



ITHAKA S+R

# Data Communities & Fostering Research Sharing

**Putting the Tri-Agency Policy into Practice**

Pillar 3 Webinar: Data Deposit

October 26, 2021



What makes  
research data  
sharing successful?



# The Data Sharing Landscape

## **Institution Driven**

Institutional repositories and data support services are increasingly available to researchers through their institution.

Collaborative groups are working to **share curation expertise** and make data sets discoverable across institutions.

## **Compliance Driven**

Funders and publishers require researchers to deposit datasets when their articles are published.

**Generalist repositories** are targeting this type of data sharing.

## **Community Driven**

Researchers form communities around the sharing and reuse of certain types of data.

Community-centric sharing usually takes place via **domain repositories**.



ITHAKA S+R

Our **Research and Teaching Support Services** program examines scholarly practices using a unique collaborative, qualitative methodology.

- Agriculture (2017)
- Art History (2013)
- Asian Studies (2018)
- Business (2019)
- **Big Data (2021)**
- Chemistry (2014)
- Civil & Environmental Engineering (2018)
- History (2012)
- Indigenous Studies (2019)
- Language and Literature (2020)
- Public Health (2017)
- Primary Sources (2020)
- Religious Studies (2017)
- **Teaching with data in the social sciences (2021)**



# By the numbers



2000+

Scholars interviewed about their information practices and research support needs



550

Librarians we've collaborated with to conduct the interviews for the research



145

Institutions in the U.S. and Canada we've partnered with to conduct the research



# Data sharing success stories

## Genetics

- GenBank (NCBI)
- Species-specific databases such as FlyBase
- Viral genomic sequencing via GISAID



## Neuroimaging

- Neuroimaging Tools and Resources Clearinghouse
- OpenNeuro
- Donders Repository



OpenNEURO





# COVID-19 data community Influenza Virus Genetics

**GISAID** (Global Initiative on Sharing All Influenza Data) is an interdisciplinary organization supporting a repository of genetic data and related projects



# Data communities

**A data community is a fluid and informal network of researchers who share and use a certain type of data.**

Most (but not all) data communities are facilitated through a website incorporating an online repository.

A data community is **not** the same as a discipline.



Thinking in terms of data communities can support sharing across institutional boundaries, mirroring how scholars do their work.



# Characteristics of successful data communities

- Bottom-up development
- Community norms
- Absence or mitigation of technical barriers



# The data community circle of life

COVID-19 has shown that new data communities can **emerge and grow** extremely quickly to meet urgent needs.

Data communities also **die**, due to lost funding or organizational support, personnel changes, or simply changes in community need.

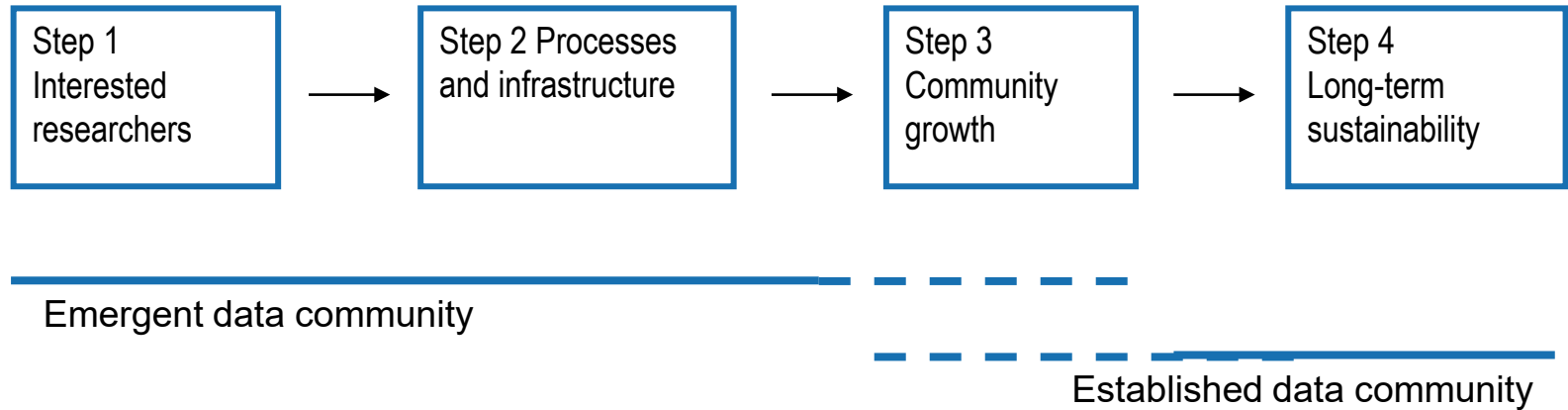
# Implications for data sharing support

- “Build it and they will come” approaches generally don’t result in data communities
- **Top-down mandates** generally don’t result in data communities
- **Institutional and generalist repositories** can provide infrastructure and curation support
- Librarians with institutional support remits have the challenge of supporting **cross-institutional** data communities

**An emergent data community** is a group of scholars who are enthusiastic about sharing and reusing a certain type of data but haven't yet fully established the necessary processes and infrastructure.



# The data community growth process





# What data communities need

- Help building or identifying existing **repository infrastructure**
- **Technical and policy advice** on metadata, vocabularies, preservation, privacy, etc.
- Guidance and advocacy for achieving organizational and financial **sustainability**
- Help **getting the word out** to researchers who might be interested in getting involved



# What's next for Ithaka S+R?

## **Pilot**

### **Data Service Assessment Tool**

Building on 2020 census of data services at U.S. institutions, we are piloting data services assessment tool to help universities evaluate and design effective data supports for researchers including those in data communities. We welcome expressions of interest.

## **Convening**

### **Leveraging Data Communities to Advance Open Science**

In collaboration with the Data Curation Network, an NSF-funded project that will bring together scientists and information technology professionals for focused discussions about initiating and sustaining data communities.

.





# Further Reading

## **Issue Brief**

Data Communities: A New Model for Supporting STEM Data Sharing

## **Research Report**

Research Data Services in US Higher Education

## **Blog Posts**

Emergent Data Community Spotlight Series



ITHAKA S+R

Thank you

**Danielle Cooper**

Manager, Collaborations and Research

@dm\_cooper

[danielle.cooper@ithaka.org](mailto:danielle.cooper@ithaka.org)