SPRINGER NATURE

springernature.com

Research Data

THE FUTURE OF FAIR

Highlights and reflections from the Better Research Through Better Data roundtable

White paper

Contents

Authors

Varsha Khodiyar (Springer Nature), **Heidi Laine** (Finnish Committee for Research Data), **David O'Brien** (International Development Research Centre), **Raul Rodriguez-Esteban** (Roche Innovation Center Basel), **Yasemin Türkyilmaz-van der Velden** (Delft University of Technology), **Grace Baynes** (Springer Nature), **Matthew Brack** (Infectious Diseases Data Observatory), **Anne Cambon-Thomsen** (CNRS, Inserm and University of Toulouse, France), **David Carr** (Wellcome Trust), **Elisa Carrus** (London South Bank University), **Louise Chisholm** (UCL), **Claudia Civai** (London South Bank University), **Maria Cruz** (NWO), **Rebecca Grant** (Springer Nature), **Elizabeth Newbold** (Science and Technology Facilities Council (UKRI)), **Fátima L. S. Nunes** (University of São Paulo)

April 2021

This white paper has been made openly available in the figshare repository.

Access white paper:
https://go.sn.pub/the-future-of-fair

DOI: 10.6084/m9.figshare.14393552

Foreword

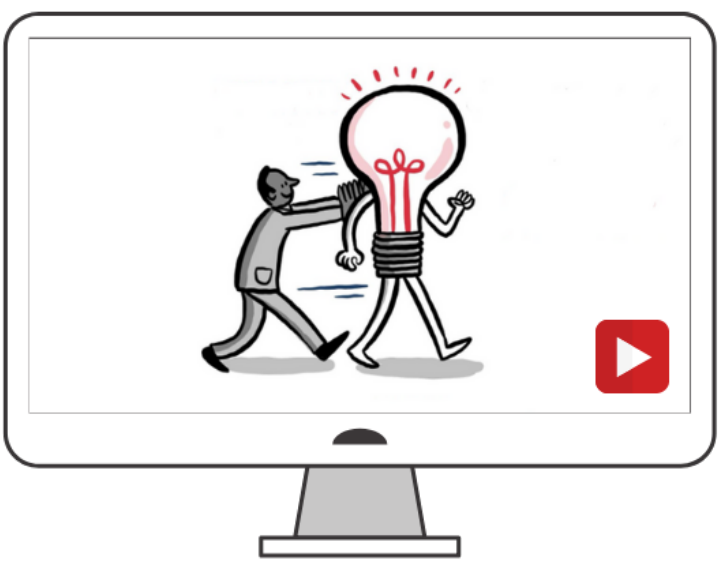
In 2021 we mark five years since publication of the FAIR data principles. The concept of FAIR data has been instrumental in bringing open science and research data to the attention of the research community, as well as to the broader group of stakeholders involved in facilitating, managing and disseminating research. In November 2020, we brought together an international cohort of research data professionals to celebrate the real-world impact of the FAIR data principles, and consider what will be next for research data and open science.

Varsha Khodiyar PhD,
Data Curation Manager,
Springer Nature

The *Better Research through Better Data: Shaping the Future of FAIR* event took place as a virtual roundtable, with an introductory keynote, smaller group discussions within five breakout sessions, followed by a panel discussion. The event was a truly global affair with attendees dialling in from North America, South America, Europe, Africa and Asia. The diversity of opinions and experiences generously shared by our attendees provided a great opportunity to consider fairness for research data in the widest possible sense.

The wide-ranging discussions which took place during the roundtable are summarised in this white paper, alongside further insights, opinions and commentaries contributed by attendees. Topics include: evaluating the impact of the FAIR data principles to date, capacity building to enable good research data management, and considerations on the fullest meaning of fair with regard to research data.

As we reach this anniversary of the FAIR data principles, it is a good time to pause and reflect. We look back to ascertain how far we have come, and look forward to where we still have work to do. I hope that the opinions and commentary in this white paper will facilitate further discussion and collaboration, as we collectively aim to move towards a fairer world in which research data play a proper part in the understanding and use of research for the benefit of all.



Highlights

The impact of the FAIR principles on a global pandemic

- The COVID-19 pandemic has strengthened the case for FAIR data sharing and significantly increased awareness and adoption of the FAIR principles at a policy level.
- Despite early signs of improvements in practices and benefits, there are still challenges. Awareness of FAIR among researchers, particularly in certain disciplines and regions, remains low. There are also implementation challenges, especially around interoperability.
- Evaluating the response to COVID-19 and impact of FAIR will be an important learning opportunity that could shape the future of how we share data and research outputs.

Fair management of genetic data and ethical considerations for research data sharing

- Research data sharing has ethical implications for study participants, as well as for researchers and reviewers.
- More than ever, the COVID-19 pandemic has highlighted the importance of the rights of the individuals behind the data, particularly as data management technology evolves.
- When it comes to genetic data management, the research community has an opportunity – and a responsibility – to act as a counterbalance to big tech, and create models for building genuine public trust.
- The evidence of benefit from sharing research data is growing, but for it to succeed, researchers, peer reviewers, and publishers have important roles to play throughout the research lifecycle.

Using Real World Data for research

- The use of Real World Data (RWD) in healthcare research is growing, as is the use of devices to gather and store huge amounts of healthcare data. Using RWD has brought benefits including a reduced trial burden on patients, but it also raises ethical and practical questions around bias, privacy and consent.
- The COVID-19 pandemic has put RWD analysis and sharing in the spotlight, highlighting the need for policy improvement and multinational collaboration. These learnings will help develop future strategies to enhance timeliness and communication, and ultimately improve patient outcomes.

Supporting roles for FAIR data

- Data professionals play an important role in training and supporting researchers to implement FAIR. The skills and roles required of these professionals should be defined by the barriers that researchers face.
- For FAIR data support to be effective, it is necessary not only to embed the FAIR principles in research education and practice, but also to reward and showcase those practicing FAIR to encourage others.
- It is important to focus on incremental steps researchers can apply in their practice to make their data as FAIR as possible.

Fair access to research data infrastructure

- The capacities of governments to build and maintain data infrastructure vary widely, meaning that access to infrastructure – and in turn, the benefits of open data – are unequally distributed across countries.
- Funding is one of the main barriers to equitable access, but sustainability is also key and this is often lacking in grant schemes.
- Addressing these challenges requires a cultural shift around the value of data sharing, ensuring best practices are embedded among early career researchers. The FAIR principles present a solution to help overcome regional and disciplinary differences.

About the Better Research Through Better Data roundtable committee

We are immensely grateful for the generous advice and guidance from the *Better Research through Better Data* programme committee. Committee members were instrumental in shaping the virtual event format, planning the breakout topics, and inviting an international cohort of research data experts and data management advocates.

Dr Varsha Khodiyar is Programme Chair of the *Better Research through Better Data* conference series. Varsha leads Springer Nature's curation team, contributes to the design, development and delivery of Springer Nature's research data training and is responsible for curating and maintaining the Scientific Data and Springer Nature recommended repository lists. Varsha is an Executive Advisor of FAIRsharing.org, a member of CODATA's International Data Policy committee and a co-author of the TRUST principles.

Varsha Khodiyar, Springer Nature

Heidi Laine is an expert in research integrity (RI) and responsible conduct of research. She is specialised in issues concerning responsible data management and RI in the open science era. She is the secretary of Finnish Committee for Research Data, and works as a Customer Solution Manager at Finnish IT Center for Science - CSC.

Heidi Laine, Finnish Committee for Research Data

David O'Brien is a Senior Program Specialist at the International Development Research Centre, Canada. He is a member of IDRC's Open Data Working Group that shaped the organisation's position on open data and the tools to support data management. Programmatically, he has developed research programmes promoting knowledge and innovation in the fields of health, climate change, economic development and science and technology policy. He is a social scientist and studied innovation and development at Wageningen University, University of Sussex, and the University of British Columbia.

David O'Brien, International Development Research Centre

Raul Rodriguez-Esteban is Principal Scientist at Roche Pharmaceuticals in Basel, Switzerland, where he works on Natural Language Processing, Machine Learning and Real World Data applied to pharmaceutical R&D. Previously, he worked as Senior Scientist in computational biology at Boehringer Ingelheim and Pfizer. He completed his PhD in machine learning applied to text mining at the laboratory of Andrey Rzhetsky at Columbia University. He has published over 40 publications, including 26 articles in international journals, and recently was a winner of the Bio-IT World Innovative Practices Award.

Raul Rodriguez-Esteban, Roche Innovation Center Basel

Dr Yasemin Türkyilmaz-van der Velden works as the Data Steward of the Mechanical, Maritime and Materials Engineering Faculty in TU Delft. She obtained her PhD degree in molecular genetics from Erasmus Medical Center Rotterdam. As a Data Steward, she supports, trains and inspires researchers to apply good practices in research data management, reproducible research, open science and research integrity. She served as the TU Delft Data Champion Community Manager and continues with her community engagement activities within the Open Science Community Delft. She is a co-chair of RDA Discipline-specific Guidance for Data Management Plans Working Group.

Yasemin Türkyilmaz-van der Velden, Delft University of Technology

Lynn Woolfrey is a data scientist with more than 20 years' experience curating and sharing microdata from African governments and research projects. She manages data operations at DataFirst, which is a unit based at the University of Cape Town that shares data and undertakes data-focused research. Lynn is responsible for Africa's only CoreTrustSeal certified open data repository. In 2012 she set up the only secure research data centre in Africa, to give researchers controlled access to sensitive or highly disaggregated government data.

Lynn Woolfrey, DataFirst, University of Cape Town

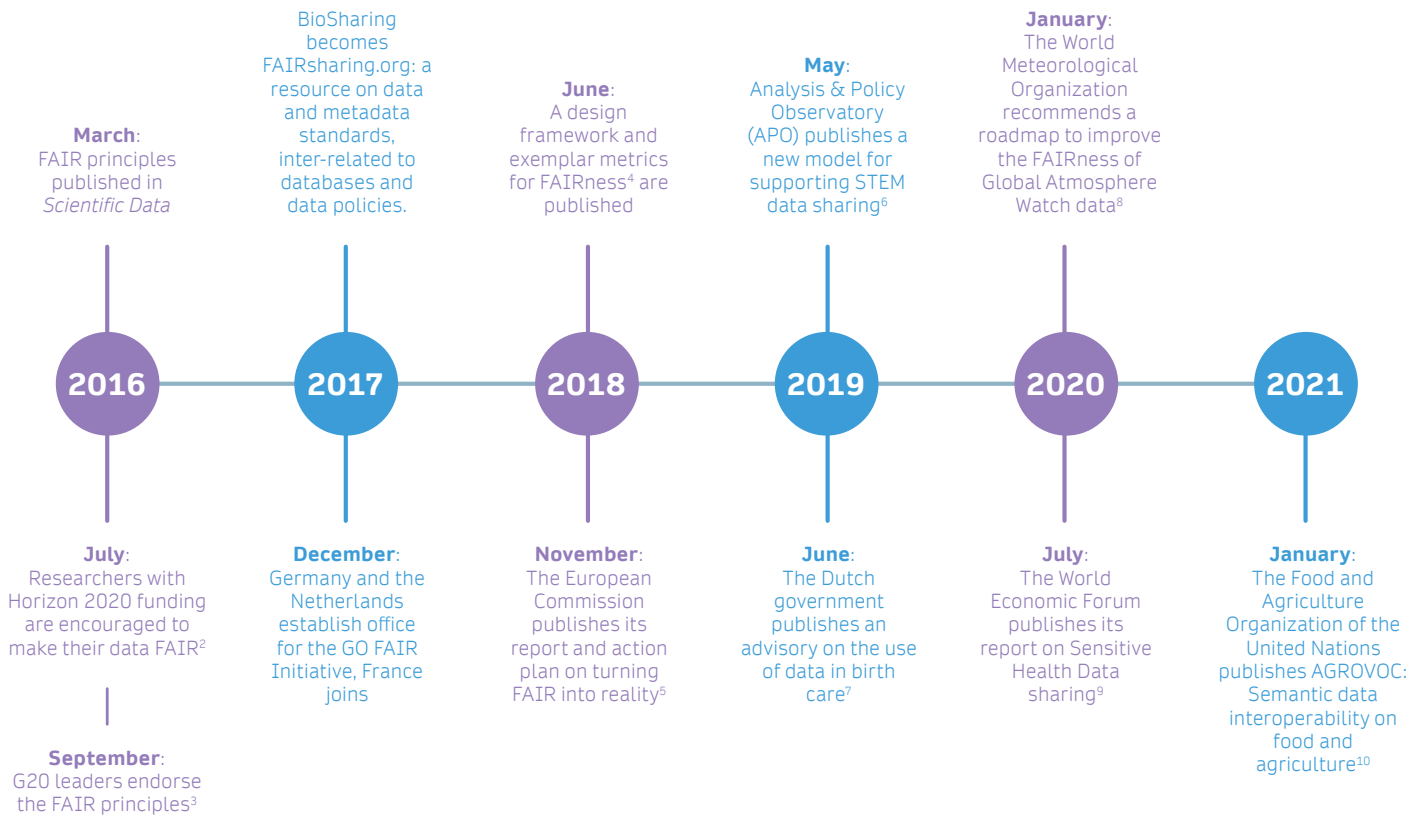
About the FAIR principles

2021 marks the fifth anniversary of the FAIR principles, which were published in *Scientific Data*[1] in 2016. The FAIR principles were developed to support the discovery and reuse of research data.

