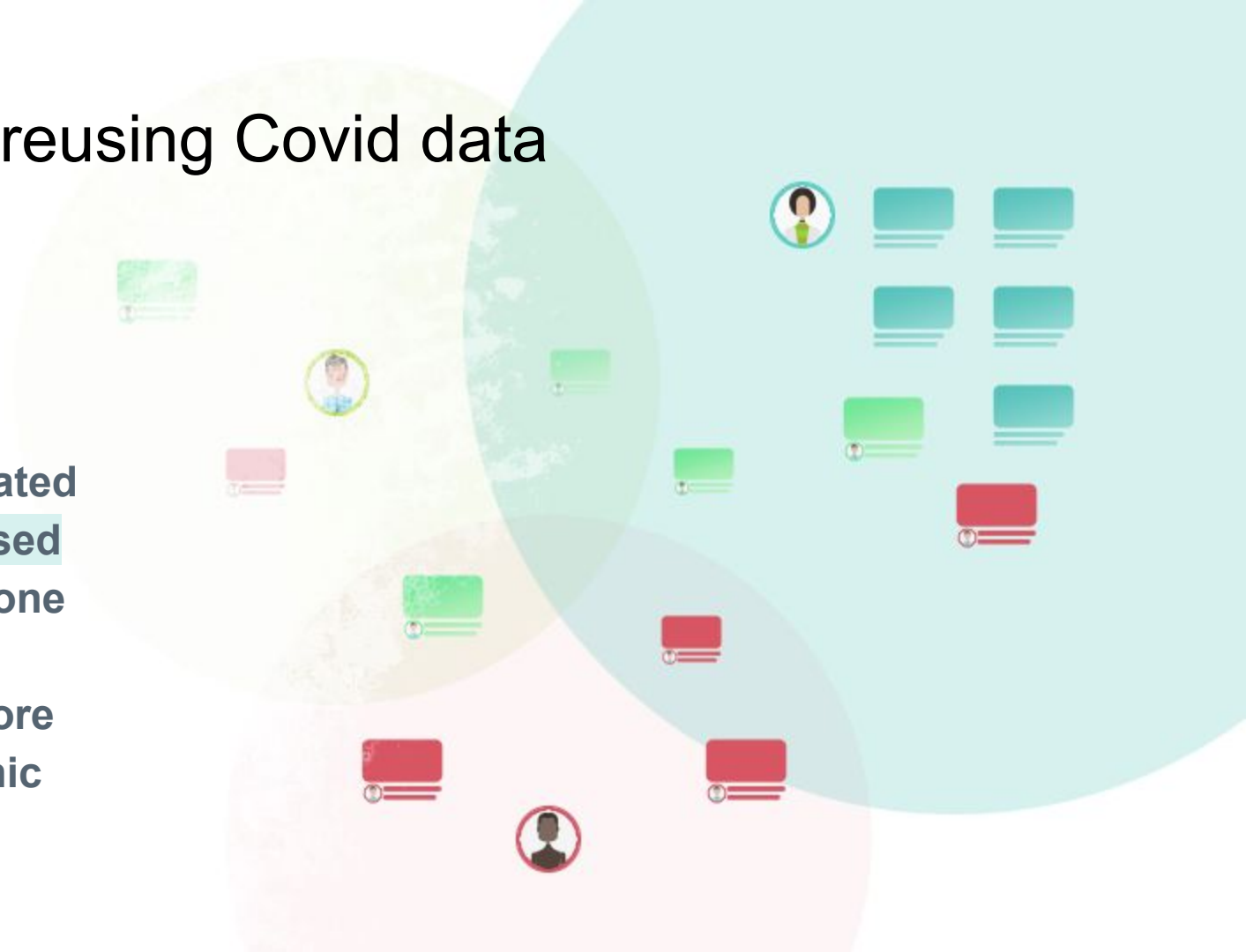
Natasha Simons' Keynote on why open data matters

" Hurdles to data sharing in the area of policy and cultural change will fall short if we do not have underpinning research infrastructure and the experts needed to run the infrastructure. "

Natasha Simons
Associate Director, Data & Services
Australian Research Data Commons (ARDC)

3 **key takeaways**

2 **There is more familiarity and compliance with the FAIR data principles than ever before**

