



The Global Dataverse Community Consortium
Supporting Dataverse repositories Around the World

Updates from the Global Dataverse Community Consortium (GDCC)

Philipp Conzett, Jon Crabtree, Jim Myers

Dataverse Community Meeting 2023
University of Minho, Braga, Portugal
June 5-7, 2023

Outline

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

Contribute to the shared notes doc!

<https://tinyurl.com/gdcc2023>

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.

Who is the GDCC?

- US! THANKS FOR CONTRIBUTING! We urge institutions that use Dataverse to join and contribute to the GDCC.
 - So far, GDCC is run as an informal light-weight organization with several WGs, one contract employee, and a steering committee:
 - Jim Myers, Senior Developer and Architect for GDCC and community contributions
 - Steering Committee:
 - Philipp Conzett from UiT The Arctic University of Norway (chair)
 - Jonathan Crabtree from Odum Institute at the University of North Carolina (treasurer)
 - Ingrid Dillo from Data Archiving and Networked Services (DANS)
 - Steven McEachern from the Australian Data Archive
 - Stefano Iacus from the Institute for Quantitative Social Science at Harvard University
- Also attending meetings: Jim Myers (GDCC) and Ceilyn Boyd (IQSS, Harvard)

Shaping the GDCC

At previous community meetings 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, and right now, here in Porto/Portugal, we are having the first Dataverse Community Meeting outside Harvard.

Jim and Jon will provide more details about GDCC activities in the past year.

Shaping the GDCC (cont.)

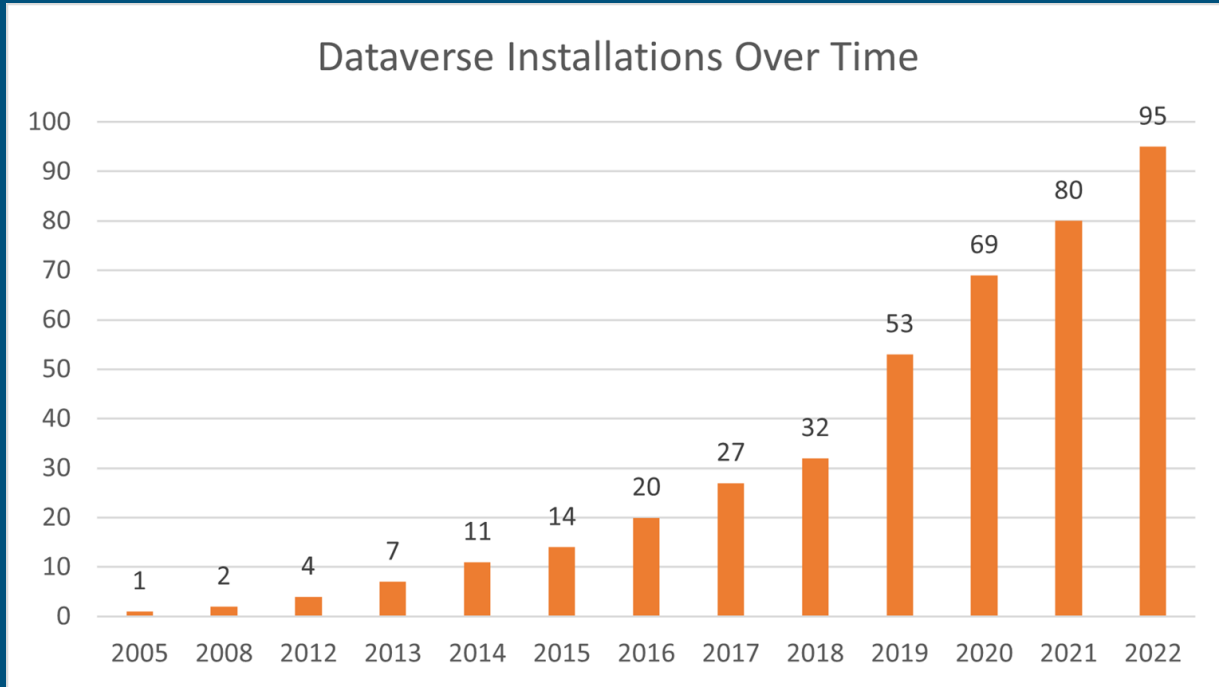
On the other hand, we're still working on more overarching aspects of the GDCC, such as:

