

The State of Open Data 2025: A Decade of Progress and Challenges

With opening remarks from Digital Science's CEO, **Daniel Hook**, and
Springer Nature's CPO, **Harsh Jegadeesan**



Contents

Opening remarks	3	Experts' voices on open data	18
Digital Science	3	Mandates and policies can help drive change but tailored support is needed	19
Springer Nature	4		
Introduction	5	Open data has long-term impact that is difficult to measure by institutions	20
		That credit gap again: A lack of recognition for researchers	22
Survey results: Longitudinal changes in open science awareness and practice	6	Interoperability and AI integration will be the next steps	24
Familiarity with FAIR principles: A tale of two cohorts	7	What do the next 10 years hold for open data?	25
Enthusiasm for open data mandates is falling globally	10		
Strong support for open practices persists	12		
The credit gap: Recognition versus reality	14	Recommendations	27
Adoption of AI tools	15	Conclusion	28
Summary of cross-cutting themes and implications based on survey data	17	Acknowledgements	29
		Appendix 1: About the State of Open Data and 2025 survey	30
		Appendix 2: About open data at Springer Nature, Figshare and Digital Science	32



DR DANIEL HOOK

Chief Executive Officer, Digital Science



A bit more than a decade ago we realised that the open data movement had the potential to be the most consequential change in how research had been conducted and shared since the advent of the research journal in the late 1600s.

It wasn't until 2016, however, that it occurred to us to establish the State of Open Data Report with the core purpose of listening to the research community and understanding its evolving relationship with data sharing against the background rise of open research.

We wanted our work to be aligned with the principles of the open research movement, and so we made the survey data public on Figshare. We also hoped that if the report were to be successful it might be repeated, and hence that we might be able to build a source of longitudinal data that might be of value to the community in tracking this formative period in the research ecosystem from both cultural and social perspectives. As a result, we specifically chose a core set of questions that we thought would endure and which could be revisited annually.

At the beginning of this journey, we knew that meaningful progress would depend not only on new policies and infrastructure, but also on understanding the lived experiences, motivations, and concerns of the researchers being asked to share their data.

We believe that the survey data and accompanying reports have given these voices a platform - capturing, year by year, how researchers felt about data sharing, what obstacles they faced, and what support they needed.

As the community's expectations have been shaped by new funder mandates, FAIR principles, and growing global momentum towards transparency, their voices have charted a steady cultural shift. Researchers have described increasing recognition of open data's value, such as its role in improving the impact and reach of research, its reproducibility and reuse, and in strengthening research integrity.

Today, the State of Open Data Report stands as a unique longitudinal record of global researcher sentiment that has evolved through a decade of rapid change. At Digital Science, we're proud to have supported the curation of this record in partnership with Springer Nature and Figshare, and to have championed researchers' voices.

By listening to the community—and reporting its views across 10 years of State of Open Data reports—we have helped highlight persistent challenges, offered recommendations for meaningful change, and demonstrated how openness has moved from aspiration to a core practice in modern research.



Today, the State of Open Data Report stands as a unique longitudinal record of global researcher sentiment that has evolved through a decade of rapid change.



SPRINGER NATURE

HARSH JEGADEESAN

Chief Publishing Officer, Springer Nature



At Springer Nature, we believe that open data is essential to accelerating discovery and driving research forward responsibly.

Over the past decade, open data has become more widespread, but progress remains uneven—partly because research is often fragmented, with data in one place, methods in another, and code somewhere else. From a decade of the State of Open Data survey and report, developed with partners Figshare and Digital Science, we have learned that while attitudes towards sharing are improving, recognition and practical support still lag behind. These insights from researchers help shape the community's understanding of challenges and opportunities in data sharing.

Building on these learnings, we listen to researchers and collaborate with our partners to better understand needs and develop solutions that support their needs. Our latest report highlights differences across regions and disciplines and shows that recognition for sharing data is still insufficient. Translating policy into practice requires collective effort. Researchers, publishers, and institutions need to work together to make data sharing simpler and more consistent through clear guidance on repositories, standards, licensing, and documentation.

Just as important is ensuring that open data is treated as a credible research output, one that is linked, citable, and visible and therefore recognised and valued as part of career progression and funding decisions.

Looking ahead, the biggest changes in open data will be by human-AI collaboration, content structured for automation and interoperability, and workflows grounded in trust and integrity. These developments will reshape how datasets are created, shared, and verified, taking openness to the next level, making it seamless, automated and trusted.



Looking ahead, the biggest changes in open data will be by human-AI collaboration, content structured for automation and interoperability, and workflows grounded in trust and integrity. These developments will reshape how datasets are created, shared, and verified, taking openness to the next level, making it seamless, automated and trusted.



Introduction

As we mark a decade of the State of Open Data reports, this milestone invites reflection on how far the movement has come, and why it matters more than ever.

What began as a push for openness through advocacy and declarations on open access, has evolved into widespread adoption, and in some countries, formal mandates, driving a global shift towards openness. Today, the FAIR principles (making data Findable, Accessible, Interoperable and Reusable) underpin these policies and practices worldwide. Sharing research data beyond researchers' immediate networks has long been seen as beneficial for

accelerating discovery, fostering collaboration, and enhancing trust and transparency for greater impact.

Despite widespread recognition of the benefits of sharing open data, significant challenges remain. As Sir Nigel Shadbolt wrote in his guest foreword to the inaugural [State of Open Data 2016 report](#), "If we are to really change custom, practice and culture in the research sector we do have to recognise that we need incentives as well as mandates."

A decade later, his comment remains as relevant as ever and it frames the questions explored in this anniversary report: How far have we come in culture and practice? Where do gaps persist? And what strategies can drive meaningful progress in the years ahead?

About the 10th anniversary report

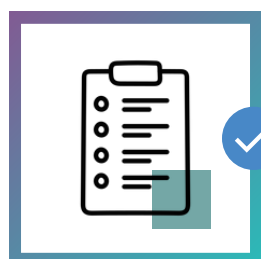
The State of Open Data is an annual collaboration between Digital Science, Springer Nature, and Figshare designed to track this transition. By surveying researchers' perspectives and experiences of open data—including motivations for sharing data, obstacles faced in making data accessible, and the perceived reliability of openly shared data—this annual report aims to provide an ongoing benchmark for change. To learn more about the impact of these reports as well as the 2025 survey methodology and demographics, see [Appendix 1](#).



Survey Insights



Expert Voices



Recommendations

Building on that foundation, this special 10th anniversary report of The State of Open Data aims to share a deeper and more holistic understanding of the barriers and potential solutions in the open sharing of research data, reflecting on the last decade as well as looking forward to the next.

The first section examines the current state of open data based on 2025 survey findings, along with a longitudinal comparison of findings for key themes. The second section brings together insights from interviews with researchers, librarians, and data experts, offering regional perspectives on open data sharing. A list of interviewees can be found in [Acknowledgements](#).

Survey results: Longitudinal changes in open science awareness and practice

This section explores the current state of open data based on 2025 survey findings and long-term trends, highlighting major shifts and ongoing challenges. The data tells a story of growing awareness and engagement with open science principles, alongside persistent barriers to implementation.

Familiarity with FAIR principles: A tale of two cohorts

Familiarity with the FAIR principles has risen sharply since the first time that this question was asked in 2018:

40.6% ↑ The 'familiar with the data principles' share climbs from 15.2% to 40.6%, overtaking 'previously heard of the data principles but I am not familiar with them' in 2024 (which sits at 38.9%).