F A I R

3 **Repositories, publishers, and institutional libraries have a key role to play in helping make data openly available**

35%
relied upon
repositories

34% upon
publishers

30% upon
institutional libraries

Key takeaways for institutions

30% of survey respondents said they would **rely upon their institutional library** for help making their research data openly available

Almost **half** of respondents share their research data in an **institutional repository** for a public audience

58% of respondents would like greater **guidance from their institution** on how to comply with their data sharing policies

The role of data curation in enhancing data and metadata quality

A day in the life of a data curator: the steps, challenges, and rewards of the data review process

Dr Connie Clare

Community Manager

Contributors:

Marta Teperek

Head, Research Data Services

Jan van der Heul

Data Curator

4TU.ResearchData

upload

review

feedback

revisions

Consolidating research data management infrastructure: a vital piece of the FAIR jigsaw & (meta)data quality improvements

Damon Strange

Digital Humanities Sustainability Project Manager

University of Oxford

"Putting all of your eggs in one basket" is an idiom with negative connotations, for example, when you're referring to personal finance or data storage practice. But in our case, for the University of Oxford's Sustainable Digital Scholarship (SDS) service, this is exactly what we are trying to do for digital research, offering digital research projects guidance, support, and a long-term home for their digital outputs. The opportunities to converge and consolidate research data management infrastructures onto managed, shared services (e.g., Figshare) are vast, but are also not without their challenges.

Findable – A very simplistic view of meeting this principle could be the simple act of hosting research data on a platform like Figshare to make it more findable (and more accessible, interoperable and reusable) than some current hosting arrangements due to native features of the platform. However, the SDS team do offer support and guidance to researchers when it comes to metadata mapping and field creation for their projects to ensure items are well-described and custom metadata is used (where relevant) to make research more discoverable.

Accessible – It is quite often the case with some research outputs that not all the data can be made fully open for reasons ranging from personal data to copyright concerns. It has been very useful to have the feature to gate certain data items behind Single Sign-On for our repository and offer varying levels of restricted access or embargo.

Interoperable – Given the fact that many of the research projects the SDS service supports are from a Humanities-leaning discipline, the range in topics and required metadata categories have been extensive. However, we continue to work toward encouraging and promoting the use of commonly used controlled vocabularies and standardizing where a standardized approach is applicable.

Reusable – Given we are currently only 9 months into our journey as a new service at the University only time will tell. However, our hope is that as we work with and onboard more projects, we can look to reuse metadata standards and techniques to yield not only efficiencies but improved clarity & quality of (meta)data.

Damon Strange

Hi, I'm Damon Strange, project manager of the University of Oxford,

Tips for engaging your researchers in open data sharing practices
– practical guidance from the University of Pretoria

Veliswa Tshetsha

Senior Coordinator:

Open Scholarship

University of Pretoria

Rosina Ramokgola

Data Curation Officer

University of Pretoria

Pfano Makhera

Metadata Specialist: Scholarly

Communications

University of Pretoria

Open Source and Open Data: Collaboration is Key

Sara Gonzales
Data Librarian
Galter Health Sciences Library & Learning Center, Northwestern University

As the world has risen to the challenge of the COVID-19 pandemic, researchers and the public alike have developed a greater appreciation for accurate and reliable open data sources. From the National Institutes of Health's Open-Access Data and Computational Resources to Address COVID-19 to the local data sources that inform our nightly news updates, open data have become a more important force in our lives than ever before. People have a stake in data and, increasingly, people are contributing their time and getting involved in developing the tools that help researchers, and the world at large, interact with that data. One way we are achieving this locally at Northwestern University is through participation in open source data repository development.