FINDABLE ACCESSIBLE INTEROPERABLE REUSABLE

FAIR timeline

Here are some of the key milestones achieved over the past five years:



1. Wilkinson et al., (2016)
2. Directorate General for Research and Innovation (European Commission) (2016)
3. European Commission (2016)
4. Wilkinson et al., (2018)
5. Directorate General for Research and Innovation (European Commission) (2018)
6. Cooper & Springer (2019)
7. De Rijksoverheid. Voor Nederland (2019)
8. World Meteorological Organization (2020)
9. World Economic Forum (2020)
10. FAO (2021)

scientific **data**

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

Article metrics as of March 2021:

Accesses	Citations	Altmetric	Tweeters	News	Policy
186K	2,322	1,808	1,348	129	35

2020 State of Open Data survey: are researchers familiar with the FAIR principles?

According to the 2020 State of Open Data survey[11] of nearly 5,000 researchers in over 190 countries:

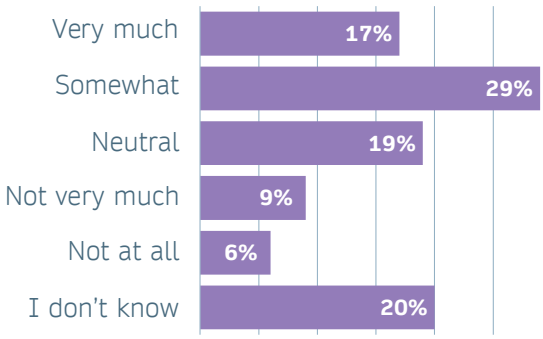
39.4% of survey respondents had never heard of the FAIR principles before taking the survey.

36.2% had heard of the FAIR principles but were not familiar with them.

24.4% were familiar with the FAIR principles.

In 2020, a total of **61%** of researchers had heard of the FAIR principles – up from **40%** in 2018.

How compliant are your data with FAIR?



Since 2017 there is a small positive trend in how well researchers view their data as complying with FAIR.

Read the 2020 State of Open Data report: https://www.springernature.com/gp/researchers/campaigns/state-of-open-data

11. Digital Science et al., (2020)

1. The impact of the FAIR principles on a global pandemic

Summary of discussion and opinion

David O'Brien, Matthew Brack, David Carr, Grace Baynes
Appreciation to Natalie Harrower, Kim Ingle, Tim Brophy and Juan Pane

On 26 December 2019 a patient experiencing respiratory difficulty was admitted to the Central Hospital of Wuhan, and others would soon follow. Within days, researchers affiliated with the hospital, the China Center for Disease Control and Fudan University concluded that this patient was suffering from a novel coronavirus. Prof. Yong-Zhen Zhang and colleagues recognised that the virus was spreading and without the genomic data, researchers could not develop diagnostic tests or treatment. On 10 January 2020, they posted the SARS-CoV2 genetic blueprint to GenBank with the notification that all were "free to download, share, use and analyze this data".

In historical perspective, this sharing of data was acted on with remarkable speed. Using the genomic data, the US Centers for Disease Control and Prevention developed a diagnostic test within ten days and by 24 February, MODERNA had manufactured its vaccine candidate[12] – one of 120 vaccine candidates that would be reported to the World Health Organization (WHO) six months after the genome was openly available. To incentivise complementary efforts, research funders and academic publishers came together to commit to support rapid sharing of research data and findings in late January.

The speed of these developments was what the WHO had hoped would become standard practice following their review of the scientific community's response to the Ebola outbreak. In their 2015 statement "Developing global norms for sharing data and results during public health emergencies", the WHO called for a break in practice: they called on researchers, funders and publishers to embrace rapid and open sharing of data and results[13]. The failure to do so during the Ebola outbreak thwarted public health efforts to contain the virus and to treat those affected[14].

The question of how to share data had garnered critical attention well before 2015 but in that year disparate efforts coalesced, and support grew for the FAIR principles. In a 2016 *Scientific Data* article, the authors of "The FAIR Guiding Principles for scientific data management and stewardship" argued that by making data Findable, Accessible, Interoperable, and Reusable, data and publications would increasingly become "accurately and appropriately found, re-used, and cited over time"[15]. As the movement for FAIR data grew, advocates pointed to the promise of FAIR data across the research spectrum, to accelerate discovery, facilitate collaboration, reduce duplication, and increase transparency and trust in science.

12. Wright (2021)
13. World Health Organization (WHO)
14. Littler et al., (2017); Modjarrad et al., (2016)
15. Wilkinson et al., (2016)

While the sharing of the genetic blueprint and use of that data was a significant early development, the breakout group participants reflected on whether the pandemic response was a tipping point for FAIR, and to identify implications for policy, practice and research.

The pandemic has made a case for data sharing and increased the need for FAIR

As the virus spread in early 2020, there was a significant and rapid response from many governments and their research funding agencies. One tracking effort identified that by October 2020, 71 funders in 28 countries had supported over 5,000 research projects totalling at least $USD 1.7 billion[16]. One novel addition to COVID-19 funding opportunities was the adoption of both FAIR and open data principles. While such requirements were increasingly common in Europe, they were new grant conditions for many funders and researchers.

Before the pandemic, some agencies had started to align their funding policies and workflows to support the WHO's new 'global norm'. In 2018, for example, a network of health research funders under the umbrella of the Global Research Collaboration for Infectious Disease Preparedness (GloPID-R) published "Principles of data sharing in public health emergencies"[17]. When funders associated with GLoPID-R launched their COVID-19 research programmes, many included requirements for data sharing and had thought through the implications, but for many funders and researchers such requirements and expectations were new.

In a related development, there was a coordinated effort with publishers to promote open access to COVID-19 research papers[18]. Over 50 publishers signed on. A *Nature* editorial announced their support on 4 February 2020: 'Calling all coronavirus researchers: keep sharing, stay open'[19].

When the roundtable convened, there were early signs that some improvements in anticipated practices and expected benefits were materialising, but that there was still a long way to go. In a comparative bibliometric study of COVID-19 and Middle East Respiratory Syndrome (MERS), the authors found 26% of COVID-19 papers were published with their data compared to 18% for MERS papers, suggesting a small but not insignificant shift in practice and/or requirements[20]. Efforts testing the power of open data platforms for discovery and collaboration grew out of existing research networks and new initiatives emerged[21]. Encouragingly, open data platforms were also utilised by health care practitioners and public agencies coordinating pandemic response. Clearly, efforts to promote open data were generating a demand but there were challenges along the way that pointed to the need for common standards.

During the panel discussion it was also noted that none of the FAIR principles necessitate data being 'open'[22] but that the best possible benefit comes when data are both FAIR and open[23]. This is particularly relevant in the context of COVID-19, where the health data of individual patients carry personal privacy concerns and emerging infections data in particular can be highly politicised, thus requiring data transfer agreements in place before sharing data. Consequently, many of the rapid data sharing models developed during public health emergencies have not been open and publicly available, but rather rely on transparent governance and appropriate data security mechanisms for equitable data access and reuse while still aligning with the FAIR principles.

One such example is the Infectious Diseases Data Observatory (IDDO). Cited by the WHO Ebola/Marburg Research and Development Roadmap as a model for collecting, standardising, and sharing clinical data under the authority of local leadership, the

None of the FAIR principles necessitate data being 'open' but the best possible benefit comes when data are both FAIR and open

16. Norton et al., (2020)
17. Global Research Collaboration for Infectious Disease Preparedness (GloPID-R)
18. See National Center for Biotechnology Information (NCBI)
19. See "Calling all coronavirus researchers" (2020)
20. Helliwell et al., (2020)
21. Tse et al., (2020)
22. Mons et al., (2017)
23. Directorate-General for Research and Innovation (European Commission) (2018)

IDDO emerging infections model for data sharing has been applied to the current pandemic response, hosting one of the largest international collections of clinical data[24] related to COVID-19, accessible to researchers according to clear data access criteria assessed by an independent and representative data access committee.

The panel agreed that the response to the pandemic has significantly increased awareness and adoption of the FAIR principles at a policy level. There was undoubtedly a push from the public health community and their insistence that pandemic research demanded a new approach. Yet, awareness of FAIR among individual researchers across disciplines globally remains low. The 2020 State of Open Data survey found that awareness of FAIR had increased to 24% of respondents (up from 15% in 2018) but that 39% had never heard of FAIR[25]. The survey was carried out from May-July in the midst of the first wave of the pandemic. There are questions whether the post-pandemic research landscape will sustain the focus on FAIR data. The panel discussion highlighted a number of implementation challenges that will need to be addressed.

The FAIR principles are simple, but implementation is not

Throughout the roundtable it was acknowledged that the promotion of FAIR data by funders, publishers and researchers had reached a new level and brought into focus a host of implementation challenges.

Efforts by funders to make it easier for grantees to make data open and FAIR[26], and publisher coordination to standardise data availability requirements, were needed and welcomed. A deepening of such efforts to provide guidance, training and support were called for. Guidance on repository selection and providing consistent metadata to make data findable, and guidance on licensing to make data accessible were some of the barriers that could be addressed with minimal effort[27]. A more intractable issue was how to address the interoperability challenge.

Among the interoperability issues discussed, the global nature of the pandemic and interdisciplinarity research were complicating factors. We've seen different countries using different measurement variables and data collection methods to track the pandemic, making it difficult to normalise data across regions and globally. This emphasises the importance of not just creating data standards within and across disciplines, but also across borders and in standardised formats. Repositories could have an important intermediary role to play in the development and dissemination of such standards within research communities, often bridging the data-sharing gap between the individual researcher collecting data, funding agencies, publishers and standards authorities.

The Research Data Alliance (RDA)'s COVID-19 working group Recommendations and Guidelines for Data Sharing[28] was highlighted as an important response as it provides granular advice to research communities on how to create FAIR data according to disciplinary best practices, and the importance of working across disciplines, borders and jurisdictions. The detailed nature of the recommendations highlights that while FAIR advances four simple principles, putting them into practice requires a dedicated effort. There will be an ongoing need to agree data standards within disciplines, across disciplines and for new research fields. The RDA report was commended for drawing attention to complementary aspects such as the importance of sharing of code and other analytical tools for data analysis, a reminder that FAIR was also to support machine-actionable research.

Participants felt the RDA guidelines, and related efforts like GoFAIR that support the adoption and promotion of FAIR data, have made an important contribution to

24. https://www.iddo.org/covid-19
25. Digital Science et al., (2020)
26. The Dutch agency ZonMw, for example, supported the creation of VODAN in a Box, a toolset to facilitate the capture of data related to virus outbreaks and the publication of metadata describing these datasets.
27. Mons et al., (2017). Also see Van Noorden (2021), as an illustration of researcher-driven calls for better access.
28. See Research Data Alliance (RDA) (2020).

identifying and contextualising COVID-19 research in a FAIR paradigm. As highlighted in parallel discussions at the roundtable, promoting broad engagement and coordination of effort will be needed to work through the mechanics of realising FAIR data.

The FAIR principles originated from, and found early support from, organisations interested in advancing e-science. While COVID-19 has brought the implementation challenges into focus, the end of the pandemic will not diminish the need for sustained collaboration across the full research landscape.

Global awareness of FAIR is a work in progress

A key theme that emerged from the discussions was the necessity to create a common global understanding and agreement around data sharing and application of the FAIR principles. While FAIR data principles have a longer historical rooting in Europe and the European Open Science Cloud has stimulated ongoing work, familiarity of, and experience in, applying FAIR practice outside of Europe is needed.

Participants from Latin America and South Africa, for example, reported that while data sharing and open data were high on the agenda in their work, the FAIR principles were not widely known. There were emergent efforts to support open data at research institutes and by some funders. The São Paulo Research Foundation (FAPESP) in Brazil had introduced open data expectations in their funding opportunities and catalysed a pioneering effort to make COVID-19 patient data openly available for research purposes[29]. In South Africa, it is becoming increasingly common for large research projects to have either local or international funders who require practices which align with FAIR research principles. From their perspective, awareness raising of what the FAIR principles are, as well as aligning them with work already happening around open data, is needed in their regions.