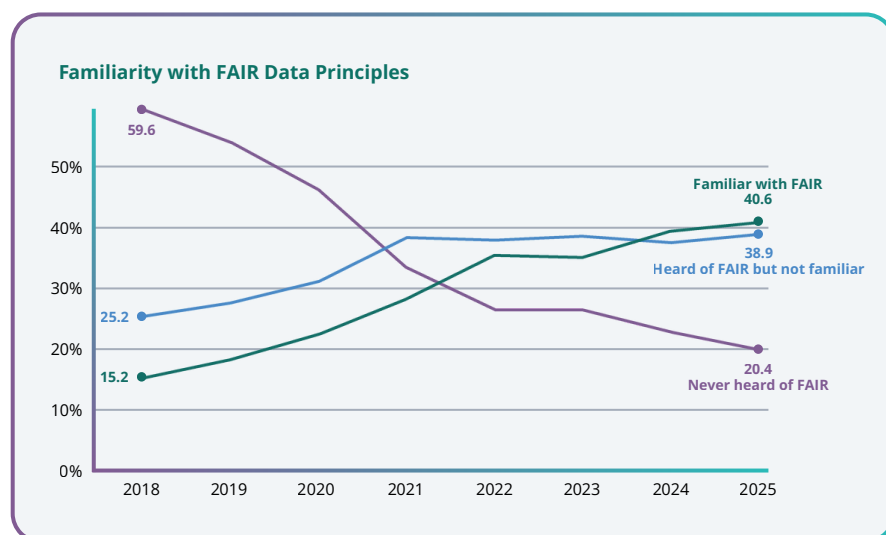
20.4% ↓ Since the FAIR awareness question was introduced in 2018, the share of respondents who had never heard of the FAIR data principles has fallen from 59.6% to 20.4% by 2025.

79.5% Today, nearly 80% have at least some awareness (up from around 40% in 2018), marking a significant shift toward engagement with these principles.



FIGURE 1

How familiar are you with the FAIR data principles (i.e. Findable, Accessible, Interoperable and Reusable)?



Researchers were asked "How familiar are you with the FAIR data principles (i.e. Findable, Accessible, Interoperable and Reusable)?" Response options included: 'Never heard of FAIR before', 'Previously heard of FAIR but not familiar with the principles', and 'Familiar with FAIR'. Yearly sample sizes: 2018 (n = 1,239), 2025 (n = 3,932).

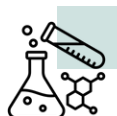
Questions relating to familiarity with the FAIR data principles were introduced in the 2018 survey. Trend data is therefore shown from 2018 onwards in this figure.

Across disciplines too, the awareness of FAIR principles has grown dramatically since 2018 (Fig. 2). In 2018, unfamiliarity was widespread: most researchers had never heard of FAIR, with rates especially high in Business/Investment (80.0%), Chemistry (64.6%) and Materials Science (65.9%).



80.0%

Business/Investment



64.6%

Chemistry



65.9%

Materials Science

By 2025, this picture has changed significantly: Unfamiliarity dropped by 40–60 percentage points across most fields. Business/Investment saw the most dramatic improvement, falling from 80.0% to 20.0%, while the Arts & Humanities nearly halved from 40.9% to 22.2%.



20.0%

Business/Investment



22.2%

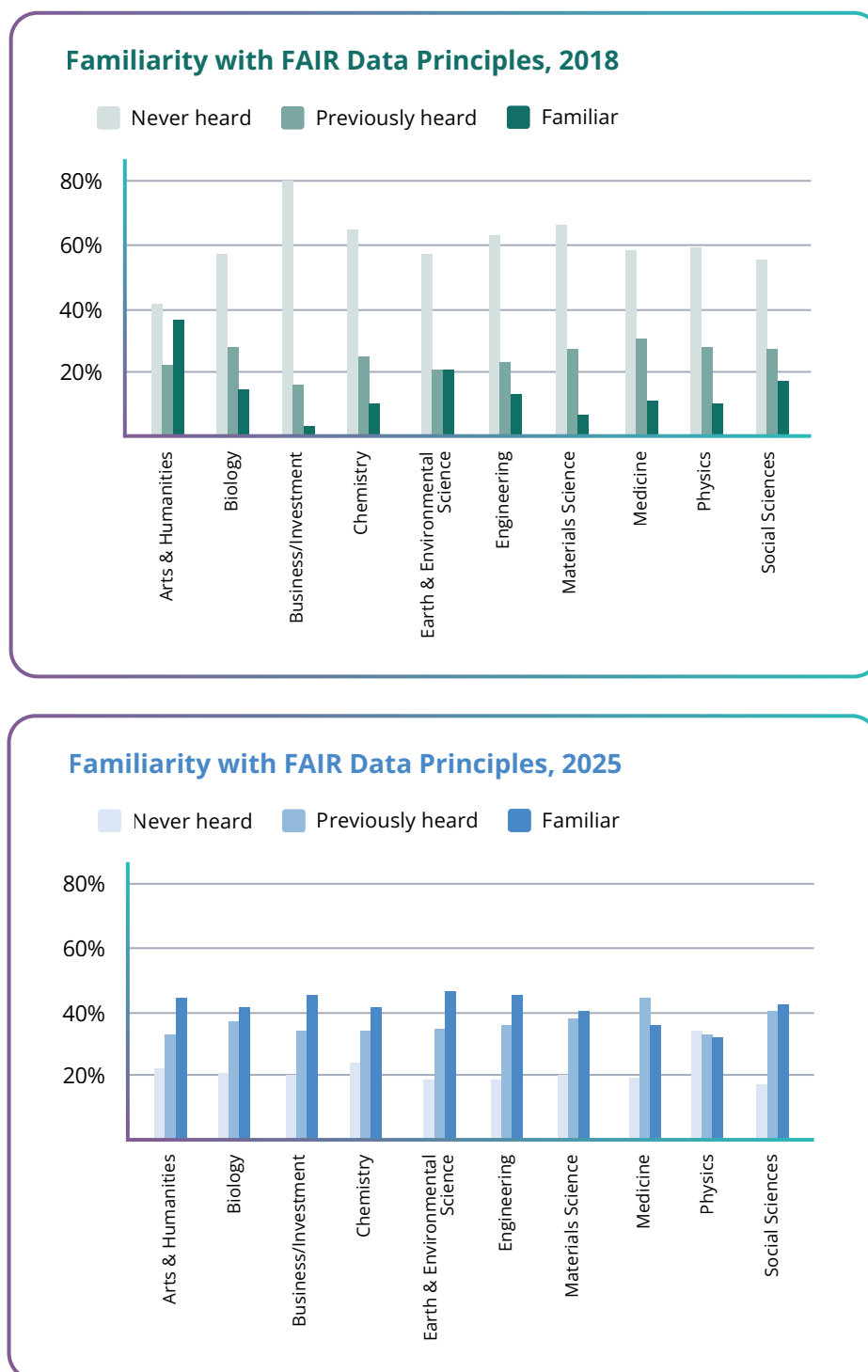
Arts & Humanities

Perhaps most significantly, the proportion of researchers who are now familiar with the FAIR principles (rather than merely having heard of them) has increased substantially.

Engineering researchers saw familiarity rise from 13.6% to 45.6%, while Biology researchers increased from 14.9% to 41.8%, signalling a shift from passive recognition to active engagement with FAIR principles.

FIGURE 2 Familiarity with FAIR data principles by subject area, 2018 & 2025

Awareness and familiarity with FAIR principles have surged across disciplines, signalling a shift towards active engagement with open science and setting the stage for how researchers view and adopt broader open practices.



Researchers were asked "How familiar are you with the FAIR data principles (i.e. Findable, Accessible, Interoperable and Reusable)?" Response options included: 'Never heard of the data principles before', 'Previously heard of the data principles but not familiar with them', and 'Familiar with the data principles'. Yearly sample sizes: 2018 (n=1,239), 2025 (n=3,932).