Sara Gonzales
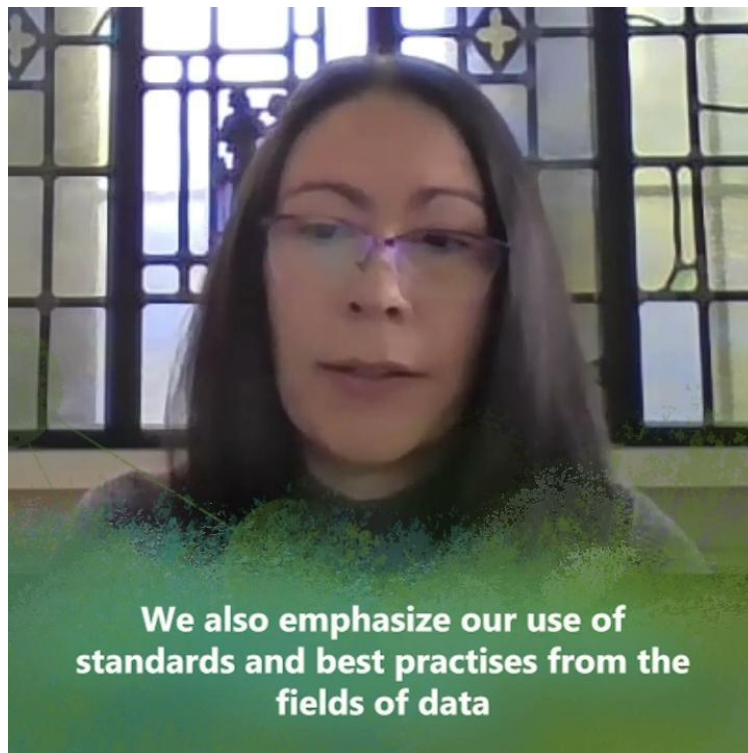
Open data and the life sciences: the turning point

Daniel Kipnis

Life Sciences Librarian, Rowan University

Key takeaways for publishers

47% of survey respondents said they would be **motivated** to share their data if there was a **journal or publisher requirement** to do so

53% of survey respondents obtained research data **collected by other research groups** from within a published research article

53% of respondents said it was extremely important that data are available from a **publicly available repository**

How publishers can uphold research quality through embedded data support

Graham Smith

Research Data Manager

Springer Nature

Scholarly publishers have a fundamental duty in upholding research quality, from editorial expertise to managing the peer review process. Research data is a growing part of Springer Nature's policies, systems and workflows and a key component of the ambition that research outputs should be openly available and reproducible. In order to uphold the quality of data alongside that of the related literature, we are building on the specialist support developed for data articles, developing processes more widely applicable across our journals.

Graham Smith
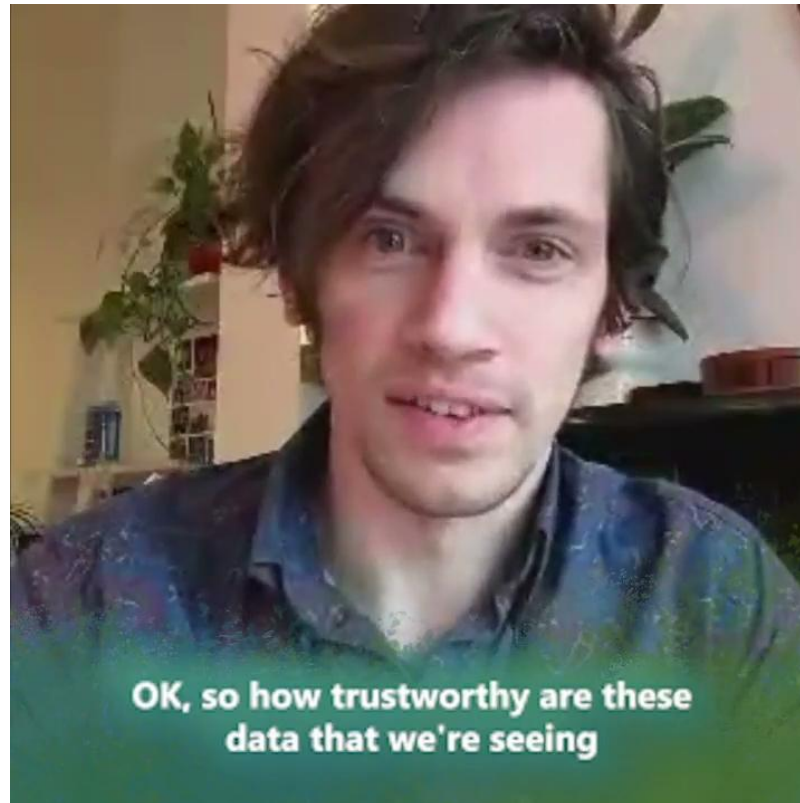
Key takeaways for funders/government agencies

