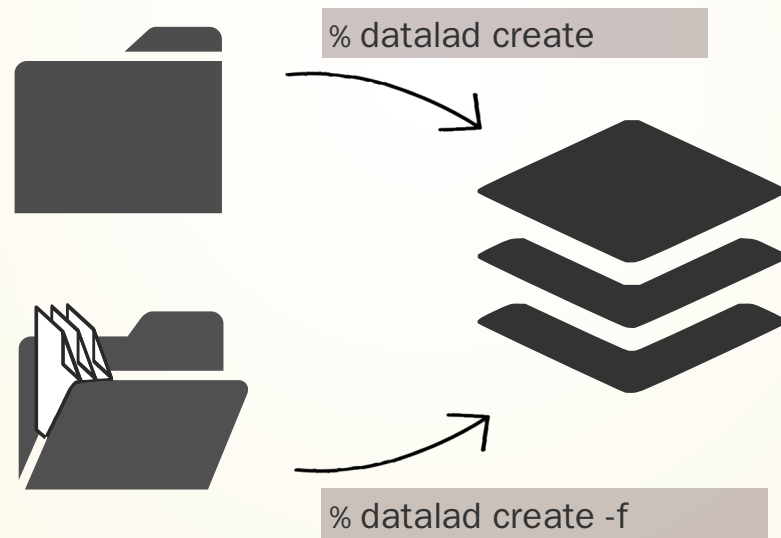


QUICK RECAP

create new, empty datasets to populate...



.... or transform existing directories into data

QUICK RECAP



Git

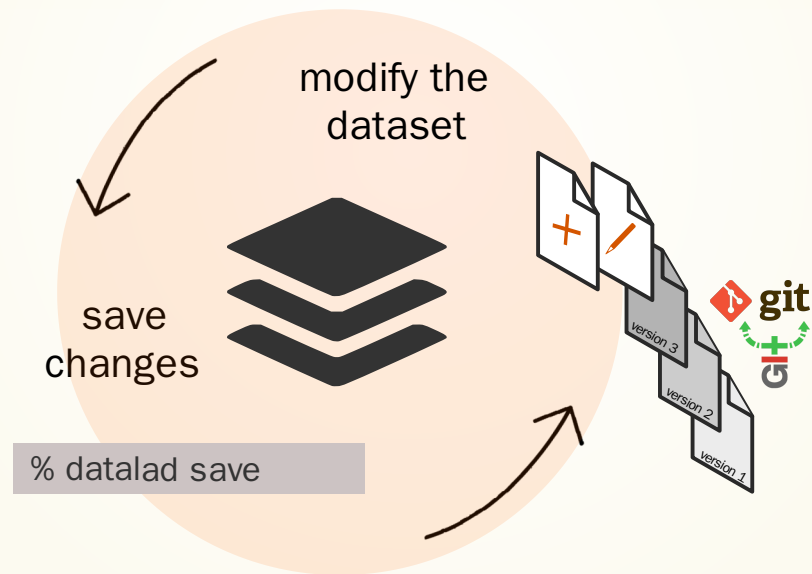
- dataset history (commit messages, run records)
- All files + content committed into Git (useful with code, text, ...)
- File identity information of all annexed files (file name, identity hash, storage locations where to retrieve it from)



git-annex

- contents of annexed files
 - organized in the "annex" or "object tree" of the dataset
- 

