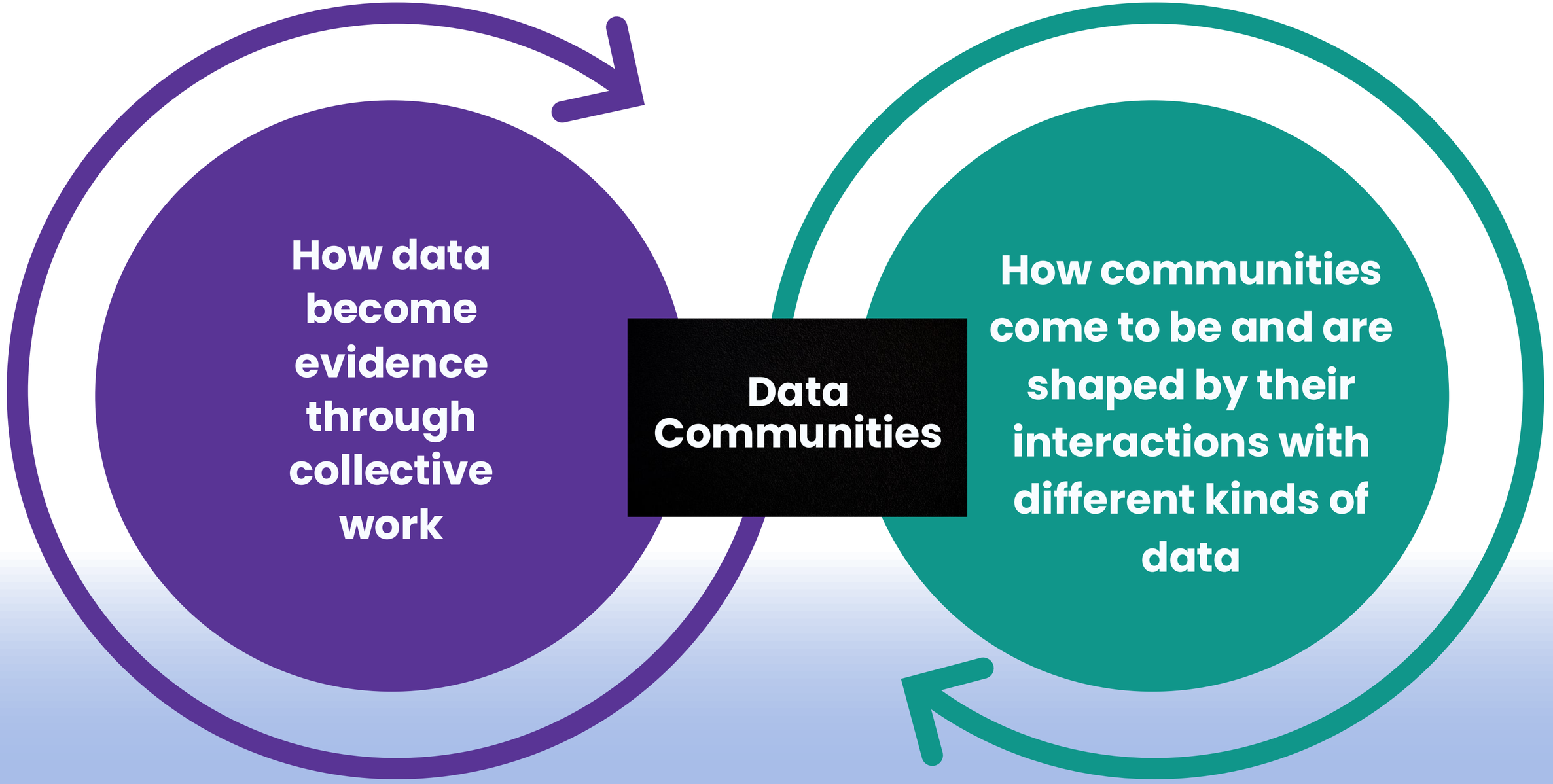
Data Communities

Emma Cavazzoni, Paola Castaño, and Federica Bocchi

Towards A Philosophy of Open Scientific Practices: Comparing Research Environments

PHIL_OS Conference

April 28, 2025


How data become evidence through collective work
Data Communities
How communities come to be and are shaped by their interactions with different kinds of data

How data become evidence through collective work

- Epistemic aims (Evidence for what?)
- Types of data/metadata
- Methods of analysis
- Standardization of protocols and practices


Formal/informal practices of coordination
Degrees of institutionalization
Individual/collective contributions
Recruitment practices
Sources of epistemic authority
How communities come to be and are shaped by their interactions with different kinds of data

DATA-TECHNOLOGY COMMUNITIES

- Productive data- and technology-intensive projects that do not converge around common agendas, adaptive problem spaces, or sets of practices.
- What, if anything, brings them together?

data-technology communities

Heterogenous groups of individuals who come together around a common interest in and use of shared data and data collection technologies, and through this common focus interact, learn from one another, and collaborate within the delimitations imposed by institutional constraints and funding structures.

DATA-TECHNOLOGY COMMUNITIES

Interactions around common data and technologies (collaborative problem-solving, requesting information, seeking advice, reusing assets, discussing).

