



Dataverse Community Meeting 2022

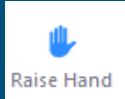
Updates from the Global Dataverse
Community Consortium (GDCC)



Zoom Etiquette during Breakouts



Keep muted during the meeting, unless you are ready to talk.



Use your raise hand icon if you have a question or want to participate. A moderator or a chair will call you when it's your turn.



When you are talking, consider turning on your video if it has been off, so others can see you and engage more easily with you.



You can also use the chat to ask questions or share some information, but keep in mind that we will be taking notes for the Breakout session in a Google Doc, and the chat is not the place to take notes.



We plan to record our Breakout session. This recording will be made public.

Outline

- Welcome (Jon)
- Introduction (Philipp)
- Sustainability and governance (Philipp)
- Membership update (Jon)
- Budget/Finances (Jon)
- Services update and overview (Jim)
- Community Discussion and Q&A

Welcome (Jon)

Introduction (Philipp)

Brief history/recap

- The Global Dataverse Community Consortium (**GDCC**) was introduced at the Dataverse Community Meeting (DCM) **2018**.
- In a breakout session, Jon Crabtree and Mercè Crosas shared initial ideas about GDCC's **governance structure** and about potential **services and experts groups** that could be coordinated by the Consortium.
- After some initial discussion, the GDCC **vision statement** has remained unchanged since 2018:

Vision Statement

The vision for the Global Dataverse Community Consortium (GDCC) is to provide **international organization** to existing community efforts and provide **collaborative venues** for institutions to leverage economies of scale in support of Dataverse repositories around the world. It is our intention that the consortium be **community driven** and ultimately the goals to be defined by the participants. We believe that promoting open and inclusive practices, formed in a **transparent environment** builds strong partnership among our members as well as with our external partner organizations. We encourage membership from organizations both large and small from diverse communities all around the globe.

Shaping the GDCC

At previous community meeting we got **valuable feedback and great ideas** about further development of GDCC.

Some of the ideas have been followed up, e.g.:

- **Coordinating community efforts** through interest groups and working groups, pull requests, resources provided through the GDCC GitHub Repository
- Supporting the community on **technical and organizational issues** through the GDCC Dataverse Repository Installation and Consulting Services, through GDCC support in discussion groups, on Slack, and other channels
- **Providing DOIs** through GDCC consortium membership in DataCite
- Organizing **community events outside Harvard and North America**, e.g., in 2020, we had the first European Dataverse Workshop, organized in Tromsø/Norway.

More details about GDCC activities in the past year later on in this session.

Shaping the GDCC (cont.)

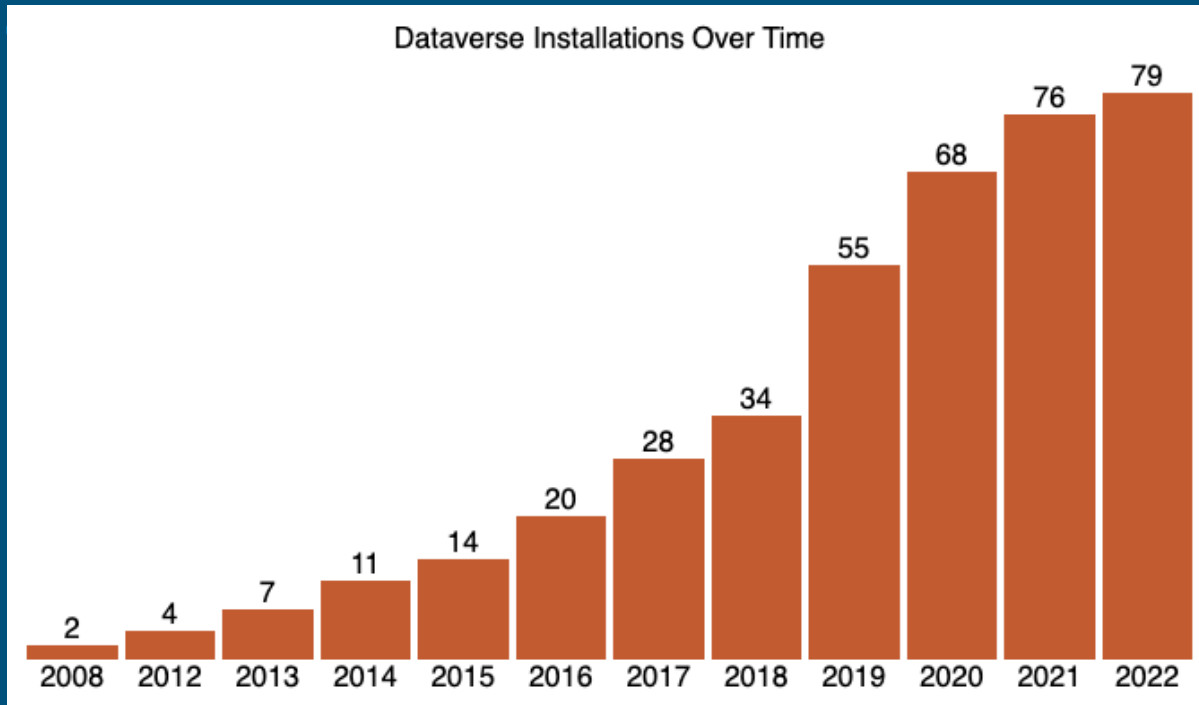
On the other hand, we're still working on more overarching aspects of the GDCC, such as:

- Governance structure of GDCC
- Sustainability of GDCC

However, in the past year, we have been putting some efforts into making further progress in this area as well.

Sustainability and governance (Philipp)

Dataverse is growing!



(Source: Phil Durbin on Dataverse Slack, June 6, 2022)

IQSS / **dataverse** Public

Issues 1.2k Pull requests 70

(Source: <https://github.com/IQSS/dataverse>, June 9, 2022)

How to meet challenges of growth?

- Need for **strengthening the sustainable development** of the Dataverse software and its ecosystem of associated tools and services
- How to further **empower the GDCC** to contribute to the Dataverse community with the support and services we need?
- First step: **survey the roadmaps and priorities** of Dataverse installations.
 - Dataverse Community Survey 2022

Dataverse Community Survey 2022

- Sent to existing and aspiring Dataverse installations in the beginning of May
- 34 responses
- 1 coordinated response per installation
- Some results* from the last part of the survey which is about sustainability and governance
- Goal of this part: figure out how important different resources are to ensure the sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools.

* All (non-personal) data will be published openly.

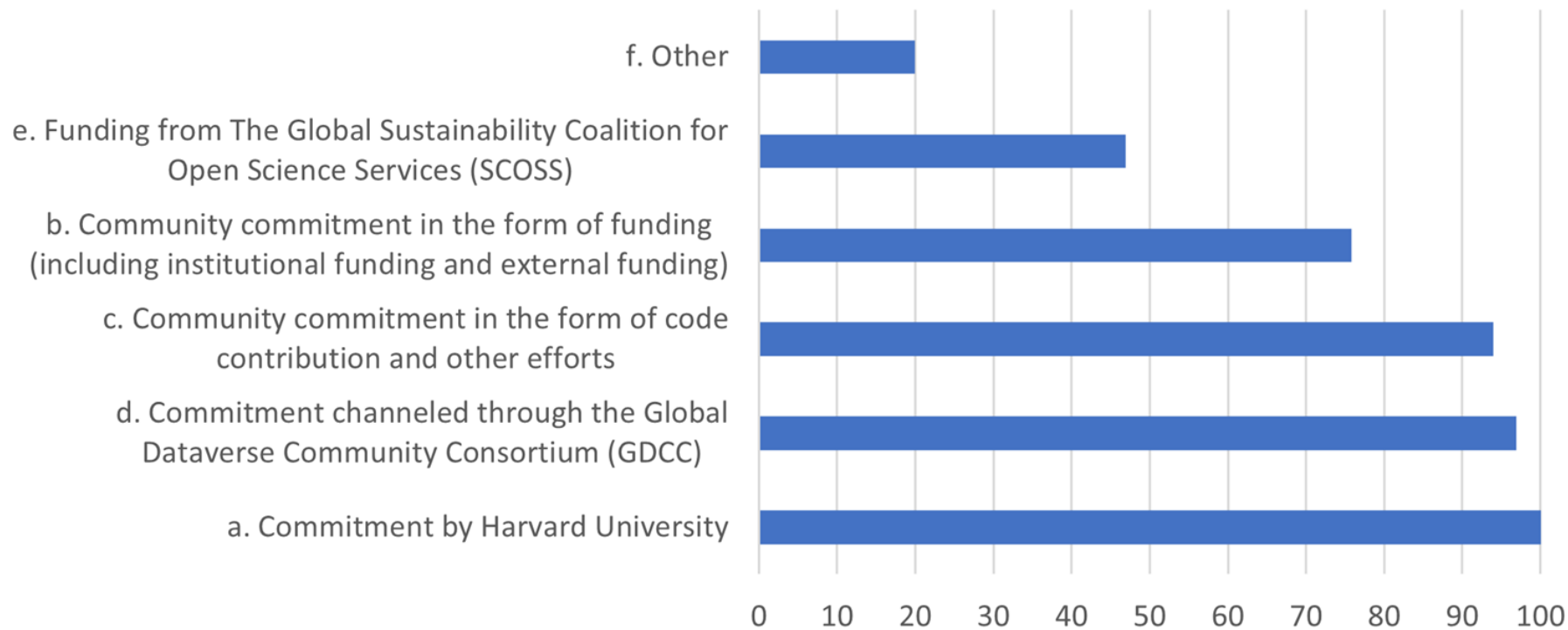
Q40. Sustainability

How important does your Dataverse installation consider the following existing and potential **sources** to ensure sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools?

Answer alternatives: Not applicable, Not important, Somewhat important, Neither nor, Important, Very important

If we merge Important and Very important, the following graph shows us the response distribution in percentages for the different sources:

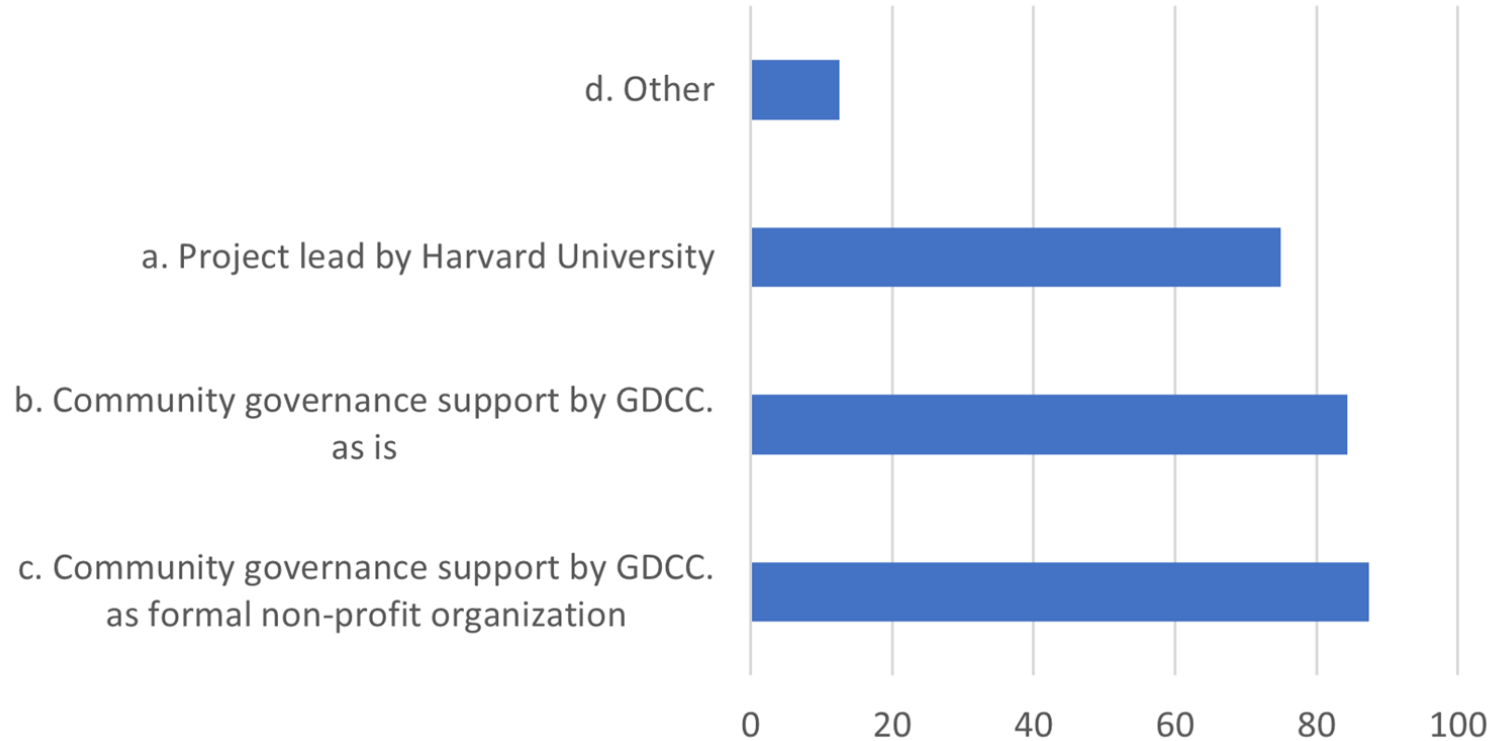
Important/Very important (%)



