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

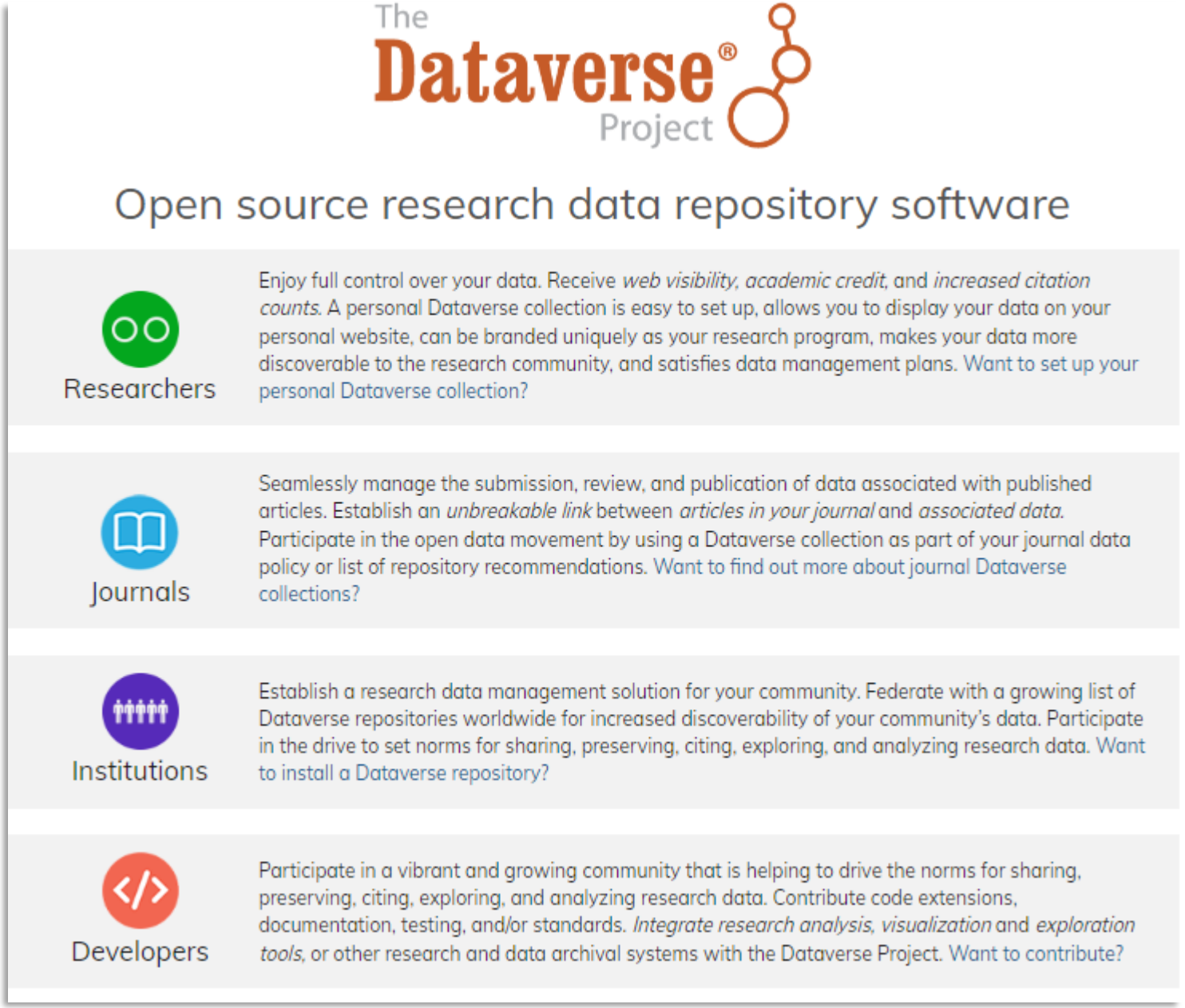
- About Dataverse
- Establishing and operating a national TDR

