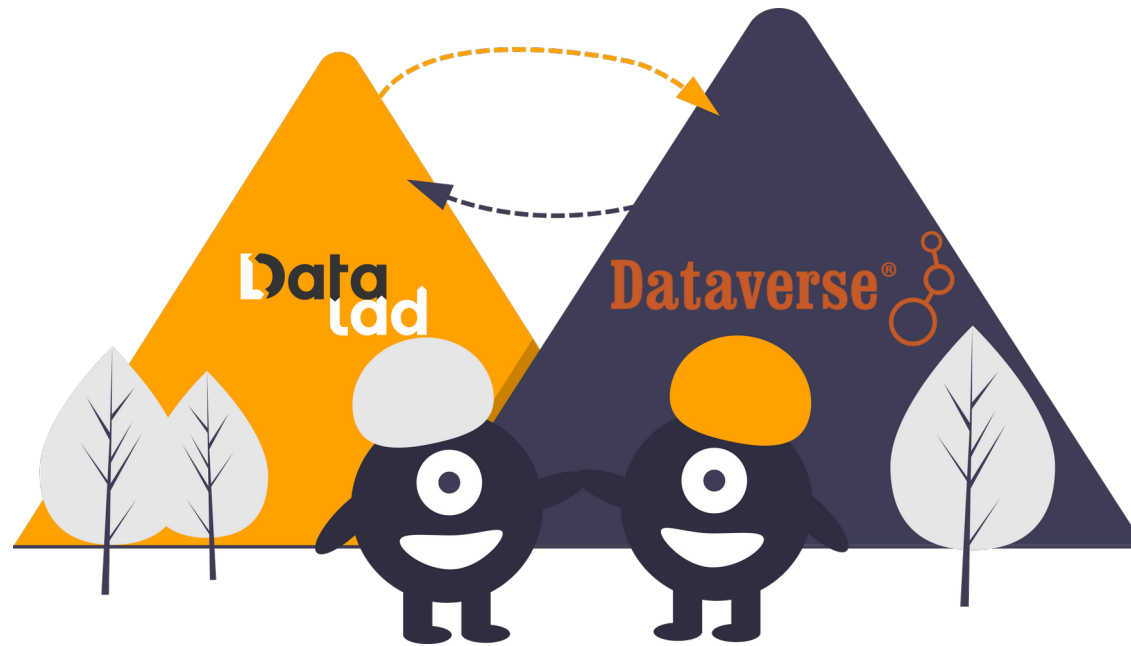


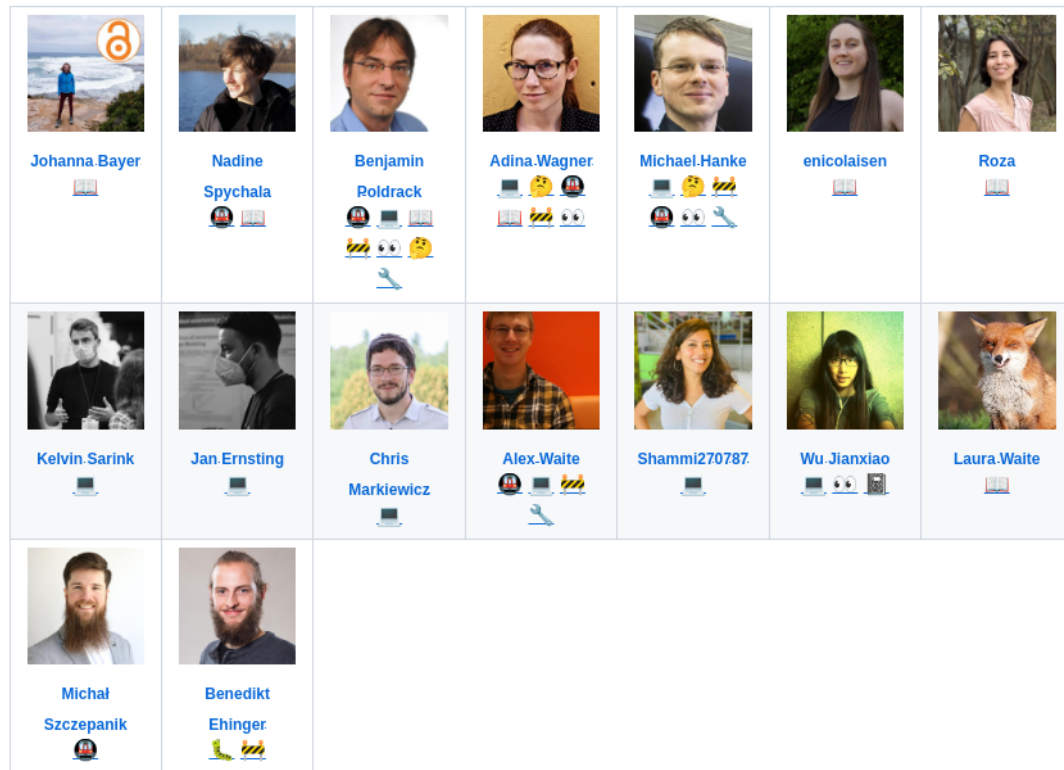
Lightweight data publishing on Jülich Data with DataLad