Q41. Governance

How important does your Dataverse installation consider the following existing and potential **governance support** for the sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools?

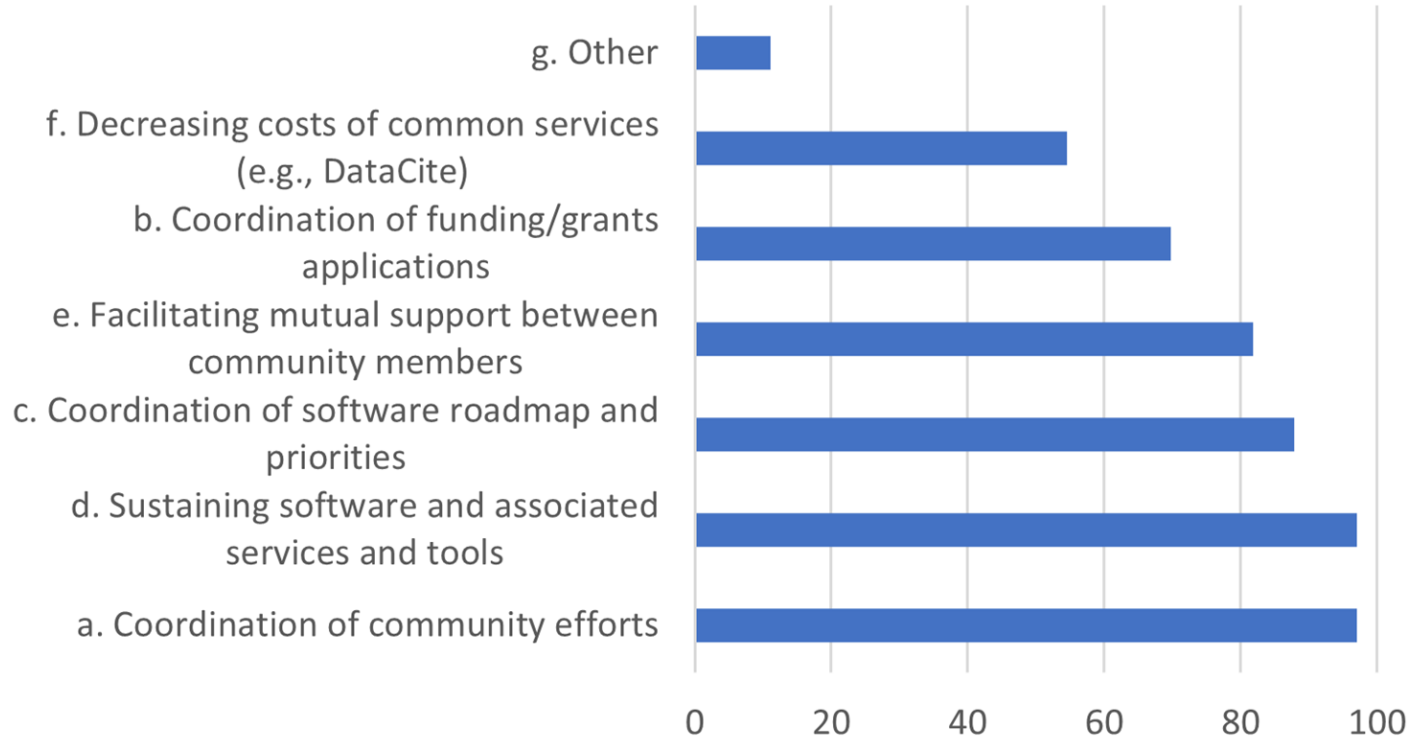
Important/Very important (%)