- Governance structure of GDCC
- Sustainability of GDCC

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

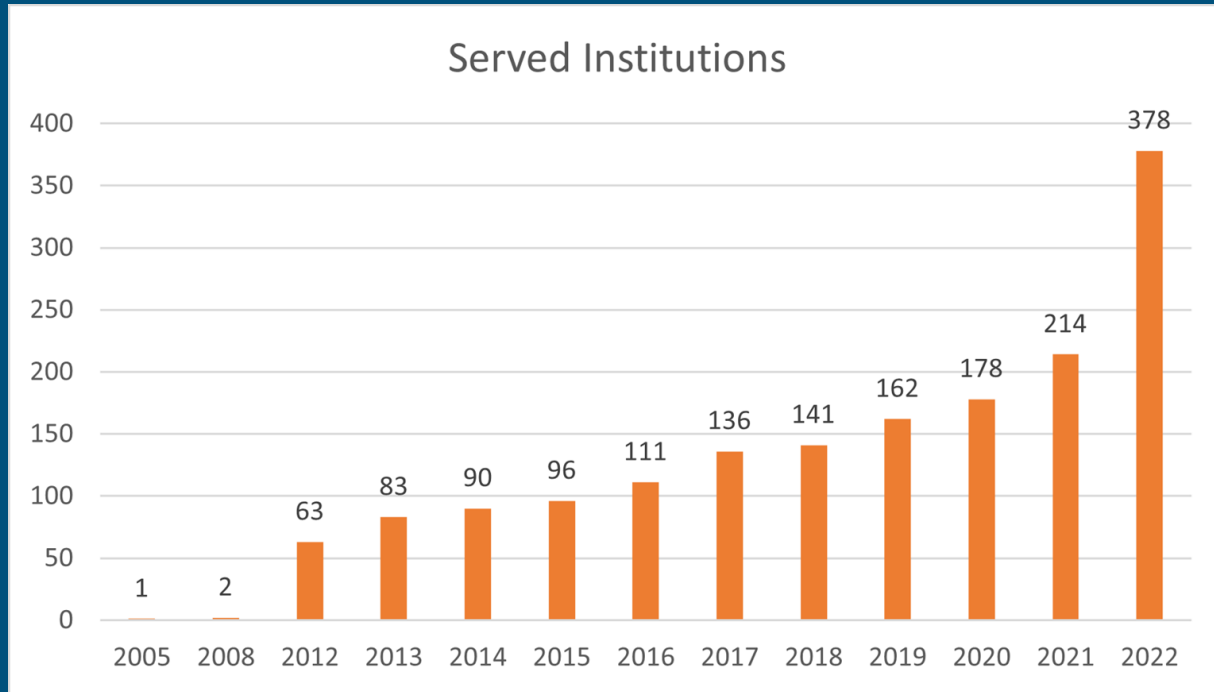
Sustainability and governance (Philipp)

Dataverse is growing!



(Based on: <https://iqss.github.io/dataverse-installations/>, May 6, 2023)

Dataverse is growing!



(Based on: <https://iqss.github.io/dataverse-installations/>, and installation web pages May 6, 2023)

Dataverse is growing!



(Source: <https://github.com/IQSS/dataverse>,
May 6, 2023)

Averaging over 430 Pull Requests merged per year!

Over 300 PRs on Dataverse and the new front-end since November 2022!

100 contributors since 2014, > $\frac{1}{3}$ active in the last year!

> 20 additional repositories with related tools, components active in the last year!

Challenges of growth

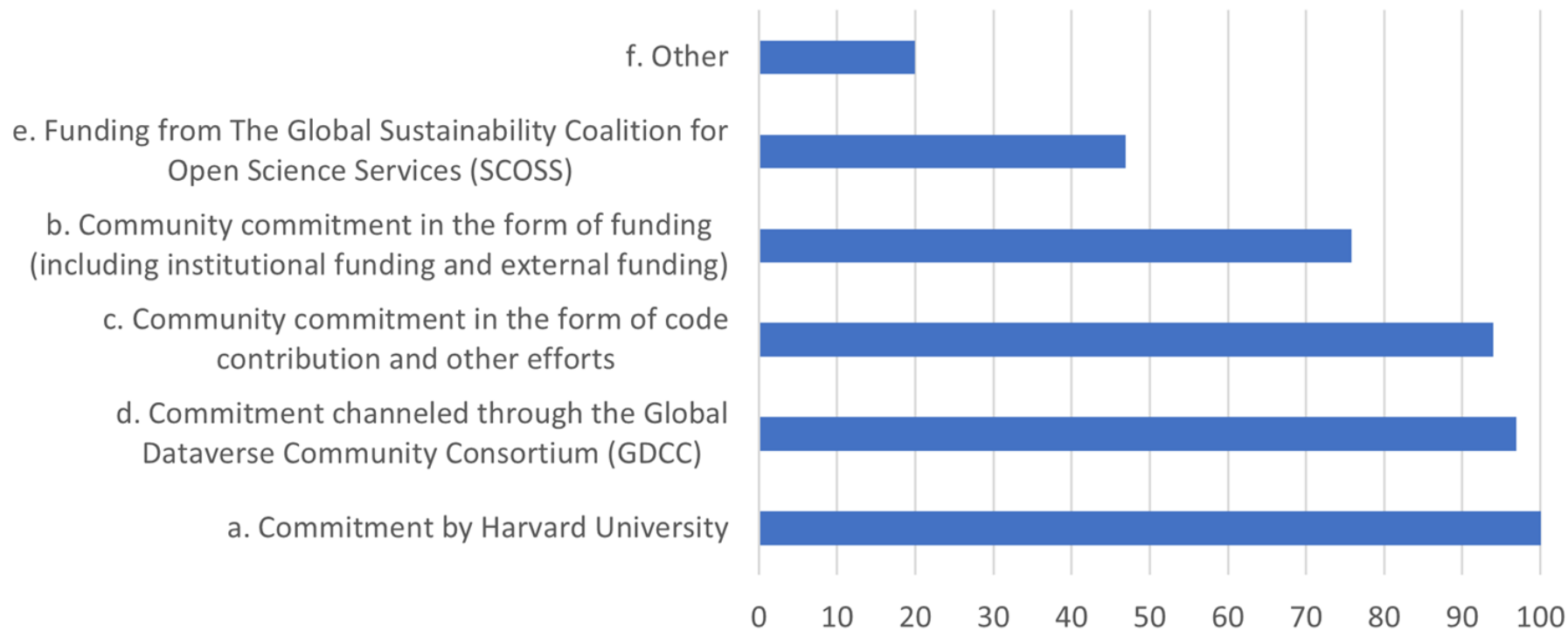
- How to make sure the Dataverse software and its ecosystem of associated tools and services are **developed and maintained in a sustainable way?**
- How to further **empower the GDCC** to contribute to the Dataverse community with the support and services we need

How to meet challenges of growth?

What have we done so far?