Acknowledgements

DataLad-dataverse contributors

Thanks goes to these wonderful people ([emoji key](#)):



This project follows the [all-contributors](#) specification. Contributions of any kind welcome!

Special thanks:

- DataLad (Hanke, Halchenko & contributors)
- Git-annex (Joey Hess & contributors)
- Psychoinformatics Group, INM-7, FZJ

Funding:

- FZJ RDM challenge 2022
- BMBF 01GQ1905
- NSF 1912266
- DFG 431549029-INF

What is DataLad

- A free and open source tool
- for decentralized (research) data management
- with a command line interface, Python API and graphical user interface
- allowing exhaustive tracking of the evolution of digital objects
- and computational provenance tracking
- to enhance modularity, portability and reproducibility
- <https://datalad.org>



What is Dataverse

- Open source research data repository software
- Hosted at > 105 institutions
- <https://dataverse.org/>

What is Jülich DATA

- An instance of Dataverse
- Institutional data repository at FZJ
- Service to index all data publications created at or in the context of FZJ
- Managed by the Central Library
- <https://data.fz-juelich.de/>



Why *lightweight* data publishing

Data Storage Linking

Research in Jülich is very diverse, as is the data generated from our activities. Ranging from as small as a few kilobyte, it's easy to reach petabyte for datasets.

Obviously, those big piles of data are likely immutable and unmovable. It makes more sense to link to such data instead of uploading it to a central storage location (unless there is a very good use case).

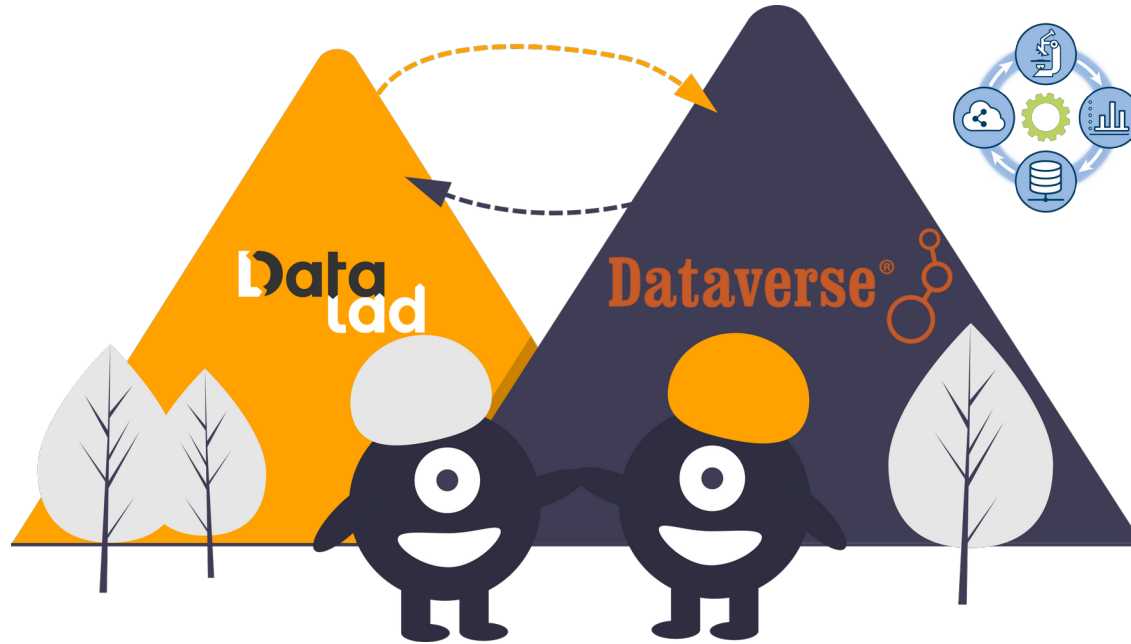
Datasets in Jülich DATA satisfy the [FZJ metadata schema](#). It contains a field *Storage Location URL*, enabling linking to individual data files, archives, snapshots, services or inventories.