Q42. Governance Support by GDCC

How important is it for your Dataverse installation that the GDCC provides / should provide the following existing and potential **types of governance support** for the sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools?

Important/Very important (%)



Strengthening sustainability

- Use the results from the Dataverse Community Survey as a starting point for more **structured discussions** about the needs and ambitions of the Dataverse installations
- **Shared vision and roadmap** for the global Dataverse community
 - Beneficial for the community at large
 - Limiting divergence across installations and community efforts is key to the continuous success of the Dataverse project and the community.

Governance of GDCC

- **Steering committee:**
Jon Crabtree (ODUM, UNC), Steve McEachern (ADA), Len Wisniewski (IQSS, Harvard), and Philipp Conzett (UiT)
- **General Advisory committee** (under development):
Mercè Crosas + To be determined (with people external to the Dataverse Community, from CODATA, RDA, other repositories or data organizations)
- **Technical Advisory committee** (under development):
Jim Myers, Gustavo Durand + To be determined (with technical people external to the community)

Governance Changes

- Philipp Conzett (UiT) will serve as Chair until next community meeting
- Jon Crabtree (ODUM, UNC), will serve as Treasurer until next community meeting

Governance - Next Steps

Formalizing Steering Committee:

- Seeking volunteers for being nominated to the steering committee
- Planning elections at next community meeting (June 2023)

Investigating organizational status for GDCC

- 501c3 in USA
- Non-profit in Europe

Need your help and feedback. Add your comments to shared notes doc and/or in the discussion part at the end of the session!

Membership update (Jon)



Currently 34 members
from 17 countries

HOME /

GDCC Members

Arizona State University	United States	Living Goods	Kenya
Australian Data Archive	Australia	Massachusetts Institute of Technology	United States
CEDIA	Ecuador	Nanyang Technological University	Singapore
Consorci de Serveis Universitaris de Catalunya	Spain	Rede Nacional de Ensino e Pesquisa (RNP)	Brazil
Consortio Madrono	Spain	SciELO Dataverse Repository	Brazil
DANS	Netherlands	Scholars Portal	Canada
DataverseNO Repository	Norway	Syracuse University	United States
French National Center for Research	France	Texas Digital Library	United States
Fudan University	China	Universidad Nacional de Rosario	Argentina
Fundacao Getulio Vargas (FGV)	Brazil	University of California Los Angeles	United States
Fundacao Oswaldo Cruz (Fiocruz)	Brazil	University of Campinas	Brazil
Gottingen eResearch Alliance	Germany	University of North Carolina Chapel Hill	United States
Harvard University	United States	University Library Stuttgart	Germany
INRAE, French National Research Institute	France	University of Minho	Portugal
Institute for Adult Learning	Singapore	University of Virginia	United States
International Centre for Research in Agroforestry	Kenya		
Istituto Italiano di Tecnologia	Italy		
Johns Hopkins University	United States		
KU Leuven	Belgium		