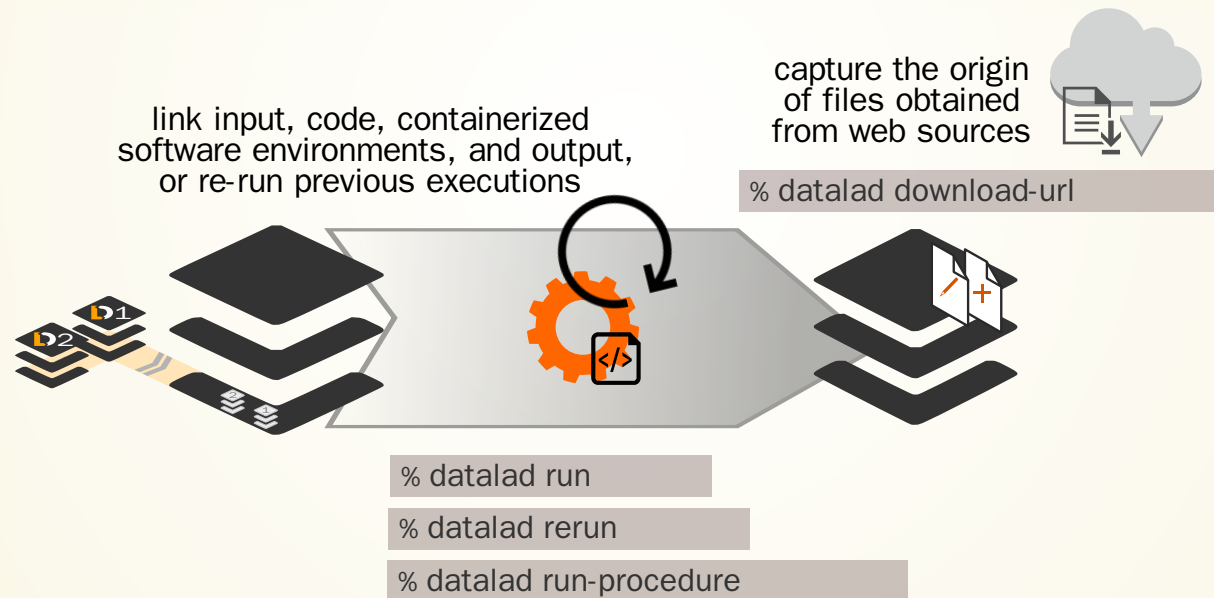
QUICK RECAP



QUICK RECAP

Date	Author	Change summary (commit message)
2020-06-05 10:58	+0200 Adina Wagner	M [master] {upstream/master} {upstream/HEAD} Merge pull request #12 from psychoinformatics-d
2020-06-05 08:24	+0200 Adina Wagner	o [finalround] {upstream/finalround} add results from computing with mean instead of median
2020-06-05 09:09	+0200 Michael Hanke	o Change wording, clarify comment
2020-06-05 07:26	+0200 Michael Hanke	M Merge remote-tracking branch 'gh-mine/finalround'
2020-05-28 16:39	+0200 Asim H Dar	o Added datalad.get() so S2SRMS() pulls data and can run standalone
2020-05-18 08:25	+0200 Adina Wagner	o {gh-asim/finalround} S2SRMS: example implementation of the S2SRMS method suggested by R2
2020-05-01 17:38	+0200 Adina Wagner	o Minor edits as suggested by reviewer 2
2020-05-29 09:04	+0200 Adina Wagner	M Merge pull request #13 from psychoinformatics-de/adswa-patch-1
2020-05-24 09:15	+0200 Adina Wagner	o {upstream/adswa-patch-1} Fix installation instructions
2020-05-24 09:53	+0200 Adina Wagner	M Merge pull request #14 from psychoinformatics-de/bf-data
2020-05-24 09:33	+0200 Adina Wagner	o [bf-data] One-time datalad import
2020-05-24 09:32	+0200 Adina Wagner	o install and get relevant subdataset data
2020-03-18 10:19	+0100 Michael Hanke	M Merge pull request #8 from psychoinformatics-de/adswa-patch-1
2019-12-19 10:22	+0100 Adina Wagner	o {gh-asim/adswa-patch-1} add sklearn to requirements
2020-03-18 10:13	+0100 Michael Hanke	o Tune new figure caption
2020-03-18 10:03	+0100 Michael Hanke	M Merge pull request #11 from ElectronicTeaCup/revision 2
2020-03-18 09:59	+0100 Adina Wagner	M [revision 2] {gh-asim/revision 2} Merge branch 'revision 2' of github.com:ElectronicTe
2020-03-18 09:58	+0100 Adina Wagner	o name parameter in caption
2020-03-18 09:59	+0100 Adina Wagner	o name parameter in caption

QUICK RECAP



BEFORE WE BEGIN...

Any left-over questions from yesterday?

DROP & REMOVE

- Try to remove (*rm*) one of the pictures in your dataset. What happens?
- Version control tools keep a revision history of your files - file contents are not actually removed when you *rm* them. Interactions with the revision history of the dataset can bring them "back to life"

DROP & REMOVE

- Clone a small example dataset to drop file contents and remove datasets:

```
$ datalad clone https://github.com/datalad-datasets/machinelearning-books.git
$ cd machinelearning-books
$ datalad get A.Shashua-Introduction_to_Machine_Learning.pdf
```

- **datalad drop** removes annexed file contents from a local dataset annex and frees up disk space. It is the antagonist of **get** (which can get files and subdatasets).

```
$ datalad drop A.Shashua-Introduction_to_Machine_Learning.pdf
drop(ok): /tmp/machinelearning-books/A.Shashua-Introduction_to_Machine_Learning.pdf (file)
[checking https://arxiv.org/pdf/0904.3664v1.pdf...]
```