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

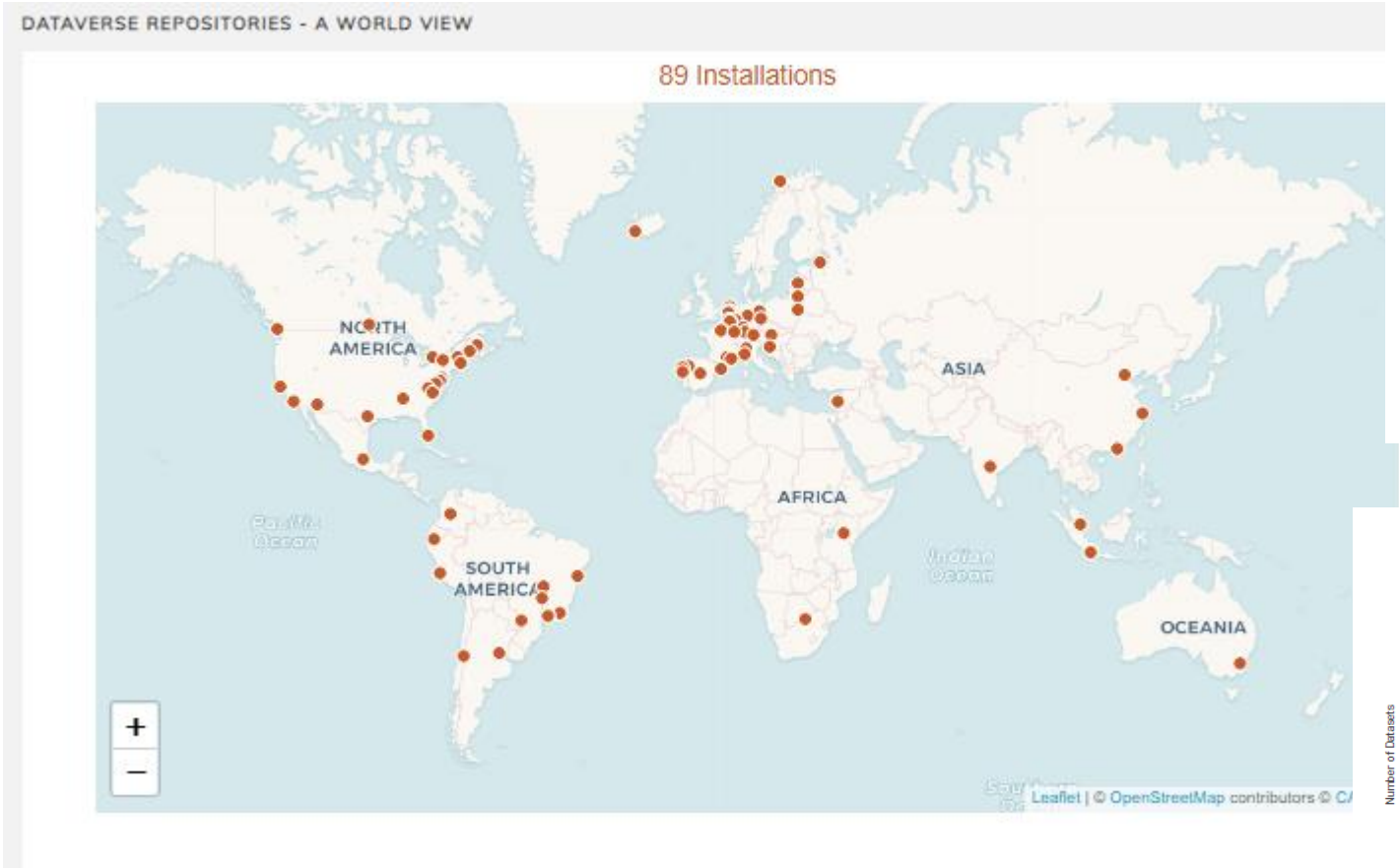


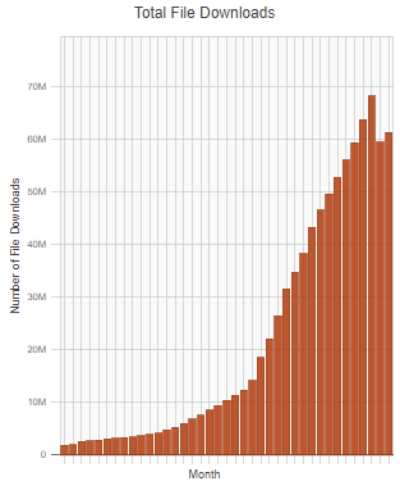
https://dataverse.org/

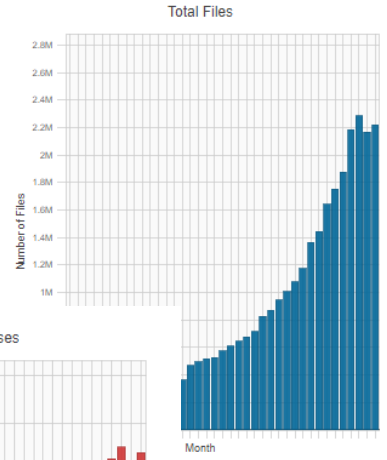
The global Dataverse community

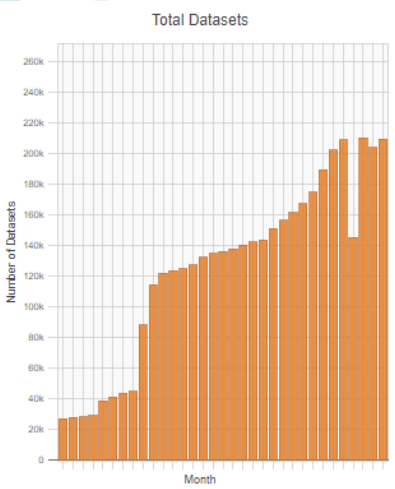
The Dataverse® Project