- **Surveyed the roadmaps and priorities** of Dataverse installations.
 - Dataverse Community Survey 2022
 - Questions on sustainability and governance
 - Large agreement in the community that
 - Harvard and GDCC are importance sources to ensure sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools;

Important/Very important (%)

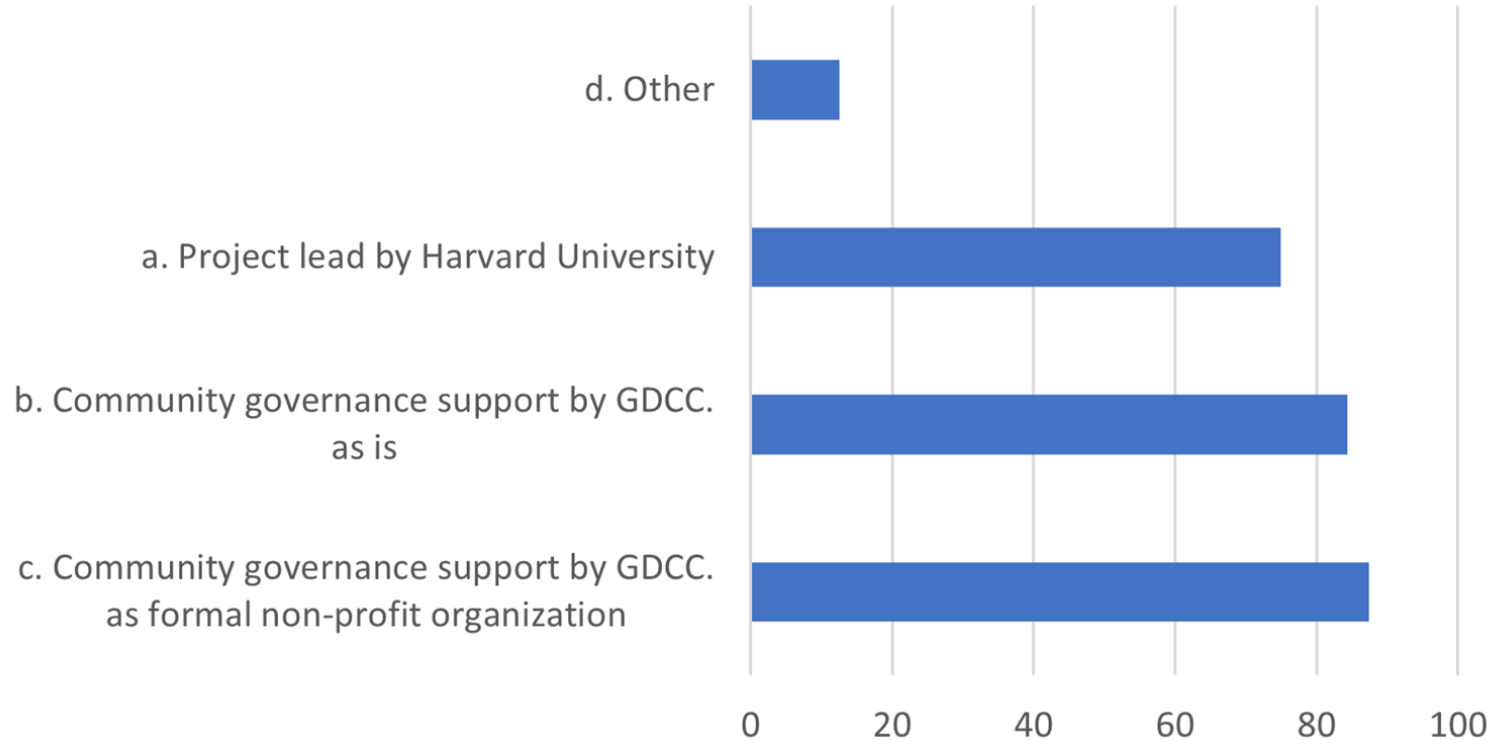


How to meet challenges of growth?

What have we done so far?

- **Surveyed the roadmaps and priorities** of Dataverse installations.
 - Dataverse Community Survey 2022
 - Questions on sustainability and governance
 - Large agreement in the community that
 - Harvard and GDCC are importance sources to ensure sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools;
 - **best governance support for the community could be achieved by GDCC as a formal, non-profit organization.**

Important/Very important (%)



How to meet challenges of growth?

What have we done so far?