52% of survey respondents said funders should make the sharing of research data part of their **requirements for awarding grants**

48% of respondents said that **funding should be withheld** (or a penalty incurred) if researchers do not share their data when the funder has mandated that they do so at the grant application stage

73% of survey respondents strongly or somewhat support the idea of **a national mandate** for making research data openly available

33% would like more **guidance** on how to comply with government policies on making research data openly available

J-STAGE Data: evidence data platform for Japan's learned society publishing

Keisuke Iida

Department for Information Infrastructure

Japan Science and Technology Agency

Interviewed and translated by: **Nobuko Miyairi**

Scholarly Communications Consultant

Japan Science and Technology Agency (hereafter JST) promotes research and development in Japan through funding basic research, commercialization of new technology, and promoting international collaboration. JST also provides a variety of information platforms and services, including J-STAGE, an electronic journal platform. In October 2019, JST commemorated the 20th anniversary of J-STAGE, which now hosts more than 3,000 journals, conference proceedings and other academic contents published in Japan. J-STAGE Data is a new data repository to make underlying data available for J-STAGE publications.

42% of respondents in the State of Open Data survey in Asia believe **funders should withhold funding from or penalize researchers** for not sharing their data if the funder has mandated that they do so at the grant application stage

Item's title

Author Name|

Submit

CHECKED
& FAIR

NIH

Health

Ivermectin: How false science created a Covid 'miracle' drug

By Rachel Schraer & Jack Goodman
BBC Reality Check

Major problems included:

- The same patient data being used multiple times for supposedly different people

- Evidence that selection of patients for test groups was not random

- Numbers unlikely to occur naturally

- Percentages calculated incorrectly

- Local health bodies unaware of the studies

How open data can help validate research and combat scientific misinformation

Prof Ginny Barbour
Co-lead, Office for Scholarly Communication,
Queensland University of Technology
and Director
Open Access Australasia

" *Open data has two important, overlapping roles to play in increasing the credibility of research: validating research, so that researchers can trust it, and combating scientific misinformation, so that wider society can trust it.* "

Many respondents are putting in the hard work of data sharing:

74% are making **data management plans**

76% curate their data for sharing

66% are **familiar with the FAIR principles** that underpin data sharing

2009

The Fourth Paradigm:
Data-Intensive Scientific Discovery.

1. Empirical Evidence - The sky is blue
2. Scientific theory - The sky is blue every day
3. Computational science - Using advanced computing capabilities to understand and solve complex problems

The goal, Dr. Gray insisted, was not to have the biggest, fastest single computer, but rather "to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other."

Implicit in the idea of a fourth paradigm is the ability, and the need, to share data.