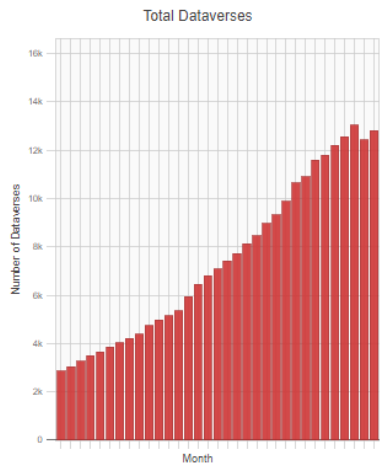
https://dataverse.org/









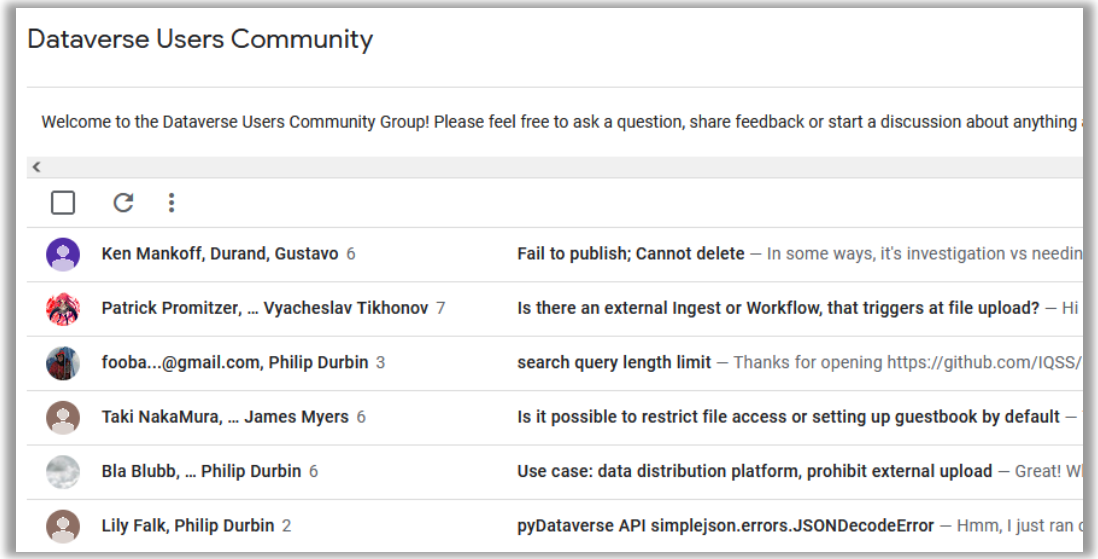


The global Dataverse community

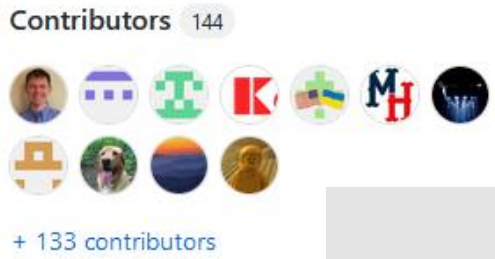
GDCC The Global Dataverse Community Consortium
Supporting Dataverse repositories Around the World