The image shows a screenshot of the 'FZJ Metadata' form. At the top left, it says 'FZJ Metadata' with a small upward-pointing arrow. Below this, there are three input fields. The first is labeled 'Institute' with a red asterisk and a help icon, and it is a dropdown menu. The second is labeled 'PoF III topic' with a red asterisk and a help icon, and it is also a dropdown menu. The third is labeled 'Storage Location URL' with a red asterisk and a help icon, and it is a text input field with the placeholder text 'Enter full URL, starting with e. g. http://, smb://, ...'. To the right of this field is a small grey button with a blue plus sign.

<https://data.fz-juelich.de/guide/juelich/data-linking.html>

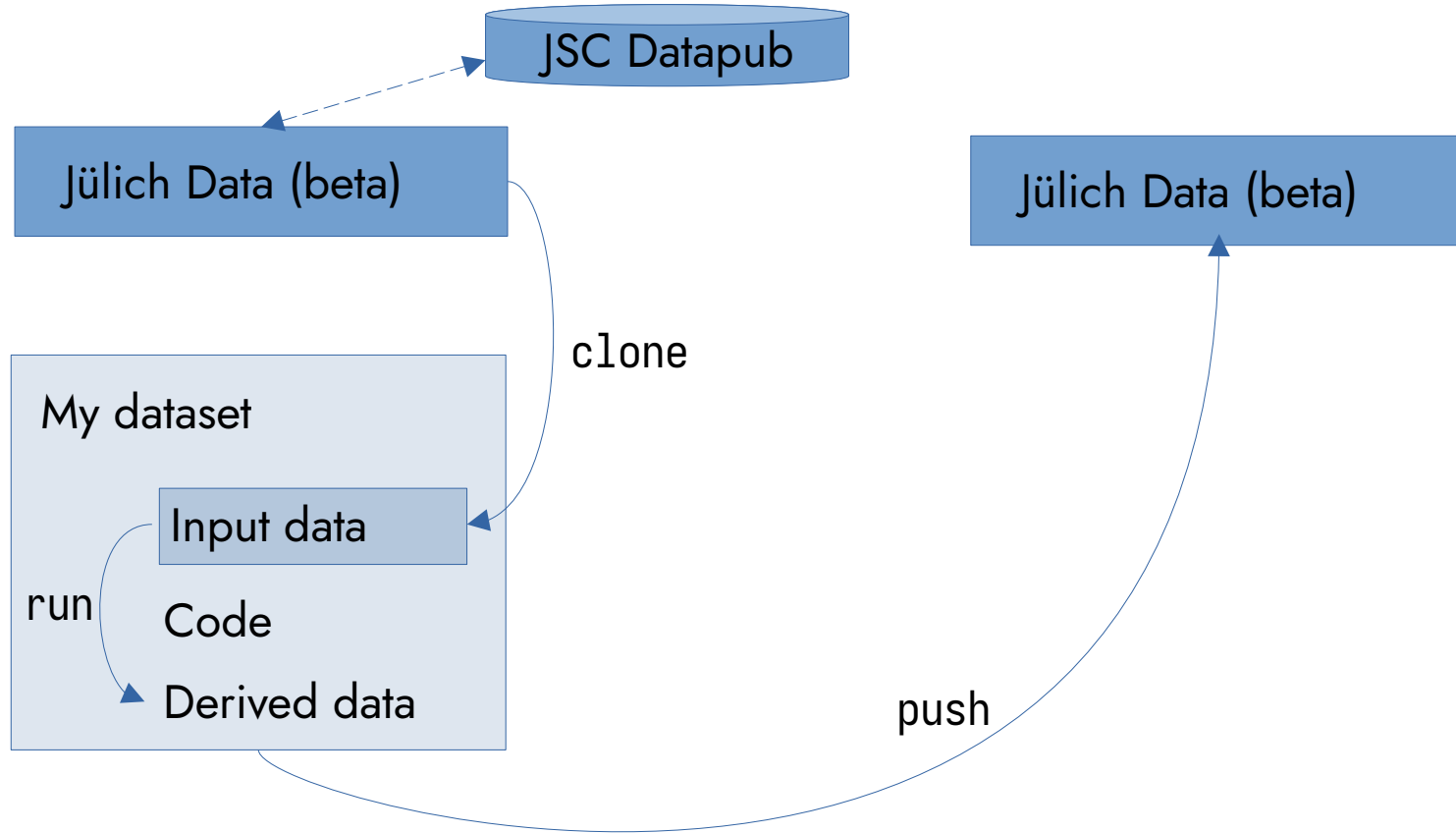
Best of both worlds



actionable data tracking with DataLad
(identity, availability, provenance)