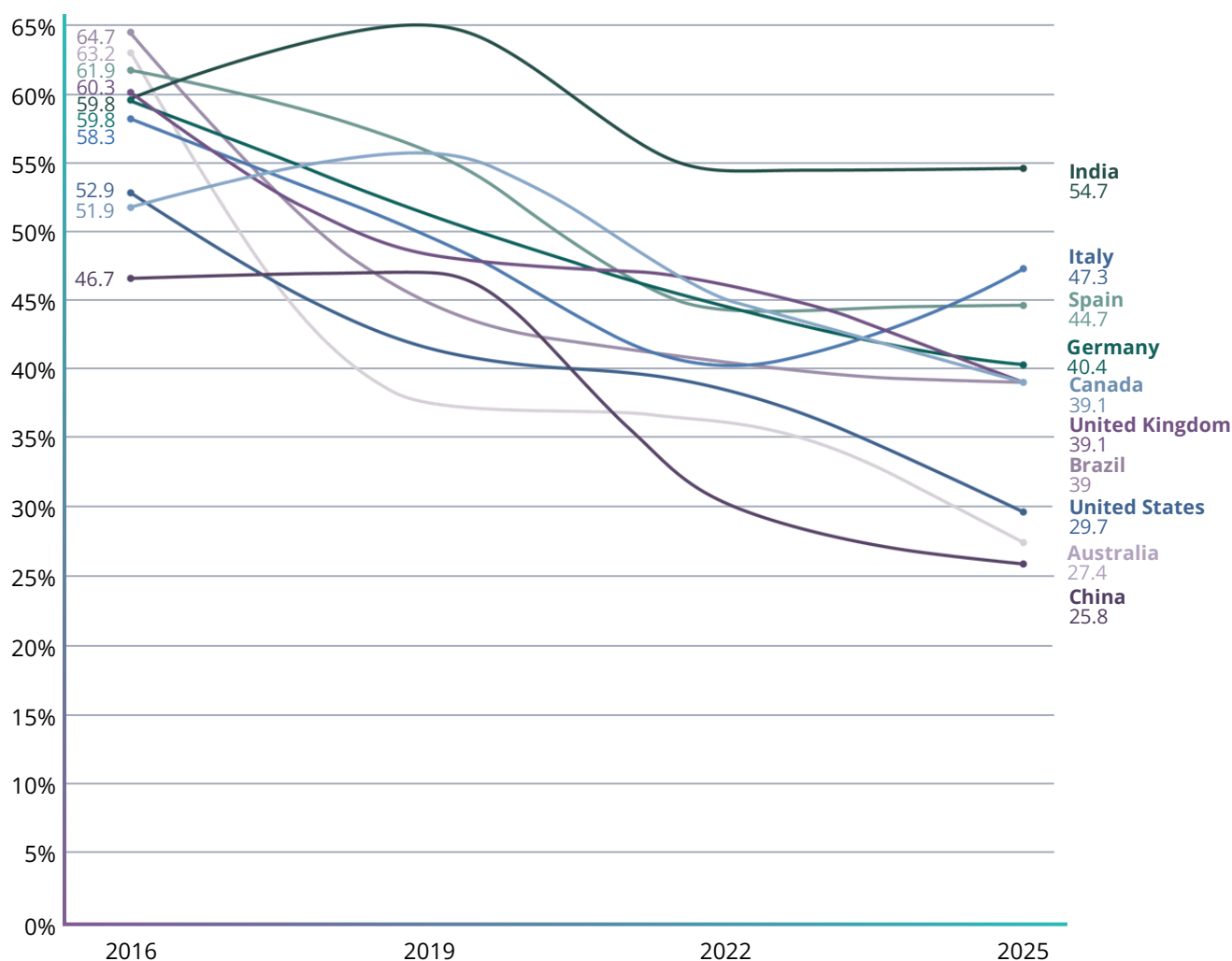
Enthusiasm for open data mandates is falling globally

Support for national mandates—policies that require researchers to share their data—has shifted both globally and regionally over the last 10 years (Fig. 3). Globally, support has remained relatively stable, but this masks significant regional disparities.

FIGURE 3

Researcher support for a national mandate requiring open research data across 10 countries, 2016-2025

Researcher support for a national mandate requiring open research data across 10 countries, 2016-2025



Respondents from Australia, Brazil, Canada, China, Germany, India, Italy, Spain, United Kingdom, and the United States were asked: "How supportive would you be of a national mandate for making research data openly available?" Responses shown as percentage of valid responses per year for those who responded 'Strongly support'. Sample sizes: 2016 (n=945), 2017 (n=1,348), 2018 (n=692), 2019 (n=2,907), 2020 (n=2,430), 2021 (n=2,303), 2022 (n=2,783), 2023 (n=3,139), 2024 (n=2,598), 2025 (n=2,410). Total longitudinal sample N=21,555.

 27.4%

Australia saw the steepest decline in the share of respondents who strongly support a national open data mandate, falling from 63.2% (2016) to 27.4% (2025), a reduction of 35.8 percentage points.

 29.7%

The United States followed a similar trajectory, with net support declining from 52.9% to 29.7%, a reduction of 23.2 percentage points.

While trajectories differ by country, strong support has softened across all ten countries since 2016. By 2025, only India remains above 50% on “strongly support”.

 54.7%

India remains the highest in 2025 at 54.7%, down from 59.8% in 2016.

A mid-range cluster has converged around ~40% in 2025:

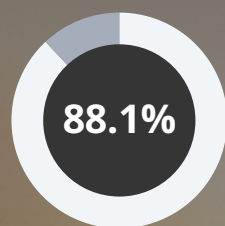

40.4%
Germany
39.1%
Canada
39.1%
UK
39.0%
Brazil **25.8%**

China (25.8%) sits at the lowest end of “strongly support” in 2025.

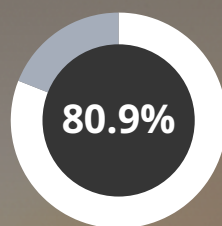
Taken together, the pattern suggests that early high-confidence endorsement of mandates may be giving way to more qualified views as implementation and compliance realities become clearer. This reinforces the need to pair policy with practical infrastructure and support that makes sharing feasible and genuinely reusable.

Strong support for open practices persists

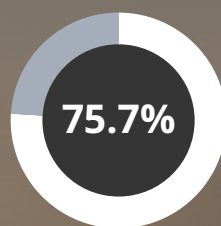
Enthusiasm for open practices remains strong, despite the declining support for mandates. This suggests that researchers continue to value openness in principle, but are increasingly critical of how openness is implemented.