https://dataversecommunity.global/

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



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



https://github.com/IQSS/dataverse

Dataverse TV

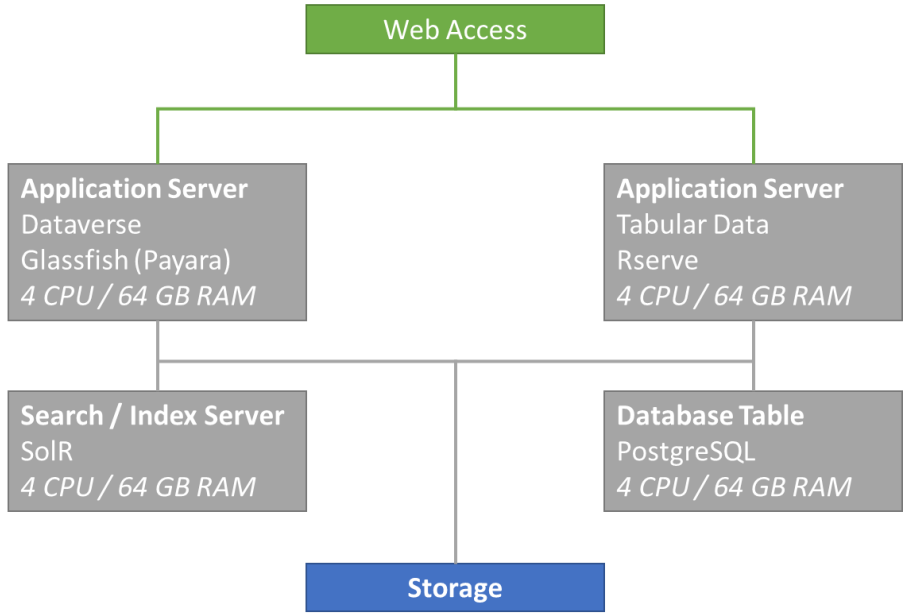
https://iqss.github.io/dataverse-tv/

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*



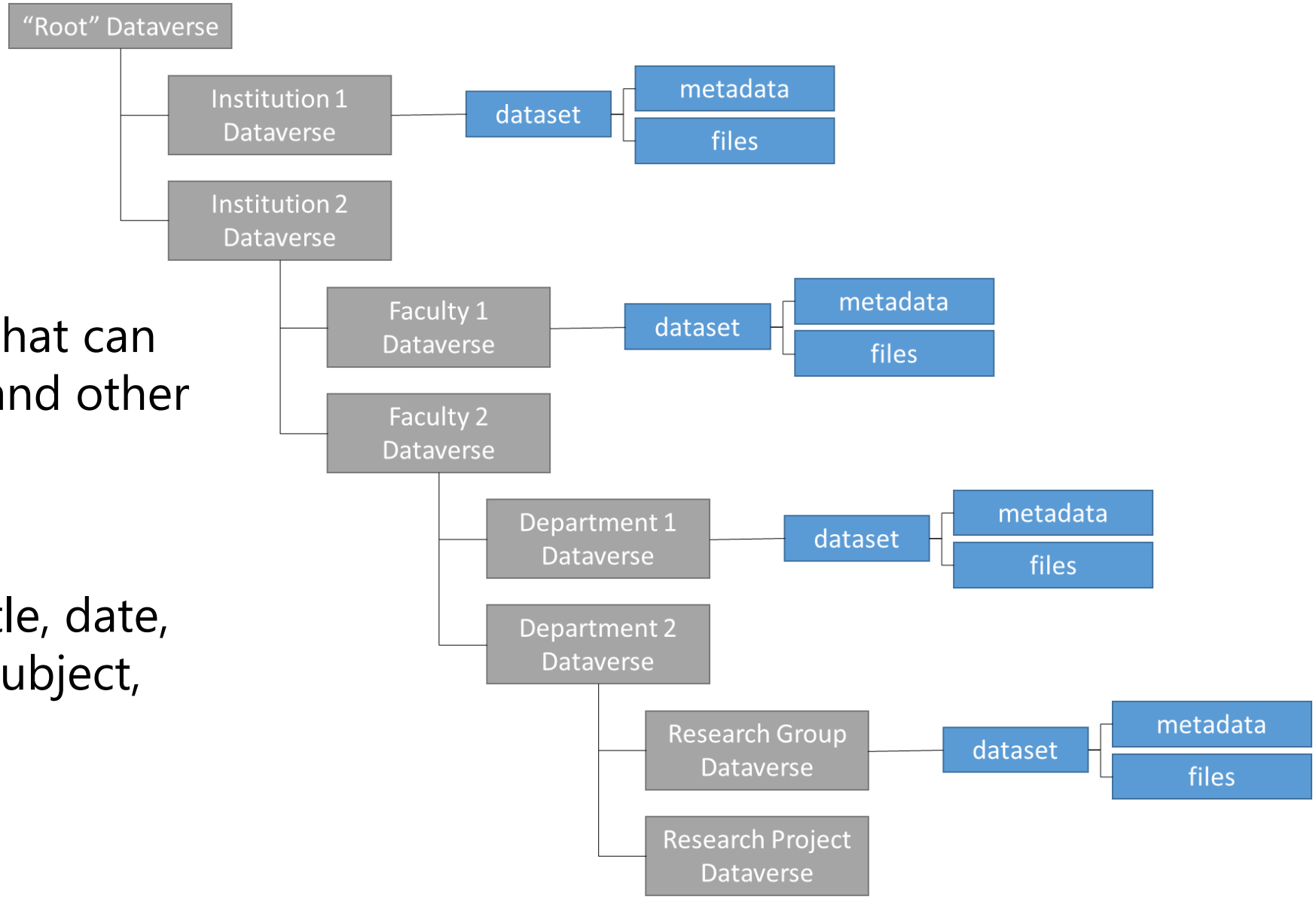
https://guides.dataverse.org/en/latest/installation/index.html

