

DeiC Dataverse

- Building a National TDR

DeiC Conference, 27-10-2022

Mikkel Ohm Søndergaard
University of Copenhagen
mikkel.ohm@adm.ku.dk

KØBENHAVNS UNIVERSITET



Agenda

- About Dataverse
- Establishing and operating a national TDR

About Dataverse

- Dataverse is an open-source platform to publish, cite, and archive research data
- Built to support multiple types of data, users, and workflows
- Main goal of core code is to focus on publishing (citing, sharing, versioning, etc.), FAIR Data principles
- Robust APIs to allow interoperability with “external tools” and other repositories / software
- Developed at Harvard’s Institute for Quantitative Social Science (IQSS) since 2006
- DeIC Dataverse – launch expected Q2 2023



Open source research data repository software



Researchers

Enjoy full control over your data. Receive *web visibility, academic credit, and increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. [Want to set up your personal Dataverse collection?](#)



Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. [Want to find out more about journal Dataverse collections?](#)



Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community’s data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. [Want to install a Dataverse repository?](#)



Developers

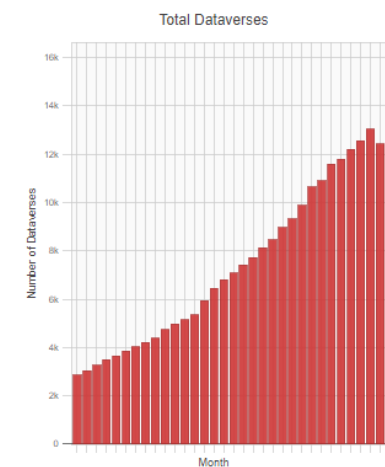
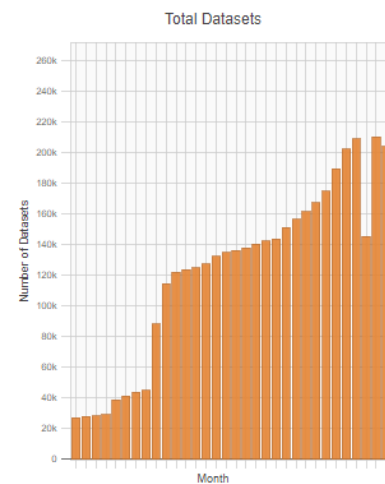
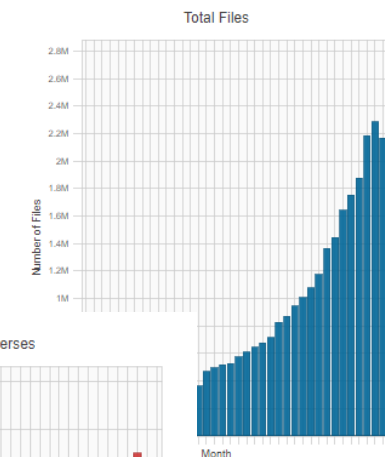
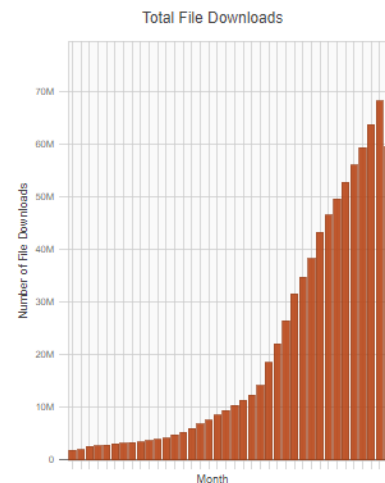
Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization and exploration tools*, or other research and data archival systems with the Dataverse Project. [Want to contribute?](#)

