



“Equitable data infrastructures to support equitable and effective pandemic intelligence”

Workshop organised by the Working Group: Collaborations and Outreach of the ITU-WHO Focus Group on Artificial intelligence for Health

**30 May 2022
13.00-15.00 (CET)**

SPEAKERS

Dr Nathalie Strub-Wourgaft: Director COVID-19 Response & Pandemic Preparedness at Drugs for Neglected Diseases initiative (DNDi); Coordinator of the ANTICOV study Consortium; one of the initiators, and a member of the Steering Committee, of the COVID-19 Clinical Research Coalition.

Dr Caesar Atuire: philosopher and bioethicist, University of Ghana, founding member of the Ghana Research Ethics Consortium, co-editor ‘Bioethics in Africa: theories and praxis’, member of the Planning Committee of the Global Forum on Bioethics in Research, 2020, and of the Ethics Working Group of the COVID-19 Clinical Research Coalition.

Professor Phaik Yeong Cheah: Professor of Global Health at the University of Oxford; founder and current head of Bioethics and Engagement at MORU Tropical Health Network, Bangkok. Chair of the Data Sharing Working Group and Member of the Ethics Working Group of the COVID-19 Clinical Research Coalition.

Dr Sumir Panji: Network (Bioinformatics) manager, H3ABioNet Pan African Bioinformatics Network for the Human Heredity and Health in Africa (H3Africa) consortium. Programme manager, eLwazi Open Data Science Platform. University of the Western Cape, South Africa

FORMAT:

80 minutes speaker presentations and Q&A, 40 minutes round-table and floor discussion.

REGISTRATION AND ACCESS TO THE MEETING

The most recent version of this file is [here](#).

To join the workshop, please register in advance [here](#).

This requires a free ITU user account available [here](#) (a user account is useful for getting access to material and events on the AI4H website and for AI4H to create a global community; [instructions](#)).

If you struggle, or know you will struggle, to get into the online workshop, please email andrew.farlow@ndm.ox.ac.uk in advance or humayra.bashir@ndm.ox.ac.uk and maryjane.nweje@ndm.ox.ac.uk on the day and we will send you a link to get in.

DRAFT CONCEPT NOTE

This is a work in progress to be refined after feedback. It is an early provisional framework to help guide a later consensus report. Please feel free to submit comments and suggest changes.¹ We plan a follow-on workshop in the next quarterly meeting of the Focus Group for AI for Health and a longer convening meeting in the spring of 2023 of a globally representative group, such that the May 2022 workshop will trigger a process that leads to a more substantial global consensus paper in spring 2023. Given the constraints of mostly-virtual meetings, we will prepare a meeting report after the 30 May workshop that blends PowerPoint material shared (permissions checked), meeting transcript, chat and other shared feedback, literature reviews and research triggered by discussions.

The need for, and the challenges of, data cooperation

The COVID-19 pandemic has illustrated that data needs to be collected and analyzed, and results shared on a local and global scale to promptly recognize, understand, prevent, and prepare for new infectious disease threats. This implies that data and the results of the analysis of data are shared potentially across organizations, cities, countries, regions, and global institutions (the exact balance between ‘data’ and ‘results of analysis’ of that data to be determined). There is broad agreement that the use of data should adhere to equitable data sharing principles. The ‘FAIR Guiding Principles for scientific data management and stewardship’² provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. These principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

Incentives to collect, generate and share data are intimately connected to who gets to use that data, for what ends, and how those giving the data are credited for their work in – and the costs they incur of – gathering and generating the data and ensuring its quality. Those who gather data on an ongoing basis and are asked to freely share it but depend on the infrastructure of others may not always feel their situation is quite so equitable after all, and their own efforts to innovate with data may not blossom if they find that their success needs to show up somewhere else and, perhaps, for someone else.

