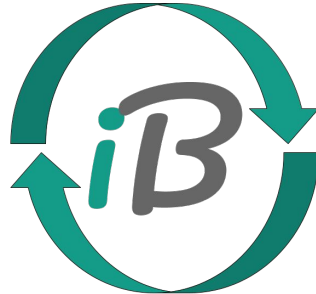


iBridges Dataverse

Christine Staiger
Utrecht University



Why integrate iRODS with Dataverse?

- iRODS systems (e.g. **Yoda**) store active research data
- **Dataverse** is widely used for data publication
- Researchers want a smooth path from **working storage** → **draft dataset**

Goal:

Enable users to *push* data from iRODS into a **draft Dataverse dataset**, while retaining enough information to ensure:

- **File integrity** (checksums)
- **Basic provenance** between the two systems

The workflow

UU Science
(Utrecht University)

DataverseNL (demo) > Utrecht University >

Contact Share

Search this dataverse... Advanced Search + Add Data

1 to 8 of 8 Results

Sort

☒ Datasets (7)

☒ Files (2)

Dataverse Category
Organization or Institution (1)

Test Draft Unpublished
Aug 1, 2025
Ich, Me; Ich, Me, 2025, "Test", <https://doi.org/10.80227/PDVNL/CIJWCQ>, DataverseNL (demo), DRAFT VERSION
bla

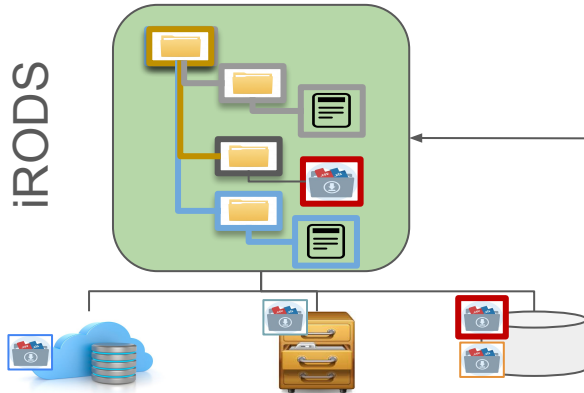
Push

- Download files from iRODS
- Upload to data to DVN dataset

Create dataset

Browse

Stage files for DVN dataset
Delete files from staging



iBridges

iBridges Plugin: Commands and GUI

- Manage Dataverse instances:
 - dv-setup: Print existing Dataverse configurations or create new ones.
 - dv-init: Store a Dataverse API token.
 - dv-switch: Switch to another Dataverse configuration by URL or alias.
- Monitoring/Managing status of drafts:
 - dv-draft: Show draft datasets and optionally delete them.
 - dv-status: Show pending uploads and draft datasets.
 - dv-cleanup: Clean up local Dataverse logs.
 - dv-add-file: Stage iRODS files for upload to a Dataverse dataset.
 - dv-rm-file: Remove a file from the staging area.
- Interaction with Dataverse:
 - dv-create-ds: Create a new Dataverse dataset.
 - dv-push: Upload staged files to Dataverse.

Advantages of the lightweight approach

- Fast to implement
- Easy to align workflow with user expectations
- No major changes required in iRODS or Dataverse
- **Checksum** verification enforced on both sides
- **GUI logging** provides a minimal provenance trail

Example log entries:

2025-07-30 17:19:28,304 - ibridges-gui - INFO - DATAVERSE: Created Dataset 10.80227/PDVNL/NIMZNA

2025-07-30 17:14:44,396 - ibridges-gui - INFO - DATAVERSE: Download /uu/home/research-christine/LICENSE
--> /Users/christine/.dvn/data/LICENSE

2025-07-30 17:14:44,612 - ibridges-gui - INFO - DATAVERSE: Upload /Users/christine/.dvn/data/LICENSE -->
10.80227/PDVNL/NIMZNA

Disadvantages of the lightweight approach

- iRODS and Dataverse **do not interact directly**
 - Data must be **downloaded and re-uploaded**, causing extra transfers
 - No real provenance metadata as required by RDM best practices
 - The relationship between the two systems is **not tracked**
 - iRODS object(s): Collection or data objects
 - Dataverse object: dataset
- They become **two independent digital objects** that evolve separately

Provenance Requirements

To maintain **digital object continuity**, the two systems must be linked:

- iRODS stores the **DOI** as soon as data is pushed to a draft dataset
- Draft datasets or files may be deleted → iRODS must detect this
- iRODS must monitor **dataset state** in Dataverse (draft → published)
- When published, iRODS stores the final DOI and updates metadata

Continuous monitoring of Dataverse may cause performance issues

Core question:

How can iRODS receive **notifications** about relevant Dataverse events?

- File added / removed
- Dataset state changed
- Metadata updated
- DOI assigned

Demo

The screenshot shows a web application interface for managing datasets. The top navigation bar includes tabs for 'Browser', 'Synchronise Data', 'Search', 'Info', 'Logs', and 'Dataverse' (which is currently selected). The main content area is divided into two panels. The left panel contains form fields for 'Dataverse URL' (with a dropdown showing 'dvn-nl -> https://demo.dataverse.nl [dataverse OK, token OK]' and buttons for 'Delete URL' and 'Add URL'), 'Drafts' (with a dropdown showing '10.80227/PDVNL/IVRVND [DRAFT]'), and 'Dataset ID' (with a text input showing '10.80227/PDVNL/IVRVND' and a 'Create New Dataset' button). Below these is a checkbox for 'Compare checksums' which is checked. A table displays a single dataset entry with columns 'Path' and 'Size'. The right panel shows a file tree structure under '/uu/home' with subdirectories 'research-christine', 'research-sane-training', 'research-training', and 'vault-training'. A '<<' button is located between the two panels. At the bottom of the left panel, there are 'Delete' and 'Upload to Dataverse' buttons. The bottom left corner shows a progress indicator at '0%'.

Dataverse URL

dvn-nl → https://demo.dataverse.nl [dataverse OK, token OK]

Delete URL Add URL

Drafts 10.80227/PDVNL/IVRVND [DRAFT]

Dataset ID 10.80227/PDVNL/IVRVND

Create New Dataset

☒ Compare checksums

	Path	Size
1	/uu/home/research-christine/pandas.csv	1351215

Delete Upload to Dataverse

0%

/uu/home

- > research-christine
- > research-sane-training
- > research-training
- > vault-training

<<