Talk outline

1. Introduction to Thesis
2. What has been done?
3. What is still to come?
4. Discussion with collaborators

Nathanael Sheehan

Thesis Plan

Part 1: The Split Understanding of Openness during Covid-19
- Regulated vs Restricted Openness ✅
- Actionability vs Accountability ✅

Part 2: Situating Openness in Diverse Research Environments
- Understandings of Openness at the European Bioinformatics Institute, United Kingdom 🛠️
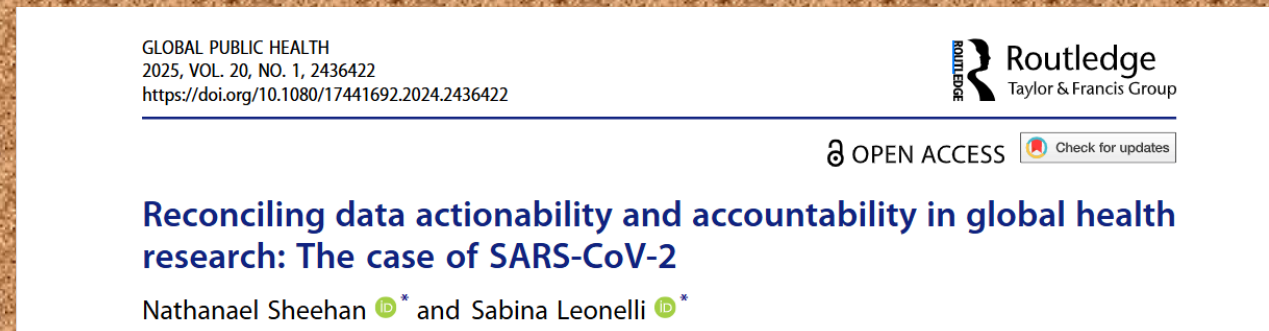- Understandings of Openness at the Center for Knowledge and Data Linkage, Brazil. 🛠️

Part 3: Reconstructing Openness for Planetary Health
- An Account of Participatory Informational Format Exclusion (PIFE) ✅
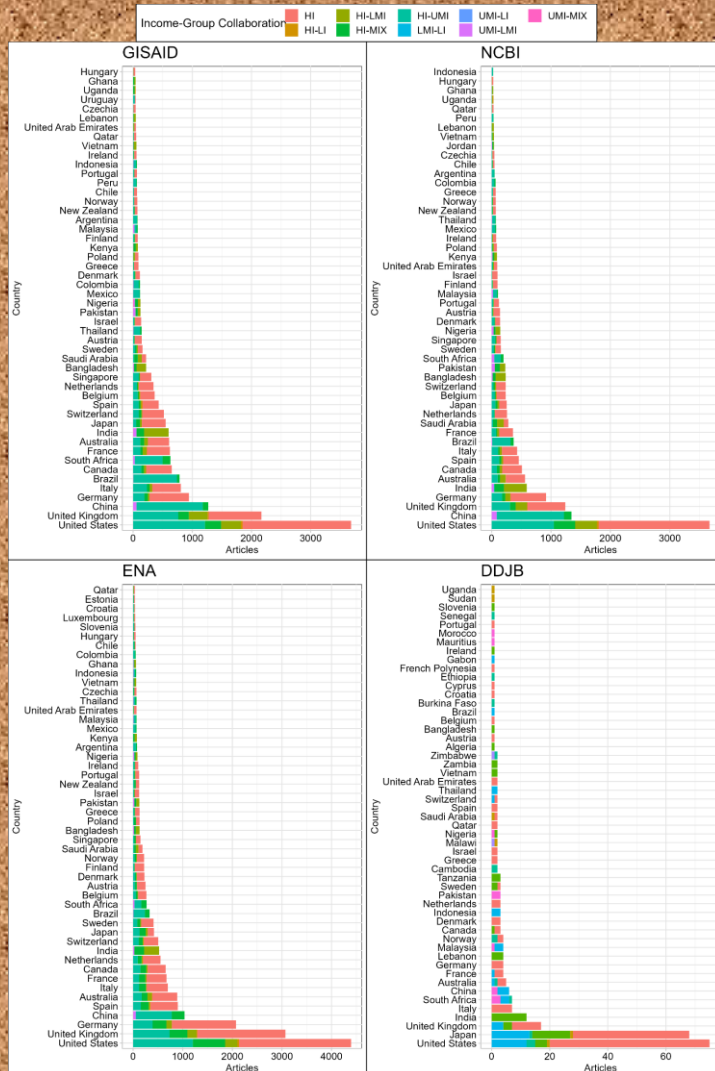- Openness in the Pluriverse 💭

Nathanael Sheehan

Unrestricted versus regulated open data governance: A bibliometric comparison of SARS-CoV-2 nucleotide sequence databases

- **Differing Governance Models:** GISAID emphasizes comprehensive global participation to track mutation patterns, whereas INSDC prioritizes data linkage and usability to support discovery of SARS-CoV-2 interactions—highlighting contrasting values in openness and representation.

- **Trends in Collaboration and Access:** The rise of hyper-authorship and both gold and green access pathways points to increasing transparency and collaboration in virology, yet paradoxically, fully open repositories like ENA may still foster restricted access and solo-authored outputs.

- **Equity and Global Representation:** Predominant collaboration among high-income countries challenges the notion of global inclusivity in open data infrastructures, raising concerns about underrepresentation of low-income regions and reinforcing systemic inequities in scientific data sharing.