The global Dataverse community

<https://dataverse.org/>

DATAVERSE REPOSITORIES - A WORLD VIEW

89 Installations



The global Dataverse community



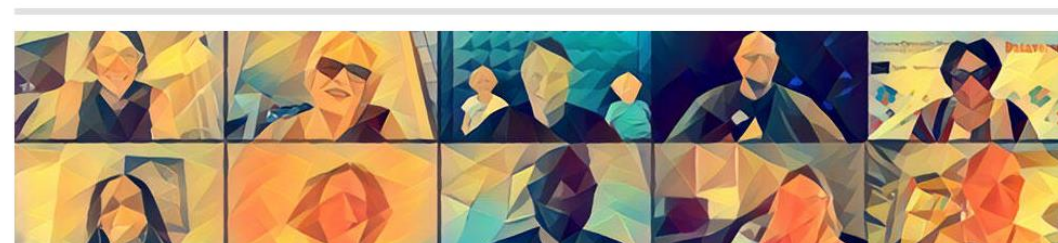
The Global Dataverse Community Consortium
Supporting Dataverse repositories Around the World

<https://dataversecommunity.global/>

Dataverse Community Meeting 2022
Held Remotely on June 14, 15, and 16



Home Agenda Speakers and Chairs Registration



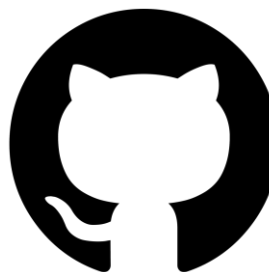
<https://projects.iq.harvard.edu/dcm2022>

Dataverse Users Community

Welcome to the Dataverse Users Community Group! Please feel free to ask a question, share feedback or start a discussion about anything.

Avatar	Author	Topic
	Ken Mankoff, Durand, Gustavo 6	Fail to publish; Cannot delete – In some ways, it's investigation vs needin
	Patrick Promitzer, ... Vyacheslav Tikhonov 7	Is there an external Ingest or Workflow, that triggers at file upload? – Hi
	fooba...@gmail.com, Philip Durbin 3	search query length limit – Thanks for opening https://github.com/IQSS/
	Taki NakaMura, ... James Myers 6	Is it possible to restrict file access or setting up guestbook by default –
	Bla Blubb, ... Philip Durbin 6	Use case: data distribution platform, prohibit external upload – Great! W
	Lily Falk, Philip Durbin 2	pyDataverse API simplejson.errors.JSONDecodeError – Hmm, I just ran c

<https://groups.google.com/g/dataverse-community>



Contributors 144



+ 133 contributors

<https://github.com/IQSS/dataverse>



<https://iqss.github.io/dataverse-tv/>

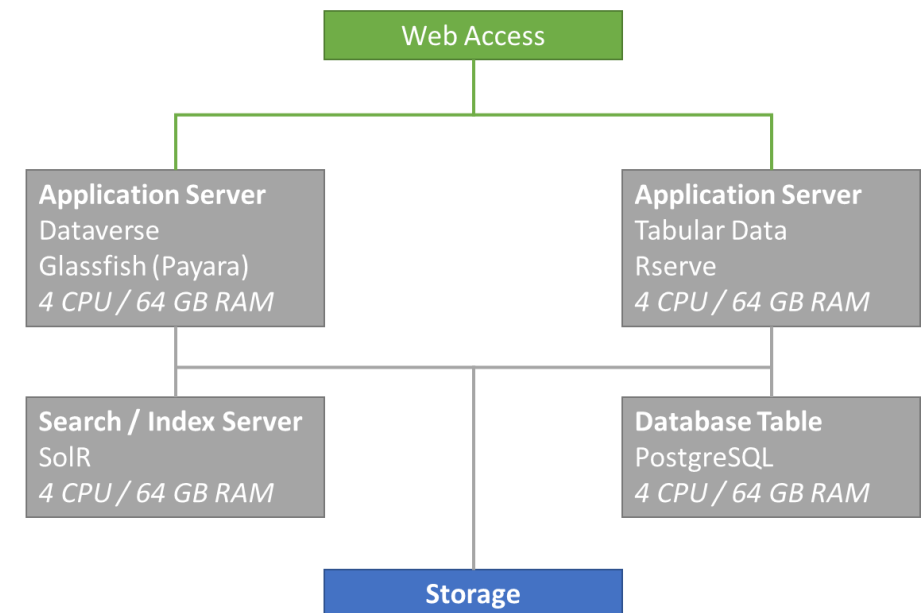
Overview of the application



<https://dataverse.org/>

The Dataverse Software is a Java Enterprise Edition (EE) web application.