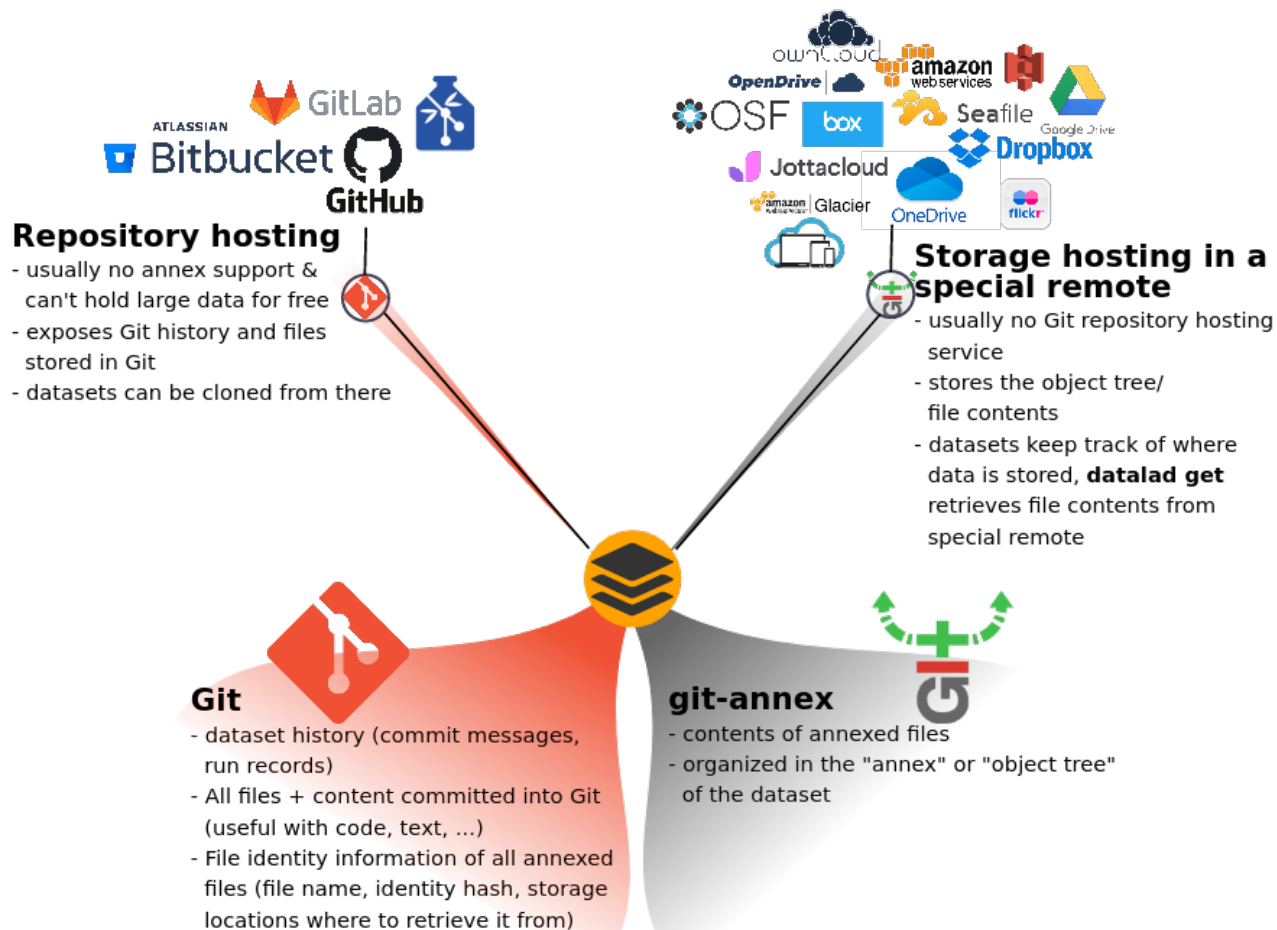
Discoverability & metadata with Jülich DATA