Collective learning around common data and technologies.

Effective collaboration around common data and technologies (relevant scientific, technical and commercial outcomes).

Enduring Questions about Scientific Collectives

- Understanding the social organization of data sharing practices both as an end in itself, and as a way to understand particular epistemic outcomes.

- Specification of the concept of “judicious connections” (Leonelli 2023) in the context of scientific collectivities working around data in a repository.

- **Boundedness:** Formal and informal practices of association (; Glaeser 2014, Kastenhofer and Molyneux-Hodgson 2021, Lounsbury et al. 2021).

- **Consensus**: Tension between conventionality and novelty / productive tradition and risky innovation (Bourdieu 1975, Kuhn [1959] 1977, Foster and Evans 2015) + combinatorial novelty (Gebhart and Funk 2020; Shi and Evans 2019).

NASA, 2025

NASA Open Science Repository Analysis Working Groups

Open Science for Life in Space

- "The AWGs aspire to scientific excellence and participation in AWGs is strictly on a volunteer basis" (AWG Charter 2025)

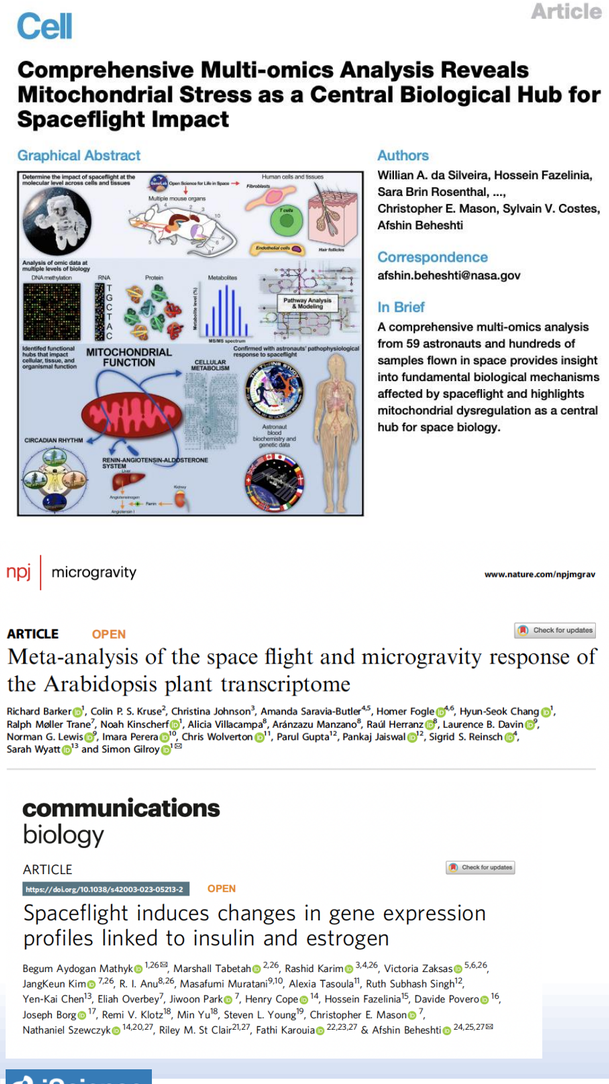
- What kind of scientific collectives are these Analysis Working Groups? What do they do?

  - *"A form of scientific crowdsourcing."*
  - *"A collaborative, large, open network of folks that, from lots of different backgrounds, are working on solving space biology-type issues."*
  - *"A good venue for people who do not have access to funding for spaceflight research."*
  - *"There are different motivations to join, but there is a weight to NASA and people join to help NASA go further into space by understanding how space affects biology."*

- What is specific to groups of people who come together to work on open data? People (data creators and users) + infrastructures (curators, computational, governance of the data)

NASA 2021

NASA OSDR AWG Workshop - American Society of Gravitational and Space Research
Washington DC, November 2023



Cell

Article

Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact

Graphical Abstract

Authors

Willian A. da Silveira, Hossein Fazelinia, Sara Brin Rosenthal, ..., Christopher E. Mason, Sylvain V. Costes, Afshin Beheshti

Correspondence

afshin.beheshti@nasa.gov

In Brief

A comprehensive multi-omics analysis from 59 astronauts and hundreds of samples flown in space provides insight into fundamental biological mechanisms affected by spaceflight and highlights mitochondrial dysregulation as a central hub for space biology.

npj | microgravity

www.nature.com/npjmgrav

ARTICLE OPEN

Check for updates

Meta-analysis of the space flight and microgravity response of the Arabidopsis plant transcriptome

Richard Barker[1], Colin P. S. Kruse[2], Christina Johnson[3], Amanda Saravia-Butler[4,5], Homer Fogle[4,6], Hyun-Seok Chang[1], Ralph Møller Trane[7], Noah Kinscherf[1], Alicia Villacampa[8], Aránzazu Manzano[8], Raúl Herranz[8], Laurence B. Davin[9], Norman G. Lewis[9], Imara Perera[10], Chris Wolverton[11], Parul Gupta[12], Pankaj Jaiswal[12], Sigrid S. Reinsch[1], Sarah Wyatt[13] and Simon Gilroy[1]✉

communications biology

ARTICLE

Check for updates

https://doi.org/10.1038/s42003-023-05213-2 OPEN

Spaceflight induces changes in gene expression profiles linked to insulin and estrogen