- Linux (RHEL)
- Payara 5 – *application server*
- Java 11
- PostgreSQL – *relational database*
- Solr – *search engine*
- Local / Swift / S3 – *file storage*



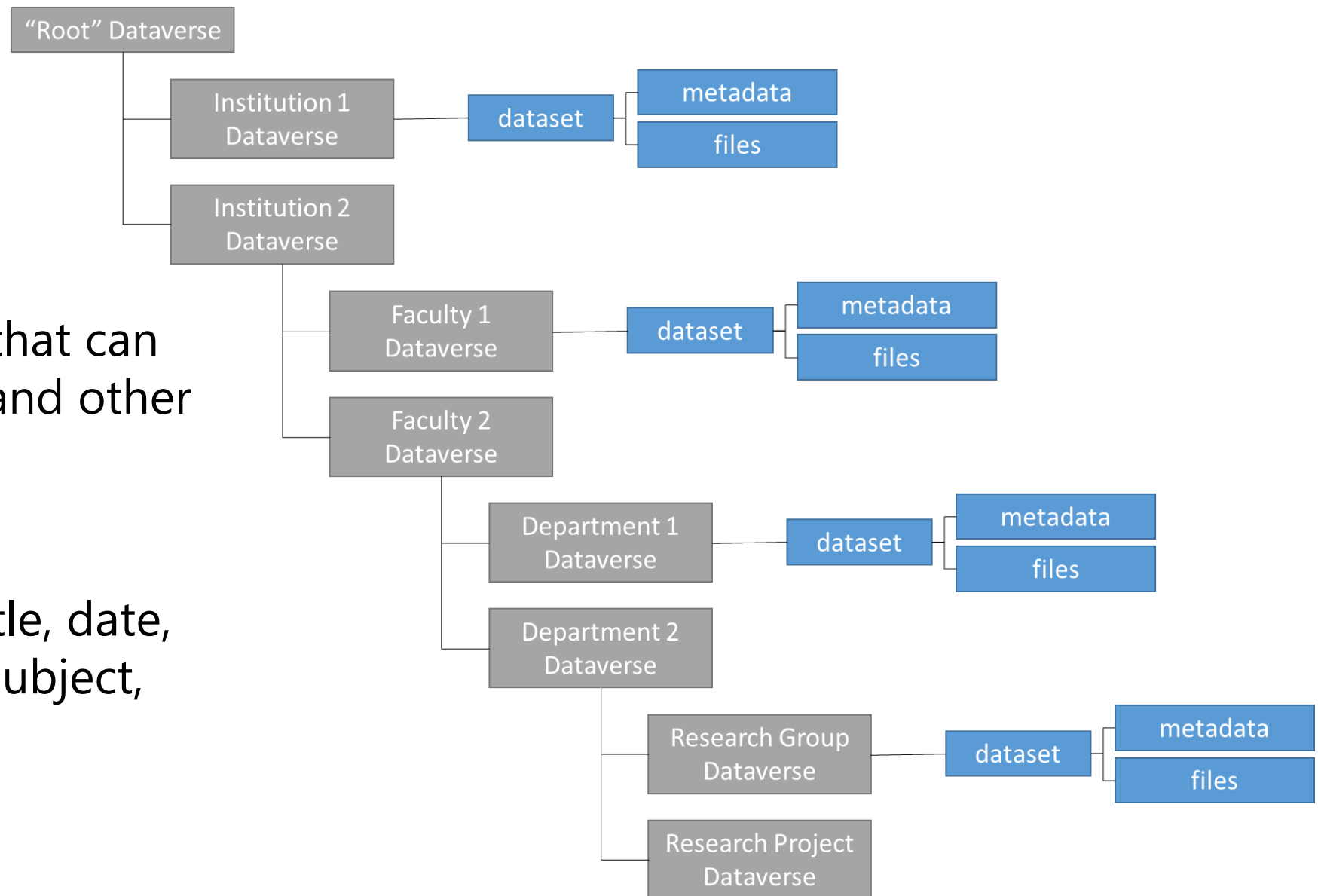
<https://guides.dataverse.org/en/latest/installation/index.html>