Software demo



Part 2: additional details

Two aspects of datasets: Git and git-annex



See:

- git-annex [special remotes](#)
- [Datalad-next](#) extension
 - uncurl remote
 - create-sibling-webdav

Publication modes: git-only (lightweight)

Contact Share

Publish Edit

Data hosted on datapub.fz-juelich.de



Draft

Unpublished

Wagner, Adina Svenja, 2023, "Data hosted on datapub.fz-juelich.de", <https://doi.org/10.0346/JUELICH-DATA-BETA/IHHXNN>, Jülich DATA - Beta Training Facility, DRAFT VERSION

Cite Dataset

Learn about Data Citation Standards.

Dataset Metrics

0 Downloads

Description I would like to get a DOI for the dataset I published to datapub.fz-juelich.de (2023-10-05)

Subject Medicine, Health and Life Sciences

Institute INM-7

Files

Metadata

Terms

Versions

Search this dataset...

Find

Upload Files

Filter by

File Type: All Access: All

Sort

1 to 2 of 2 Files

Edit Files

Download

XDLRA-2D--2D-refs
annex/3f7/4a3/
Plain Text - 24 B - Oct 5, 2023 - 0 Downloads
SHA-256: 380be5b79e5e9b51bccfea655b6217d46ef87f3de270ff04a7da19e4607dd93e

Download

XDLRA-2D--2D-repo-2D-export
annex/eb3/ca0/
Plain Text - 14.3 KB - Oct 5, 2023 - 0 Downloads
SHA-256: 2cdd983f26911aad811c18c08b3c82fcccdf19f128b26e936b4f76c53cbd26b0

Download

Publication modes: annex (default)













Files Metadata Terms Versions

Change View Table Tree

Search this dataset...

Filter by

1 to 3 of 3 Files

<input type="checkbox"/>		MD5E-2Ds6-2D-2Dd577273ff885c3f84dadb8578bb41399 annex/4f5/973/ Plain Text - 6 B Deposited Mar 16, 2023 MD5: d57...399 	 
<input type="checkbox"/>		XDLRA-2D-2Drefs annex/3f7/4a3/ Plain Text - 145 B Deposited Mar 16, 2023 MD5: eb0...765 	 
<input type="checkbox"/>		XDLRA-2D-2Drepo-2Dexport annex/eb3/ca0/ Unknown - 19.2 KB Deposited Mar 16, 2023 MD5: 6f2...19e 	 

Publication modes: filetree

Files **Metadata** Terms Versions







Change View Table Tree

Search this dataset...

Filter by

File Type: All Access: All

1 to 6 of 6 Files

<input type="checkbox"/>	 config _datalad/ Plain Text - 63 B Deposited Mar 16, 2023 MD5: 9eb...dd5	<input type="button" value="Download"/> <input type="button" value="More"/>
<input type="checkbox"/>	 my-2Dfile Plain Text - 6 B Deposited Mar 16, 2023 MD5: d57...399	<input type="button" value="Download"/> <input type="button" value="More"/>
<input type="checkbox"/>	 refs _datalad/dotgit/ Plain Text - 145 B Deposited Mar 16, 2023 MD5: 6a2...1aa	<input type="button" value="Download"/> <input type="button" value="More"/>
<input type="checkbox"/>	 repo.zip _datalad/dotgit/ Unknown - 20.7 KB Deposited Mar 16, 2023 MD5: 054...417	<input type="button" value="Download"/> <input type="button" value="More"/>
<input type="checkbox"/>	 _gitattributes _datalad/ Plain Text - 32 B Deposited Mar 16, 2023 MD5: 82a...811	<input type="button" value="Download"/> <input type="button" value="More"/>
<input type="checkbox"/>	 _gitattributes Plain Text - 55 B Deposited Mar 16, 2023 MD5: 5e8...538	<input type="button" value="Download"/> <input type="button" value="More"/>

See also

- [datalad.org](https://data-lad.org)
- handbook.datalad.org
- docs.datalad.org/projects/dataverse
- data.fz-juelich.de/guide/juelich/
- dataverse.org



18.-20. Apr 2024

Düsseldorf



Technologies for distributed data management – Call for contributions open

<https://distribits.live> -- fosstodon.org/@distribits/