- But: Default safety checks require that dropped files can be re-obtained to prevent accidental data loss. **git annex whereis** reports all registered locations of a file's content
- **drop** does not only operate on individual annexed files, but also directories, or globs, and it can uninstall subdatasets:

```
$ datalad clone https://github.com/datalad-datasets/human-connectome-project-openaccess.git
$ cd human-connectome-project-openaccess
$ datalad get -n HCP1200/996782
$ datalad drop --what all HCP1200/996782
```

DROP & REMOVE

- **datalad remove** removes complete dataset or dataset hierarchies and leaves no trace of them. It is the antagonist to **clone**.

```
# The command operates outside of the to-be-removed dataset!  
$ datalad remove -d . machinelearning-books  
uninstall(ok): /tmp/machinelearning-books (dataset)
```

- But: Default safety checks require that it could be re-cloned in its most recent version from other places, i.e., that there is a *sibling* that has all revisions that exist locally **datalad siblings** reports all registered siblings of a dataset.

DROP & REMOVE

- Create a dataset from scratch and add a file

```
$ datalad create local-dataset
$ cd local-dataset
$ echo "This file content will only exist locally" > local-file.txt
$ datalad save -m "Added a file without remote content availability"
```

- **datalad drop** refuses to remove annexed file contents if it can't verify that **datalad get** could re-retrieve it

```
$ datalad drop local-file.txt
$ drop(error): local-file.txt (file) [unsafe; Could only verify the existence of 0 out of 1 necessary copy;
(Note that these git remotes have annex-ignore set: origin upstream);
(Use --reckless availability to override this check, or adjust numcopies.)]
```

- Adding **--reckless availability** overrides this check

```
$ datalad drop local-file.txt --reckless availability
```

- Be mindful that **drop** will only operate on the most recent version of a file - past versions may still exist afterwards unless you drop them specifically. **git annex unused** can identify all files that are left behind

DROP & REMOVE

- **datalad remove** refuses to remove datasets without an up-to-date *sibling*

```
$ datalad remove -d local-dataset
uninstall(error): . (dataset) [to-be-dropped dataset has revisions that are not available at
sibling. Use `datalad push --to ...` to push these before dropping the local
or ignore via `--reckless availability`. Unique revisions: ['main']]
```

- Adding **--reckless availability** overrides this check

```
$ datalad remove -d local-dataset --reckless availability
```

REMOVING WRONGLY

- Using a file browser or command line calls like `rm -rf` on datasets is doomed to fail. Recreate the local dataset we just removed:

```
$ datalad create local-dataset
$ cd local-dataset
$ echo "This file content will only exist locally" > local-file.txt
$ datalad save -m "Added a file without remote content availability"
```

- Removing it the wrong way causes chaos and leaves an usable dataset corpse behind:

```
$ rm -rf local-dataset
rm: cannot remove 'local-dataset/.git/annex/objects/Kj/44/MD5E-s42--8f008874ab52d0ff02a5bbd0174ac95e.txt/MD5E-s42--8f008874ab52d0ff02a5bbd0174ac95e.txt': Permission denied
```

- The dataset can't be fixed, but to remove the corpse `chmod` (change file mode bits) it (i.e., make it writable)

```
$ chmod +w -R local-dataset
$ rm -rf local-dataset
```