Budget/finances (Jon)

Current Budget Numbers (June 2022)

Income \$ 179,463.33 *** including unpaid invoices

Expenses \$ 175,535.00

Current Accounts Receivables \$51,637 *** Unpaid memberships or contracts

Current Balance \$3,928.33 ***This includes ARs

***Fiscal Year July 2021-June 2022

Unpaid Memberships

Total Invoices Billed \$17,000.00

Total Invoices Paid \$10,948.00

Balance -\$6,052.00

Texas Digital Library **

Fudan University **

UCLA **

Johns Hopkins University

International Centre for Research in Agroforestry (ICRAF)

Australian Data Archive

RNP Brazil **

Living Goods **

INRAE French National Research Institute

University of Minho **

CORPORACION ECUATORIANA PARA EL DESARROLLO DE LA INVESTIGACION Y LA ACADEMIA -CEDIA:

Unpaid DataCite Invoices

Total Invoices Billed To Members \$ 29,540.00

Total Invoices Paid by Members \$22,700.00

Account Receivables from Members Due \$6,840.00

Fees Paid to DataCite \$24,735.00

Balance -\$2035.00

Unpaid

Syracuse University

Johns Hopkins University

International Centre for Research in Agroforestry (ICRAF)

RNP Brazil

Living Goods

University of Minho **

CORPORACION ECUATORIANA PARA EL DESARROLLO DE LA INVESTIGACION Y LA ACADEMIA -CEDIA:

Proposed Changes to Membership Fees 2023

Propose membership fee \$550

Due January 1st each year

Fees for wire transfer and credit card included in base membership

Proposed GDCC 2023 DOI Fee Schedule

Organizations using less than 2000 DOIs/yr	\$1,140
Organizations using 2000-9999 DOIs/yr	\$1,700
Organizations using 10,000- 99,999 DOIs/yr	\$4,800
Organizations using over 100,000 DOIs/yr	\$5,700

*** These include bank fees

*** Same as 2022

Services update and overview (Jim)

GDCC's Range of Community Services

A broad range of activities (augmenting IQSS's efforts):

- Supporting coordination of community development efforts
- Providing design guidance
- Initiating strategic redesign discussions/documents
- Development/maintenance related to community contributions/issues/projects
- Outreach/support in using new features, managing Dataverse installations

Community Coordination

Community coordination and unification efforts

- Community meetings
- GDCC Git Repositories (previewers, languages, DVUploader, anisible, pyDataverse, ext. vocabs, ...)
- GDCC Jenkins, AWS Infrastructure
- GDCC Working groups - slack, docs, meetings

Community Collaboration Projects

Installation Training and Support Projects

Texas Digital Library Installation and Programming Support

Istituto Italiano di Tecnologia Installation and Upgrade Support

“Coming Soon Johns Hopkins University Upgrade Support”

Community Collaboration Projects

Custom programming and consultations

French National Center for Research, SciencesPo Dataverse

Data Archiving and Networking Services (DANS)

Texas Digital Library (TDL)

Harvard Data Commons Project Support

”Coming Soon Harvard NIH Grant Support”

GDCC/Community Development

- Lots of Caveats in trying to quantify impact and differentiate between orgs/projects. That said, in the last year:
 - Range from major features (embargo, multiple license) to areas of focus (i18n, accessibility) to one shot bug fixes/updates
 - Pull Requests/Code contributions (not counting previewers /metrics /ansible /languages/ vocabs /other repos)
 - ~30% of completed from GDCC/Community
 - ~50% of open PRs from GDCC/community
 - >50% of commits from GDCC/Community
 - Security - community key in identifying/reporting/assessing/fixing issues

More information

Website: <https://dataversecommunity.global/> with more information, including

- how to join GDCC
- how to apply for DOIs through the GDCC DataCite consortium
- other services
- contact addresses

Reach out to us if you want to be added to GDCC/Dataverse Slack!

Community Discussion and Q&A