There is no one-size-fits-all approach to responding to infectious disease epidemics. A successful prevention, preparedness, and response system will come from a combination of reliable disease surveillance, robust health systems, responsive medical innovation, equitable access to medicine and health care, a pre-emptive focus on social determinants of health, and tackling risk factors from One Health and planetary health perspectives. Any initiative that seeks to isolate a subset of ‘pandemic only’ issues – especially if tech-heavy pandemic warning mechanisms – will struggle to sustain engagement once the current urgency has passed. To be sustainable for the long haul, new initiatives will need to be synergistic with, and not aggravate, already ongoing initiatives and networks of global health researchers and practitioners, and truly serve those researchers and practitioners on their own terms and the populations they serve. New initiatives will need to start by first asking researchers, especially in resource-limited settings, what their needs are, listening to the challenges they face in using data platforms and clouds and sharing data, and uncovering what they feel comfortable with in terms

¹ andrew.farlow@ndm.ox.ac.uk.

² Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>. See also <https://www.go-fair.org/implementation-networks/>

of sorting, managing, and analyzing data, and not imposing what some distant ‘others’ think is needed.

When building pandemic-threat response mechanisms on the foundations of broader health system strengthening, data sharing principles need to be backed by data infrastructures that are also fair and equitable. What, then, are fair principles for the data architecture itself? And what might such architecture look like anyway? How centralized? How distributed and federated? In this workshop we will explore data infrastructures that handle data according to FAIR data principles, and that support, in ways that are fair and equitable, local capacity strengthening and the careers of researchers in low-resource settings.

The notion of a ‘Trust architecture’

One growing tech trend – much talked about in the context of pandemic intelligence – is that of trust architectures. Interest in these is a product of rapid advances in cybersecurity in response to rising cybercrime. In 2019, over 8.5 billion data records were compromised globally. As AI gets ever more sophisticated and powerful, those who seek to protect data are locked in an arms race with those who seek to steal and misuse data; the concept of ‘trust architecture’ is one response. Paradoxically, trust architectures are for situations where trust is lacking. The parties – including countries if the context is pandemic intelligence – do not need to trust each other; the ‘trust’ is built into the design of the trust architecture by using, for example, distributed ledgers such as blockchain. Efforts to build trust across parties, including countries, is replaced by efforts to build trust in the trust architecture and those who run it.

Trust is still needed. All must trust that the trust architecture works, that it will not suffer breaches of security measures supposedly to protect it, that the way in which it functions has been designed to capture all possible eventualities, that its priorities over how data is used are not skewed preferentially to some, such as the rich, over others, such as the poor, that power over how data is used is not abused, and that it is GDPR (General Data Protection Regulation) compliant. This is a tall order. The difficulty of providing such a function may depend on how much the trust architecture is expected to handle (is it the raw data or the results of analysis and advice coming in from a network such that the original data stays in the hands of those who generated it?), and on the breadth of coverage of the trust architecture (does it operate globally and in interventionist and centralised fashion or does it have more limited functions and is federated because little data analysis is taking place centrally?; does it cover all data science innovation or leave the innovation for others to do?).

With so many different platforms already out there, not everybody is going to change their practices to work anyway from a single platform. Instead of expending energy on trying to centralise, it might be better to find out what are the commonalities across what already exists and making what already exists interoperable (such as with common metadata models that make data for exploring particular problems more discoverable). Helping existing groups to talk much more to each other, and building capacity and skill sets in all places and especially in resource-low places, may yield the highest payoff.

Critically for those seeking to improve pandemic intelligence, issues regarding the trust architecture can be separated somewhat from issues regarding innovation in AI and use of data. The trust architecture, a global public good perhaps, might be provided and maintained by one or a few entities, but innovation take place across a highly distributed network of many entities. Indeed, to achieve maximum value for those using the trust architecture, we may need to significantly invest in and strengthen capacity across global networks of data innovators.

A rapidly evolving technological landscape

Technology is changing rapidly. In the coming decades, nearly all major businesses and organisations will be using distributed IT infrastructures – hybrid-cloud or multi-cloud platforms – with data and processing handled in the cloud and results quickly made accessible to devices and practitioners on the ground. The shift to next-generation computing – including quantum AI and neural networks and machine learning that will write code and create new software – will lead to software applications far more powerful and capable than anything we have today and will unlock unprecedented opportunities to find answers to previously intractable problems. The opportunity for distributed learning will see machine learning across networks without the need to take data and, at least in theory, put local voices in control. Yet this is in the future. Till then, what can those do who live in parts of the world where few clouds are based or where there is no budget for analysis and storage on the cloud?

AI is in its early days. As AI becomes ever more sophisticated, it's potential for developing new data-rich, AI-driven applications for tackling global health challenges will be greater if the expertise and efforts to develop, implement, scale, and defuse it are distributed widely. Indeed, the second-round boosting effects to local science communities from their own technological innovation favours a federated approach to data science innovation for health and more funding into local innovation capacity across networks. The question becomes how to square this with more centralised components such as hubs. Given that technology is making it ever more possible to do distributed learning across networks and given that many researchers/practitioners across those networks are in constant dialogue anyway, some clear articulation is needed of the value-add contribution of a hub or hubs. And, if a hub or hubs, how many hubs, what is or are their remits, and how are incentives created to do a good job of being a hub?

R&D

Lack of availability and use of interventions has seriously hindered our capacity to effectively prevent and curb epidemics. If we are to achieve more responsive medical innovation and equitable access to medicine and health care, a particular priority is to replace the current largely fragmented non-collaborative R&D ecosystem with a more robust globally coordinated end-to-end health innovation system for infectious disease epidemics, that will drive the development, registration, availability, and equitable use of pharmaceutical products.³ This can only be achieved if it is owned by – and grounded in the legitimacy of – those with long histories of doing global health research in low-resource settings and if it is based on a public-health-orientated governance mechanism. There is little explanation so far of how collective intelligence and coordination of the whole R&D and health delivery ecosystem will be achieved, the role of any trust architecture, and what the exact roles of hubs and federated networks will be in this. Who takes responsibility, is accountable, takes credit or blame? What are the incentives and modes of accountability? A global approach based on open science will require those working in local science ecosystems surrounding hubs and across networks to adopt an open science approach for themselves and not to preach the virtues of open science and 'sharing' for everyone else while avoiding application of such principles to themselves.

³ Piero Olliaro, P., and Torreele, E., "Global challenges in preparedness and response to epidemic infectious diseases", Volume 30, issue 5, p1801-1809, May 04, 2022, DOI: <https://doi.org/10.1016/j.ymthe.2022.02.022>.

Networks, hubs, and funding flows

We have FAIR data sharing principles. We now need to work out how to make data platforms that are also truly fair. What is the technology for doing this? Is it with federated data-sharing and open science? Is it in a mix of hubs and networks or entirely based on networks with no dominant hubs? Are hubs responsible for parts of the infrastructure only? Do hubs leave data largely in the hands of those who generate the data so they can use and develop their expertise to work with it, or do hubs agglomerate data and expertise themselves? Is the ‘shared experience’ and effort to build pandemic intelligence found at a hub or distributed across a more federated network? What is the best hub/network balance to support global capacity strengthening and to create the political imperative to bolster pandemic intelligence efforts? What are the bounds delineating the responsibilities and accountability of hubs? What is the governance structure? Where exactly is the decision-making taking place and how is power and influence over decisions and priorities distributed?

Equity is driven by financial flows that create data infrastructure. We have an opportunity to reframe thinking about funding to fill important gaps, including surveillance and timely sharing of information on circulating pathogens and R&D, to create a network of activity funded from the bottom up to meet activity fed by funding that is more top-down, to create a level playing field so that countries from all corners of the world feel comfortable joining a global effort. This is no easy feat. One funding model is to channel funds to activities across networks through a hub or hubs. Another funding model would treat a hub or hubs as providers of limited-function global public goods, maybe as little as a coordinating and enabling role to help bring groups together (even then, most groups are already talking with each other), funded from one source, and to treat the funding of all other activities through networks as quite separate. The latter (separation) model might enhance the incentive of a hub or hubs to deliver, and bolster accountability. Incidentally, given the huge collective payoff, it is to the advantage of high-income countries to invest in the capacity strengthening of low-income countries even from a purely selfish point of view. Given the early days of some of these new pandemic initiatives, we have a one-off opportunity to rigorously question the funding and incentives structures being created.