"Dataverse"

- collection/folder that can contain datasets and other Dataverses

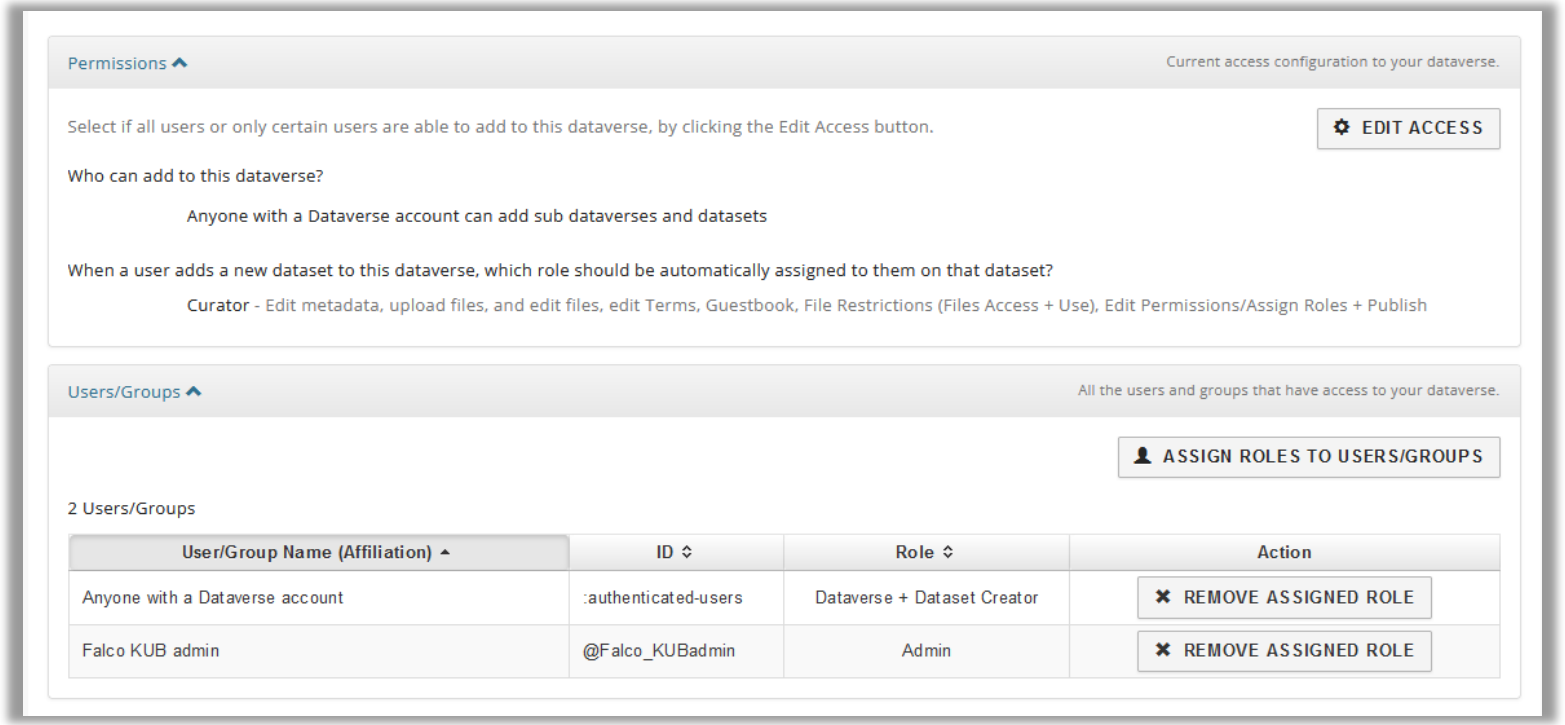
"Dataset"