The pandemic could shape the future adoption of FAIR

Data holds the potential to inform decision-making and the pandemic has brought into relief the importance of open data and the concerted effort by many to make research outputs FAIR. The panel felt the breadth of application during the pandemic could illustrate the impact of FAIR data by providing case studies that FAIR accelerates discovery, promotes transparency, reduces duplication, and aids replication. There are FAIR converts as well as sceptics and the panel hoped to see research take advantage of empirical opportunities. For example, studies could investigate whether open data accelerated the development of diagnostic and treatment options, if commitment to early data publishing and transparency promoted good science, and whether researchers and practitioners used published data to inform their work.

Answers to such questions could send important signals throughout the research system. Early adopters may be convinced by their efforts or garner insights that might refine their approach. For those who experimented with the FAIR principles or stood on the sidelines, evidence from implementation could be impactful research in shaping the research system. What's clear is that a retrospective on the response to COVID-19 and impact of FAIR will be an important learning opportunity and could shape the future of how we share data and research outputs.

A retrospective on the response to COVID-19 and impact of FAIR will be an important learning opportunity and could shape the future of how we share data and research outputs

29. Mello et al., (2020)

2. Fair management of genetic data and ethical considerations for research data sharing

Summary of discussion

The ethical implications of research data sharing are many and varied. Much of the focus in this area has been on the rights of study participants, with questions about consent and privacy very much at the forefront.

For example, what obligations do researchers have towards patients when sharing or reusing clinical data beyond the primary research study? And do researchers have an ethical responsibility to use clinical data as widely as possible to ensure better health outcomes for all patients, or are the rights of the individual paramount?

However, there are also ethical implications for the academic world. Can data sharing disadvantage researchers who are less well-resourced or well-established, for example? We know from surveys of researchers that being 'scooped' – sharing data and not being credited in discoveries based on that data – is a major barrier for research data sharing.

In addition, when we know reviewers are already pushed for time, is it ethical to ask them to evaluate the data underlying claims made in an article, in addition to assessing the manuscript? And on the other hand, is it ethical for reviewers to evaluate a manuscript without first assessing the data set underlying the claims being made?

The rights of patients in clinical data sharing

It was felt by participants in the discussion that this was an area without one obvious solution. One of the biggest challenges raised was the fact that the consent required from patients and legalities of sharing the data varied widely depending on the country and jurisdiction where the data are gathered.

One participant gave the example of biobanks in Finland. Under current Finnish biobank law dating to 2012, participants have the right to know which studies their sample has been assigned to, and to know about research results. They also have the right to decide whether to be informed or not about research findings that may be relevant to their health.

But recent discussions on this topic have raised questions about whether informed consent should also include information about data security and threats such as hacking. For example, a major Finnish research initiative using genetic biobank data from half a million participants has come under criticism for using Google's cloud storage.

Overall, it was felt that the most important consideration was for the people whose data are being used to know how they are being used. Patients should have a means of knowing what collections their data are in and what they are being used for.

The most important consideration is for the people whose data are being used to know how they are being used

This raised further questions in the discussion around the practicalities of how this should be done. One participant asked whether it was a role publishers could take on. While another raised the concept of 'data trusts', where the data owners effectively own the data in the data repositories.

This idea was met enthusiastically by discussion participants, who felt it could be helpful in providing more transparency between the researcher and the subject. It was pointed out that it is often not possible at the point of consent to specify to the subject exactly how and with whom the data will be shared in the future. Participants asked whether a data trust could potentially help to keep the data owners fully aware of how their data was being used.

Another key challenge raised in the discussion was around acknowledging the rights of the collective, as well as the individual. For example, if the data set identifies a particular community, what collective rights does that community have over the data?

The 'fairness' of data sharing within the research community

When it came to discussing the ethics of data sharing within the research community, the most prominent concern was recognition. In other words, ensuring that researchers who share their data are adequately recognised for their contributions to any discoveries made from that data.

It was felt that ensuring those recognition structures are in place would reduce the fear that many researchers have around being 'scooped'. One of the discussion participants raised an example published in *Data Science Journal* of researchers taking a radical real-time approach to sharing their data, in order to overcome the fear of being scooped[30].

Cost and access to the right repositories was another issue raised around the fairness of data sharing. If researchers lack the resources to pay for their data to be stored in a repository, then they are reliant on repositories that are freely available to use. While there are a growing number of these, they are not always relevant or appropriate. Therefore, it was felt that imposing the same data sharing requirements on all researchers could discriminate against those with fewer resources.

The ethics of data peer review

Another ethical issue within the research community which discussion participants dwelt on with interest was peer review.

In general, participants felt that if data are shared, they also need to be checked. However, there was also acknowledgement that this would lead to an even greater burden on peer reviewers, who are often already pressed for time. Some participants were concerned that if data peer review was included in the peer review process, it would result in journals struggling to find any peer reviewers willing to take on the task.

In contrast, however, one participant felt that peer reviewers have the right to access the underlying data for a study – particularly when that data will be published alongside the paper. In their opinion, if data was available at publication, it should also be available at peer review and might actually make it easier for the reviewer to assess the paper.

30. Laine (2017)

The consensus was that peer review of data was important – particularly in cases where the data was to be shared at publication – and that the way to ensure this happened would be to give reviewers greater recognition.

Opinion: Fair management of genetic data

Heidi Laine

The FAIR principles are for machine-friendly data management. When humans are the sources of data, especially when the data are about health, we need human-friendly data management. Instead of creating fair principles to complement the FAIR principles, genetic data stewards can tap into already established principles for research ethics and research integrity. Processes for obtaining informed consent from genetic data donors can act both as a way to implement fairness and test it.

The coronavirus pandemic, which is at the time of writing still very much ongoing, has put a spotlight on the need for interoperable human health data. The FAIR principles for data management offer a framework for creating a common interoperable data resource, that can be efficiently mined for insight and solutions during this uniquely global crisis. But as the pressure for ever wider access and broader pooling of human health data mounts (and will undoubtedly continue to do so also in the post-pandemic world), the research community can't put aside the rights of the individuals behind the data. FAIR for machines doesn't matter, if it's not also fair for humans.

Even in the middle of a race to solve a global pandemic, the stewards of genetic data, such as biobanks, need to continue to address the question of what fair data management means for the people whose genetic fingerprints are stored in their samples and data. Asking and answering once, or even a few times, is not enough. As data management technologies evolve, so do the definitions of FAIR, and fair. What we need is a perpetual motion machine syncing FAIR to fair, and vice versa.

As data management technologies evolve, so do the definitions of FAIR, and fair. What we need is a perpetual motion machine syncing FAIR to fair, and vice versa

While the exact manifestations of both FAIRness and fairness need to be updated according to available technologies, there are certain unchanging core principles to both. For FAIR they are of course findability, accessibility, interoperability, and reusability. The equivalent principles of fair data management could be for example honesty, security, opacity, and understandability. Or something similar. I'm not suggesting that we try to coin these terms down like has been done with FAIR. Catchy acronyms can be handy, and FAIR has already been followed or complemented by at least TRUST, and CARE, but the existing body of thought on research integrity and research ethics offers useful conceptual tools as it is, see for example The Singapore Statement on Research Integrity, or the European Code of Conduct for Research Integrity. I find it most constructive to take these existing frameworks for responsible research and apply them to fair data management. It is about time we stop treating data management as an art of its own, and something that is addressed only in DMP's and data policies, instead of the actual research plan, or methods sections of articles.

You get what you measure, so in addition to principles, fair data management needs indicators. If we were to create a fairness test for genetic data, mimicking the FAIR maturity evaluations that check the machine readability of data, the practices and processes in place for informed consent would offer the ideal pressure point for testing.

Informed consent is one of the main ethical guarantees in medical and health sciences. The concept of informed consent was developed after World War II, with the goal to never again allow in the name of science criminal acts like those committed by certain scientists, working within the national socialist regime, to defenceless prisoners. Practices for establishing informed consent vary, but follow roughly three ideals: voluntarity, preferably professed through a deliberate and documented act of opting-in, sufficient information about the study in question, and the ability of the consentée to understand the information given.

Currently the information offered to biobank donors in my home country Finland, in the actual consent documentation is general and describes the different research cases for which the data are likely to be used, as well as giving reassurances about privacy. Data management is rarely addressed directly. Fortunately most Finnish biobanks offer additional information on their websites to those who want to dig deeper. The massive FinnGen project on personalised medicine, that uses data from most Finnish biobanks, addresses the issue of data security in its FAQ section as follows:

> *Information security is managed with utmost care, and the data of the participants is coded for processing. In the FinnGen research (as in all biobank projects) the participants' data are coded and processed in a high information security environment, the access rights of which are carefully monitored. The FinnGen study has nominated persons responsible for information security and data protection, whose job is to supervise the information security of the research subjects. We have also had an outside party specialised in the matter to make an assessment on the data protection and security of the project. The risk for misuse of information is very minor*[31].

While the Finnish biobanks offer data donors information that is accurate, relevant and honest, as exemplified by the FinnGen description above, it is not very detailed, and doesn't allow thorough evaluation of their data management practices. This is understandable, since the biobanks are walking a fine line between being too scarce and general with the information, and exhausting the donors with overly technical minutiae. But I still feel more could be done to communicate information about data management, especially from the point of view of risk management. The lack of focus on data security, which is from the data donors' point of view the most relevant part of the data management, at least to me begs the question, is there enough understanding within the biobanks themselves about data security, risks, and how to prepare for the worst. All banks can be robbed, even biobanks.

Mainstream media discourse about risks related to genetic data tend to revolve around the possibility that employers or insurance companies could start ranking their customers based on their heretic susceptibility to this or that condition. That is not a data security risk, but a risk related to regulation and how we use genetic information in our societies. When it comes to data breaches genetic data are by nature relatively safe from misuse. It is certainly sensitive as it deals with people's health, but analysis to the level of identifying individuals requires specific expertise that not every hacker out there is in possession of. However, what almost every hacker is capable of, is orchestrating a blackmail scheme.

The danger with this was last manifested in Finland during late 2020, when an online psychotherapy service lost the data of tens of thousands of individual customers, who then received demands for ransom from the hacker/hackers, or else their sensitive therapy data would be published online (it had likely already exchanged hands in the dark web at this point, possibly several times). It is very hard to estimate the actual harm that was caused by this breach to the data subjects, but for the rest of their lives

31. FinnGen

they live with the knowledge that this intimate information is out there, perhaps waiting to resurface at the most inopportune moment. On a societal level the worst part of the hacking was perhaps the humiliating inability of the company and official authorities to do anything to mitigate the situation. When data are lost, they are lost forever. When trust is lost, it can thankfully be repaired, but only with time and labour.

What if the data in question had been genetic data instead of therapy diaries? I don't think it would have made much difference from the point of view of the victims, however many experts would have explained on news shows that genetic data is not easily identifiable. No one wants their genetic information to become a commodity on Tor.

People are expected to not understand data management and data security issues. But the pandemic has shown that thousands of people can turn into amateur virologists overnight, debating about R0 values on Twitter like it's their job. I'm not suggesting we should wish for a suitable global crisis to get people excited about data management, but if we are serious about the informed part in informed consent, we need to make more of an effort in educating ourselves about genetic data management and its perils. We, the society, and we, the research community. When the world around us gets more complicated, we learn new things. That has been the homo sapiens way since ca. 300,000 BCE.

It's hard to learn new information if it doesn't tickle your interest. Data management solutions can be described using different methods such as visualisations and animations in order to make them more easily understandable. It should not be enough to deliver legally waterproof but mind numbingly boring jargon, like the big tech companies do whenever you create a new profile in an online service or start using a new smart product (or both, as it has become virtually impossible to use any digital device without a personal profile attached to it).

Informing the public about information security and personal data management would pay off also beyond biobanks and management of genetic data. Informed consent concerning the handling of digital personal data shouldn't be an isolated event, that comes across once in a person's lifespan. The part of the research community dealing with personal data has the opportunity, I dare even say the responsibility, to try to act as a counterbalance to big tech, and create models for building genuine trust with the public, instead of just relying on people looking the other way in exchange for convenience, and doing at all times the bare minimum decreed by regulation.

When data are lost, they are lost forever. When trust is lost, it can thankfully be repaired, but only with time and labour

Opinion: Ethical considerations for research data sharing

Anne Cambon-Thomsen, Elisa Carrus, Claudia Civai, Fátima L. S. Nunes, Varsha Khodiyar

When it comes to data sharing and ethical considerations, there are many aspects which could be explored. Here we expand on the discussions which took place in breakout five of the *Better Research through Better Data* roundtable event. We can map the discussions to a typical research study lifecycle; considering data sharing before/during data collection via the clinical data use case, data sharing during the literature peer review process, and ending with data sharing considerations after research study publication.