Research incentives

When designing any new data architecture, we need to think what success looks like for researchers working under the rules of that architecture. The current system of research incentives needs to change to be more fit for purpose according to how technology is evolving, especially for young researchers whose careers can only advance if they get credit for their contributions. In current academic publications, leading authorship is one of the few ways to recognize scientific contributions. Global pandemic intelligence requires data sharing where such immediate and individual recognition is less obviously feasible. It is unclear what other types of contributions will be accepted as such by the scientific community to incentivize open data sharing. Younger people are at the heart of breaking with past research paradigms, but it is still the case that individual researchers need to get credit for their efforts. A wider vision involves health academies and training too.

Case studies and lessons to apply

We do not face a blank slate. We do not need to reinvent the wheel. We need to make the most of what we already have. Much is being learned across numerous current initiatives, many more in LMICs than is often appreciated, with Africa especially rich in recent examples. Many of the lessons can fruitfully be applied to new initiatives to improve pandemic intelligence. To reap maximum benefits, what lessons can be learned from such initiatives and applied in

current efforts to improve pandemic intelligence and how might all this prior and ongoing activity be supported to be a key part of the solutions we seek?⁴

- Covid 19 Clinical Research Coalition, <https://covid19crc.org>
 - See also <https://covid19crc.org/videos/2022/effective-equitable-and-efficient-sharing-of-individual-level-data-recommendations-solutions-and-ways-forward/>
- DNDi ANTICOV pandemic preparedness platform supporting preparedness and rapid response to emerging infectious diseases via the development of tools with a focus on low- and middle-income countries (starting with Africa).
 - <https://dndi.org/wp-content/uploads/2021/08/DNDi-COVID-19-Policy-Report-2021.pdf>
- GloPID-R, network with extensive work on data sharing, <https://www.glopid-r.org/our-work/data-sharing/>
- H3ABioNet - Pan African Bioinformatics Network, <https://www.h3abionet.org> for the Human Heredity and Health in Africa (H3Africa) consortium. 28 partner Institutions in 14 African countries aimed at developing bioinformatics capacity and infrastructure within Africa while providing bioinformatics support to the joint NIH and Wellcome Trust funded H3Africa Initiative.
- H3Africa The Human Heredity and Health in Africa Initiative, <https://h3africa.org>
 - see <https://h3africa.org/index.php/2022/03/30/african-data-science-research-receives-major-funding-from-the-national-institutes-of-health-usa/>
 - See 'lessons for pandemic hubs?' (see comment/news such as [How a South African community's request for its genetic data raises questions about ethical and equitable research](#))
- VODAN Africa, <https://www.vodan-totafrica.info/index.php> "Fighting the COVID-19 with FAIR Data" "[VODAN Africa To Deploy Digital Health Systems Across Africa](#)" "mission of VODAN-Africa to generate continuous, real-time, high velocity clinical observational patient data from resource-limited communities that have not been well represented in digital health data. The key feature is that the data produced remains in the health facility only. It will not leave the health facility. Since the data is machine-actionable the input of the data only happens once; in the deployable architecture, the data is used for four parallel use cases...This is an ethical data infrastructure, fully compliant with the guidelines of FAIR Data, GDPR, and local regulation... able to pick up pandemic trends early, localize them and work with local health authorities to act on the data..." "This African-led initiative holds amazing potential for science and is a breakthrough for ethical FAIR AI Ready data pipelines."
- NIH/Fogarty programme: Harnessing Data Science for Health Discovery and Innovation in Africa, <https://commonfund.nih.gov/africadata/fundedresearch>
- 'INFORM Africa' (Role of Data Streams in Informing Infection Dynamics in Africa)
 - scientists in Nigeria, South Africa, and Zambia, University of Maryland USA "using big data from SARS CoV-2 and HIV syndemics with the overall goal to provide population-based information on transmission dynamics to inform interventions and as a cornerstone for future pandemic preparedness."
 - 'INFORM Africa' is an opportunity to consolidate on gains from other pan African consortiums led by African scientists such as H3Africa on health issues of significant public health challenge to the African continent. African scientists

⁴ Since this file might be shared and it is very challenging to keep up and be fair to all who might be involved, the names of individuals are not listed but kept in a separate file.

and public health leaders should benefit from the immense strength that comes from the current era of sharing big data freely and openly for insights to address health issues and outbreaks.

- Institute of Human Virology in Nigeria (IHVN) lead INFORM Africa
- <https://www.fic.nih.gov/Grants/Search/Pages/u54tw012041.aspx>
- <https://www.fic.nih.gov/Programs/Pages/harnessing-data-science-for-health-discovery-innovation-africa.aspx>
- <https://www.ncbi.nlm.nih.gov/books/NBK321550/>
- [German West-Africa Centre for Global Health and Pandemic Prevention \(G-WAC\)](#), involving KNUST-College of Health Sciences in Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana,
- African Open Science Platform (managed by the NRF in South Africa and supported by the Committee on Data (CODATA) of the International Science Council, based in Paris). According to its [launch document](#), the platform will be ‘bringing scientists and non-scientists together as knowledge partners in open networks of collaborative learning and problem solving’. The Platform’s building blocks include a federated hardware, communications and software infrastructure, including policies and enabling practices to support open science in the digital era. Strands of activity include: A federated network of computational facilities and services; Software tools and advice on policies and practices of research data management; A Data Science and AI Institute at the cutting edge of data analytics; Priority application programmes: e.g. cities, disease, biosphere, agriculture; A Network for Education and Skills in data and information; A Network for Open Science Access and Dialogue.
 - <https://council.science/current/blog/african-open-science-platform-takes-shape/>
 - <https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/african-open-science/>
 - <https://council.science/current/news/the-national-research-foundation-of-south-africa-to-host-the-african-open-science-platform/>
- Committee on Data (CODATA) of the International Science Council <https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/african-open-science/>
- [The Global Health Network](#)
- Oxford Modelling for Global Health group based at the Big Data Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford; new MSc programme “Modelling for Global Health” at the Nuffield Department of Medicine to train future cohorts of global health modellers; new consortium of international modellers from over ten countries spanning four continents, set up in March 2020 to address urgent in-country requests for modelling support during the global COVID-19 pandemic.
- Knowledge Network on Innovation and Access to Medicines, <https://www.knowledgeportalia.org>
- Global Health Strategy Group for Pandemic Lessons and Future Pandemic Preparedness, University of Oxford
- WHO
- Rockefeller Foundation, see recent Bellagio workshop on data-sharing in clinical trials (suggesting cross pollination of lessons from clinical trials?)
- Various Academies of Science

- Oxford Centre for Global Health Research, with mission “to be committed partners in world-leading global research networks that promote equitable knowledge sharing to advance and protect the health of the most vulnerable”... including by “ Building partnerships to tackle global health challenges especially those affecting the most disadvantaged,” and by “Advancing equitable international research collaborations and the methods and platforms that support these.” See also amongst others:
 - Epidemic Diseases Research Group Oxford (ERGO) Which is reducing the health and socioeconomic impact of epidemic infections through patient-centred research and the development of methodological approaches suited to epidemic settings.
 - International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) a global federation of clinical research networks, providing a proficient, coordinated, and agile research response to outbreak-prone infectious diseases.
 - Africa Oxford Initiative (AFOX), a cross-university platform based at the University of Oxford with the aim of facilitating equitable and sustainable collaborations between researchers based at the University of Oxford and African universities, as well as increasing the number of African students pursuing postgraduate degrees in Oxford.
- Institute for Ethics in AI, University of Oxford
- ITU-WHO Focus Group on AI for Health
 - Working Group - Collaborations and Outreach (WG-CO)
 - Working group Ethical considerations on AI for health (WG-Ethics)
 - Topic Group for Outbreak Detection (TG-Outbreaks)

A Bibliography of interesting publications and activity (please suggest additions)

Articles, Research on Data Sharing

- **Equitable data sharing in epidemics and pandemics**
<https://bmcomedethics.biomedcentral.com/articles/10.1186/s12910-021-00701-8#Sec17> : Pratt, B., Bull, S. Equitable data sharing in epidemics and pandemics. *BMC Med Ethics* **22**, 136 (2021). <https://doi.org/10.1186/s12910-021-00701-8> -utility prioritized over equity in terms of data sharing during epidemics and pandemics
- **Data sharing and the public interest in a Digital Pandemic-**
 Staunton, Ciara: *Data Sharing and the Public Interest in a Digital Pandemic**, *VerfBlog*, 2021/3/20, <https://verfassungsblog.de/data-sharing-and-the-public-interest-in-a-digital-pandemic/>, DOI: [10.17176/20210320-151158-0](https://doi.org/10.17176/20210320-151158-0) : **Data and governance, equity, stewardship, right to privacy**
- **Trust, risk, and the challenge of information sharing during a health emergency**
 Lencucha, R., Bandara, S. Trust, risk, and the challenge of information sharing during a health emergency. *Global Health* **17**, 21 (2021). <https://doi.org/10.1186/s12992-021-00673-9>
- **Fostering global data sharing: highlighting the recommendations of the Research Data Alliance COVID-19 working group**

Austin CC, Bernier A, Bezuidenhout L, et al. Fostering global data sharing: highlighting the recommendations of the Research Data Alliance COVID-19 working group. *Wellcome Open Res.* 2021;5:267. Published 2021 May 26. doi:10.12688/wellcomeopenres.16378.2

- **Solutions to COVID-19 Data Sharing**
- **Data Governance : Organizing data for trustworthy Artificial Intelligence**
Marijn Janssen, Paul Brous, Elsa Estevez, Luis S. Barbosa, Tomasz Janowski, Data governance: Organizing data for trustworthy Artificial Intelligence, *Government Information Quarterly*, Volume 37, Issue 3, 2020, 101493, ISSN 0740-624X, <https://doi.org/10.1016/j.giq.2020.101493>.
- Policy Opportunities to enhance sharing of Pandemic Research
<https://doi.org/10.1126/science.abb9342>
- **What Can We Learn from Data-Sharing During the COVID-19 Pandemic?**
<https://www.appliedclinicaltrials.com/view/what-can-we-learn-from-data-sharing-during-the-covid-19-pandemic->

Data sharing and the global south:

- **Narratives and counternarratives on Data Sharing in Africa**,
<https://doi.org/10.48550/arXiv.2103.01168> : most datasets are from African communities however, challenges related to data sharing employ a deficit narrative and are championed by global north voices. Power imbalances that underpin inequities in data sharing
- **Open data sharing and the global south: who benefits?**
<http://dx.doi.org/10.1126/science.aap8395> :
- **A widening data divide: COVID-19 and the Global South**
<https://www.opendemocracy.net/en/openmovements/widening-data-divide-covid-19-and-global-south/>

Reports:

- Data sharing in public health emergencies (Wellcome, Glpid-R, Dfid):
<https://www.glopid-r.org/wp-content/uploads/2017/02/data-sharing-in-public-health-emergencies-case-studies-workshop-reportv2.pdf>
- Data Protection in the time of the pandemic (Centre for civil society and governance, Centre for information policy leadership):
https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_hku_roundtable_series_report_-_data_protection_in_the_time_of_the_pandemic_22_april_2021_.pdf
- Ethical sharing of health research data in low- and middle-income countries: views of research stakeholders (Wellcome Trust) :
<https://cms.wellcome.org/sites/default/files/ethical-sharing-of-health-research-data-in-low-and-middle-income-countries-phrdf-2014.pdf>
- World development report 2021- Building Trust (World Bank):
<https://wdr2021.worldbank.org/stories/building-trust/>
- Data for international health emergencies: governance, operations and skills (Science academics of G7 nations):
<https://royalsociety.org/-/media/about-us/international/g-science-statements/G7-data-for-international-health-emergencies-31-03-2021.pdf>