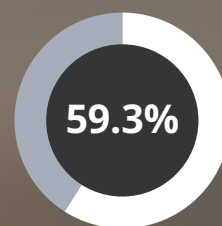
Support for open access



Support for open data



Enthusiasm for open peer review

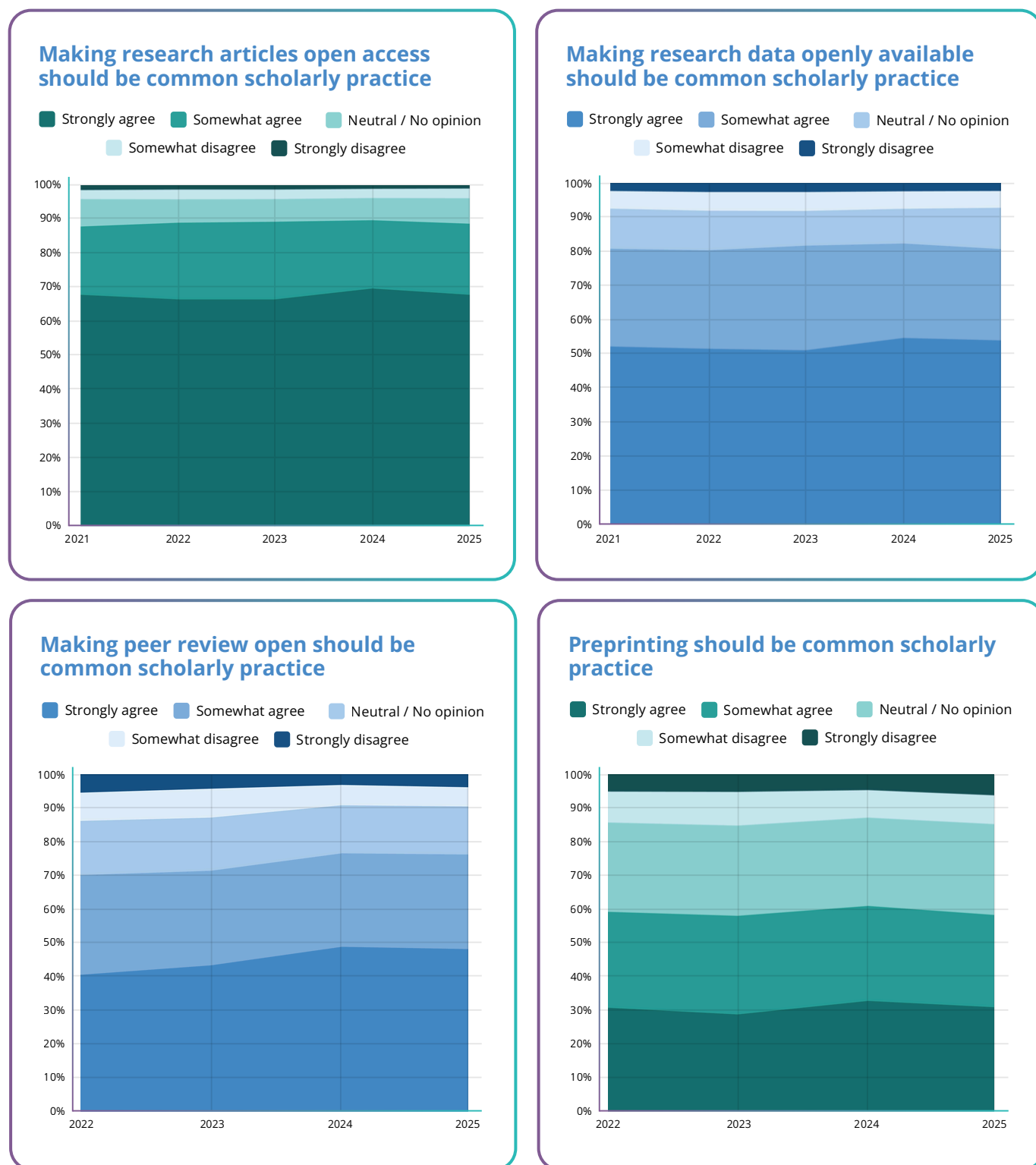


Support for preprinting

Support for open access (88.1%) and open data (80.9%) is robust, contrasting with a slightly lower enthusiasm for open peer review (75.7%) and notable hesitation around preprinting (59.3%), likely shaped by disciplinary norms and recent concerns about large-scale AI-generated content. These patterns underscore that practical and cultural factors, not mandates alone, drive adoption of specific open practices.



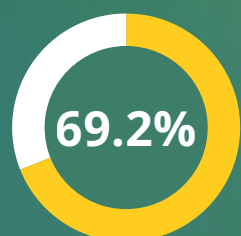
FIGURE 4 Researcher attitudes towards open science practices, 2021-2025



Percentage of respondent agreement/disagreement that each practice should be common in scholarly communication. Respondents were asked "To what extent do you agree/disagree with the following statements?" on a 5-point Likert scale ('Strongly agree', 'Somewhat agree', 'Neutral / No opinion', 'Somewhat disagree', 'Strongly disagree'). Questions on open peer review and preprinting were introduced in 2022. Annual sample sizes: 2021 (n=4,491), 2022 (n=6,104), 2023 (n=6,091), 2024 (n=4,454), 2025 (n=3,959).

These questions were introduced in different survey years. Trend data is shown from the first year each question was included in the survey.

The credit gap: Recognition versus reality



Say they receive
too little credit

Strong support for open practices does not eliminate barriers, particularly when recognition for data sharing remains limited. Fig. 5 highlights a persistent concern about credit attribution for data sharing. When we began to ask this question in 2020, 77.9% of respondents felt that researchers received too little credit for sharing data, compared to only 22.1% who believed credit was sufficient. While this gap has narrowed by 2025 (69.2% vs 30.7%), the overwhelming majority of researchers still perceive a fundamental misalignment between the effort required for data sharing and the professional recognition received.

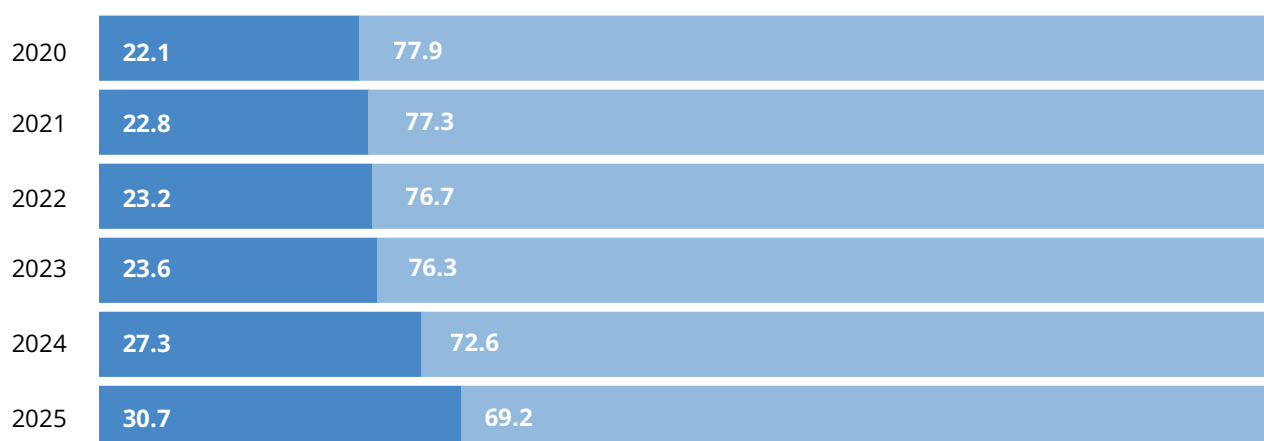
This "**credit gap**" represents one of the most significant barriers to widespread adoption of open science practices. The modest improvement over five years suggests that while awareness of the issue has grown, systemic changes to reward structures in academia have been slow to materialise.

FIGURE 5

Perceptions of credit for data sharing, 2020-2025

Do you think researchers currently get sufficient credit for sharing data?

■ Yes or too much credit (%) ■ Too little credit (%)



Researchers were asked "Do you think researchers currently get sufficient credit for sharing data?". The bar chart displays the percentage responding 'Yes' and 'They receive too much credit' versus 'No, they receive too little credit' for each survey year. Longitudinal data from six annual State of Open Data surveys (N=28,584). Yearly sample sizes: 2020 (n=4,945), 2021 (n=4,491), 2022 (n=6,104), 2023 (n=6,091), 2024 (n=3,721), 2025 (n=3,232).

Adoption of AI tools



As open data practices have evolved, so too have our survey questions. In response to the growing interest in AI, we introduced a question on the adoption of AI tools in 2024.

What was tentative then looks very different in 2025: barriers such as low awareness and lack of consideration have given way to active experimentation and regular use. Because this question is relatively new, the 2025 results should be viewed as movement from a 2024 baseline rather than a long-term trend.



2024 2025
22.1% - 31.9% 
Started using or use AI regularly

The most dramatic shift can be seen in data processing. 'Active use' (those who have started using or use AI regularly) climbed from 22.1% to 31.9%—an increase of nearly 10 percentage points in a single year—driven primarily by a surge in 'started using' responses (14.4% to 21.7%). At the same time, the pool of those 'not aware' or 'aware but not considering' contracted meaningfully (-9 percentage points combined), indicating that hesitation is rapidly diminishing.



2024 2025
16.1% - 25.1% 
Rise in the use of AI in metadata creation

The use of AI in metadata creation exhibits a comparable trajectory. 'Active use' rose from 16.1% to 25.1% (+9 percentage points), alongside a marked reduction in unfamiliarity of 15.1% to 9.7%. This suggests researchers are moving beyond passive awareness to practical engagement with AI for description and documentation tasks.



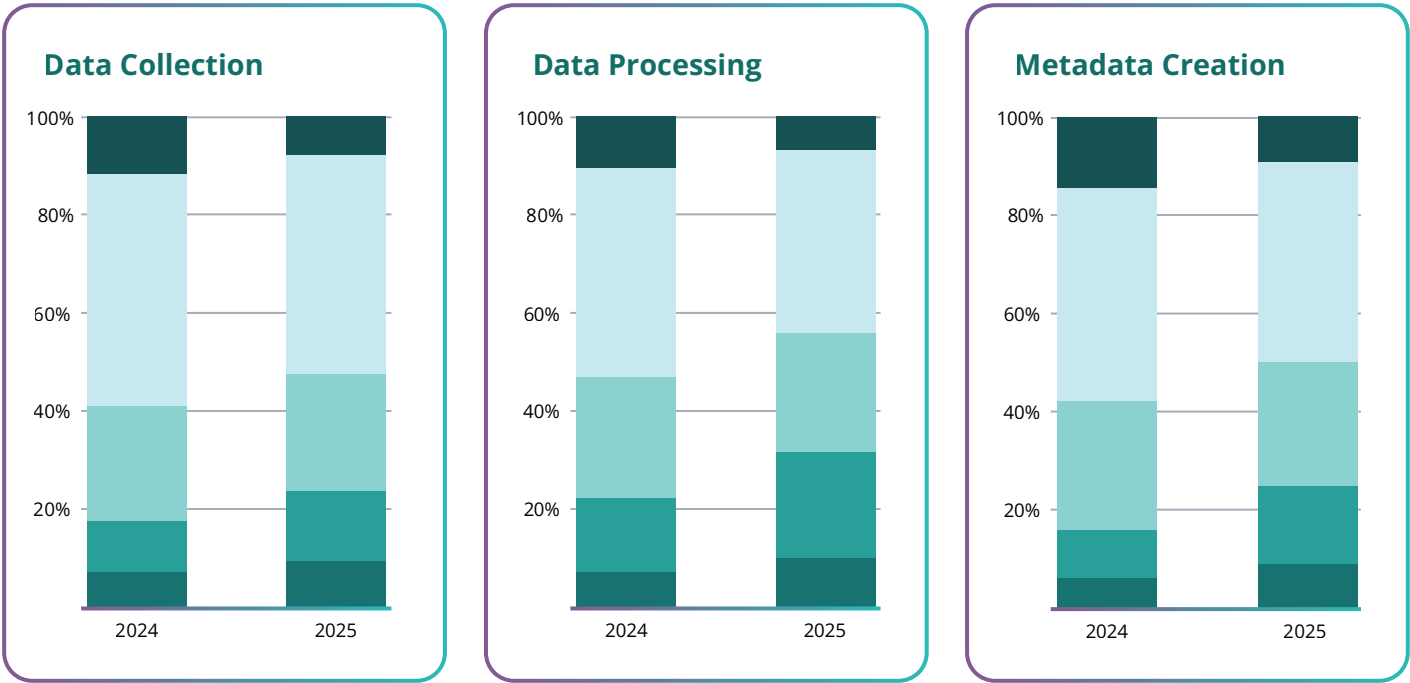
2024 2025
15.1% - 9.7% 
Reduction in unfamiliarity

Even the uptake of AI for data collection, which is often slower to change, shows clear momentum. Active use increased from 17.2% to 23.5%, with the responses ‘not aware / don’t know’ dropping into single digits (12.0% to 8.2%).

Between 2024 and 2025, AI tools shifted from peripheral curiosity to integral parts of research workflows, with the strongest gains in data processing, notable advances in metadata, and steady progress in collection. As adoption grows, ensuring transparency, reproducibility, and responsible use is critical—trends that may also support more FAIR-aligned practices, especially in metadata standardisation and repository integration.

FIGURE 6
Use of AI tools for data collection, 2024–2025

■ Yes, using it regularly ■ Yes, I've started using it ■ No, but I've considered it
■ I'm aware of these tools but haven't considered it ■ Not aware / don't know



Researchers were asked: "Are you using ChatGPT or similar AI tools for data collection, processing or metadata creation?" Longitudinal data from two annual State of Open Data surveys (N=8,386). Yearly sample sizes: 2024 (n=4,454), 2025 (n=3,932).



Summary of cross-cutting themes and implications based on survey data

The data reveals three critical insights about the current state of open science:

1 Knowledge without support

Awareness of open science principles has grown, but lack of credit and resources still hinder implementation. Mandates alone are not enough—practical support is essential.

2 Regional divergence

Support for data mandates varies widely, showing that one-size-fits-all strategies are ineffective; regional approaches that account for local research cultures and infrastructures could work better.

3 Disciplinary disparities

STEM fields lead in FAIR adoption while the Humanities and Social Sciences lag, underscoring the need for discipline-specific support and training.

Overall, progress is clear, but structural and cultural barriers remain. The next decade must focus on incentives, recognition, and resources to turn awareness into action—topics which are explored by our experts in the next section.

Explore the interactive survey dashboard:
<https://stateofopendata.com>

Experts' voices on open data

While the first section explores global trends and longitudinal data, the next brings together perspectives from researchers, librarians, and data experts to show how policies translate into practice. Their insights reveal cultural, infrastructural, and disciplinary nuances behind the numbers, and what the next decade of open data may require. The full list of experts can be found in the [Acknowledgements](#).

Mandates and policies can help drive change but tailored support is needed

In line with the survey responses showing waning support for national mandates in some regions (Fig. 3), our experts agreed that mandates cannot be the sole solution to encourage open data sharing, as there are implementation challenges.

Over the past decade, our experts noted that both funder mandates and grassroots advocacy have driven open data adoption, making it a standard practice for many researchers and institutions. Funder mandates and expectations have been key to creating this growth for those researchers and institutions who were not early adopters.

Infrastructure has evolved alongside policy. In China, Dawei Zhang notes rapid progress in data-sharing platforms and standards: 20 national scientific data centres (including one on the topic of materials corrosion where he works) and later a national materials big data centre have been established to support multiple industries, laying the foundation for data standards, AI-ready data, and more intelligent R&D. This diffusion of standards helps explain why familiarity with the [FAIR principles](#) has increased across disciplines (Fig. 2).

Dawei Zhang

Deputy Director of the National Materials Corrosion and Protection Data Centre, Professor at the School of Advanced Materials Innovation, University of Science and Technology Beijing, China



In Africa, local initiatives are driving practical solutions for sustainable data management.

Joy Owango: *"Data management and infrastructure in research has always been expensive [...] so, the rise of open infrastructure has been key. Let's not reinvent the wheel. The [Africa PID Alliance](#) uses existing open infrastructure that is tried and tested—we have incrementally innovated it so that it suits our community."*

While many regions struggle with translating policy into reusable datasets, the Global South is taking a proactive approach by developing policies tailored to indigenous knowledge and practices. Open science can be seen as an opportunity to protect and preserve these traditions.

Joy Owango

Founding Director of Training Centre in Communication (TCC Africa), Kenya



However, alongside the need for new infrastructure, a new challenge has emerged. Despite support for open science practices (Fig. 4), some platforms host poorly structured datasets where standards are inconsistently defined.

Melissa Haendel: *“When National Institutes of Health sharing policies came into place, what we started to see was data dumping grounds. People are sharing their data, but they're not necessarily making it reusable.”*

Melissa Haendel

PhD, FACMI, Sarah Graham Kenan Distinguished Professor – University of North Carolina at Chapel Hill



Turning policy into genuinely reusable datasets remains challenging, making quality, documentation, and effective infrastructure critical (Fig. 4). Regional differences in mandate support (Fig. 3) show that locally governed infrastructures—not one-size-fits-all solutions—are essential. And while open data is now part of academic workflows, these challenges will persist unless addressed.



Brian Nosek: *“The support that's really needed is: how does a researcher make it easy for another person to know their data as well as they do? [...] There is enormous capability, for example, in the academic librarian community that knows how to do these sorts of things well.”*

Brian Nosek

Co-founder and Executive Director of the Center for Open Science, Psychology Professor at the University of Virginia, USA

Open data has long-term impact that is difficult to measure by institutions

Institutions need to demonstrate the impact and return on investment of open data to justify infrastructure costs, meet funder expectations, and align with performance metrics such as rankings and visibility. Evidence of value can also drive cultural change, helping embed open science into research assessment and reward systems while highlighting the cost of inaction, such as duplicated research and lost opportunities.

Hilary Hanahoe: *“An institution that wants to promote open science or open data is expected to do the data management, wrangling, and cleaning themselves. The productivity of a researcher is reduced significantly. There is definitely a return on investment, but it's hard to demonstrate, and people like to put a dollar or a euro on it.”*

Hilary Hanahoe

Secretary General of the Research Data Alliance (RDA)



Open data delivers benefits over decades, making its impact hard to capture in the short term, especially for institutions focused on immediate metrics. Our experts emphasised that its true value emerges over time, which poses a challenge for current assessment methods.

Brian Nosek: *“Fundamental research is on a very long timescale. How did we get to GPS today? We couldn't have anticipated it 60 years ago. [...] I put open science into that same frame. This data is shared. Someone might re-analyze it tomorrow, and we get the value of that reanalysis in a week, in a month, in a year, whatever. But some of it may not be revisited for a decade.”*

Su Nee Goh highlighted the importance of recognising diverse types of impact, giving an example of the [PATHOS](#) project, which measures academic, societal, and economic impact, rather than relying solely on citation metrics. Measuring these broader outcomes is key to demonstrating the value of open data and communicating its long-term potential.

Frameworks and self-assessment tools can play a critical role in helping institutions strengthen their open science policies, build a supportive ecosystem, evaluate their progress, and align with global best practices.

Su Nee Goh: *“I strongly encourage funders and institutions to look at frameworks and tools. For example, the Digital Curation Centre's [RISE framework](#); [Research Infrastructure Self-Evaluation](#). There is also the [Research Data Management Framework for Institutions](#) created by Australian Research Data Commons, the [OpenAIRE Open Science Policy Checklist for Research Performing Organisations](#), their research funding organizer tools, and the [UNESCO Checklist](#) for universities on implementing the UNESCO recommendation on open science and guidance for funders.”*



Su Nee Goh

Deputy Director and Lead of the Open Science and Research Services Team at the Nanyang Technological University Library, Singapore

The understanding of the broad and long-term impact of open data is beneficial to institutions, funders and society, but researchers also need their time and effort reflected in research assessment.

That credit gap again: a lack of recognition for researchers

Responses to the survey over the past five years were clear: the vast majority of researchers continue to perceive a fundamental disconnect between the effort required to share data and the professional recognition they receive (Fig. 5). On this, our expert contributors are in agreement with the survey respondents.

Researchers need their work in creating FAIR open data to become integrated into research assessment systems so it is recognised. Achieving this requires senior leaders to understand that while open data sharing demands resources, it delivers significant long-term benefits.

Joy Owango: *“Researchers get massive recognition and financial incentives for publishing in top tier journals. University leadership needs to do the same for open science, because this is also going to help early career researchers. It has to start from the government because I can assure you an institution will be hard pressed to start giving recognition.”*

Authorship, citations, and impact factors remain the main ways to recognise contributions to an area of study, often overlooking the contribution of datasets. Acknowledgement of this type of contribution can help to build community and trust within a group of researchers.

Lucy Lu Wang: *“A lot of dataset work is treated as less rigorous and less of a contribution than other aspects of research, like developing new methodology, or a new discovery.”*

Lucy Lu Wang

NLP and Health Informatics researcher,
Assistant Professor at the University of
Washington Information School, USA



The concept of small or nano-attributions is especially interesting in medical research for rare diseases, where the instances are geographically widespread, and data ontologies are built from a wide range of practitioners and researchers working with patients.

Melissa Haendel: *“The Mondo Disease Ontology collates a lot of different resources, especially for genetic and rare diseases. There are hundreds of people who have contributed, and we track every single contribution on our GitHub repository. By giving people the opportunity to be authors on a paper we can really change the incentive structure.”*

Institutions can help normalise and celebrate open data by recognising researchers' efforts and sharing best practices. Initiatives like the Singapore Open Research Awards show how formal recognition can drive cultural change.

Su Nee Goh: *“Winners get to present their open science practices at the Singapore Open Research Conference. We had a digital humanities award recipient who was able to document how he collected his data. At every point where the data can be found, he showed how to reuse that data, what were the steps he took, what software he used to arrive at the derivative of that original data.*

We encourage our researchers to share the kind of impact they have achieved with open science. Some of them show their GitHub or their data repository numbers, some include impact narratives. They could use a combination of both to articulate the value-add they have brought to the discipline.”

Closing this credit gap is the first step towards a research culture that fosters a more connected ecosystem, where Findable, Accessible, Interoperable, and Reusable data, supported by intelligent tools, can amplify the impact of every researcher's work.



Interoperability and AI integration will be the next steps

Another challenge lies in making open data not just available, but interoperable, so that datasets can work together seamlessly. As we saw in Fig. 6, active use of [AI tools](#) increased significantly compared to 2024. AI could facilitate reformatting datasets, but this depends on clear, practical standards that promote best practice without discouraging data sharing. Speed is also a consideration—researchers want access to data as soon as they read a paper, not months later.



Lucy Lu Wang: *"In a lot of the machine learning conferences, there are pretty strong norms around releasing code and datasets prior to review, at the same time as readers have access to the paper. That's been a nice shift as it gets at the difference between just releasing the data, and actually making it useable."*

AI could also play a role in finding and fixing errors that prevent interoperability from working as smoothly as possible.

Dawei Zhang: *"In the future, there could be tools or models that could automatically identify data discrepancies, correct them and improve them with minimal human intervention. I think that will really help with interoperability."*

Harnessing AI and interoperability will make data more usable and efficient, but sustaining progress will require addressing deeper structural and cultural challenges.

What do the next 10 years hold for open data?

Our survey captures past experiences and the current state of open data sharing but what strategies can drive meaningful progress in the decade ahead? Our experts spoke about the importance of data protection, research security, and training.

One focus area will be to use data protection strategies to resolve tensions with openness, especially when collaborating with other regions.



Joy Owango: *"We have a few countries [in Africa] that have data protection strategies. We need to look at data within institutions and come up with policies on how you can share and manage the kind of tools you can use, so that if you're having a research collaboration, it's very clear that data stays here."*

Look at Timbuktu, in Mali. When that area blew up, people had to fight to protect those texts, because they go back a thousand years. We need to protect our indigenous knowledge. And how do we do that? You digitize it. Then you assign the identifiers that are resolvable, and also look at various ways in making that information FAIR."

Data privacy and protection are also vital when working with some of the most sensitive data about an individual, the state of their health.

Melissa Haendel: *"We just don't even have answers to the most basic questions about rare disease patients. They're just completely off the radar for the way in which we capture information about them in the context of clinical care. How do we connect the dots between basic research data, lived experience, and individual personalized data that you would never see in a clinical context?"*

Research security is an increasingly critical issue for institutions, requiring a careful balance with openness. Data can be provided more promptly where there is a data management plan in place to make it available, considering ethical and security concerns ahead of time.

Su Nee Goh: *"It's not just about FAIR now, the university is also looking at how to make it FAIRER, where E stands for ethical and R stands for responsible. We are looking at protecting research because of the more stringent cybersecurity requirements at the national and university level. [...] In 2024, the G7 developed and endorsed a set of best practices for secure and open research. [...] Research security is a constantly evolving area, and the open science practices will need to adapt to the new risk pertaining to research security in an appropriate way."*

FAIR → FAIRER

Expertise can be developed through training and resources that already exist to guide the scholarly community, such as the [Transparency and Openness Promotion Guidelines](#) mentioned by Su Nee Goh. The next step is to make sure that their use is widely adopted.

Building on earlier observations, upskilling in AI for both researchers and librarians is also an opportunity that should be considered.

Hilary Hanahoe: *"I think in the next 10 years we need investment in skills. We have a real opportunity to show the true value of open data with AI. High quality, well managed data is really crucial for accurate, reproducible, and robust AI driven insights."*

Brian Nosek: *"Open data and AI can progress hand in hand. There's a reciprocal benefit to AI training on a wide variety of data, and researchers using AI tools."*

While progress in managing open data is encouraging, the ultimate goal remains clear: research should advance knowledge and deliver societal benefit.

Hilary Hanahoe: *"I represent a community of people who believe that finding these solutions will actually improve the world, and therefore they can give something back. The goal of public good and benefits for society is so rewarding. I would advise people to invest in understanding the landscape and identifying means of synchronizing, cooperating and collaborating with others."*

The message from our experts is clear: advancing open data is a shared responsibility, demanding cooperation across stakeholders, regions, disciplines, and institutions.

Recommendations

Based on insights from the survey and from expert interviews, three key actions are highlighted below to accelerate progress in open data.

1

Reform research assessment so that data sharing is given the credit it is due

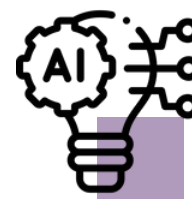
To enable this, datasets need to be citable, cited, machine-readable, discoverable, and measurable. Funders and academic institutions need to align policies that reward researchers and promote culture change through incentivising excellence in data sharing—in terms of quality as well as quantity.



2

Invest in practical, AI-enabled solutions that make sharing easier, faster, and more interoperable

Shift from policy-heavy approaches to concrete solutions: strengthen journal-repository integration, adopt common metadata and identifier standards, and use responsible AI to support metadata creation, quality checks, and format normalisation. Interoperable, human- and machine-actionable repositories, combined with clear guardrails for AI use, will raise data quality, reduce researcher burden, and accelerate reuse across disciplines.



3

Coordinate regionally and by discipline to build fit-for-purpose repositories and support systems

Develop shared infrastructures, standards, and training that reflect local contexts, disciplinary norms, and data sovereignty needs. Regional networks and domain-specific stewardship models help smaller institutions participate on equal terms, strengthen community governance, and ensure that open data practices are inclusive, sustainable, and aligned with cultural, ethical, and regulatory expectations.



Conclusion

Ten years of the State of Open Data survey reveal a clear trajectory: open data has moved from aspiration to expectation, yet the journey from policy to practice remains incomplete. Awareness of FAIR principles and enthusiasm for openness have grown substantially, but structural barriers, particularly around recognition, resourcing, and technical support, continue to hinder progress.

These findings underscore that success depends on making openness practical and rewarding. Integrated workflows, machine-actionable links, and structured data statements are proven enablers, while regional networks and locally governed infrastructures show how global principles can translate into real-world impact.

Emerging technologies such as AI offer powerful opportunities to improve quality and interoperability, provided they are deployed responsibly and transparently.

Looking ahead, the challenge is not convincing researchers of the value of open data—it is ensuring that sharing and reuse are supported, incentivised, and embedded into everyday research culture.

The next decade must focus on operationalising these principles so that openness is not an extra burden but the default, efficient, and most beneficial path for research and society.



Acknowledgements

Thanks are extended to every researcher who responded to this survey over the last 10 years, as well as contributors to all of the State of Open Data reports.

Interviewees

We would like to thank the following experts for generously sharing their time and insights for this report.



Brian Nosek

Co-founder and Executive Director of the Center for Open Science, Psychology Professor at the University of Virginia, USA



Dawei Zhang

Deputy Director of the National Materials Corrosion and Protection Data Centre, Professor at the School of Advanced Materials Innovation, University of Science and Technology Beijing, China



Hilary Hanahoe

Secretary General of the Research Data Alliance (RDA)



Joy Owango

Founding Director of Training Centre in Communication (TCC Africa), Kenya



Lucy Lu Wang

NLP and Health Informatics researcher, Assistant Professor at the University of Washington Information School, USA



Melissa Haendel

PhD, FACMI, Sarah Graham Kenan Distinguished Professor - University of North Carolina at Chapel Hill



Su Nee Goh*

Deputy Director and Lead of the Open Science and Research Services Team at the Nanyang Technological University Library, Singapore

**Views from Su Nee Goh are personal and do not reflect the position of Nanyang Technological University Library.*

Authors

- **Mark Hahnel**
VP of Open Research
Digital Science
- **Megan Toogood**
Senior Associate
Maverick Publishing Consultants
- **Paul Jarvis**
Senior Research Analyst
Springer Nature
- **Christina Emery**
Head of Thought Leadership Programmes
Springer Nature

Editor

- **Melissa Lapine**
Product Marketing Manager
Digital Science

Reviewers

- **Graham Smith**
Springer Nature
- **Hannah Tippet Simpson**
Springer Nature
- **Ed Gerstner**
Springer Nature
- **David Ellis**
Digital Science

APPENDIX 1

About the State of Open Data and 2025 survey

Since its launch 10 years ago, the State of Open Data survey has gathered insights from over 43,000 researchers across 212 countries and territories, making it the longest-running longitudinal study on this topic.

As open data has shifted from theory to practice, our approach to gathering, analysing and presenting findings has evolved alongside it. In 2023, we introduced a supplementary report, [From Theory to Practice](#), which expanded on the quantitative results of our years of surveys through a series of real-life perspectives and viewpoints offered by diverse members of the research community. And in 2024, we developed a new type of data analysis to help us better understand the gap [between policy and practice](#) in data sharing. Our focus has both expanded and sharpened: mapping global patterns while incorporating commentary from regional experts in order to better understand variation and nuance across contexts.

The survey reports have been collectively viewed 354,000 times, with 44,000 downloads from over 170 countries, and cited in more than 100 scholarly works. They have been referenced in influential policy documents from organisations including the World Health Organization, the European Union, the Organisation for Economic Co-operation and Development, and the Australian Academy of Science. The series has also received notable media coverage in outlets such as *Science*, *Times Higher Education*, and *Inside Higher Ed*. Beyond readership and citations, the State of Open Data Report has fostered direct engagement with the research data community—through keynote speeches, panel sessions and discussion forums. These activities underscore the report's role as both a trusted resource and a catalyst for global dialogue on open data.



2025 survey methodology and demographics

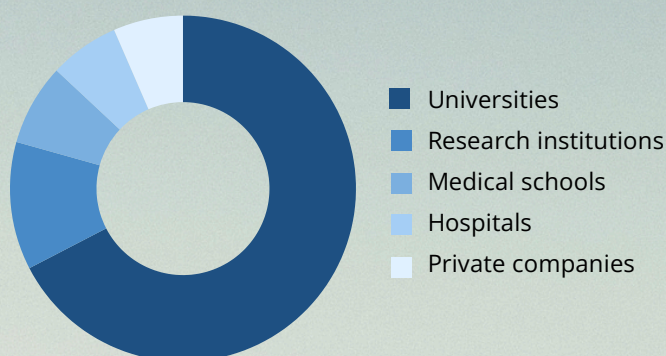
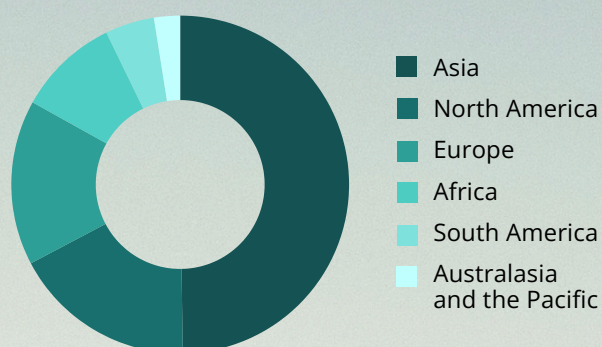
The 2025 survey ran from 28 April to 15 July 2025 and received 4,774 qualified respondents from 151 countries. It was offered in five languages (English, Simplified Chinese, Japanese, French, and German), and included both quantitative and qualitative questions exploring key themes discussed in this

and previous reports. Over the past decade, the survey has maintained a consistent core question set to support longitudinal analysis, while evolving to include topical additions such as AI.



2025 respondent breakdown

Recruitment channels included email, social media, and website intercepts, with fraud prevention measures in place. Respondents represented diverse regions (Asia 49.7%, North America 17.5%, Europe 15.8%, Africa 9.7%, South America 4.7%, Australasia and the Pacific 2.5%) and affiliations (62.0% universities, 10.9% research institutions, 6.5% medical schools, 6.1% hospitals, 5.9% private companies), government / local government 3.4%, other 5.1%. Early career researchers accounted for 44.0%, continuing a rise seen in recent years.



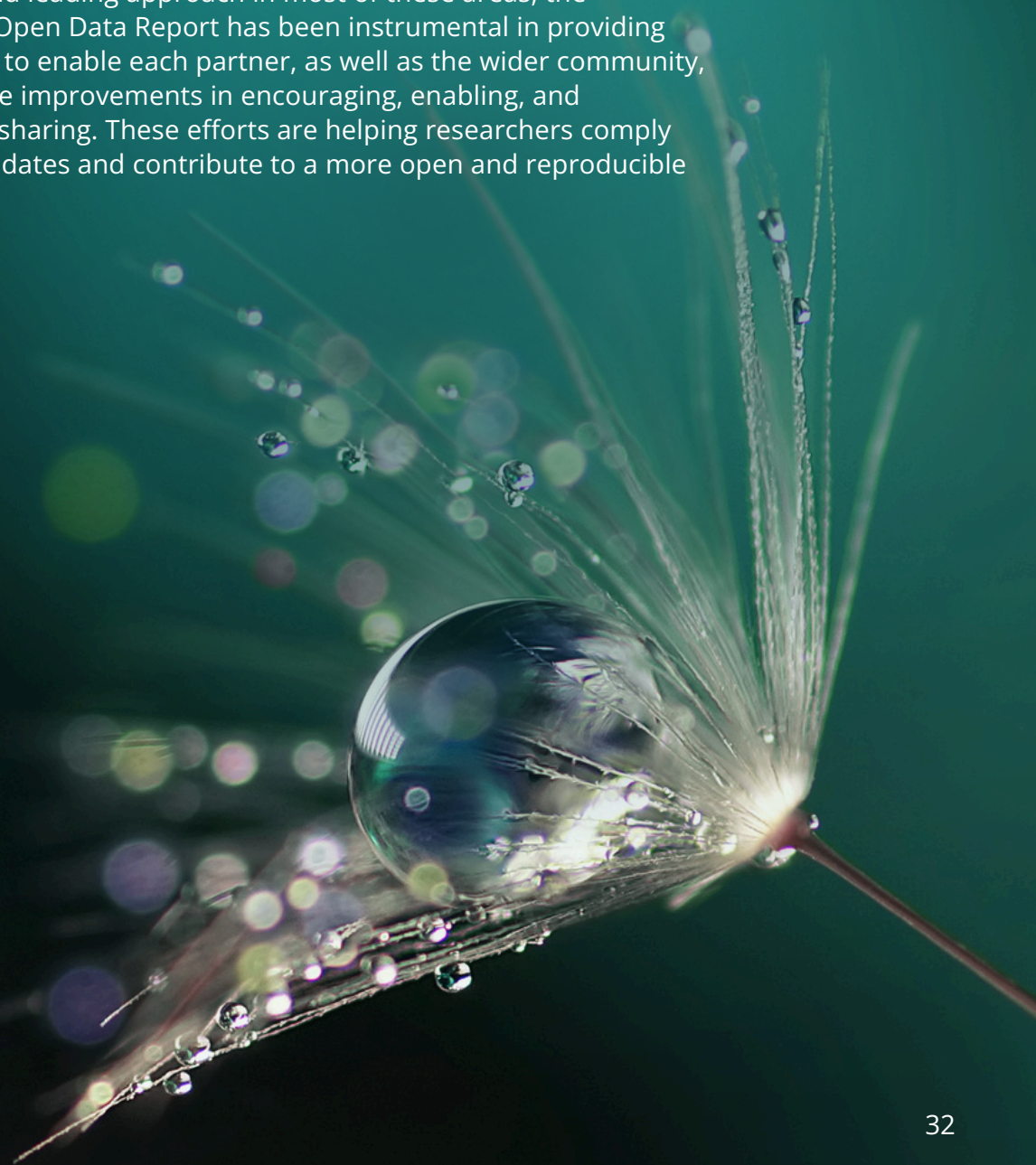
Access to survey data and dashboard

- Survey questions and data from 2025 are available on Figshare: <https://doi.org/10.6084/m9.figshare.30763190>
- Explore trends over time via the interactive dashboard: <https://stateofopendata.com>

APPENDIX 2

About open data at Digital Science, Springer Nature and Figshare

Partners Digital Science, Springer Nature and Figshare have been working together for some time to explore, address and implement changes that have a direct impact on the adoption of and awareness of open research practice. Taking a pioneering and leading approach in most of these areas, the longstanding State of Open Data Report has been instrumental in providing shared understanding to enable each partner, as well as the wider community, to better make tangible improvements in encouraging, enabling, and supporting open data sharing. These efforts are helping researchers comply with data-sharing mandates and contribute to a more open and reproducible research ecosystem.



DIGITAL science

In 2024, Digital Science deepened its leadership in making open research, and especially open data, the default. By pledging support for the [Barcelona Declaration on Open Research Information](#), the company affirmed that open research information underpins integrity, accessibility, and equity.

It also introduced four [Digital Science Open Principles](#) to guide how it advocates for, and encourages, data sharing across the research ecosystem:

1. Community ownership
2. Participating in open infrastructure
3. Stakeholders' primacy
4. Establishing trust

To enable adoption at scale, Digital Science partnered with Wellcome, the Kahle Austin Foundation and others to launch the IOI Fund for Network Adoption, directing support to shared, open data infrastructure. With a specific emphasis on Africa and Latin America, this initiative helps democratise the tools and networks that make sustained data sharing and reuse possible in under-resourced research communities.

Digital Science turned commitment into measurable progress. With the UK Reproducibility Network (UKRN), it powered the Open Research Indicator Pilots from 2023 to 2024 by supplying Dimensions and Figshare data and analytics across 15 institutions and five solution providers, testing practical indicators for data availability, FAIR data, preregistration, and CRediT across more than 27,000 research outputs. In 2025 it broadened community access by relaunching the Scientometric Researcher Access to Data (SRAD) program, giving researchers no cost access to Altmetric and Dimensions for noncommercial projects and accelerating the creation of open, transparent indicators of research and data sharing.

SPRINGER NATURE

Data Availability Statements

In early 2023, Springer Nature adopted a single research data policy across all journals. This requires all original research articles to include a [Data Availability Statement \(DAS\)](#) and strongly encourages the sharing of underlying data, establishing clearer and more consistent expectations for authors. The rollout was a major step in aligning our open science ambitions with those of funders, governments and institutions, but it also highlighted a key challenge: DAS remaining incomplete or unclear. Springer Nature recognised that whilst policies can set expectations, they do not always make it easier for researchers to meet them.

In order to support its authors, Springer Nature continues to explore the development of responsible AI-powered tools, in line with its community, that provide rapid detailed guidance and checks on research data sharing, repositories and linking, pairing technology with policy and editorial expertise.

Enabling data sharing in submission workflows

Even with clear policies in place, many researchers still face practical barriers to sharing their data. To help address this, Springer Nature integrated the Figshare repository directly into the submission systems of several journals. Data remain private until publication, are accessible to editors and reviewers throughout peer review, and are published with a digital object identifier (DOI) to ensure long-term preservation and credit.

The integration has shown strong uptake. The fact that ~10% of authors across the Nature journals have chosen to use this entirely optional solution is a huge success in behaviour change. This indicates that reducing the effort by making data deposition a simple, built-in step of submission encourages broader participation and significantly increases the amount of data being shared and made available for reuse by the wider community. These outcomes demonstrate the impact publishers can have in supporting researchers to adopt open research practices by providing the right infrastructure, helping to normalise data sharing as a standard part of the research process.

Engagement with the wider research community

Springer Nature continues to engage with a wide range of initiatives across the open research landscape, including the [Open Science Monitoring Initiative](#), the [UK Reproducibility Network](#), the [Committee on Data of the International Science Council](#), and [UNESCO](#). These partnerships enable us to align with global priorities, contribute to shaping best practices, and work collectively to address barriers to data sharing and reuse.

A core part of the approach is putting researchers at the centre of these conversations. Their perspectives are essential to ensuring that open science policies, infrastructure, and tools are relevant, practical and effective.

A recent example of this approach was a joint workshop with EMBO Press, which brought together internal teams from Springer Nature and EMBO Fellows to explore in part how AI-powered tools can support data sharing and reproducibility. Researchers provided actionable feedback on existing tools and pitched new ideas for future development ranging from AI-enhanced peer review to personalised research discovery tools. These types of collaboration not only help inform product roadmaps but also challenge our assumptions and broaden the conversation about how meaningful support can be provided for open science.

Springer Nature has leveraged this knowledge as the starting point for discussions with universities and funders. Insights have been shared about the attitudes of researchers globally and regionally to sharing their data—particularly when it comes to the support they receive to comply with open data mandates (or absence thereof), and the need to include greater recognition for open science practices in research assessment.





Figshare, an open data repository offered by Digital Science, has evolved from a single public platform to institutionally branded repositories that connect researchers with research data managers for curation and adherence to FAIR metadata standards. Institutional repositories for data, code, and nontraditional research outputs have been available since 2015; from 2020 many deployments expanded to include journal publications and theses. Today, Figshare supports more than 130 academic and government research institutions.

Through the NIH Office of Data Science Strategy's Generalist Repository Ecosystem Initiative (GREI), Figshare is helping set shared standards for interoperable metadata, discovery, metrics, and repository workflows that put the NIH Data Management and Sharing Policy into practice. In Year 3 (2024-2025), Figshare implemented COUNTER-aligned usage tracking via DataCite Usage Tracker, adopted FAIR Signposting Level 1 on item pages, and advanced an item-type metadata refactor guided by DataCite recommendations. GREI also released citable practitioner resources, including a guide for incorporating generalist repositories into DMS plans and a data-submission checklist, turning policy into day-to-day workflows for researchers and librarians and positioning generalist repositories as core infrastructure for transparency and reproducibility in the NIH ecosystem.

Across South Africa, the Figshare network demonstrates a nationally distributed model for generalist repository infrastructure. Twenty-two institutions, sixteen live and six forthcoming, contribute to a regional discovery hub that surfaces diverse research outputs. This coordinated adoption shows how shared platforms and common metadata practices convert open-research policy into web-visible, citable outputs across domains from cultural heritage to digital archives.



A collaboration between Digital Science,
Springer Nature, and Figshare