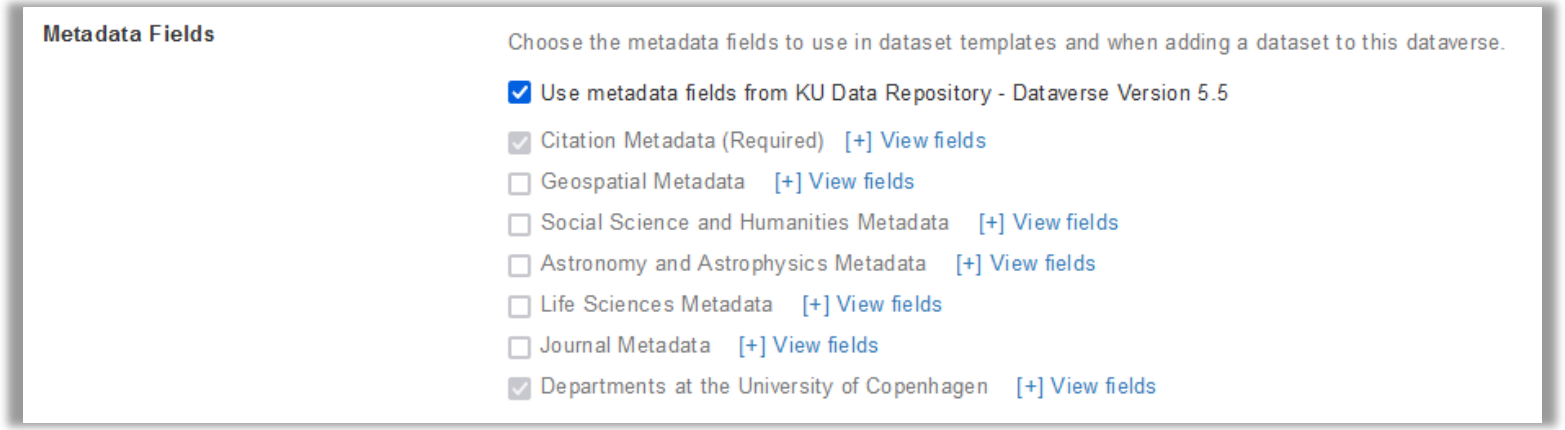
- Metadata (DOI, title, date, author, abstract, subject, terms of use, …)
- Files (optional)



Each Dataverse can be administrated separately:

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





Datasets can be shared with a "private URL" before publication.

Metadata are always openly available after publication.

Acces 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:

Containig design, development, documentation and technical implementation of the DeiC Dataverse platform and technical infrastructure to fit demands of a national TDR.
Establishing the neccessary 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.
Develoing processes for operations, maintenance and 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

Build and development:

Technical development

Governance structures and models

Test

30/06

Go-Live DeiC Dataverse 1.0: