The Global Health Network

Webinar Report

"COVID-19 focused platforms and registries for sharing participant-level clinical-epidemiological, OMICs, and imaging data"

Contents

1. Introduction

On the 14th of July 2021 The Global Health Network supported the virtual webinar 'COVID-19 focused platforms and registries for sharing participant-level clinical-epidemiological, OMICs, and imaging data' in conjunction with TDR, the Special Programme for Research and Training in Tropical Diseases, and the COVID-19 Clinical Research Coalition.

In the research and public health response to COVID-19, there has been a rush to share data, marked by an explosion of population- and discipline-specific registries. Data sharing platforms, which generally predate the COVID-19 pandemic, have been expanded to include different types of COVID-19-related data.

In this webinar, speakers presented a comprehensive, living overview of COVID-19-related platforms and registries for sharing participant-level clinical-epidemiological, OMICs, and imaging data and reviewed the interoperability of data sharing efforts and how these initiatives map to best practice for

ethical, equitable, and effective data sharing and application of the FAIR Principles for managing data resources.

Chair:

- **Phaik Yeong Cheah** - Associate Professor at the University of Oxford and chair of the COVID-19 Clinical Research Coalition Data Sharing Working Group.

Speakers:

- **Lauren Maxwell**, PhD, MPH
  Senior Researcher at Heidelberg University's Institute for Global Health, TDR, and the WHO Department of Sexual and Reproductive Health and Research
- **Susanna-Assunta Sansone**, PhD
  Associate Director of the Oxford e-Research centre, and Associate Professor at the University of Oxford.
- **Fruzsina Molnar Gabor**, PhD
  Legal scientist and research group leader at the BioQuant Centre and the Academy of Sciences and Humanities, in Heidelberg, Germany
- **Abha Saxena**
  Former lead for Global Health Ethics at WHO, now an independent bioethics advisor to international research groups and Universities. Has a special interest in ethics of infectious disease outbreaks, new technologies and health systems research.
- **John Amuasi**
  Leader of the Global Health and ID Research team at the Kumasi Centre for Collaborative Research (KCCR) in Ghana.
- **Yo Yehudi**
  Open Source Technology Lead, Data for Science and Health
- **Jeremy Geelen**
  Senior advisor at the Canadian Institutes of Health Research (CIHR). Current chair of the GloPID-R Data Sharing WG

2. Registration

A total of 568 people registered for this event, from 81 countries; on the day 217 people attended the webinar:

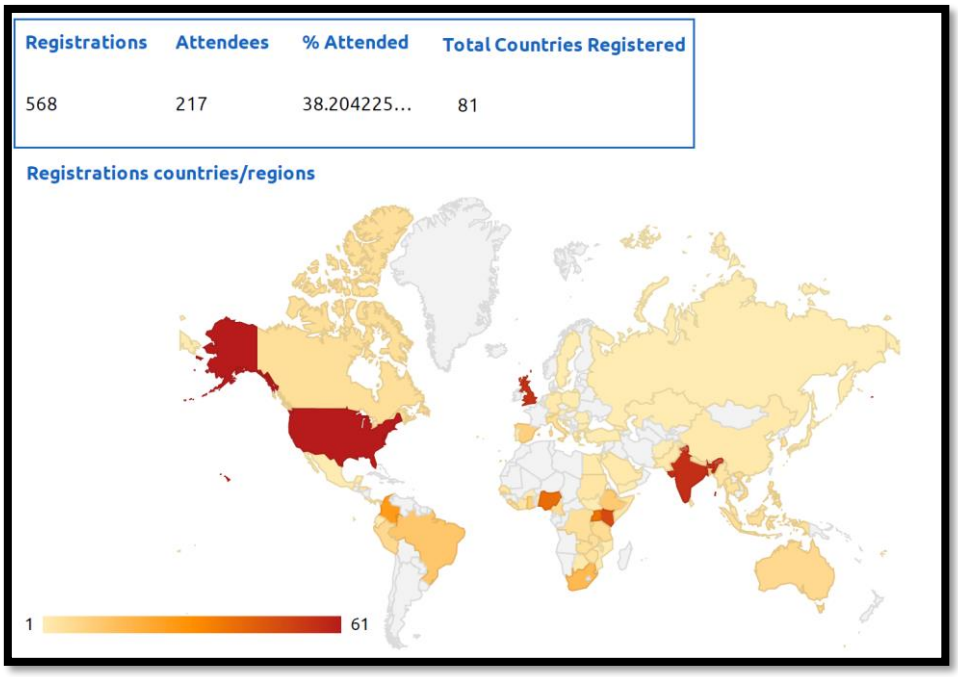

Figure 1. Heat map showing geographical distribution of registrants, covering 81 countries across the world. The scale bar shows how colour corresponds to number of registrants from each country.