"Dataverse"

- collection/folder that can contain datasets and other Dataverses

"Dataset"

- Metadata (DOI, title, date, author, abstract, subject, terms of use, ...)
- Files (optional)



Each Dataverse can be administrated separately:

- Roles and permissions
- Curation workflow
- Metadata templates
- Search facets
- Guestbooks

Permissions Current access configuration to your dataverse.

Select if all users or only certain users are able to add to this dataverse, by clicking the Edit Access button. [EDIT ACCESS](#)

Who can add to this dataverse?

Anyone with a Dataverse account can add sub dataverses and datasets

When a user adds a new dataset to this dataverse, which role should be automatically assigned to them on that dataset?

Curator - Edit metadata, upload files, and edit files, edit Terms, Guestbook, File Restrictions (Files Access + Use), Edit Permissions/Assign Roles + Publish

Users/Groups All the users and groups that have access to your dataverse.

[ASSIGN ROLES TO USERS/GROUPS](#)

2 Users/Groups

User/Group Name (Affiliation) ^	ID v	Role v	Action
Anyone with a Dataverse account	:authenticated-users	Dataverse + Dataset Creator	✕ REMOVE ASSIGNED ROLE
Falco KUB admin	@Falco_KUBadmin	Admin	✕ REMOVE ASSIGNED ROLE

Metadata Fields

Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

- Use metadata fields from KU Data Repository - Dataverse Version 5.5
- Citation Metadata (Required) [\[+\] View fields](#)
- Geospatial Metadata [\[+\] View fields](#)
- Social Science and Humanities Metadata [\[+\] View fields](#)
- Astronomy and Astrophysics Metadata [\[+\] View fields](#)
- Life Sciences Metadata [\[+\] View fields](#)
- Journal Metadata [\[+\] View fields](#)
- Departments at the University of Copenhagen [\[+\] View fields](#)

Datasets can be shared with a “private URL” before publication.

Metadata are always openly available after publication.

Access to each individual file in a public dataset can be:

- Open
- Restricted
- Embargoed
- Closed



Use cases

- Publication of data underlying a submitted manuscript
- Archival of data from a PhD project
- Publication of metadata for sensitive data
- Publication of data with restricted access to files
- Publication of metadata for physical objects
- Registration of data stored elsewhere
- Automated processes for file upload and metadata creation

Establishing a national TDR

- DeiC call for a national solution
- Strategy for national cooperation on digital research infrastructure

Project purpose

- The purpose of the project is to establish a trusted data repository available to all researchers and Ph.D.- students across the universities in Denmark
- And hereby offer an easy and flexible way for researchers and Ph.D.- students to publish, share and archive their research data so they fit the criteria for open and FAIR data

3 main deliverables

Establishing the platform:

Containing design, development, documentation and technical implementation of the DeiC Dataverse platform and technical infrastructure to fit demands of a national TDR.
Establishing the necessary linkage to the national solution for storage.

Establishing the governance model:

Containing organizational structure for operation and support with a central back office and local front offices. Developing processes for operations, maintenance and further development of the platform.

Implementation:

... of the governance model and go-live of the platform.
Training of front officers.
Go-live procedure (Check and Approve)

- The project does not deliver end-user implementation!

Project timeline

Establishing the project:

- Contractual negotiations
- System design
- Basic infrastructure

01/08

01/01

30/06

Go-Live DeiC Dataverse 1.0:

Build and development:

- Technical development
- Governance structures and models
- Test