- **Surveyed the roadmaps and priorities** of Dataverse installations.
 - Dataverse Community Survey 2022
 - Questions on sustainability and governance
 - Large agreement in the community that
 - Harvard and GDCC are importance sources to ensure sustainable development and maintenance of the Dataverse software and the ecosystem of associated services and tools;
 - best governance support for the community could be achieved by GDCC as a formal, non-profit organization.
 - **Survey report available at <https://doi.org/10.7557/7.6872>.**
 - **Raw data available at <https://doi.org/10.18710/UOC8CP>.**

How to meet challenges of growth?

What have we done so far?

- Surveyed the roadmaps and priorities of Dataverse installations.
- **Applied for funding to professionalize the GDCC and to transform it into a non-profit organization. So far not successful.**

How to meet challenges of growth?

How to move on?

- Need to take a step back and include more members of the community to work on the sustainability of GDCC, or more generally the Dataverse software and its ecosystem of integrated services and tools.

How to meet challenges of growth?

How to move on?

- Need to take a step back and include more members of the community to work on the sustainability of GDCC, or more generally the Dataverse software and its ecosystem of associated tools and services. This includes
 - Assessing and reassessing the **needs of the community**.
 - Addressing **different aspects of open-source software sustainability** in a structured way, e.g., as described in the Sustainability Wheel created by the It Takes a Village project:

- Governance model
- Roles and responsibilities
- Rules for participation
- Strategic and tactical decision making
- Communication

- Facilitating and fostering involvement within the community
- Encouraging users to become stakeholders and champions
- Communication and outreach



- Architecture choices
- Rules for contribution
- Maintenance
- Development
- Technical debt
- Refactoring
- Re-architecture

- Human resources
- Organizational homes with fiscal stewardship
- Financial resources

Figure: Arp, Laurie Gemmill & Megan Forbes. 2022. It Takes a Village: Open Source Software Sustainability. LYRASIS. Used under a CC BY 4.0 license.

How to meet challenges of growth?

What are we suggesting we should do next?

- Establish a **working group** dealing with the sustainability of Dataverse / the community.
- Join the Working Group session on **Wednesday 15:00-16:00** to discuss how to get started with work group!
- Add your comments to the [shared notes doc for the WG](#) and/or speak up in the discussion part at the end of the session.

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
- Adding to the Core Team (Jim Myers, Don Sizemore, Oliver Bertuch), reviewing community PRs, helping with CI, infrastructure
- Outreach/support in using new features, managing Dataverse installations

Community Coordination

Community coordination and unification efforts

- Community meetings
- Providing/maintaining/contributing to GDCC Git Repositories (dvwebloader, exporters, easyDataverse, xoai, previewers, languages, DVUploader, dataverse-ansible, pyDataverse, ext. vocabs, ...)
- GDCC Jenkins, AWS Infrastructure
- GDCC Working groups - slack, design 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

IAL Singapore

Johns Hopkins University Upgrade Support

Initial discussions with US Census, US DOE, AUSSDA, Yale, ...

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

DataverseNO

Johns Hopkins University

Harvard NIH Grant

Australian Data Archive (ADA)

GDCC/Community Development

- Lots of Caveats in trying to quantify impact and differentiate between orgs/projects. That said, in the last year:
 - Range from new features (Plug-in Exporter support, Improved Performance for Large Datasets, PermaLinks, Signposting, Globus/remote storage support, Folder/Tag sorting of files, Folder upload, signedURLs for API, Protected Metadata) to areas of focus (i18n, accessibility, archiving) to one shot bug fixes/updates
 - Pull Requests/Code contributions (not counting other repositories, e.g. SPA, previewers, metrics, ansible, languages, vocabs, dvwebloader, DVUploader, ...)
 - ~2/3 of completed from GDCC/Community (27 people)
 - ~85% of open PRs from GDCC/Community
 - ~2/3 of commits from GDCC/Community (~<1/3 contracted, ~>1/3 volunteered) (~50% counting SPA)
 - Security - community key in identifying/reporting/assessing/fixing issues

Membership update (Jon)