When registering for the webinar, attendees were asked the question 'What do you see as the chief priority for improving data sharing in advance of the next epidemic?'. The 97 responses were categorised, and the following bar chart represents the number of responses in each category. The need for **Interoperability and Collaboration**, as well as **Accessibility** and **Capacity Building**, were particularly emphasised:

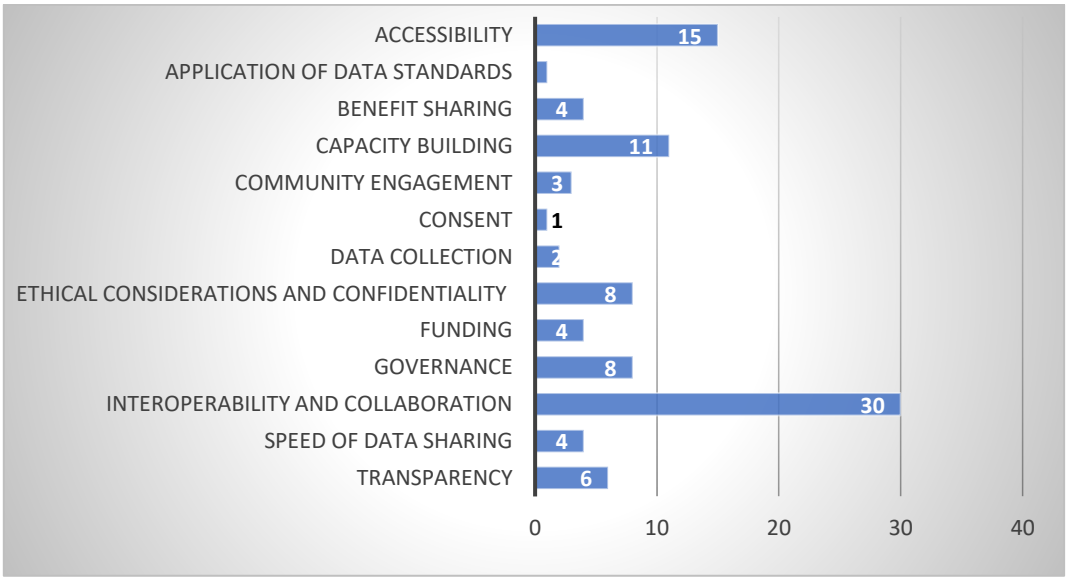

Figure 2. Bar chart categorising the 97 responses submitted by registrants to the question 'What do you see as the chief priority for improving data sharing in advance of the next epidemic?'

3. Summary of individual panellist presentations:

COVID-19-related Platforms & Registries

***Lauren Maxwell**, PhD, MPH*
Senior Researcher at Heidelberg University's Institute for Global Health, TDR, and the WHO Department of Sexual and Reproductive Health and Research

Work on the Platforms & Registries spreadsheet has been ongoing since the summer of 2020; the tool aims to protect, harmonise and share COVID-19 related participant level data. Key questions at the start of the project were: How are data being shared? What data are being shared? How FAIR are these resources? How do resources address legal & ethical concerns, benefit sharing & community engagement?

Monthly systematic searches were used to find English resources, and natural language processing to find non-English language resources. Using this method, they found 68 COVID-19 related platforms & registries and 11+ COVID-19 related catalogues of platforms or data lakes. Most repositories did not adhere well to the FAIR principles (Findable, Accessible, Interoperable, Reusable), many did not require patient consent, and few of the registries included benefit sharing with the contributing studies or engaged with the community from which the data was taken.

Data sharing efforts were often siloed by data type, comorbidity/coinfection, and by special populations. Regarding data type, most registries collected clinical or epidemiological data; the only 2 that did not collect this data type instead collected imaging data. Many platforms collected only human OMICs data or only pathogen OMICs data, while just 4 platforms collected data linking clinical and human OMICS data. Again, only 4 platforms collected data linking clinical, human OMICs and pathogen OMICs data, despite the importance of linking these data in the emergence of and reaction to variants of concern. Siloing by comorbidity/coinfection was especially common in those registries which primarily collected clinical or epidemiological data.

The numbers of registries siloed by the following comorbidities are shown in parentheses: cancer (6), blood conditions (3), cardiovascular system (3), skin conditions (7), rheumatic disease (3), digestive system (3), liver disease (2), Multiple Sclerosis (2). 6 registries siloed data by special population: 4 as a paediatric population and 2 as a pregnant population.

They next examined the governance of these platforms and registries. Most (12) platforms and 1 registry were open access, and many registries (16) and platforms (6) had a Data Access Committee. In 1 registry and 4 platforms the data providers themselves managed access to the data. Importantly, 17 registries did not have plans to share participant level data while the data access mechanism was unclear for 8 registries and 1 platform

With regards to benefit and credit sharing, most registries had data dashboards. These, help return some benefit to the contributing researchers and participating communities. Furthermore, most platforms required citation of the registry or data providers but only 8 registries and 4 platforms mentioned community engagement

The location of resources for sharing COVID-19 participant-level data located were investigated. The majority of resources for collecting and harmonising data are in Europe and North America. Most platforms deal with high dimensional data types (such as human or pathogen OMICs) and require larger investments: these generally predated the pandemic. On the other hand, most registries were created during the pandemic, and were generally limited to clinical data.

To assess the FAIRness of these resources, the following 4 criteria were considered:

- **Findable**: Have they been assigned a DOI?
- **Accessible**: Are their directions for how to acce3ss the data available on the website?
- **Interoperable**: Is the community-developed standard applied for the participant-level data or the metadata?
- **Reusable**: Do they have a data usage license or agreement in place for using that data?

This project helped assign a DOI to 19 registries and 12 platforms to help these resources become more FAIR. In total, 3 registries and 7 platforms adhered to all 4 FAIR criteria whereas 14 registries and 4 platforms adhered to none of the FAIR criteria. Those registries which collect data for specific comorbidities or primarily collect clinical or epidemiological data tended not to be very FAIR. There was a general trend in that those platforms that predated the pandemic and may be limited to data types but not to comorbidities tended to be more FAIR

The next steps: **challenges** identified and *approaches* that could be taken (these challenges will be addressed in more depth in a paper that will be written on this data):

- **Interoperability & accessibility of HER**
  - *Expand efforts to link EHR through shared standards*
  - *Make EHR accessible through shielded platform-based approaches*
- **Interoperability of EHR & research data**
  - *Apply standards to observational research that are closely related to or the same as EHR data standards (e.g. SNOMED, LOINC)*
- **Resources siloed by data type, comorbidity, body system**
  - *Address root causes of data silos (e.g., different teams, different storage, legal barriers, re-identification concerns, lack of consent)*
- **Interoperability of platforms & registries**
  - *Open access tools and guidance for meta harmonization across standards*
- **Interoperability of governance structures**
  - *Guidance, nuanced methods of exploring possibility of re-identification*
- **Not very FAIR data sharing resources**
  - *Focus on key components of FAIRness*
  - *Provide support, foster accountability*
- **Benefit sharing & community engagement**
  - *Provide support, foster accountability*
- **Competition between data sharing resources**
  - *Incentivize cooperation and collaboration between these platforms*

Data Policies and Data Sharing to Support COVID-19 Research: the role of FAIRsharing

***Susanna-Assunta Sansone**, PhD*
Associate Director of the Oxford e-Research centre, and Associate Professor at the University of Oxford.

The FAIR principles are a set of guidelines to ensure data is Findable, Accessible, Interoperable and Reusable. Repositories have a very important role in ensuring FAIR data sharing – they need to implement globally unique and persistent identifiers, richer metadata provenance, clear terms of access and use, community-defined terminologies and community-defined descriptive metadata.

FAIR data sharing is essential at all times, but especially during a pandemic: the OECD stressed the importance of having a network of interlinked repositories with compatible standards to enable FAIR data sharing. The FAIRsharing initiative aims to map the landscape of repository and platform data resources, understanding if and how they implement standards such as terminology, unique identifiers and common descriptors. It also examines the communication and interactions between different data repositories and platforms, as well as between these resources and data policies as well as community standards. FAIRsharing are therefore mapping the data sharing landscape and promoting the value of over 3,500 repositories, standards and policies. This helps both the *producer* of a resource (to make that resource more findable, adopted and cited) and the *consumer* of a resource (to access, select and use sharing resources with greater ease).

The following multistep approach is one example of how the interactions between these resources can be mapped.
First, you can look for relationships between repositories within the collection, then you can examine the relationships between resources in the collection with any repositories outside the collection. Next, you can then determine the relationships the repositories within the collection have with the standards they implement; you can also see which standards are shared by many repositories. You can also look at the policies that recommend the repositories; this tells you which repositories are well known to, and adopted by, the policy makers, and those resources which are not so well known to the policy makers.

Why are clearer and stronger data policies so important?
Firstly, they strengthen the Design and Access (DSA) statements to ensure access and reusability of the data (datasets, software, code and other digital resources) underlying COVID-19 research. Secondly, harmonisation of the data guidelines could ensure policies become clearer on how COVID-19 data should be made available via the appropriate repositories.
FAIRsharing is also involved in the COVID-19 Rapid Review initiative: This is an initiative that brings several open access publishers together with other scholarly organisations to ensure that both preprint papers and articles are rapidly reviewed It also has a working group focussing on data, which primarily focusses on strengthening the data availability statements of these publishers and journals, so that the data is FAIRly available in standards-implementing public repositories.

Data Protection Challenges & Solutions

***Fruzsina Molnar Gabor**, PhD*
Legal scientist and research group leader at the BioQuant Centre and the Academy of Sciences and Humanities, in Heidelberg, Germany

The General Data Protection Regulation (GDPR) laws in the European Union (EU) has been a subject of concern when it comes to international data sharing in the time of COVID-19 research. This presentation reflected on the legal implications of these GDPR laws in several steps in the data sharing pathways. While specific to the EU's GDPR laws, the points of the presentation were generally applicable to other data protection laws

The 'Territorial Applicability' of data protection laws is one of the primary questions: GDPR laws define a very broad scope of application, covering any personal data being processed by persons within the EU. GDPR also has a broad scope of applicability when it comes to exporting data outside of the EU, and thus has an 'extraterritorial effect'. Other data protection laws will also have such (possibly contradictory) extraterritorial effects; thus, it will be important to harmonise the communication of these applicable laws. There are also questions specific to platforms and registries – if data is accessed from one of these but is not physically moved then it may not technically involve international data transfer, thus laws such as GDPR may not be applicable.

The material scope of application of data protection laws is another concern. The legal context of data sharing will depend on the type and anonymity of the data. While personal data may not identify a particular person, if different data is also used, or if metadata is subtracted and combined, the data may become personally identifiable. The technical and legal terms of anonymisation must be differentiated – the legal term means that the data subject cannot be identified through reasonable means and does not refer simply to a more technical factual anonymity. This 'anonymity threshold', however, is difficult to define

The role of data subjects in the context of data protection laws is also evolving. Below are some examples of changes, and the effects that these are having. These changes include novel investigation methods such as genomics, new actors in international data sharing and research that may not have any personal connection to the data subjects, and the changing healthcare relevance on an individual level. Some effects that these changes may have include a 'justificatory role' for the data subject in which retrospective use of data may mean that the consent gathered from the patient is no longer relevant to what the data is being used for. Furthermore, an 'assessment/evaluation role' may be needed to ensure the data subjects continue to be allowed to assess and evaluate how their data is being used. Furthermore, a 'controlling role', such that the role of the data subjects in controlling the use of their data is maintained.

The role of data *users* is also changing. Those researchers responsible for using and sharing data within a research institution will increasingly find themselves in a position of responsibility. These researchers should include a data protection perspective when deciding the purpose of, and methods for, accessing and using the data.

As fields such as scientific research are given legal privileges when it comes to data sharing, it is important to ask how we define scientific research. Based on GDPR rules, any method-driven way of

deriving new knowledge counts as scientific research – thus these privileges currently apply to all kinds of scientific research (both commercial and otherwise). Thus, this definition seems to be sufficient for now.

Given that there are so many data protection laws, different legal positions and interests must be balanced when applying them to specific individual cases. Neither privacy nor scientific research should be guaranteed without limitations; these interests need to be balanced, with technical data security measures ensuring privacy but also enabling scientific research to be carried out. It is therefore important to be aware of the need to translate these legal data security measures into the act of legal and fair data interoperability. It is important to note that there is no standardised way of balancing rights and interest; this must be done on an individual case basis application

GDPR defines scientific research in the public interest, contributing to the legal privileges surrounding data use in health research. It is, of course, hard to define the public interest – it is a notion that needs to be assessed with social and cultural perspectives (proportionality). In the context of infectious diseases, however, it is clear that scientific research is in the public interest. Once the guiding principle of 'public interest' is defined, the definition can be used to define and apply the data protection regulations applying to relevant scientific research

Needs and further research:

- Sectoral rules on a legislative level and the level of law application
- Standardised weighting of privacy and scientific interests
- Safe data spaces – instead of moving data around, we could provide secure access to data sharing registries and platforms
- Technical and legal interoperability – these are strongly connected and should be developed in tandem
- Public interest – we could better define public interest and increase awareness and recognition of it
- Connection between research and healthcare could be improved on a public and individual level

Ethical Issues in Data Sharing

Abha Saxena

Former lead for Global Health Ethics at WHO, now an independent bioethics advisor to international research groups and Universities. Has a special interest in ethics of infectious disease outbreaks, new technologies and health systems research.

This presentation was about the ethical concerns surrounding platforms and registries sharing data ranging from clinical and epidemiological data to pathogen related data to genetic data. The ethical concerns are dependent on data type, the context in which it is collected, how it is collected and who it is collected by.

To whom do ethical obligations and responsibilities accrue? As well as the data collectors and the data users, other stakeholders also have ethical obligations. These include anyone with a vested interested in the data: policy makers, data repositories, funders, journal editors and Universities/research organisations. There therefore needs to be an effective ethics and governance structure which balances the competing needs of these stakeholders, although ethical guidelines for the development and use of data platforms do not yet exist.

A recent systematic review has identified four principles relevant to the ethical sharing of data:

- **Societal benefits and value** – this can be operationalised by implementation of both the FAIR principles, and of peer reviewing and scientific oversight to ensure quality and scientific validity
- **Distribution of risks, benefits and burden** – to individuals, to populations, for data collectors, for data users
- **Respect for individuals and groups** – Autonomy (consent), privacy, confidentiality and data security are all concerns that need to be addressed
- **Public trust, integrity, solidarity, transparency, accountability** – effective stakeholder engagement and governance systems are required to ensure data is being used in the right way, and is also recognised as being so

In this presentation, 3 of the more 'problematic' points were focussed on

Accessibility & Benefit Sharing: As one of the FAIR principles, data accessibility is vital for equitable data sharing. Accessibility refers to the ability to gain access to data within a searchable database that arises from standardised data sharing procedures. Given that global databases tend to establish connections with LMICs, access issues are related to both the global databases' access to data originating from LMICs, but also access of researchers in LMICs to the data within the global databases. While both parties may suffer as a consequence, the relationship is often skewed in the direction of power and/or resource imbalances. From an ethical perspective, benefits should be put in place in contexts in which such a skewed relationship exists, ensuring that data providers are afforded appropriate recognition/attribution in publication of research conducted by researchers in LMICs who provide data to global registries. Furthermore, ownership rights to the data should be drawn up and upheld. The concept of 'Real' capacity building is also important – capacity is often raised only for collecting and sharing data, but not for using this data; accessibility of data not only means that data is available, but that those that access it have the capacity to be able to understand and use the data, using appropriate software and digital tools; they should also have the ability to use their findings to inform development of diagnostic tools, therapeutics, vaccines etc. Funders, sponsors and national governments will play an important role in developing such real capacity in LMICs

Consent: Appropriate consent responds to the principle of autonomy as well as engendering trust and trustworthiness, but issues surrounding consent are one of the more prominent obstacles in data sharing. Researchers need to think about how the potential future use of data for public health purposes could be used, and should modify consent processes accordingly. Furthermore, ethics committees should familiarise themselves with such considerations, ensuring that they are somewhat flexible and allow the broad use of data in consent processes. More generally, a broad and continuing stakeholder engagement prior to data collection can aid the development of appropriate consent mechanisms and engender trust of researchers in ethical committees and in the communities

Governance Systems: Data registries and platforms should have an explicit, well-defined governance system (that incorporates ethics committees and is informed by stakeholder engagement) in place that promotes integrity, solidarity, accountability and transparency. Such a governance system could outline responsibilities of various partners, establish measures of accountability, monitor and audit the registry or platform and ensure compliance with existing ethical and legal requirements, as well

as with collaborators' ethical arrangements. Furthermore, a Governance Committee could oversee policy development and implementation.

Research Ethics Committees (RECs) play a vital role in maintaining ethical data sharing standards, although they do have limitations. RECs need to be open and flexible, and aware of the current research architecture; they should also be willing to understand and implement consent models that respond to the needs of researchers and further the public health mandate of research. One concern is the capacity of RECs: while this may limit them in some contexts, at a *minimum* they should have the capacity to ask the right questions and be willing to refer any issues to data governance committees

ALERRT's COVID-19 Clinical Characterisation Protocol and data sharing

John Amuasi

Leader of the Global Health and ID Research team at the Kumasi Centre for Collaborative Research (KCCR) in Ghana.

Clinical Characterisation Protocols (CCPs) are so important because they ensure data collection is standardised, allowing us to interconnect data coming from different sources. Once data is interconnect, we can have comparable data pools and much larger datasets rather than many smaller pools of data that cannot be compared. These larger datasets allow much more impactful conclusions to be drawn from the data, regarding diagnostic tools, treatment options, disease characterisation etc, depending on the type of data

ALERRT – COVID-19 CPP implementation in Africa

ALERRT (the African coaLition for Epidemic Research, Response and Training) has partners across all parts of Africa, and has been coordinating regional responses to the COVID-19 pandemic. ALERRT has been running a customised CPP to suit the disease and research context in Africa to collect standardised data in many different African countries. Tools are freely available on the TGHN ALERRT hub.

Given that the clinical presentation of COVID-19 is constantly evolving and changing, it is important that research is done in all affected areas, as conclusions drawn from data from one continent may not directly translate to another. Flexibility in the ALERRT CPP protocol allows different sites and countries to collect different levels of information from basic sociodemographic patient information to signs and symptoms, to laboratory-derived biological data from blood/urine samples. Within this framework, studies looking at coinfections/efficacy of therapeutic agents such as steroids/long COVID can be nested.

For the CCP data collection, researchers collect data using either paper or electronic Case Report Forms (p/eCRFs). If pCRFs are used, the responses are later converted into eCRF form, The eCRF responses are loaded onto a centralised or decentralised REDCap server (see below). Stakeholders other than the research team such as the ministry of health are then given access to the data as needed.

ALERRT CPP – current numbers of records by country:

- Senegal – 1432

- Cameroon – 409
- Ghana – 3360
- Uganda – 387
- Kenya, Guinea, DRC – data collection will start in September

While there are disparities between the numbers of records, the numbers do not show the type of records. In Cameroon, the records are very detailed, consisting of serially collected samples and follow ups of patients over a long period of time. In Ghana, most records contain basic clinical data collected from patients during their stay in a health facility. Thus, one challenge is harmonising the different level and quantity of data collected

Challenges:

Challenges faced when collecting and managing this data are listed below; most have arisen due to the sudden need for, and time pressured nature of, the research thanks to the abruptness of the onset of the pandemic

- Ethical clearance delays
- Implementation and training
- Political challenges
- Monitoring
- Resource mobilisation
- Tools and technology
- Quick and relevant outputs
- Race against time

In a pandemic context, elements which make data sharing easier may need to be traded off for flexibility. While ensuring research is carried out, challenges arise when data needs to be retrospectively harmonised and compared. In the ALERRT CCP, three options were presented to participants in the protocol:

- A fully centralized data management system where the REDCap was set up for them with their data going into a central repository, with externally set up security arrangements such as passwords.
- Building and managing their own database but benefitting from a central hosting option with access management to REDCap software being managed centrally.
- Building their own databases and managing it independently, receiving support (such as XML files or any other relevant information) when appropriate.

The challenge arising from this flexibility was making sure there is harmonization across all sites and data levels, enabling a meta-analysis of the data.

Another challenge, is *real* capacity building. Even between sites in Africa, let alone between Africa and Europe/America, there is not yet truly equitable data sharing. Supporting the skills that allow African researchers to make the most of the data is important. For example, ALERRT have an agreement with the Medical Research Council in Gambia that provides researchers with the skills needed to display CPP data graphically, enabling conclusions drawn from data to be shared with stakeholders and authorities within the country. This lets data-derived conclusions to directly influence policy.

Wellcome & COVID-19 Data Sharing

Yo Yehudi

Open Source Technology Lead, Data for Science and Health

Wellcome requires that all SARS-CoV-2 genomic data be shared publicly, and is co-deposited in the following 2 places to ensure accessibility is maximised: INSDC (International Nucleotide Sequence Database Collaboration) and GISAID repositories. Embargoes on COVID-19 data are mostly avoided, but have a maximum of 3 months if they are well-justified

Wellcome, along with Mastercard and the Bill & Melinda Gates Foundation among other funders, are supporting the COVID Therapeutics Accelerator: This aims to gather evidence for new and existing treatments for COVID-19. Those receiving grants are also given access to guidance in the form of Outputs Management Plans – these ensure data sharing is effective. For example, IDDO and Vivli have both been awarded grants by the CTA: these platforms are increasing the accessibility of their data, ensuring that it adheres to the FAIR principles and increases credit for the contributors of the data.

Data Sharing and COVID-19: A Funder Policy Perspective

Jeremy Geelen

Senior advisor at the Canadian Institutes of Health Research (CIHR). Current chair of the GloPID-R Data Sharing Working Group.

This presentation dealt with funders policies through the lens of the policies of the Canadian Institute of Health's research. From regional to national to international levels, the last few years has seen the publication of various frameworks concerning FAIR data sharing

The COVID-19 pandemic has stimulated the development of even more of these data sharing frameworks, and those developed during the pandemic tend to be more explicit. Examples include:

- The *Joint Statement on Sharing Research Data and Findings relevant to the COVID-19 Outbreak*: This builds on the *Statement on Data Sharing in Public Health Emergencies, 2016.* The Joint Statement is led by the Wellcome trust and now has over 160 signatories including publishers, institutions and other funders. It ensures that '*research findings and data relevant to this outbreak are shared rapidly and openly to inform the public health response and help save lives*'
- National Science and Technology Advisors – *Call for Open Access to COVID-19 Publications* (*and underlying data*): This statement is a call from technology advisers from 15 countries as well as the European Commission for all COVID-19 related publications to be open access immediately upon publication, as well as for all data to become openly available.

These guidelines have had a significant impact on research funder policy requirements; data and publications must be open access in most cases. As this is the first instance in which data sharing requirements have been so broad, such guidelines therefore raise novel questions for funders, including:

- Questions related to research practices
    - Have researchers complied with requirements?
    - Where are researchers sharing their data?
    - Are researchers publishing data according to the FAIR principles?

- Questions related to platforms and registries
    - How has the platform/registry landscape developed during COVID-19?
    - How to platforms and registries support access, interoperability, and governance?

TGHN's review of platforms and registries supporting COVID-19 data sharing allows such questions to be tackled

4. Summary of Q&A

Policies are focusing increasingly on sharing of code as well as data, but is the infrastructure available for code to be shared - do the resources analysed by Susanna and colleagues also enable code to be archived? *Yes, we also try to collect, describe and register resources that store code, software etc.* see https://fairsharing.org/search/?q=software
Data sharing has a vast level of approval and hence it's never easy to confidently share the data even though its necessary as a preventative approach *Yes, this is in fact the case, and researchers probably need more guidance on whether data sharing can happen, for which data and how.*
Susanna-Assunta could you expand more on your work to strengthen DAS please? *live answered. please, see more details at* https://oaspa.org/data-deposition-required-for-all-c19-rapid-review-publishers/
John - have you had any pushback at all from researchers who did not want to share data from their patients e.g. so they can publish their own research papers? Or are practical problems with data gathering and integration more important? '*This is something that we struggle with all the time. In a pandemic there is evolving knowledge and information. We agreed to share our data but conduct the data analysis in such a way that it does not neglect country-specific nuances and interests. This has meant that some level of dilution of the metadata presented at a 'macro level' has to happen. This kind of issue is certainly a problem; the earlier these concerns are discussed, the better. There is also the argument about who publishes first – the consortium, metadata publications or country specific publications? A timeline is one way forward – for examples both parties agree that if a country-specific publication is not published by a particular date then the consortium level data can be published*.'- **John**
How can we best ensure that structures are in place to gather data from infected persons in the community as well as hospitalised cases during the early phase of a future outbreak, wherever in the world it occurs? '*In general, during normal times without an ongoing pandemic it is important to have extensive stakeholder and public engagement to discuss the benefits of data sharing and how it can be useful. Secondly, engendering trust in the communities from which data is gathered is very important – by showing that the results and conclusions of the research carried out using their data comes back to help them these communities will be more willing to contribute.'* – **Saxena**

How may we add this data sharing and FAIR principles to protocols and part of ethics board reviews?

'*While this is partly a concern, many of the FAIR principles are technical issues and ethics committees may not have the capacity to review this; Data Access Committees or Data Review Committees should be established instead. Ethics boards should be aware of such principles and committees; they should be willing and confident in delegating such concerns to these more specialised committees*.' – **Saxena**

'*As well as what has already been said, prestructuring different data management plans to deal with these cases may be useful, as COVID-19 has confronted us with lots of data protection issues that could be prospectively clarified. More general issues around the FAIR principles and the law, ethics of data sharing and the law, whether ethical committees should have a say on data protection laws etc need more research; many such questions could be answered in a standardised manner with only small adaptations depending on the context*.' - **Fruzsina**

Can a researcher share the finding to the respondents after analysis?

it is complicated, because sharing the findings back with the participants conflicts with the principle of privacy and researchers should plan a-priori whether or not individual level findings will be shared.

What are some actionable items concerning the platforms/registries tool that attendees could take away from today's webinar?

'*One of the things we hope will happen is that the resource will generate more research. As Fruzsina was saying we could think more deeply about how GDPR and other data protection laws has led to this siloing of data types – how have these laws driven the separation of clinical data from OMICs data from imaging data etc. The resource could be used to pursue this research, ensuring we are ready for the next pandemic or wave of COVID-19*
Also, the interoperability of governance and legal frameworks is key – platforms and registries themselves have said that they need more information on what effective benefit sharing, governance and community engagement would look like. That is a conversation that we could have, using this resource, to develop community-driven ways forward in terms of governance and benefit sharing' – **Lauren**

Our desire in supporting this and other work around data sharing is to try to push through new solutions to the numerous challenges raised today. This work has shown that there is no shortage of sharing initiatives; we just need to maximise the utility of these platforms and registries. Greater interoperability of both the data and the governance processes stand out from today's session' **– Rob**

'We've talked a lot about the available data and the challenges there are. But when we do reuse data we need to make sure the data source is being cited, with people sharing the credit' **– Yo**

'To echo Rob's thoughts, the funders have a fundamental interest and understanding of the environment that they've helped develop with respect to data sharing; as well as improving the data sharing practices within the COVID-19 pandemic, this tool can be used to improve data sharing more generally.' **– Jeremy**

It's not just about data, but about the code, the software etc – on the FAIRsharing website we also have this type of registry; it's important that policies direct users to these as well as the data **- Susana**

5. Call to Action and Next steps

If you have enjoyed this webinar and would like more information on Data Sharing, The Global Health Network has a Data Sharing Course: Expectation from health research funders, regulatory agencies, and journals for sharing of de-identified individual-level health research data is now increasing. Arguments in favour of data sharing include maximising the utility of the data, improving research transparency and allowing confirmation of the interpretation of results, with the overall goal of improving science and health. However, the volume of data shared remains low. This has been partially attributed to lack of data management capacity and lack of knowledge of how and where to share data. This course aims to fill this gap by giving you basic and practical guidance in data management and sharing.
Each module has been written by an expert in the subject matter and has been peer-reviewed. This course is aimed at early career researchers, postgraduate students and clinical research support staff who are collecting, managing and using health data.

Over the next few days please send in your comments and feedback on this workshop. You can get in touch here: info@theglobalhealthnetwork.org

6. Registration and attendee report

Countries - Top 20

Registrations	Count	Attendees	Count
The United States	61	The United Kingdom	41
India	56	The United States	27
The United Kingdom	55	India	17
Kenya	49	Kenya	14
Nigeria	40	Uganda	13
Uganda	38	Nigeria	10
Colombia	27	Colombia	9
South Africa	17	South Africa	6
Brazil	13	Brazil	6
Ethiopia	12	Canada	4
Ghana	10	Vietnam	4
Spain	10	Tanzania	4
Peru	7	Indonesia	3
Australia	7	Bangladesh	3
Bangladesh	7	Belgium	3
Vietnam	6	Spain	3
Tanzania	6	Ghana	3
Sierra Leone	6	Australia	3
Cameroon	6	The Philippines	3
Italy	5	Myanmar	3

Attendee institution (top 20 recorded)

Academia (university, college, ...)	96
Government research organisation	16
Other research organisation	10
Hospital (Public)	10
Non-government organisation (NGO)	10
Other	8
Government Ministry	6
Hospital (Private)	6
Industry (including Pharma)	5
Consultancy	4
Self-employed	3
Commercial Research Organisation	3
Public Health institute	2
Academia (university, college, ...), Hospital (Public)	2
Government research organisation, Public Health institute	2
Community Health Centre/Facility	2
Hospital (Public), Other research organisation	2
Government Ministry, Hospital (Public)	1
International organisation (IGO), Other research organisation	1

Attendee job title (top 20 recorded)

Research investigator	29
Student	28
Member of a research team	26
Other	13
Academic (teacher in university or other institute of higher education)	12
Laboratory team	12
Working for a research funding organisation	8
Medical Doctor	8
Nurse / Midwife / public health professional	6
Working in research policy/policymaker	6
Other clinical role	5
Academic (teacher in university or other institute of higher education), Medic...	3
Member of a research team, Student	3
Member of a research team, Nurse / Midwife / public health professional	2
Research investigator, Medical Doctor	2
Member of a research team, Medical Doctor	2
Medical Doctor, Member of a research team	2
Academic (teacher in university or other institute of higher education), Resea...	2
Academic (teacher in university or other institute of higher education), Labora...	2

Upon registration, attendees were also asked if they had any questions for the panellists. While there was unfortunately not enough time to answer these during the webinar, many were addressed in the presentations or in the live Q&A.

What can funders of research do to make data sharing happen?
How do we find out more about funding to conduct research on these datasets?
how to incentivize academic researchers to share their data
What are best approaches to ensure respect for views of individuals whose data are shared by others?
What do you consider to be examples of best practice in access to clinical data?
Given all the nationalism seen with this pandemic and vaccine inequity, what is the benefit of sharing data for LMICs?
what are the updates currently on data sharing on COVID-19?
How can one handle a non-responding Collaborator after 90% work on data has been completed?
How do you think research can be promoted at the local, governmental, ministerial and public policy levels in organizations?
How are you going to ensure that process of data sharing will not be abused?
Where can I access data on relative humidity from the onset of the Pandemic?
Why did they decide to do this talk/webinar, was there a gap in data sharing organisations/countries
How should collaborating research teams deal with cross-border data sharing in the context of country-specific regulations?
Ethics and transparency of maintaining databases
According to the experience, what would be the best strategy to face the next pandemic?
To what extent does this covid variants affecting data collection. Has the vaccine affected data collected so far
How have various sectors managed to clean data despite data coming from different sources e.g. County and National?
What do you suggest to maintain a quality of data in developing countries like Ethiopia where data is collected mostly manually?
What barriers, if any, does information governance pose to data access?
How to bring legislature of different countries in the same platform?