Lab 1

Lab 2

Lab 3

NEW
TREATMENT
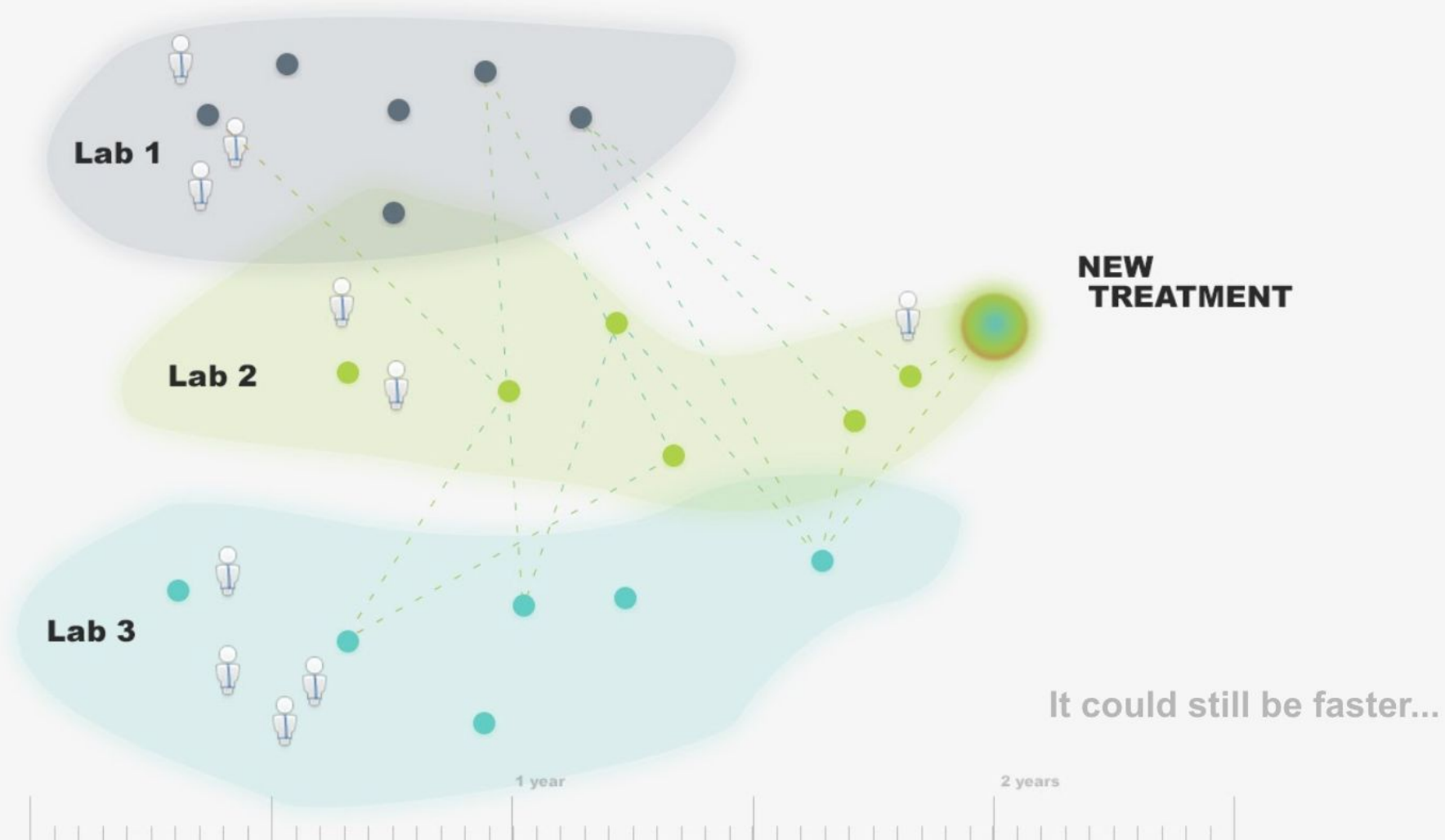
It could still be faster...

1 year

2 years

"The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another" Jim Gray

FAIR-ly provide:

- Access to literature
- Ability to publish literature
- Access to data
- Ability to publish data

Tech

People

Culture

Repositories
Libraries
Archives
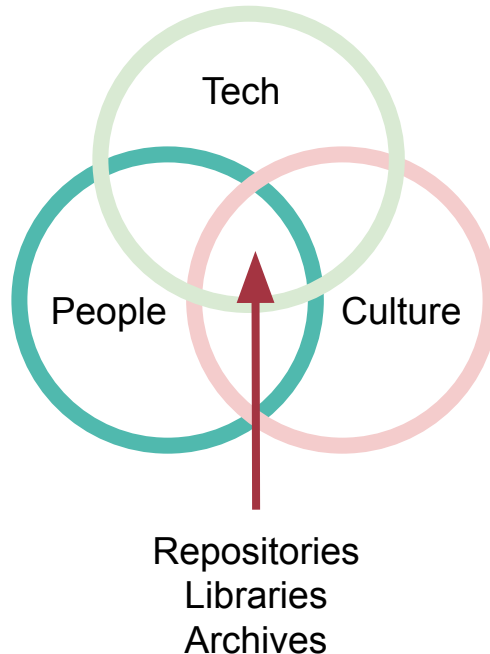
Mark Hahnel

Founder & CEO, Figshare
mark@figshare.com
@markhahnel
@figshare

Read the full State of Open Data 2021

doi.org/10.6084/m9.figshare.17061347

or **knowledge.figshare.com/sood**

10 years of fig**share**