Currently 35 members
from 17 countries

HOME /

GDCC Members

Arizona State University	United States
Australian Data Archive	Australia
CEDIA	Ecuador
Consorci de Serveis Universitaris de Catalunya	Spain
DANS	Netherlands
DataverseNO Repository	Norway
Deloitte Research & Insights	USA
French National Center for Research	France
Fudan University	China
Fundacao Getulio Vargas (FGV)	Brazil
Fundacao Oswaldo Cruz (Fiocruz)	Brazil
Gottingen eResearch Alliance	Germany
Harvard University	United States
INRAE, French National Research Institute	France
Institute for Adult Learning	Singapore

International Centre for Research in Agroforestry	Kenya
Istituto Italiano di Tecnologia	Italy
Johns Hopkins University	United States
KU Leuven	Belgium
Living Goods	Kenya
Massachusetts Institute of Technology	United States
Nanyang Technological University	Singapore
Rede Nacional de Ensino e Pesquisa (RNP)	Brazil
SciELO Dataverse Repository	Brazil
Scholars Portal	Canada
Syracuse University	United States
Texas Digital Library	United States
Universidad Nacional de Rosario	Argentina
University of California Los Angeles	United States
University of Campinas	Brazil
University of North Carolina Chapel Hill	United States
University Library Stuttgart	Germany
University of Minho	Portugal
University of Virginia	United States
Yale University	USA

Budget/finances (Jon)

2022 Summary Budget Numbers

2022 Income \$ 191,110.01

2022 Expenses \$ 186,689.06

2022 End of Year Accounts Receivables \$2,140 *** Unpaid memberships or contracts

2022 End of Year Balance \$6,560.95 ***This includes ARs

2022 Memberships Total 34

Total Invoices Billed \$17,000.00

Total Invoices Paid \$16,000.00

Balance -\$1,000.00

Non-Payment

International Centre for Research in Agroforestry (ICRAF)

Australian Data Archive

2022 DataCite Invoices

Total Invoices Billed To Members \$ 29,540.00

Total Invoices Paid by Members \$28,400.00

Account Receivables from Members Due \$1,140.00

Fees Paid to DataCite \$24,735.00

Balance \$3,665.00

Unpaid

International Centre for Research in Agroforestry (ICRAF)

2023 Summary Budget Numbers Thru May 31

2023 Income \$ 121,636.50

2023 Expenses \$ 111,085.70

2023 Current Accounts Receivables \$18,538.50 Unpaid memberships or DataCite

2023 Current Balance \$33,510.25 ***This includes ARs and \$4,420.95 carry over from 2022

***Some additional contract ARs exist and under investigation

2023 Memberships Total 35

Total Invoices Billed \$17,500.00

Total Invoices Paid \$9,461.50

Balance -\$8,038.50

2023 DataCite Invoices

Total Invoices Billed To Members \$ 24,720.00

Total Invoices Paid by Members \$13,920.00

Account Receivables from Members Due \$10,800.00

Fees Paid to DataCite \$20,520.00

Balance -\$6,600 until ARs collected

Contracting Service June 2022-May 2023

Invoiced \$199,867.50 in contracts and services

Received \$141,055.50 in payments

Current Contract Account Receivables \$58,812***

*** We are still tracking some things down and I think some have been paid but not credited to the correct account due to issues at UNC.***

Proposed Changes to Membership Fees 2024

Propose membership fee \$550

Due January 1st each year

Fees for wire transfer and credit card included in base membership

Proposed GDCC 2024 DOI Fee Schedule

Organizations using less than 2000 DOIs/yr	\$1,200
Organizations using 2000-9999 DOIs/yr	\$1,800
Organizations using 10,000- 99,999 DOIs/yr	\$4,900
Organizations using over 100,000 DOIs/yr	\$5,800

*** These include bank fees

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!

Join us in the WG session on **Wednesday 15:00-16:00** to get started with the Dataverse Sustainability WG!

Community Discussion and Q&A