Sharing clinical data

We are far from a consensus on how best to share clinical data. Questions as basic as "What can we effectively share?" and "How can one ensure patient privacy and confidentiality?" or "What is the definition of anonymisation and the adequate methods to achieve it?" do not have an universally accepted answer. Neither have we as a community reached any consensus on the obligations and ethical responsibilities of researchers using patient data. However, the sheer volume of peer-reviewed literature on this topic demonstrates that this is a very well explored issue.

Considering just one challenge, preserving patient privacy whilst facilitating the use of data, we have seen a number of technical solutions being proposed, for example deep neural networks[32] and blockchain[33]. Multiple proposals and commentaries on principles, responsibilities, and recommendations[34] are also evident. Clinical data sharing is vital for research in all biomedical fields, and this is reflected in the literature with experiences drawn from genomics[35], psychopathology[36], pharmacology[37], and the current COVID-19 pandemic[38].

Analysing the code of ethics from 191 organisations, Vitolla *et al*[39] confirm what can be considered as public knowledge, that *"despite the global relevance of ethics, the drafting processes, and the contents of ethical codes differ considerably between countries"*. This aptly demonstrates that one of the biggest challenges for clinical data sharing is that legalities of sharing this type of data vary widely depending on the country and jurisdiction where the data were gathered[40].

Perhaps there are lessons to be learned for digitally-held patient data from biobanks which hold physical patient samples. For example Finnish biobanks are heavily regulated, and provide clear guidance that participants have the right to know which studies their sample has been assigned to, and to know about research results of the resultant studies[41]. In Brazil, researchers are obligated to provide study participants with a nominated contact to answer doubts and communicate any issues arising from the research study[42]. However, a recent Brazilian law (similar to European GDPR) differentiates private and anonymised data, establishing more flexible rules for the latter. In a recent study[43] with 700 German patients, 93% showed strong willingness to give broad consent for secondary data use and almost 76% declared strong approval of abolishing patient consent. A major point is the source of the clinical data, patient care (health records) or research (clinical trial or other biomedical research) as regulations often differ. Tools providing information to researchers who wish to share sensitive data or to use sensitive data are being produced, such as in the European Open Science Cloud. As a matter of fact in such a complex landscape, it is as important to facilitate easy and rapid access to clearly described applicable rules and regulations than to reach consensus, which is often a long term endeavour at international level.

Differences among rules and laws are becoming more evident with the increasing amount of research carried out in international consortia. Nevertheless, Kalkman *et al*[44] analysed ethical guidelines for international health research and discovered considerable convergence in themes related to "societal benefits; distribution of risks, benefits and burdens; respect for individuals and groups; and public trust and engagement".

Independent of the locality or goals of the research study, we can then establish that we have begun to research a consensus in some areas of clinical data sharing: (1) respect for the study participant is the main requirement; (2) transparency between the researcher and the subject is necessary; (3) although it is often not possible at the point of consent to specify to the participant exactly how and with whom the data will be

One of the biggest challenges is that legalities of sharing clinical data vary widely depending on the country and jurisdiction where the data were gathered

32. Beaulieu-Jones et al., (2019)
33. Luo et al., (2019)
34. Kalkman et al., (2019)
35. Byrd et al., (2020), Shahin et al., (2020), Riggs et al., (2019)
36. Krypotos et al., (2019)
37. Burton et al., (2020)
38. Moorthy et al., (2020)
39. Vitolla et al., (2021)
40. Vitolla et al., (2021)
41. Biobank
42. Castro et al., (2020), Ministério da Saúde
43. Richter et al., (2019)
44. Kalkman et al., (2019)

shared in the future, this can be made clear in the participant consent document; (4) anonymisation is a requirement in all cases in order to ensure the subject's privacy.

Sharing data with peer reviewers

The process of peer review for research literature has been a vital and important part of scholarly publishing since the middle of the 20th century[45]. Over the last decade the importance of peer reviewing has been in the spotlight with calls for greater recognition and credit for reviewers[46]. Whether we can then reasonably add to the 'peer reviewer's burden' with the data underlying the claims being made in a manuscript is currently under debate. Data journals such as *Scientific Data* provide clear guidance to reviewers on assessing the data being presented[47]. However reviewers of data papers are not expected to form a judgement on the novelty of potential impact of the data, as data papers are assessed on research soundness. So for traditional research articles where reviewers are asked to consider novelty and value to the field, is it also reasonable to ask those same reviewers to assess the underlying data?

On the other hand it seems ethically ambiguous to review a research manuscript without an assessment of the robustness of the evidence supporting the claims being made. In recent years the importance of trust in scientists and the scientific process has been laid bare, with global challenges such as climate change requiring co-ordinated action from policymakers worldwide. There has to be a level of trust in the underlying data and evidence which are showing the need for change in policies and practice. The coronavirus pandemic has brought into view yet another spotlight on the challenges faced by governments and health officials when populations distrust or misunderstand research findings. Public understanding of the science on issues such as mask wearing, social distancing and vaccination have important factors in compliance with governmental guidance. Public trust in research is an important and significant challenge, and the authors are of the opinion that a part of this is ensuring that research has been validated as robust and trustworthy, prior to being formally published. This does not deny the value of preprints, especially in urgent situations such as the COVID-19 pandemic, but transparency and clarity about their status as non-peer reviewed articles is paramount, as well as facilitation of preprint peer open comments. Ensuring the scientific evaluation processes are communicated widely and in an accessible manner to the general population has never been so important.

Sharing data with the research community

Many journals and funding bodies now require researchers to share their data and their data management plans, and there are clear benefits of engaging with these practices, such as transparency and accelerating research progress. Data sharing has been perceived increasingly more favourably in recent years, though the key obstacles of insufficient time, lack of academic incentives and financial resources persist[48]. For example, researchers who do not have the funding to meet data sharing requirements will be required to find a suitable freely available repository, which may offer limited options[49]. Even when financial resources are available, data sharing and management is time-consuming and the majority of researchers have reported a lack of sufficient organisational support for data management with a clear need for training to be implemented, especially in academic institutions[50]. For the reasons above, it was felt that it would be unfair to impose the same data sharing requirements unconditionally as this would discriminate against those with fewer resources. Additionally, any data sharing requirements should accommodate training and support needs.

Whilst data sharing and data reuse are viewed positively, a prominent concern is also that of recognition, consistently voiced by over 90% of researchers[51]. A particular issue was felt around recognition from any discoveries that are made from the shared data.

The coronavirus pandemic has brought into view yet another spotlight on the challenges faced by governments and health officials when populations distrust or misunderstand research findings

45. Burnham (1990)
46. Glonti et al., (2019), Publons
47. https://www.nature.com/sdata/policies/for-referees
48. Tenopir et al., (2011), Tenopir et al., (2015), Tenopir et al., (2020)
49. Tenopir et al., (2020)
50. Tenopir et al., (2011)
51. Tenopir et al., (2015), Tenopir et al., (2011)

Other concerns may also exacerbate data sharing engagement; for example, researchers may feel that their research will be 'scooped', with the majority reporting the need to publish first as one the main barriers to data sharing[52]. There are additional concerns over the reuse of data, specifically data misinterpretation, poor quality of data, and data misuse, i.e., used for purposes not initially intended. Therefore, to foster data sharing practices more widely, a system of incentives must be provided to authors (e.g. citation, co-authorship, system of credits), and as suggested by Tenopir *et al*[53], embedded in organisational systems of rewards.

To foster data sharing practices more widely, a system of incentives must be provided to authors, and embedded in organisational systems of rewards

Considerations for the future

The evidence of benefit from sharing research data is growing, with demonstrable benefits for individual researchers, for scientific progress, and for wider society[54]. For data sharing at all points in the research study life cycle to succeed, researchers need to remain mindful of the two groups which need to buy into the data sharing ethos, the researchers themselves and the participants of the studies. Once a decision has been made to share study data, researchers may then require practical guidance on data handling, processing and archiving, which needs to be accompanied by a system of incentives for this work. Peer reviewers have an important role to promote, encourage and enforce as needed the normal data sharing practices in their disciplines. Journals and publishers have an important role to play in research dissemination to ensure that data underlying a research paper have been appropriately referenced, and that these are fed into the nascent data citation tracking infrastructure[55].

52. Tenopir et al., (2011)
53. Tenopir et al., (2015)
54. Colavizza et al., (2020), Markowetz (2015), "EconOmics" (2013)
55. Khan et al., (2020), Cousijn et al., (2018)

3. Using Real World Data for research

Summary of discussion

Only a few years ago, Real World Data (RWD) were not considered useful for healthcare research. This was mainly due to concerns around bias which, it was felt, rendered the data unusable for decisions such as which medicines to approve.

However, in 2016, the Federal Drug Administration (FDA) in the US implemented new guidelines based on the 21st Century Cures Act. These guidelines allowed companies to submit data from real-world settings to support the approval of new drugs and also to satisfy post-approval requirements.

Since then, there have been at least 15 cases in which RWD have been used in approvals of medicines and also in approving new indications for existing medicines. And this shift in policy has in turn prompted increased interest in gathering RWD.

There are two areas in particular where the use of RWD in research is growing at pace. The first is population genomics databases, some of which can be used by industry, while others are for purely academic purposes. The second is electronic health records, many of which are now being made available for use outside of hospitals and other medical settings.

At the same time, the use of mobile devices, wearables and other biosensors to gather and store huge amounts of health-related data has been rapidly accelerating. These data hold the potential to allow us to better design and conduct clinical trials and studies in the healthcare setting to answer questions previously thought infeasible.

However, these developments bring with them a whole host of ethical and practical questions. Should hospitals and clinics be profiting from sharing patient data? Is patient consent being properly obtained? Is de-identification effective enough to ensure patient anonymity? And should patients be compensated for the use of their data, either financially or otherwise?

Bias in Real World Data

In the context of the discussion, RWD were defined as patient data derived outside of the clinical trials environment. One concern raised early on about this type of data was how representative it can really be. Would it overrepresent certain ethnic groups, for example? However, it was argued among participants that this could be equally true for some clinical trials based on limited populations. Thus, RWD could in fact help mitigate the bias of existing clinical trial results.

Participants noted that one use of RWD in research is machine learning and developing AI systems. A serious issue, therefore, regarding the bias of the data is that it could become inherently built into the systems that use them, but without the bias being adequately acknowledged or accounted for.

Balancing progress against privacy

Perhaps the biggest issue identified by all participants in the discussion was the balance of the potential progress to be made by using the proliferation of data now available, against the need to protect patient privacy.

A study published in *Nature Communications*[56] in 2019 highlighted the fact that even anonymising data did not necessarily lead to protection of patients' privacy. And the discussion participants were keen to point out that in many instances it is possible to track down individuals or groups based on indicators such as georeferencing or by cross-referencing of databases.

Concerns for patients raised in the discussion around this topic ranged from possible discrimination – for example, health insurance companies discriminating based on locality – through to deportation, in the case of a study on the health of illegal immigrants.

Consent and public trust

It was also raised that, in many instances, patients may be unaware to what extent their data are being used. The example of the Mayo Clinic in the US[57] demonstrates that patients can be entirely ignorant of how their healthcare records are being shared. The example of a company building an app and making money on the basis of using patient data was discussed, in contrast to researchers using the data for purely academic purposes.

At the other extreme, in some cases patients are being actively incentivised to share their data. PicnicHealth – which offers patients money in exchange for their data – raised different ethical considerations. Will those who are particularly in need of money be more likely to share their information? This would, in turn, create another bias in the data collection.

One participant gave the example of national repositories that already exist across Scandinavia for storing patient data and different health outcomes. This raised the question of whether the responsibility for collecting this type of data should ultimately be with national governments. However, it was agreed that public willingness to participate in and share data with the government could be widely different in different countries.

Ultimately, it was felt that the majority of patients do not understand how their data will be used, even when they sign a consent form. Participants asked whether the key issue here was actually transparency and whether achieving the highest level of transparency would be preferable to a guarantee of a particular level of identifiability. A trusted partner could help patients understand the consequences of sharing their data, as well as the benefits.

A trusted partner could help patients understand the consequences of sharing their data, as well as the benefits

Opinion: Real World Data sharing in the time of Covid

Raul Rodriguez-Esteban

The analysis of RWD has been gaining importance over the past few years. It represents a new iteration in the science of observational studies[58], bringing the added novelty of large datasets, new analytical tools and a changing regulatory landscape. At its core, it offers an attractive proposition towards a more consequential use of existing healthcare data and a reduction of trial burden on patients. Successful examples of the application of RWD analysis have led to reducing the need for control groups in clinical trials, particularly for rare diseases, oncology or small patient subpopulations; delivering

56. Rocher et al., (2019)
57. Ross (2020)
58. Makady et al., (2017)

insights regarding new treatment concepts or regimes; and addressing medical questions for which running a clinical trial would be unfeasible or too costly[59].