Nathanael Sheehan

Reconciling data actionability and accountability in global health research: The case of SARS-CoV-2

actionability = *capacity to repurpose data for novel goals, methods of analysis and research settings.*

accountability =*responsiveness of data work and infrastructures to a wide variety of prospective users and situations*
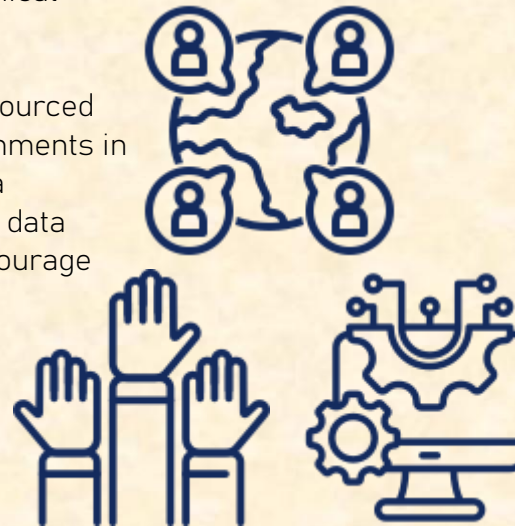
Actionability & Accountability

≠

A trade off

Data diversity:
- Ensure geographical heterogeneity of contributors
- Support low-resourced research environments in contributing data
- Foster variety of data formats and encourage rich metadata

Visible infrastructure:
- Make data curation and processing work visible through well-structured documentation
- Provide opportunities for both face-to-face and distributed interactions (synchronous and asynchronous) for training and participation
- Attend and organize domain-related conferences to present and discuss infrastructural developments

Participative governance:
- Involve domain experts across a variety of locations and languages
- Develop user-friendly mechanisms to collect feedback from the community
- Require user registration
- Support low-resourced research environments in designing and governing prospective data use policies

Nathanael Sheehan

Part 2: Situating Openness in Diverse Research Environments

While part one of my thesis used methods from computational social science and philosophy of science...

Part two of my thesis instead draws on methods from sociology and anthropology e.g. ethnography and semi-structured interviews.

The aim of this is to understand on the ground a situated account of openness in two different contexts, geographies and environments.

Nathanael Sheehan

A History and Sociology of Openness at the European Nucleotide Archive.

ENA
European Nucleotide Archive

3 Months Ethnography

24 Semi-structured Interviews

37 Primary Source Texts

43 Years of Archive Development
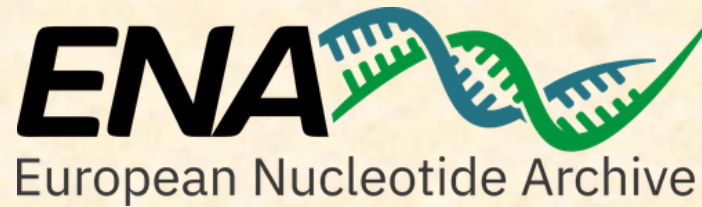
- Oldest and largest nucleotide archive in the world
- Its platform enables the management, sharing, integration, archiving and dissemination of sequence data.
- Key open infrastructure for COVID-19 response
- Is part of broader institutional networks e.g. European Bioinformatics Institute -> Wellcome Genome Campus -> Elixr Infrastructure -> Global Core Biodata Resource

a)   what is the ENA's historical commitment to unrestricted openness?
&
b)   how do scientists at the ENA conceptualise and enact Open Science (OS)?

Nathanael Sheehan

A History and Sociology of Openness at the European Nucleotide Archive.

Interview Themes:

- **Understandings and Motivations for Open Science**
- **Data Management and Curation (Or, The Many Faces of Data)**
- **Collaboration, Community, and Policy Ecosystems**
- **Tensions, Barriers, and the Future of Openness at ENA**

"storage is cheaper and cheaper... it is very often still more expensive to find out which datasets you should discard than just buying a new disk"

"US government shutdowns threaten the very existence of scientific infrastructures"

"Visa applications are taking quite long, so we can't even guarantee they will get it."

"People may mean very different things with metadata"

"If somebody in Latin America were to generate data, you know, that's their data. That's their country's kind of like sovereignty... is that data personal to the country or personal to them or their institution"

Unrestricted openness was not an initial feature of the ENA (in fact users of data had to pay a subscription in the early phases of the infrastructure). This transition was less a mere technical solution than a political commitment to the seamless circulation of biological knowledge in line with free software movements of a similar time.

Nathanael Sheehan

A Narrative of the Political Sociology on Openness at the Center for Data and Knowledge Integration for Health (Cidacs)

cidacs
Centro de Integração de Dados
e Conhecimentos para Saúde
FIOCRUZ

- First Latin American research lab to own a supercomputer facility.
- Links public administrate data to study the impact of environmental, health and social policy.
- Initially funded for the creation of the "100 million cohort".
- Part of broader institutional networks e.g. FioCruz

a) What are CIDACS views on Open Science?
&
b) How can trusted environments and data visitation be an alternative model to complete openness?

4 Months Ethnography

20 Semi-structured Interviews

- being honest about my lack of understanding in certain topics like epidemiology,
- attempting to speak the Portuguese even at the expense of embarrassing myself,
- learning about key theories and figures in the Bahian canon such as "collective health" and Milton Santos,
- not using the language of calling Brazil somewhere from the global south or a developing country,
- collaborating with researchers based on our own skills in order to develop ideas collectively,
- presenting work with slides in Portuguese ,
- ensuring regular check-ins and assessments of my relationships with members at CIDACS.

Nathanael Sheehan

Now a conversation with my collaborators from CIDACS

Mariara Pitta Lima

Bethonia Almeida

Drawings by Erwin Dasapta

Nathanael Sheehan