Begum Aydogan Mathyk[1,26]✉, Marshall Tabetah[2,26], Rashid Karim[3,4,26], Victoria Zaksas[5,6,26], JangKeun Kim[7,26], R. I. Anu[8,26], Masafumi Muratani[9,10], Alexia Tasoula[11], Ruth Subhash Singh[12], Yen-Kai Chen[13], Eliah Overbey[7], Jiwoon Park[7], Henry Cope[14], Hossein Fazelinia[15], Davide Povero[16], Joseph Borg[17], Remi V. Klotz[18], Min Yu[18], Steven L. Young[19], Christopher E. Mason[7], Nathaniel Szewczyk[14,20,27], Riley M. St Clair[21,27], Fathi Karouia[22,23,27] & Afshin Beheshti[24,25,27]✉

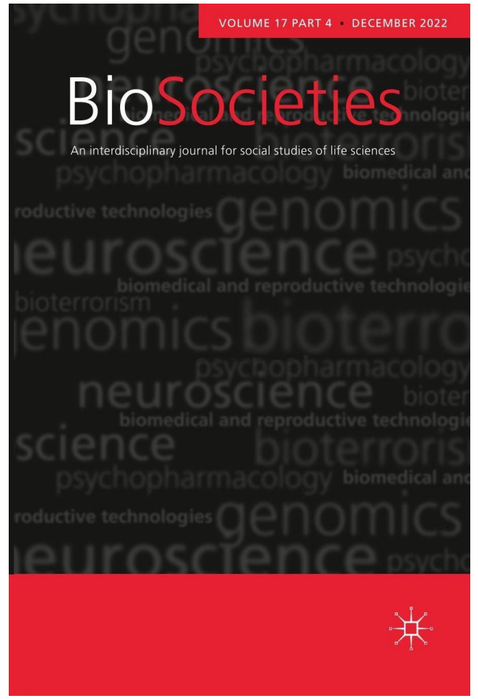
iScience

NASA GeneLab RNA-seq consensus pipeline: Standardized processing of short-read RNA-seq data

Eliah G. Overbey [1 39], Amanda M. Saravia-Butler [2 3 39], Zhe Zhang [4], Komal S. Rathi [4], Homer Fogle [5 3], Willian A. da Silveira [6], Richard J. Barker [7], Joseph J. Bass [8], Afshin Beheshti [37 38], Daniel C. Berrios [3], Elizabeth A. Blaber [9], Egle Cekanaviciute [3], Helio A. Costa [10], Laurence B. Davin [11], Kathleen M. Fisch [12], Samrawit G. Gebre [3 37], Matthew Geniza [13], Rachel Gilbert [14], Simon Gilroy [7], Gary Hardiman [6 15]... Jonathan M. Galazka [3 40] ✉

- A data community is not a bounded collective but a node of interactions → nature of those interactions as they appear in key outcomes (publications).
- Data communities are not necessarily ‘grassroot’ efforts. Mix of top-down/formal and bottom-up/informal practices.
- Closer look at epistemic outcomes of scientific communities created around data resources contrast to large-scale analysis of scientific publications → Maturity of areas of consensus in the field + ‘outsiders’ actively contributing to establish the baselines of a field (templates for data processing, batch effects to compare datasets, replication of protocols) while opening new domains of investigation.
- Durability of these networks depend on their ability for future data generation.

Special Issue in *BioSocieties*



"Turning Biodiversity Data into Evidence: The Role of Protocols and Data Communities"
Federica Bocchi (University of Copenhagen)

"Fostering data communities - perspective from a data archive service provider"
Francesca Morselli (TU Delft), Cees Hof (DANS), Andrea Scharnhorst (DANS)

"Studying data communities: Analytical dimensions from and for empirical research"
Kathleen Gregory (Leiden University) and Sarah R. Davies (Vienna university)

"The Cost of (Data) Community: Error and Repair in Data Processing Pipelines"
Kathryne Metcalf (University of California, San Diego)

"Data-technology communities: collaboration and diversity in data- and technology-intensive multidisciplinary research"
Emma Cavazzoni (TU Munich)

“From Concerting Expertise to Building a Community around Space Biology Data: NASA GeneLab’s Analysis Working Groups”
Paola Castaño (Exeter University)

“It’s a People Thing”: Reimagining Communities of Practice in Biodiversity Data Portals”
Zoe Nyssa (Purdue University), Beckett Sterner (Arizona State University), and Ute Brady (Arizona State University)

"The Future of FAIR Data is FAIR Data Communities"
Commentary by Sabina Leonelli (TU Munich)

Acknowledgments

The Philosophy of Open Science for Diverse Research Environments is funded by the European Research Council (ERC) under the Horizon 2020 Research and Innovation programme [Grant Agreement 101001145]

Thanks to our collaborators at NASA OSDR and the Haly.Id project.