The emergence of RWD analysis as a new trend in healthcare data analysis predates the COVID-19 pandemic. However, as in other areas, the pandemic has accelerated its development and raised its profile. Limitations in mobility and postponement of healthcare procedures, whether due to government-mandated restrictions or fear towards infection, have slowed the pace of clinical trials, including patient recruitment, increasing the need for efficient leveraging of existing healthcare data[60]. The use of RWD has been key in the discovery of co-medications associated with differential outcomes in patients affected by the disease[61], in the assessment of patient-specific risk factors[62], and in the study of the real-world effectiveness of new interventions, such as the vaccination of country-wide populations and at-risk populations[63]. Additionally, debates around RWD analysis that would normally have stayed within scientific circles have played out over news headlines[64]. In this context, the need for well-designed RWD studies[65] paired with rigorous gold-standard clinical trials has been made plain, highlighting the importance of both.

The spotlight on the value of RWD analysis has also led to increased attention being paid to the importance of RWD sharing[66], and particularly on timely and transparent sharing to support public health and medical decision making, from test results to hospitalisation statistics to the monitoring of excess deaths and vaccine inoculations. The rapid sharing of results and data in non-peer-reviewed venues such as preprint servers has both catalysed advances in RWD analysis of the disease and attracted attention towards resources that were relatively obscure to the general public before the pandemic. This has resulted in the public being educated to some extent in RWD topics such as the difference between clinical trial performance of medical advances and their real-world performance, from diagnostic tests to medicines to vaccines.

The pandemic has laid bare and magnified bottlenecks in RWD sharing that predate the pandemic and whose amelioration has taken on renewed urgency. This has been acknowledged by regulatory bodies, such as the FDA, which early on highlighted the importance of RWD analysis to combat the pandemic[67]. In response they created a COVID-19 Evidence Accelerator initiative[68] to foster research on RWD that spearheaded the COVID-19 Real World Data (RWD) Data Elements Harmonization Project[69], which proposed common data models for COVID-19 RWD. Another example of government initiative in RWD sharing is the COVID-19 laboratory reporting standard from the CDC, created under the umbrella of the US CARES Act[70] for the harmonised reporting of diagnostic test results. Furthermore, certain countries such as Korea have been able to develop centralised repositories for the investigation of RWD associated with the disease[71].

The virus does not stop at borders, thus multinational RWD sharing collaborations have been crucial to combat the pandemic. One example is the European Union (EU) interoperability gateway that has allowed the interconnection of contact tracing and warning apps across EU countries[72]. International players in the healthcare industry have created a COVID-19 Research Database[73] that "enables public health and policy researchers to use real-world data to better understand and combat the COVID-19 pandemic." Another area of important cross-national RWD exchange has been that of viral sequencing data sharing. During the pandemic, viral sequencing has experienced a brisk increase due to concerns about new virus variants associated with more aggressive phenotypes and faster propagation. The open access repository GISAID (Global Initiative on Sharing All Influenza Data), started in 2008 with the purpose of sharing influenza virus sequencing data, has become the central international repository for

The virus does not stop at borders, thus multinational RWD sharing collaborations have been crucial to combat the pandemic

59. Rogers et al., (2021), Bolislis et al., (2020)
60. Narozniak (2020)
61. Halpin et al., (2020), Hughes et al., (2020)
62. Schwab et al., (2021), Gadgeel et al., (2020)
63. Chodick et al., (2021)
64. Dolgin (2020)
65. Franklin et al., (2021)
66. "Transparency during global health emergencies" (2020)
67. Sutter (2020)
68. https://evidenceaccelerator.org/
69. https://www.fda.gov/drugs/coronavirus-covid-19-drugs/covid-19-real-world-data-rwd-data-elements-harmonization-project
70. Section 18115. https://www.congress.gov/116/bills/hr748/BILLS-116hr748enr.pdf
71. Rho et al., (2021)
72. European Commission (2020)
73. Adams (2020)

SARS-CoV-2 sequencing data and experienced extraordinary growth. Its success is likely based on its data sharing approach, which allows the "sharing of genetic data to meet emergency situations, without infringing intellectual property rights"[74]. Due to its centralised nature, it offers the advantage of a singular data model but data completeness is an important challenge[75], as patient characteristics are missing in many entries. Moreover, there is high variability in the degree of data contribution by country[76]. More recently, the Global Health Initiative has created a data repository for virus variant tracking from over 100 countries[77].

The pandemic has left unresolved legal and ethical aspects concerning data sharing consent by patients; the regulatory framework for commercial sharing and use of RWD; and the role of public institutions as both guardians of patients' health data and stakeholders in the development of treatments. For example, the governments of Israel and Iceland negotiated with the companies Pfizer and BioNTech the provision of RWD associated with vaccination outcomes in exchange for greater access to vaccines. Such an agreement would not have been acceptable for other countries, such as EU member states, as expressed by the president of the European Commission[78].

Learnings from the experience of RWD sharing during the pandemic will help develop future strategies to enhance timeliness, clear communication with the public and stakeholders, support for exchanges and standards by regulatory bodies, and international initiatives. Improvements in these areas should feed back into greater future interest in utilising and generating RWD. The FAIR principles (Findable, Accessible, Interoperable, Reusable) maintain their currency as a reference for data sharing within a context of preservation of privacy and intellectual property rights.

The pandemic has brought renewed attention to the biomedical sciences, and to RWD analysis in particular. A post-pandemic world will hopefully maintain its focus on furthering the feasibility and the impact of RWD analysis for the ultimate improvement of patient outcomes.

Learnings from the experience of RWD sharing during the pandemic will help develop future strategies to enhance timeliness, clear communication with the public and stakeholders, support for exchanges and standards by regulatory bodies, and international initiatives

74. Shu & McCauley (2017)
75. Velazquez et al., (2020)
76. Chen et al., (2021)
77. Maxmen (2021)
78. Finke (2021)

4. Supporting roles for FAIR data

Summary of discussion

Shortly after the publication of the original FAIR principles, it became obvious that the application of those principles in practice would require a wide range of new competencies, skills and knowledge within the research community.

This has led to many organisations recruiting, or looking to recruit, data professionals such as data managers or data stewards. These roles, while sharing titles, can vary from organisation to organisation and often include responsibilities around training and supporting researchers in implementing FAIR.

Meanwhile, it has also become clear that research data and software are intrinsically linked, so having data skills and competencies alone is not enough. This has resulted in a need for organisations to have not only data managers and data stewards, but also research software engineers and possibly other roles to support their researchers effectively.

As these are new professions, there's not always a consensus on the exact responsibilities and tasks of these roles. Is every need accounted for or are we still missing vital support for implementing the FAIR principles? What should the responsibilities of these roles be? And what barriers are there to ensuring that the FAIR principles become 'the norm' for researchers?

Defining roles and responsibilities

One of the key issues raised in the discussion was defining the roles needed to support the implementation of FAIR. It was felt by the group that while new roles such as data managers and data stewards now existed in research organisations and institutions, the actual responsibilities of these roles differed widely.

Some participants suggested that rather than looking at whether we have the right roles in place, we should instead be considering what skillsets are needed. And in turn, whether this means creating entirely new roles or upskilling existing roles. As well as stewardship and management of data, participants felt that support would be needed for researchers in areas such as legal advice – for example, in understanding and meeting standards of themes such as GDPR and software licensing.

Ultimately, however, it was felt that the skills and roles required to implement FAIR needed to be defined by what barriers exist. Ideally, the needs of researchers in an institutional, regional and/or discipline-specific context should be taken into account to determine the roles needed.

Fear of misuse of data was given as an example of a common issue cited by researchers. Participants asked, can tools, resources or support roles help with this or is this a cultural problem stemming from scientific training?

The example of Delft University of Technology was touched upon during the discussion. There, data stewards are embedded in each faculty. They can't get involved at the project level as there is not enough resource for this, but their main role involves helping with funders' and publishers' policies, showing researchers how to make a start with FAIR data principles, and providing training to the researchers.

Even with this level of data stewardship, it was acknowledged that they're still lacking more discipline-specific, hands-on support. There was also a concern raised around the career track for these new data-related roles – at the moment, people taking up these roles don't have a clear career progression path.

Education and building FAIR into research practices

Participants were broadly in agreement that the only way that the FAIR principles can be implemented effectively is if they become part of research best practice – in other words, there's a need to embed the principles in how research is being produced.

The only way that the FAIR principles can be implemented effectively is if they become part of research best practice – in other words, there's a need to embed the principles in how research is being produced

More than one participant pointed out that there is still a lack of knowledge in the research community about what FAIR is and why it's needed. However, while it was felt that the importance of the FAIR principles needed to be embedded in the research culture, participants didn't feel that researchers necessarily needed to know the principles themselves.

Instead, what was felt to be required were practical translations of the principles that could be applied in different disciplines. In other words, tangible actions that researchers can take.

The discussion touched upon the fact that offering support to researchers is not enough. Instead, they need to know that data management is part of responsible research. And this understanding needs to be built into the research curriculum from the beginning.

Participants agreed that it was better for researchers to start using the FAIR principles in their data collection from as early as possible in their careers. This would mean that they're applying them before they've built up a body of research and associated data. However, it was also discussed that it was important to make sure researchers know it's never too late to apply the concepts to their work.

Overall, there was general consensus that there needs to be support and stewardship to help researchers understand how they can better manage data, as well as helping them to understand the difference between FAIR versus open data. In general, participants felt that the main requirement was to make it as easy as possible for researchers to do the 'right' thing.

Rewarding and recognising practicing the FAIR principles

Participants also raised the question of whether there are incentives to encourage the use of FAIR. It's currently considered as "nice to have" but often not as a requirement in funding, appointment, and promotion considerations. Although there are some examples of open science practices being considered as a part of promotion criteria, still these are not comparable to publishing in high impact journals or getting a large funding grant. However, one participant did flag that at their institution open science practices are part of the promotions criteria, with researchers needing to demonstrate how they've embedded open practices in their work, in order to become a professor.

Community versus individual responsibility

Many participants felt that placing the responsibility for FAIR practices on the shoulders

56. Rocher et al., (2019)
57. Ross (2020)
58. Makady et al., (2017)

of individual researchers was not necessarily the right approach. The discussion led to the idea that at least some of the work to implement consistent standards and approaches, for example, needs to be carried out at a discipline or community level.

The field of astrophysics was cited as a good example of where data sharing and common data standards are relatively well established. This was felt to be community-driven, in the sense that the field is very international with shared equipment and facilities – for example, the Hubble Telescope, which manages data collected independently of individual researchers.

Role models and case studies are needed to set an example

Participants agreed that the discussion had made apparent how complex this particular area is and how much is still needed to make the FAIR principles part of research practice.

Role models and case studies of researchers and organisations that have embedded the FAIR principles into their research practices were felt to have an important role to play in this. It was felt that by providing clear examples, relevant to different disciplines, researchers would be able to see the tangible benefits of what could otherwise be a rather nebulous concept.

The conclusion of the discussion was that it was important to show that it's not an 'all or nothing' approach – researchers can take small steps towards better data management and sharing. As one participant put it, "We need to find a way of making implementation achievable through practical bitesize actions."

Opinion

Authors in alphabetical order:
Louise Chisholm, Maria Cruz, Rebecca Grant, Elizabeth Newbold, Yasemin Türkyilmaz-van der Velden

The FAIR principles have received worldwide attention and support since their publication in 2016[79]. Since then, an increasing number of research funders, institutions, universities and publishers have developed policies that encourage if not mandate researchers to follow FAIR data principles. However, application of the FAIR principles in practice requires a wide range of competencies, skills and knowledge as well as research infrastructure and support services. In this breakout session, we looked into the current research support landscape and tried to address the most urgent needs and biggest challenges when it comes to supporting roles for FAIR data and how to tackle these remembering that FAIR is a journey and not a binary state. In this opinion piece, we revisit the main topics discussed during the breakout session and focus on possible next steps for tackling some of the remaining challenges.

Defining roles and responsibilities

Organisations such as Delft University of Technology have already embedded data stewards in their faculties[80,81] while many others are looking for ways to implement this role in their organisations. Yet, as this is a new profession, there is not always consensus on the exact responsibilities and tasks of this role. Additionally, there are a lot of discussions around how to ensure visibility and recognition as well as capacity building and staff retention. Therefore there is a need for professionalisation of these roles and

79. Wilkinson et al., (2016)
80. Teperek et al., (2018)
81. Plomp et al., (2019)

development of career progression tracks. Currently there are ongoing efforts to establish these at the national[82,83] and international levels[84].

