

How we organize Dataverse community efforts

Collaborative strategies for effective community participation

Lightning Talk at the Birdaro Training Program
October 9, 2025

Presenting:
Ceylin Boyd, Harvard University
Philipp Conzett, UiT The Arctic University of Norway

Contributed to the slides:
Jim Myers, Global Dataverse Community Consortium



OUTLINE



- ❖ **About the Dataverse Project and GDCC**
- ❖ **Three types of community efforts**
 - **Community Meetings**
 - **Working Groups**
 - **Community Calls**
- ❖ **Summary**
- ❖ **Acknowledgements**

ABOUT THE PROJECT

ABOUT THE DATAVERSE PROJECT



Open-Source Repository Software

Dataverse is open-source software that enables sharing, citation, and preservation of research data.

Global and Multidisciplinary Support

It supports data sharing across diverse disciplines and institutions worldwide.

Enhancing Accessibility and Reproducibility

The project aims to improve data accessibility and reproducibility in scholarly research.

ABOUT THE GLOBAL DATAVERSE COMMUNITY CONSORTIUM (GDCC)



Community Support and Coordination

GDCC was established in 2018 to support the DataVerse community by coordinating and facilitating collaboration, community efforts, and resources worldwide.

Governance and Strategic Direction

GDCC provides governance and strategic guidance to keep the DataVerse Project aligned with user needs from a diverse community.

Collaborative Venue

GDCC provides a collaborative venue for institutions to leverage economies of scale in support of DataVerse repositories around the world, e.g., through the GDCC DataCite consortium membership that provides DOIs for DataVerse installations.



THREE TYPES OF COMMUNITY EFFORTS

DATAVERSE COMMUNITY MEETINGS

Annual event since 2015, evolving from Harvard-hosted to globally distributed (Portugal, Mexico, North Carolina).

Main venue for facilitating discussing past, ongoing, and future community efforts, and for fostering networking and collaboration.

Organized by a **rotating team**: local hosts + other (more experienced) community members to support newcomers.

Planning starts with a **community survey** to identify relevant topics (e.g., Metadata, Large Data Support, AI, Governance).

Open call for contributions in diverse formats: talks, workshops, roundtables, demos, posters, WG sessions, etc.

Challenges: engaging newcomers, travel limitations, and inclusivity in hybrid formats.



WORKING GROUPS

Current WGs:

- Containerization
- Documentation
- Large Data Support
- pyDataverse

WGs form organically when community members identify a shared need or goal.

Led by **volunteer coordinators**, ensuring alignment with broader community priorities.

Open and inclusive membership, welcoming diverse contributors.

Common characteristics: Regular meetings, progress documentation, and updates via mailing lists and community events.

Challenges include **limited capacity**, shifting leadership, and extended timelines.



WORKING GROUPS:

Examples

Current WGs:

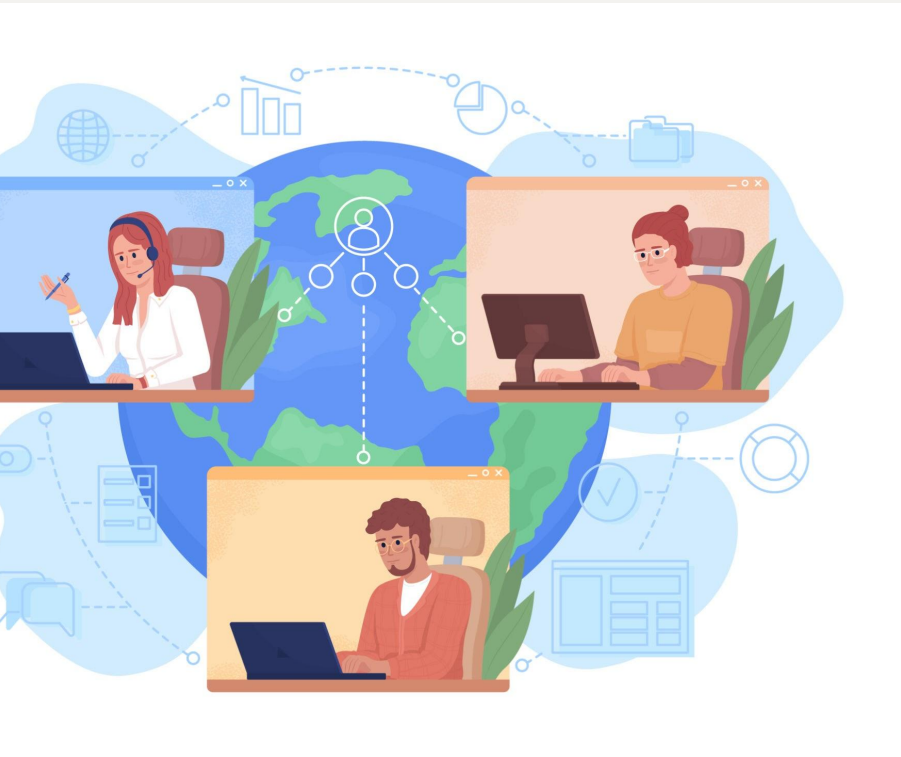
- Containerization
- Documentation
- Large Data Support
- pyDataverse

Sustainability WG – initiated by GDCC with a [mandate/charter](#), guided by the It Takes a Village framework, produced a [public report](#) with strategic recommendations.

External Working Vocabulary WG - defined a mechanism to use community vocabularies and PIDs in Dataverse, created a [design document](#) to reach consensus, [developed software](#) and a [paper](#), now used with ORCID, ROR, keywords, etc



COMMUNITY CALLS



Regular virtual calls open to all community members.

Platform for sharing updates, discussing issues, and showcasing new developments.

Encourages cross-institutional collaboration and informal knowledge exchange.

Often features guest presentations, demos, and Q&A sessions.

Archives and upcoming call info available at dataverse.org/community-calls.

...

SUMMARY

SUMMARY



The Dataverse community uses **different types of efforts to scale community contributions**:

- **Community Meetings:** Annual, evolving format that fosters collaboration, learning, and shared ownership.
- **Working Groups:** Flexible, goal-driven teams with open participation and volunteer leadership.
- **Community Calls:** Regular, inclusive virtual space for updates, discussions, and engagement.

Core Principles: Inclusivity, transparency, shared responsibility, and adaptability.

Challenges: Capacity, inclusivity in hybrid formats, and onboarding new contributors.

An aerial photograph of a large, open plaza paved with light-colored square tiles. Numerous small figures of people are scattered across the plaza, connected by thin black lines that form a complex network diagram. The lines represent connections between individuals, illustrating a community or organizational structure. The overall scene is bright and open, with shadows cast by the people.

ACKNOWLEDGEMENTS

Gratitude to Project Teams

Acknowledging the Dataverse Project team and GDCC members for their leadership and ongoing contributions.

Institutional Support

Harvard University and global community institutions provide crucial support to the Dataverse Project.

Training Program Recognition

Special thanks to Birdaro Training Program organizers for offering a platform to share knowledge and learn.

RESOURCES

[The Dataverse Project](#)

[The Global Dataverse Community Consortium \(GDCC\)](#)

[Dataverse Community Meetings](#)

[Dataverse Community Calls](#)

[Dataverse Working Groups](#)

