

Applying intelligent open science to combat future pandemics

16th May 2023

Lucia Loffreda, Eleanor Cox, Rob Johnson

Investigating open data sharing during COVID-19

Our study was commissioned by the UK's Department for Business, Energy and Industrial Strategy, and aimed to strengthen recommendations made as part of the G7 Research Compact.



Identify **barriers** to open data sharing during the pandemic



Identify **enablers** of open data sharing during the pandemic



Highlight **lessons learned** to support responses to future pandemics

Evidence review

295 sources

We identified a total of 295 sources via a mixture of Google searches and snowball sampling. An additional 120 sources were reviewed based on a similar study on open research practices

24 interviewees

Interviewees were identified via desk-based research, snowball sampling and a text processing algorithm. Interviews comprised of a series of 60-minute one-to-one with experts in genomic data sharing, and public health policymakers.

18 peer reviewers

Peer review was carried out by international experts and members of the G7 Open Science Working Group.

Pathogen data is critical in pandemics to support public health responses

1. Identify

Detect, investigate, monitor and control the virus, including its mutations

4. Understand

Help to trace the source and transmission of outbreaks

2. Understand

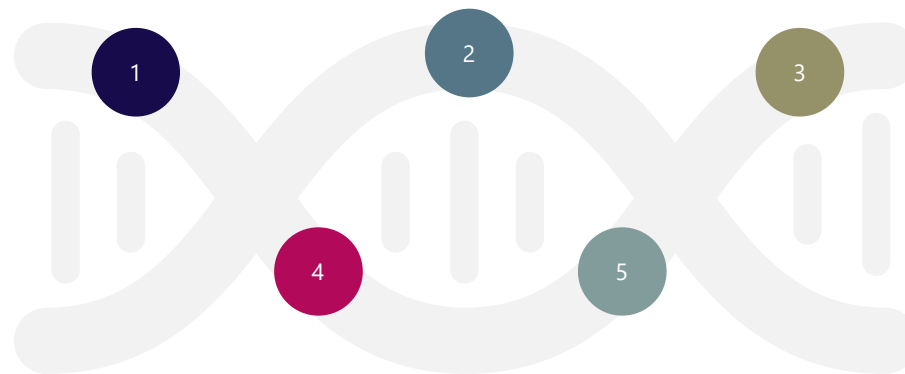
Disease surveillance to understand how people become exposed to the virus

3. Trace

Help trace the source and transmission of outbreaks

5. Learn

Learn about virus evolution such as the development of variants of concern



Data sharing yielded mixed results, but certain factors contributed to its success



Timeliness

Data sharing was delayed for various reasons, meaning that sequencing data was shared widely but too late.



Standardisation

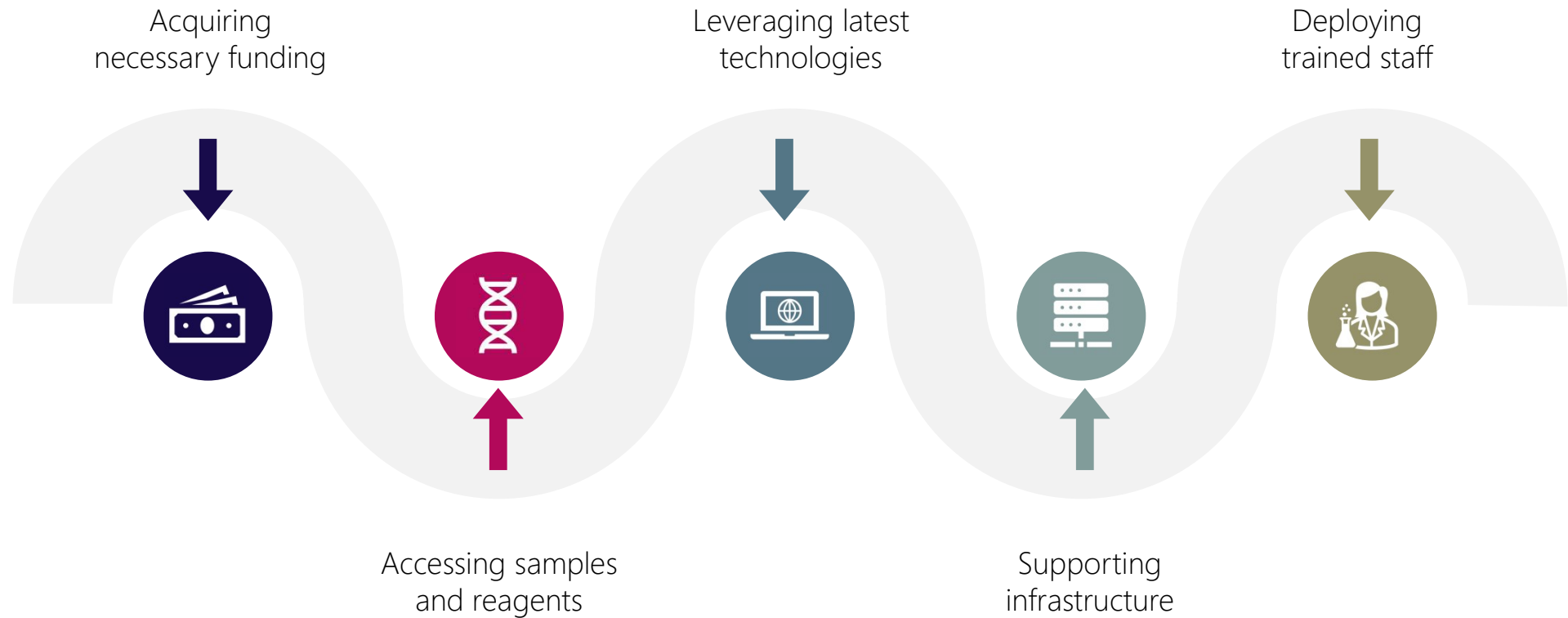
Variations in data quality, format, and metadata standards presented barriers to effective sharing across the globe.



Openness

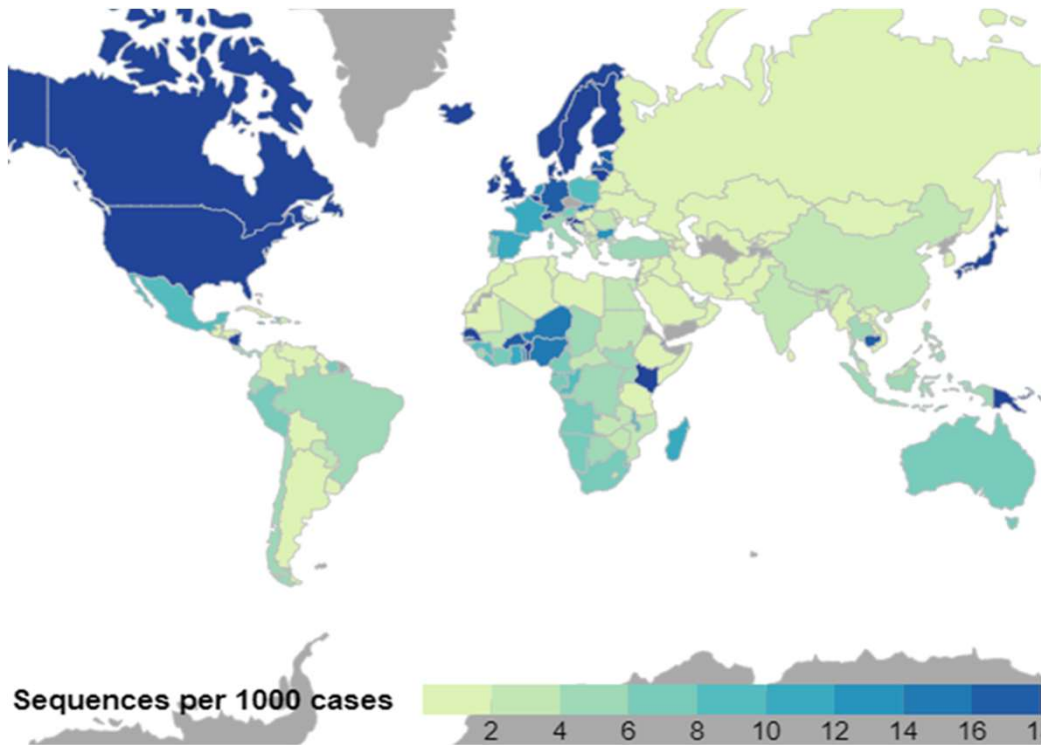
Open access repositories provided greater benefits for re-use and integration than restricted repositories.

Lesson 1: Invest for the long term



"It will be very good for the world if, even in developing countries, we could have a continuous source of funding for studies on infectious diseases and emergence... This will have connections with the sharing of data as one of the requirements [of funding] could be that you have to share your data."
Academic (Interviewee)

Lesson 2: Take a global perspective



- Many countries in the global south lack capacity and investment, and so are unable to carry out large scale sequencing efforts.
- Researchers in the global south are more concerned about their data being 'scooped' by labs in higher-income countries if they upload sequence data to open data repositories.
- Some countries have faced severe political and economic consequences as a result of sharing data openly (e.g. South Africa and Botswana sharing the Omicron variant).

"If the world keeps punishing Africa for the discovery of Omicron and 'global health scientists' keep taking the data, who will share early data again?"
Professor Tulio de Oliveira (via [Twitter](#))

Lesson 3: Create incentives for equitable sharing



Recognise that recognition and reward are key drivers for data generators



Review data access agreements to promote equitable data sharing in emergency situations



Enable collaborations between researchers and public health officials

"In pathogen genomic data very specifically it's clear to me that the predominant paradigm is inadequate. We need the data to be available way before any associated publications."

Policymaker (Interviewee)

Lesson 4: Adapt to changing circumstances



- Supporting academics and public health organisations to work collaboratively is necessary.
- This will rely on the capture of high-quality metadata behind datasets that follow the FAIR principles.
- Technical and legislative solutions also need to adapt to respond to emergency scenarios.

We are in this bubble of open science and [clinical labs] are in their own bubble. Breaking those silos within science is a tremendous amount of work, and a much bigger issue than I ever anticipated."
Infrastructure provider (Interviewee)



Lesson 5: Move beyond current sharing paradigms

“I think the largest lesson that I've learnt from this situation is that you need to pull in a diverse set of voices... We have to think about how the least among us are going to benefit from [sequencing] and why they might care about it. And if we fail at that task, then we will fail to get representative data every time.”

ACADEMIC EXPERT (Interviewee)

Overview of lessons learnt



Invest for the long term

Developing sustainable funding and infrastructure with transparent standards.



Take a global perspective

Generating capacity in all parts of the world, without risking economic and political consequences.



Incentivising equitable data sharing

Rewarding data generators and cross-boundary collaboration.



Adapting to changing circumstances

Open infrastructure needs to identify and adapt to emerging requirements and sharing sensitive data.



Move beyond current sharing paradigms

A new globalised approach to funding, governance and benefits sharing is needed.

Thank you

Lucia Loffreda, Ellie Cox



lucia.loffreda@research-consulting.com

eleanor.cox@research-consulting.com



www.research-consulting.com



[@rschconsulting](https://twitter.com/rschconsulting)