Furthermore, supporting FAIR data requires diverse skill sets in a variety of support roles within the research ecosystem. It is not enough for organisations to hire only data experts such as data stewards. It is also necessary to equip existing professionals with the knowledge to be able to support FAIR data. These professionals include data librarians, data curators and other roles in research infrastructure as well as legal experts who can help with relevant legislation (including data protection and software licensing) and Professional Research Investment and Strategy Managers (or PRISMS)[85].

Education and building FAIR into research practices

For realising the FAIR principles in research practice, providing FAIR data support alone is not enough. It is essential that responsible data management is seen as a fundamental element of research practice. Making data FAIR is not something that can be easily done at the end of a research project if no thought and planning has been put into it. Even if the necessary infrastructure, tools and expert support are all available, if data has not been stored in the right format, with adequate file naming and documentation, it would be difficult to achieve all of this at the very last moment. It is essential that research is planned from the very beginning in a way that good data management practices are applied at every stage during the research life cycle[86]. It is therefore important to start training researchers in FAIR data practices as early as possible by embedding the required knowledge and skills in higher education curricula.

This does not mean that every researcher needs to know the detailed definitions corresponding to each individual letter of the FAIR acronym. For most researchers, it is much more relevant if the FAIR principles are translated into practical actions that they can apply to the data in their own discipline. The Top 10 FAIR Data & Software Things[87], which provide brief guides for various disciplines ranging from nanotechnology and astronomy to music and humanities, set a great example for translating FAIR into practical actions.

Rewarding and recognising practising the FAIR principles

Policy changes like Horizon Europe's support for "Open data by default"[88] introduce new obligations for funded researchers which must be complied with. However, outside of policy obligations, the rewards and incentives for FAIR data sharing have not yet been widely embraced. Researchers are not necessarily incentivised to share their research data in the first place: less than half of journal authors deposit their data into repositories[89], and nearly 60% of researchers still believe that they do not receive sufficient credit for sharing their data[90]. If researchers do not feel they are rewarded for data sharing, why should they undertake additional work to produce FAIR data?

Researchers report that the types of recognition or rewards that they value for data sharing include full data citations when their datasets are reused; and increased impact and visibility of their research[91]. We can already demonstrate the positive impact that data sharing can have: an average 25% increase in citations for papers which share data[92]. But can we demonstrate the additional benefits that come from FAIR data?

Evidence of tangible benefits, rewards and recognition for researchers who share FAIR data can help to move us on from policy compliance and mandated data sharing activities. Case studies or larger scale studies like Colavizza *et al*[93] , can provide evidence of benefits; while other reward mechanisms such as Open Data Badges are also being explored[94]. The Declaration on Research Assessment (DORA) has also developed case studies which show how institutions have improved their evaluation of scholarly research

For most researchers, it is much more relevant if the FAIR principles are translated into practical actions that they can apply to the data in their own discipline

89. Stuart et al., (2018)
90. Digital Science et al., (2020)
91. Digital Science et al., (2020)
92. Colavizza et al., (2020)
93. Colavizza et al., (2020)
94. Center for Open Science (COS) and Pearce & Grant (2020)
82. Jetten et al., (2021)
83. Wildgaard et al., (2020)
84. Research Data Alliance (RDA)
85. PRISMs are managers of large research, training, and network grants in Higher Education Institutions. See https://www.pris-managers.ac.uk/
86. Network of the National Library of Medicine (NNLM)
87. Martinez et al., (2019)
88. Directorate-General for Research and Innovation (European Commission) (2016)

outputs[95]. For example, University College London has adopted an academic career assessment framework which specifically embraces 'excellence' and 'quality' in an open science environment, which includes FAIR data[96].

Community versus individual responsibility

In 2020, nearly 40% of surveyed researchers stated that they had never heard of the FAIR principles[97]. Even for those who had heard of the principles, only 25% of researchers felt they were "familiar" with them. The FAIR principles are comprehensive but technical, describing practices which touch on metadata, persistent identifiers, access protocols and repositories. Can we assume that even those researchers who are "familiar" with FAIR really understand its practical application?

The FAIR principles themselves indicate the responsibilities of communities as well as individual researchers to create and share FAIR data. For example, the "Reusable" goal of ensuring that "(Meta)data meet domain-relevant community standards" can only be met where such standards exist. It is therefore the work of a community to develop and disseminate new standards which are applicable to their data, where none exist. Such communities might consist of researchers in a particular discipline, as well as those working and collaborating across labs and facilities.

Role models and case studies are needed to set an example

Academic role models can advocate the adoption of the FAIR principles, they can demonstrate how it can support their career aspirations by increasing research impact and new opportunities. Role models are also needed for specialist staff who have roles supporting FAIR data (such as data stewardship) which showcase new career pathways. Funders, research organisations and academic societies could champion these role models (see European Group of FAIR Champions[98], Data Champions of TU Delft[99],[100] Cambridge University[101], EPFL[102] and University of Melbourne[103]) to demonstrate that they, and their adoption of the FAIR principles, are valued within the research community.

For researchers adopting FAIR approaches for the first time it can be difficult to understand where to begin. Making data FAIR should be considered to be an incremental process. To reduce the barriers, the FAIR principles need to be accessible and tailored to different disciplines. Subject-specific case studies can demonstrate the steps researchers can take to adopt the FAIR principles. These can also be used to address discipline-specific challenges and cultural concerns. Sharing case studies in conferences and doctoral training programmes can help disseminate tangible and achievable actions and the benefits of adopting the FAIR principles.

The potential to use data to develop new insights might be limited if the academic community does not have metadata and other standards agreed. Interoperability of standards across disciplines can also enable unique research opportunities. Case studies can demystify how academic communities have developed metadata and other data standards and highlight the impact such standards can have across the discipline. They can help demonstrate to funders and academic communities, which do not currently have agreed standards, why investing in their development and dissemination would be worthwhile. The celebration of standards, research datasets and databases as research outputs – as they are by the Hidden REF[104] competition or by funding initiatives such as the Wellcome Open Research Fund[105] and the NWO Open Science Fund[106] – can also raise awareness of how they contribute to the success of research.

Evidence of tangible benefits, rewards and recognition for researchers who share FAIR data can help to move us on from policy compliance and mandated data sharing activities

95. Declaration on Research Assessment (DORA)
96. Declaration on Research
97. Digital Science et al., (2020)
98. https://www.fairsfair.eu/advisory-board/egfc
99. https://osc-delft.github.io/initiatives
100. Clare (2019)
101. https://www.data.cam.ac.uk/intro-data-champions
102. https://www.epfl.ch/campus/library/services/services-researchers/rdm-contacts-communities/epfl-data-champions/
103. https://library.unimelb.edu.au/digital-stewardship/data-champions
104. https://hidden-ref.org/
105. https://wellcome.org/grant-funding/schemes/open-research-fund
106. https://www.nwo.nl/en/researchprogrammes/open-science/open-science-fund

Conclusion

As portrayed by the variety of the topics that were discussed during the session, supporting FAIR data requires many different components and considerations. For FAIR data support to be effective, it is necessary not only to embed the FAIR principles in research education and practice but also to reward and recognise those practicing FAIR. Showcasing role models and examples where the FAIR principles have been implemented is important as those would encourage others to practise FAIR. Finally, instead of seeing practising FAIR as a binary state, it is important to focus on incremental steps researchers can apply in their practise to make their data as FAIR as possible.

Instead of seeing practising FAIR as a binary state, it is important to focus on incremental steps researchers can apply in their practice to make their data as FAIR as possible

Author Contributions
All authors are listed alphabetically. Contributions, in accordance with the CRediT taxonomy:

Author	Role
Louise Chisholm	Conceptualisation (equal), Writing – original draft (equal)
Maria Cruz	Conceptualisation (equal), Writing – review & editing (equal)
Rebecca Grant	Conceptualisation (equal), Writing – original draft (equal)
Elizabeth Newbold	Conceptualisation (equal), Writing – review & editing (equal)
Yasemin Türkyilmaz-van der Velden	Conceptualisation (lead), Writing – original draft (equal), Writing – review & editing (supporting)

5. Fair access to research data infrastructure

Summary of discussion

Research data infrastructure is often thought of as technologies, but it also includes institutions, policies, legislation and human resources. The capacities of different governments to build and maintain this data infrastructure varies widely, meaning that access to infrastructure is unequally distributed. In turn, this implies that the benefits of open data are also unequally distributed across countries.

A clear example of these inequities is the geographical distribution of certified research data repositories with either the World Data System certification, the Data Seal of Approval, or the CoreTrustSeal. While the Global North has multiple repositories in a number of regions that meet these criteria, there is just one on the entire African continent.

Many low-income countries (LICs) and low-and-middle-income countries (LMICs) have collected quite substantial amounts of data, but these datasets are not generally in the public domain or even the research domain. There are many reasons for this, including the absence of institution and country-level policies for data exchange, limited funding, and skills and technology restrictions.

So, what data governance issues are relevant to data policy-making in LICs and LMICs? What institutional infrastructure could benefit research data sharing in these countries? And what are the main technological challenges for better data management in these countries?

Sustainability and funding

There was consensus among participants in the discussion that funding was, without question, one of the main barriers to equitable access to research data infrastructure. Without appropriate funding, setting up data policies, standards, and technologies becomes near-impossible, thereby creating inequality of access.

However, hand-in-hand with funding comes the issue of sustainability, which was raised during the discussion several times. Participants highlighted the fact that funding to set up infrastructure is not enough. There needs to be support and funding available to ensure the uptake, use, and long-term management of any new infrastructure. It was felt that sustainability was often lacking in current grant schemes.

Applying standards across countries and disciplines

There was much discussion about the need to create appropriate data standards – ones that would work both across disciplines and countries. One of the key challenges flagged in this area was language barriers and how to overcome these to ensure people had equal access to, and understanding of, shared data.

Another key challenge identified was whether it was possible to balance the need for standards that are general enough to be used across disciplines, against ensuring the

data had enough specificity to be useful to others. Participants were split as to how feasible this was, with some feeling that general standards such as metadata standards could be easy to apply across various disciplines.

One of the problems with applying standards across disciplines was felt to be that many disciplines had already created their own standards. This, in turn, has led to different research communities favouring certain standards and being less open to the idea of applying a more general standard that could cross disciplinary boundaries.

There was general agreement among participants that solutions should centre around FAIR models of data governance and that the aim should be to see the FAIR principles universally applied. While constraints may be localised, the FAIR principles were seen to present a solution which has been built to be region and discipline agnostic.

A culture shift is required

The issue of cultural attitudes among researchers towards open and FAIR data was also raised as part of the discussion.

More than one participant expressed their concerns that while funding was an important requirement to put the right data infrastructure in place, there was also work to do to change attitudes in order to ensure new infrastructure was used. It was felt there needed to be a belief among the research communities in different countries that there was real value in sharing data.

Participants agreed it would be important to embed data sharing practices among students and early career researchers in order for them to become a standard part of the research process.

The power of funders and publishers

The idea of impressing the value of data sharing on researchers led participants to the subject of funders and publishers. All felt that these groups – funders, in particular – have the power to influence the value placed on data sharing by the research community.

The example was given of one funder implementing new requirements around sharing data for grantees and how this had led to a significant surge in interest in data sharing among researchers in that region.

Technology and skills constraints

When it came to the specific issues faced by LMICs and LICs around technology, it was felt there were constraints in a number of areas – not least, internet access and even access to electricity.

Data science skills development was also felt to be an urgent need, in order to ensure that researchers could make use of technology when it was available.

There were also felt to be opportunities, however. For example, because of the current lack of established trustworthy data repositories, there are no legacy systems to retrofit to meet the FAIR principles. And it was felt that establishing certified local repositories could encourage risk-averse research teams to deposit data for sharing.

Ultimately, it was agreed that the challenges around technology and skills development inevitably came back to the problems raised earlier in the discussion around funding and sustainability.

One of the problems with applying standards across disciplines was felt to be that many disciplines had already created their own standards. This, in turn, has led to different research communities favouring certain standards and being less open to the idea of applying a more general standard that could cross disciplinary boundaries

References

Adams, B., 2020. Life science companies combine to form COVID-19 research database. Fierce Biotech. URL https://www.fiercebiotech.com/biotech/life-science-companies-combine-to-form-covid-19-research-database (accessed 15.03.21)

ALLEA, 2017. The European Code of Conduct for Research Integrity. ALLEA All European Academies, Berlin, Germany. URL https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf (accessed 24.03.21)

Arts, D. 2019. FAIR Data: 7 initiatives you should know about. URL https://researchdata.springernature.com/posts/51916-fair-data-7-initiatives-you-should-know-about# (accessed 24.03.21)

Beaulieu-Jones, B. K. et al., 2019. Privacy-Preserving Generative Deep Neural Networks Support Clinical Data Sharing. Circ: Cardiovascular Quality and Outcomes. 12. https://doi.org/10.1161/CIRCOUTCOMES.118.005122

Beskow, L., Dombeck, C., Thompson, C. et al., 2015. Informed consent for biobanking: consensus-based guidelines for adequate comprehension. Genet Med 17, 226–233. https://doi.org/10.1038/gim.2014.102

Biobank. Laws and regulation. URL https://www.biopankki.fi/en/laws-and-regulation/ (accessed 12.03.21)

Bolislis, W.R., Fay, M., Kühler, T.C., 2020. Use of Real-world Data for New Drug Applications and Line Extensions. Clin Ther. May; 42(5):926-938. https://doi.org/10.1016/j.clinthera.2020.03.006

Burnham, J. C., 1990. The Evolution of Editorial Peer Review. JAMA. 263, 1323. doi:10.1001/jama.1990.03440100023003

Burton, J., Bhattacharya, S., Romero, K. & Conrado, D. J., 2020. Open Data for Clinical Pharmacology. Clin. Pharmacol. Ther. 107, 703–706. https://doi.org/10.1002/cpt.1779

Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X. & Greene, C. S., 2020. Responsible, practical genomic data sharing that accelerates research. Nat Rev Genet 21. 615–629. https://doi.org/10.1038/s41576-020-0257-5

Calling all coronavirus researchers: keep sharing, stay open. 2020. Nature. 578, 7. https://doi.org/10.1038/d41586-020-00307-x

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M., 2020. The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), p.43. http://doi.org/10.5334/dsj-2020-043

Castro, C. F. de, Quintana, A. M., Olesiak, L. da R. & München, M. A. B., 2020. Termo de consentimento livre e esclarecido na assistência à saúde. Rev. Bioét. 28, 522–530. https://doi.org/10.1590/1983-80422020283416

Center for Open Science (COS). Open Science Badges enhance openness, a core value of scientific practice. URL https://www.cos.io/initiatives/badges (accessed 09.03.21)

Chen, A.T., Altschuler, K., Zhan, S.H., Chan, Y.A., Deverman, B.E., 2021. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. eLife. 2021;10:e63409. https://doi.org/10.7554/eLife.63409

Chodick, G., Tene, L., Patalon, T., Gazit, S., Ben Tov, A., Cohen, D., Muhsen, K., 2021. The effectiveness of the first dose of BNT162b2 vaccine in reducing SARS-CoV-2 infection 13-24 days after immunization: real-world evidence. medRxiv. 2021.01.27.21250612. https://doi.org/10.1101/2021.01.27.21250612

Clare, C., 2019. The Real World of Research Data. Zenodo. http://doi.org/10.5281/zenodo.3584373

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., McGillivray, B., 2020. The citation advantage of linking publications to research data. PLOS One. https://doi.org/10.1371/journal.pone.0230416

Cooper, D. & Springer, R., 2019. Data communities: a new model for supporting STEM data sharing. APO. https://doi.org/10.18665/sr.311396

Cousijn, H. et al., 2018. A data citation roadmap for scientific publishers. Sci Data. 5, 180259. https://doi.org/10.1038/sdata.2018.259

De Rijksoverheid. Voor Nederland. 2019. Adviesrapport secundair gebruik van data in de geboortezorg: een schets van contouren en optimalisatie van een datalandschap. URL https://www.rijksoverheid.nl/documenten/rapporten/2019/06/30/adviesrapport-secundair-gebruik-van-data-in-de-geboortezorg (accessed 24.03.21)

Declaration on Research Assessment (DORA). Reimagining academic assessment: stories of innovation and change. URL https://sfdora.org/dora-case-studies/ (accessed 19.02.21)

Declaration on Research Assessment (DORA). 2020. Case Study: University College London. URL https://sfdora.org/case-study/university-college-london/ (accessed 12.02.21)

Digital Science, Hahnel, M., McIntosh, B., Borrelli., L., Hyndman, A., Baynes, G., Crosas, M., et al., 2020. The State of Open Data 2020. figshare. https://doi.org/10.6084/m9.figshare.13227875.v2

Directorate General for Research and Innovation (European Commission). 2016. H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. URL https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 19.02.21)

Directorate General for Research and Innovation (European Commission). 2018. Turning FAIR into reality. https://doi.org/10.2777/1524

Dolgin, E., 2020. Core Concept: The pandemic is prompting widespread use-and misuse-of real-world data. Proc Natl Acad Sci U S A. Nov 10;117(45):27754-27758. https://www.pnas.org/content/117/45/27754

Dutch Research Council (NWO). Open Science Fund. URL https://www.nwo.nl/en/researchprogrammes/open-science/open-science-fund (accessed 17.02.21)

EconOmics. 2013. Nat Genet. 45, 715. https://doi.org/10.1038/ng.2688

EPFL. EPFL Data Champions. URL https://www.epfl.ch/campus/library/services/services-researchers/rdm-contacts-communities/epfl-data-champions/ (accessed 19.02.21)

European Commission. 2016. G20 Leaders' Communique Hangzhou Summit. URL https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967 (accessed 17.03.21)

European Commission. 2020. Coronavirus: EU interoperability gateway goes live, first contact tracing and warning apps linked to the system. URL https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1904 (accessed 15.03.21)

FAO. 2021. AGROVOC – Semantic data interoperability on food and agriculture. https://doi.org/10.4060/cb2838en

FAIRsFAIR. European Group of FAIR Champions. URL https://www.fairsfair.eu/advisory-board/egfc (accessed 19.02.21)

Finke, B., 2021. Ursula von der Leyen im Gespräch. "Das hätten wir früher machen können." Süddeutsche Zeitung. URL https://www.sueddeutsche.de/politik/von-der-leyen-interview-impfstoff-1.5196520 (accessed 15.03.21)

FinnGen. Questions and answers. URL https://www.finngen.fi/en/questions-and-answers (accessed 08.03.21)

Franklin, J.M., Lin, K.J., Gatto, N., Rassen, J., Glynn, R.J., Schneeweiss, S., 2021. Real-world evidence for assessing pharmaceutical treatments in the context of COVID-19. Clin Pharmacol Ther. Feb 2. https://doi.org/10.1002/cpt.2185

Gadgeel, S.M., Thompson, M.A., Izano, M.A., Hwang, C., Mikkelsen, T., Weese, J.L., Wolf, F.M., Schrag, A., Walters, S., Singh, H., Hirsch, J., Brown, T.D., Kluetz, P.G., 2020. Abstract S10-02: Using real-world data (RWD) from an integrated platform for rapid analysis of patients with cancer with and without COVID-19 across distinct health systems. AACR. doi: 10.1158/1557-3265.COVID-19-S10-02

Global Research Collaboration for Infectious Disease Preparedness (GloPID-R). Our Preparedness Work. URL: https://www.glopid-r.org/our-work/data-sharing/ (accessed 05.03.21)

Glonti, K. et al., 2019. A scoping review on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. BMC Medicine. 17, 118. https://doi.org/10.1186/s12916-019-1347-0

Halpin, D.M.G., Singh, D., Hadfield, R.M., 2020. Inhaled corticosteroids and COVID-19: a systematic review and clinical perspective. Eur Respir J. May 7;55(5):2001009. https://doi.org/10.1183/13993003.01009-2020

Helliwell, J.A., Bolton, W.S., Burke, J.R., Tiernan, J.P., Jayne, D.G., Chapman, S.J., 2020. Global academic response to COVID-19: Cross-sectional study. Learned Publishing. https://doi.org/10.1002/leap.1317

Hidden REF, The. URL https://hidden-ref.org/ (accessed 11.02.21)

Hughes, R., Whitley, L., Fitovski, K., Schneble, H.M., Muros, E., Sauter, A., Craveiro, L., Dillon, P., Bonati, U., Jessop, N., Pedotti, R., Koendgen, H., 2020. COVID-19 in ocrelizumab-treated people with multiple sclerosis. Mult Scler Relat Disord. Dec 30;49:102725. https://dx.doi.org/10.1016%2Fj.msard.2020.102725

Jetten, M., Grootveld, M., Mordant, A., Jansen, M., Bloemers, M., Miedema, M., & van Gelder, C.W.G., 2021. Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship. Zenodo. http://doi.org/10.5281/zenodo.4486423

Kalkman, S., Mostert, M., Gerlinger, C., van Delden, J. J. M. & van Thiel, G. J. M. W., 2019. Responsible data sharing in international health research: a systematic review of principles and norms. BMC Med Ethics. 20, 21. https://doi.org/10.1186/s12910-019-0359-9

Khan, N., Pink, C. J. & Thelwall, M. 2020. Identifying Data Sharing and Reuse with Scholix: Potentials and Limitations. Patterns. 1, 100007. https://doi.org/10.1016/j.patter.2020.100007

Krypotos, A.-M., Klugkist, I., Mertens, G. & Engelhard, I. M., 2019. A step-by-step guide on preregistration and effective data sharing for psychopathology research. Journal of Abnormal Psychology. 128, 517–527. https://doi.org/10.1037/abn0000424

Laine, H., 2017. Afraid of Scooping – Case Study on Researcher Strategies against Fear of Scooping in the Context of Open Science. Data Science Journal. 16, p.29. http://doi.org/10.5334/dsj-2017-029

Lin, D., Crabtree, J., Dillo, I. et al. 2020. The TRUST Principles for digital repositories. Sci Data 7, 144. https://doi.org/10.1038/s41597-020-0486-7

Littler, K., et al., 2017. Progress in promoting data sharing in public health emergencies. Bulletin of the World Health Organization. 95, 243 – 243. http://dx.doi.org/10.2471/BLT.17.192096

Luo, Y., Jin, H. & Li, P. A., 2019. Blockchain Future for Secure Clinical Data Sharing: A Position Paper. Proceedings of the ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization - SDN-NFVSec.19 23–27. doi:10.1145/3309194.3309198

Maguire, B., McLean, A., Rashan, S., et al., 2020. Baseline results of a living systematic review for COVID-19 clinical trial registrations. Wellcome Open Res. 5:116. https://doi.org/10.12688/wellcomeopenres.15933.1

Makady, A., de Boer, A., Hillege, H., Klungel, O., Goettsch, W., (on behalf of GetReal Work Package 1), 2017. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. Value Health. Jul-Aug; 20 (7) :858-865. https://doi.org/10.1016/j.jval.2017.03.008

Markowetz, F., 2015. Five selfish reasons to work reproducibly. Genome Biol. 16, 274. https://doi.org/10.1186/s13059-015-0850-7

Martinez, P.A., Erdmann, C., Simons, N., Reid, O., Labou, S., Johnson, R., … Fankhauser, E., 2019. Top 10 FAIR Data & Software Things. Zenodo. http://doi.org/10.5281/zenodo.3409968

Maxmen, A., 2021. Massive Google-funded COVID database will track variants and immunity. Nature. https://doi.org/10.1038/d41586-021-00490-5

McDermott, J. & Hardeman, M. 2018. Increasing Your Research's Exposure on Figshare Using the FAIR Data Principles. figshare. https://doi.org/10.6084/m9.figshare.7429559.v2

Mello, L.E., Suman, A., Medeiros, C.B., Prado, C.A., Rizzatti, E.G., Nunes, F.L.S., … Carvalho, R. 2020. Opening Brazilian COVID-19 patient data to support world research on pandemics. Zenodo. http://doi.org/10.5281/zenodo.3966427

Ministério da Saúde. URL https://bvsms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html (accessed 12.03.21)

Modjarrad, K., Moorthy, VS., Millett, P., Gsell, PS., Roth, C., et al., 2016. Developing Global Norms for Sharing Data and Results during Public Health Emergencies. PLOS Medicine. 13(1): e1001935. https://doi.org/10.1371/journal.pmed.1001935

Mons, B., Neylon, C., Velterop, J., et al., 2017. Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud. Information Services & Use. 49 – 56. https://doi.org/10.3233/ISU-170824

Moorthy, V., Henao Restrepo, A. M., Preziosi, M.-P. & Swaminathan, S., 2020. Data sharing for novel coronavirus (COVID-19). Bull. World Health Organ. 98, 150–150. http://dx.doi.org/10.2471/BLT.20.251561

Narozniak, R., 2020. COVID-19 Pandemic Highlights the Importance of Real World Data in Oncology. Oncology Nursing News. URL https://www.oncnursingnews.com/view/covid19-pandemic-highlights-the-importance-of-real-world-data-in-oncology (accessed 12.03.21)

National Center for Biotechnology Information (NCBI). Public Health Emergency COVID-19 Initiative. URL https://www.ncbi.nlm.nih.gov/pmc/about/covid-19/ (accessed 05.03.21)

Network of the National Library of Medicine (NNLM). Data Life Cycles. URL https://nnlm.gov/data/data-life-cycles (accessed 19.02.21)

Norton, A., Bucher, A., Antonio, E., et al., 2020. A living mapping review for COVID-19 funded research projects: three-month update [version 2; peer review: 2 approved]. Wellcome Open Res. 5:209. https://doi.org/10.12688/wellcomeopenres.16259.2

Open Science Community Delft (OSCD). OSC Delft Initiatives. URL https://osc-delft.github.io/initiatives (accessed 19.02.21)

Pearce, R., & Grant, R., 2020. Do Open Data Badges Influence Author Behaviour? A Case Study at Springer Nature. International Journal of Digital Curation. https://doi.org/10.2218/ijdc.v15i1.694

Pisani, E., Ghataure, A., Merson, L., 2018. Data sharing in public health emergencies: A study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats. Wellcome Trust. https://doi.org/10.6084/m9.figshare.5897608.v1

Plomp, E., Dintzner, N., Teperek, M., & Dunning, A., 2019. Cultural Obstacles to Research Data Management and Sharing at TU Delft. Insights. 32 (1): 29. http://doi.org/10.1629/uksg.484

PRISM network. About PRISMs. URL https://www.pris-managers.ac.uk/ (accessed 11.02.21)

Psychotherapy centre's database hacked, patient info held ransom. 2020. YLE News. URL https://yle.fi/uutiset/osasto/news/psychotherapy_centres_database_hacked_patient_info_held_ransom/11605460 (accessed 24.03.21)

Publons. URL http://publons.com (accessed 12.03.21)

Research Data Alliance (RDA). 2020. The final version of the RDA COVID-19 Recommendations and Guidelines for Data Sharing, published 30 June 2020. URL https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-1 (accessed 05.03.21)

Research Data Alliance (RDA). Professionalising Data Stewardship IG. URL https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig (accessed 11.02.21)

Resnik, D.B. J.D. Ph.D. & Shamoo, A.E. Ph.D., 2011. The Singapore Statement on Research Integrity. Accountability in Research. 18:2, 71-75. https://doi.org/10.1080/08989621.2011.557296

Rho, Y., Cho, D.Y., Son, Y., Lee, Y.J., Kim, J.W., Lee, H.J., You, S.C., Park, R.W., Lee, J.Y., 2021. COVID-19 International Collaborative Research by the Health Insurance Review and Assessment Service Using Its Nationwide Real-world Data: Database, Outcomes, and Implications. J Prev Med Public Health. 54(1):8-16. https://doi.org/10.3961/jpmph.20.616

Richter, G. et al., 2019. Patient views on research use of clinical data without consent: Legal, but also acceptable? Eur J Hum Genet. 27, 841–847. https://doi.org/10.1038/s41431-019-0340-6

Riggs, E.R., Azzariti, D.R., Niehaus, A. et al. 2019. Development of a consent resource for genomic data sharing in the clinical setting. Genet Med 21, 81–88. https://doi.org/10.1038/s41436-018-0017-5

Rocher, L., Hendrickx, J.M. & de Montjoye, YA. 2019. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 10, 3069. https://doi.org/10.1038/s41467-019-10933-3

Rogers, J.R., Lee, J., Zhou, Z., Cheung, Y.K., Hripcsak, G., Weng, C., 2021. Contemporary use of real-world data for clinical trial conduct in the United States: a scoping review. J Am Med Inform Assoc. Jan 15;28(1):144-154. https://doi.org/10.1093/jamia/ocaa224

Ross, L., 2020. At Mayo Clinic, sharing patient data with companies fuels AI innovation — and concerns about consent. STAT. URL https://www.statnews.com/2020/06/03/mayo-clinic-patient-data-fuels-artificial-intelligence-consent-concerns/ (accessed 08.03.21)

Saxén, H., 2020. Asiantuntijanäkemys biopankkiluonnoksen kohdista bioeettisestä näkökulmasta sosiaali- ja terveysministeriölle. URL https://stm.fi/documents/1271139/2013549/Asiantuntijanäkemys+biopankkilakiluonnoksen+kohdista+bioeettisestä+näkökulmasta/f5e8ce26-e137-bd8b-2041-d0af871ac640/Asiantuntijanäkemys+biopankkilakiluonnoksen+kohdista+bioeettisestä+näkökulmasta.pdf (accessed 24.03.21)

Scientific Data. For Referees. URL https://www.nature.com/sdata/policies/for-referees (accessed 12.03.21)

Schwab, P., Mehrjou, A., Parbhoo, S., Celi, L.A., Hetzel, J., Hofer, M., Schölkopf, B., Bauer, S., 2012. Real-time prediction of COVID-19 related mortality using electronic health records. Nat Commun. Feb 16;12(1):1058. https://doi.org/10.1038/s41467-020-20816-7

Shahin, M. H., et al., 2020. Open Data Revolution in Clinical Research: Opportunities and Challenges. Clin Transl Sci. 13, 665–674. https://doi.org/10.1111/cts.12756

Shu Y., McCauley, J., 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 30;22(13):30494. https://dx.doi.org/10.2807%2F1560-7917.ES.2017.22.13.30494

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., et al., 2018. Whitepaper: Practical challenges for researchers in data sharing. figshare. https://doi.org/10.6084/m9.figshare.5975011.v1

Sutter, S., 2020. Real-World Evidence On COVID-19: US FDA Approaching With 'Sense Of Urgency.' Pink Sheet. URL https://pink.pharmaintelligence.informa.com/PS142063/RealWorld-Evidence-On-COVID19-US-FDA-Approaching-With-Sense-Of-Urgency (accessed 15.03.21)

Tenopir, C., et al., 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE. 6, e21101. https://doi.org/10.1371/journal.pone.0021101

Tenopir, C. et al., 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLoS ONE. 10, e0134826. https://doi.org/10.1371/journal.pone.0134826

Tenopir, C. et al., 2020. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. PLoS ONE. 15, e0229003. https://doi.org/10.1371/journal.pone.0134826

Teperek, M., et al., 2018. Data Stewardship Addressing Disciplinary Data Management Needs. International Journal of Digital Curation. 13, no. 1: 141–49. https://doi.org/10.2218/ijdc.v13i1.604

Transparency during global health emergencies. 2020. The Lancet Digital Health. Sep;2(9):e441. https://doi.org/10.1016/S2589-7500(20)30198-9

Tse, E.G., Klug, D.M., Todd, M.H., 2020. Open science approaches to COVID-19. F1000Research. 9:1043. https://doi.org/10.12688/f1000research.26084.1

University of Cambridge. Data Champions. URL https://www.data.cam.ac.uk/intro-data-champions (accessed 19.02.21)

University of Melbourne, The. Data Champions Network. URL https://library.unimelb.edu.au/digital-stewardship/data-champions (accessed 19.02.21)

Van Noorden, R., 2021. Scientists call for fully open sharing of coronavirus genome data. Nature. 590, 195-196. https://doi.org/10.1038/d41586-021-00305-7

Velazquez A., Bustria M., Ouyang, Y., Moshiri, N., 2020. An analysis of clinical and geographical metadata of over 75,000 records in the GISAID COVID-19 database. medRxiv. 2020.09.22.20199497. doi: https://doi.org/10.1101/2020.09.22.20199497

Vines, T.H., et al., 2013. The Availability of Research Data Declines Rapidly with Article Age. Current Biology. 24, 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Vitolla, F., Raimo, N., Rubino, M. & Garegnani, G. M., 2021. Do cultural differences impact ethical issues? Exploring the relationship between national culture and quality of code of ethics. Journal of International Management. 27, 100823. https://doi.org/10.1016/j.intman.2021.100823

Wellcome Trust. Open Research Fund. URL https://wellcome.org/grant-funding/schemes/open-research-fund (accessed 17.02.21)

Wellcome Trust. 2020a. Publishers make coronavirus (COVID-19) content freely available and reusable. URL https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable (accessed 05.03.21)

Wellcome Trust. 2020b. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. URL https://wellcome.org/coronavirus-covid-19/open-data (accessed 05.03.21)

Wildgaard, L., Vlachos, E., Nondal, L., Larsen, A.V., & Svendsen, M., 2020. National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum) (Version 1). Zenodo. http://doi.org/10.5281/zenodo.3609516

Wilkinson, M., Dumontier, M., Aalbersberg, I., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M., Sansone, S.A., Schultes, E. et al. 2018. A design framework and exemplar metrics for FAIRness. Sci Data 5, 180118. https://doi.org/10.1038/sdata.2018.118

Wilkinson, M.D., Dumontier, M., Sansone, S.A. et al., 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. Sci Data 6, 174. https://doi.org/10.1038/s41597-019-0184-5

World Economic Forum. 2020. Sharing Sensitive Health Data in a Federated Data Consortium Model: An Eight-Step Guide. URL https://www.weforum.org/reports/sharing-sensitive-health-data-in-a-federated-data-consortium-model-an-eight-step-guide (accessed 22.03.21)

World Health Organization (WHO). Developing global norms for sharing data and results during public health emergencies. URL http://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/ (accessed 05.03.21)

World Meteorological Organization (WMO). 2020. 20th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Measurement Techniques (GGMT-2019). URL https://library.wmo.int/index.php?id=21758&lvl=notice_display#.YFt3fa_7SUl (accessed 24.03.21)

Wright, L., 2021. 'The Plague Year', The New Yorker, pp20-59.

Xu, B., Gutierrez, B., Mekaru, S., et al., 2020. Epidemiological data from the COVID-19 outbreak, real-time case information. Sci Data. 7, 106. https://doi.org/10.1038/s41597-020-0448-0

How Springer Nature is supporting FAIR data

Data sharing helps speed up the pace of discovery and increases the benefit of research to society. We believe that research data should be FAIR, and in the words of the European Commission 'as open as possible, as closed as necessary'[107]. As a proactive partner of the research community, Springer Nature is committed to supporting researchers in sharing their data in line with the FAIR principles.

Research data training

Addressing the challenges researchers encounter around data sharing, these in-person and online workshops include an introduction to the practical applications of the FAIR data principles. Part of Nature Research Academies, our workshops offer trusted quality and a professional perspective on research.

Learn more: https://go.sn.pub/research-data-training

Research data help desk

Our expert team has put together a collection of free articles on the most commonly asked questions about research data.

Learn more: https://go.sn.pub/research-data-help-desk

Data publishing

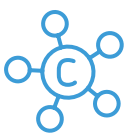
Data publishing provides an opportunity for researchers to increase the findability and accessibility of their datasets via a peer-reviewed publication.

Learn more: https://go.sn.pub/research-data-publishing

Recommended repositories list

Data which are formally archived in a repository are more likely to be accessible for the longer term[108]. They are also easier to find, to reuse and to cite. To help researchers select the right repository for their data, we have put together a list of recommended repositories.

Learn more: https://go.sn.pub/recommended-repositories



Research Data Community

The Research Data Community is a space for those working with research data to discuss ideas, find support, and share best practice.

Join now: researchdata.springernature.com

107. Directorate-General for Research and Innovation (European Commission) (2016)

108. Vines et al., (2013)

SPRINGER NATURE

springernature.com

Around our complex and interconnected world, the research community is advancing discovery for all of us. These illustrations celebrate some of the great minds who have helped advance discovery through history.

Antarctica meltdown could double sea level rise

Researchers at Pennsylvania State University have been considering how quickly a glacial ice melt in Antarctica would raise sea levels. By updating models with new discoveries and comparing them with past sea-level rise events they predict that a melting Antarctica could raise oceans by more than 3 feet by the end of the century if greenhouse gas emissions continued unabated, roughly doubling previous total sea-level rise estimates. Rising seas could put many of the world's coastlines underwater or at risk of flooding and storm surges.

Credit / Copyright: Shutterstock

For information: Source: Nature

Follow @SpringerNature

Acknowledgements

Fair access to research data infrastructure: **Lynn Woolfrey** (DataFirst, University of Cape Town)

The impact of the FAIR principles on a global pandemic: **Natalie Harrower** (Digital Repository of Ireland), **Kim Ingle** (SALDRU, University of Cape Town), **Tim Brophy** (SALDRU, University of Cape Town), **Juan Pane** (Centro de Desarrollo Sostenible).

Springer Nature: Mithu Lucraft, Isabel Roth, Katie Baker, Christina Emery, Vipin Kumar Singh

Arkbright: Kamal Arkinstall

Gee St Design: Rob Dicks