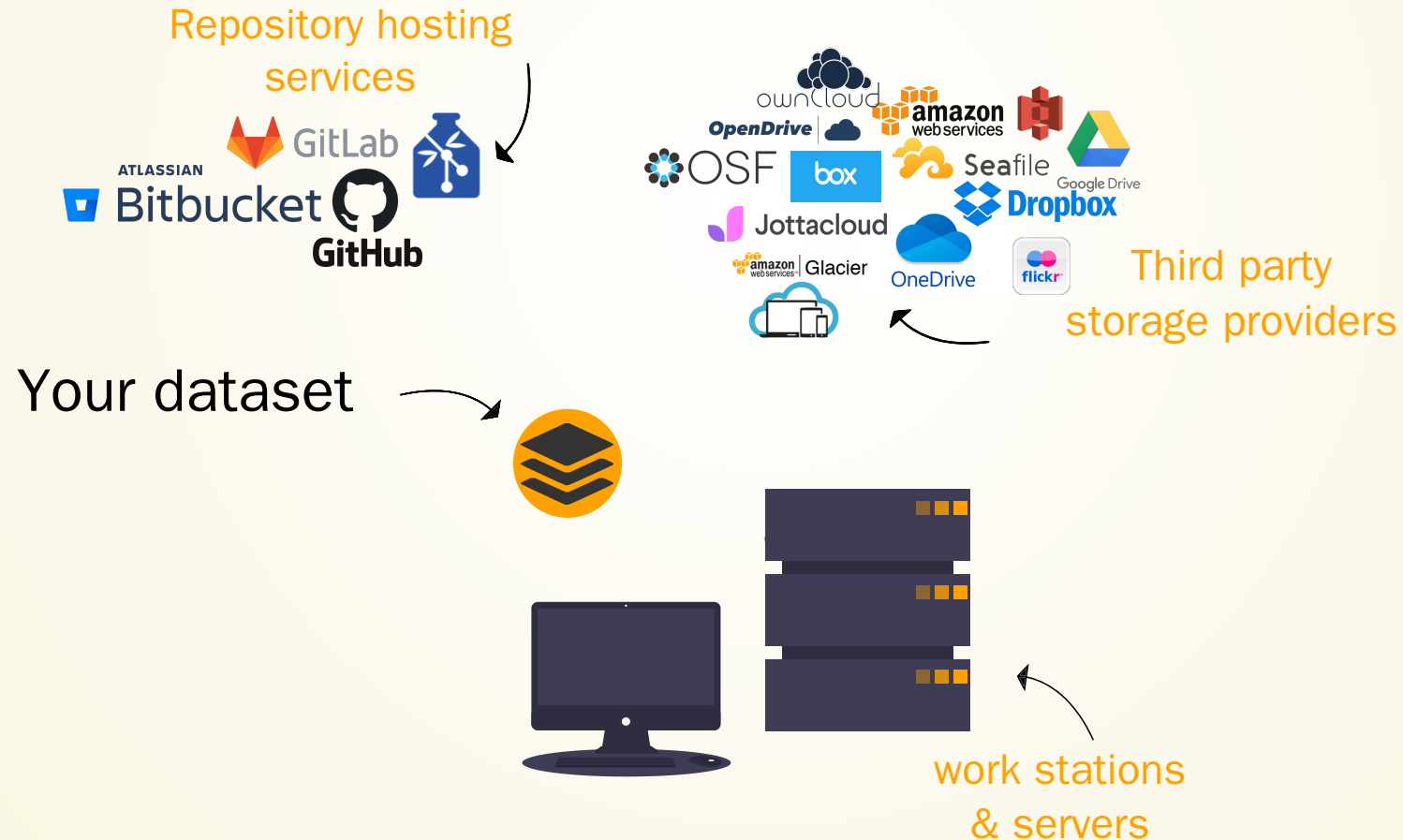
PUBLISHING DATASETS

How to share your work with others

Repository hosting services, siblings, and `datalad push`

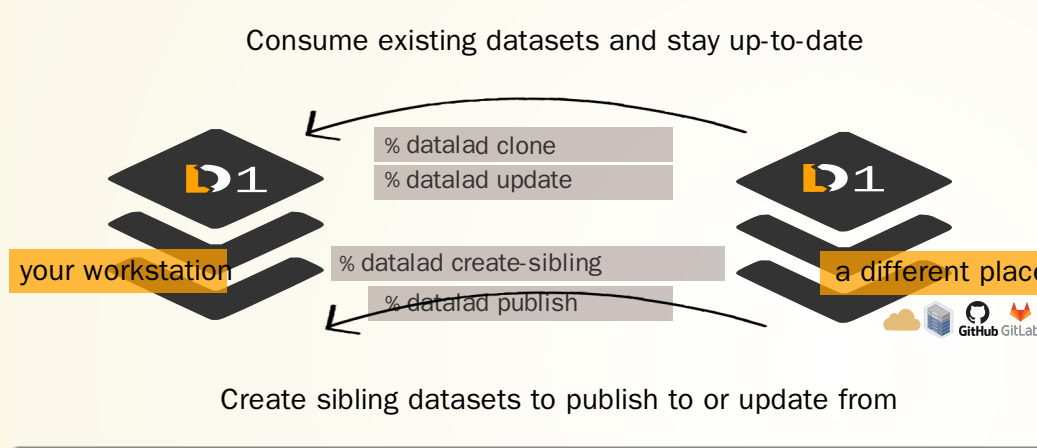
PUBLISHING DATASETS

I have a dataset on my computer. How can I share it, or collaborate on it?



"SHARE DATA LIKE SOURCE CODE"

- Datasets can be cloned, pushed, and updated from and to **local** and **remote** paths, remote **hosting services**, external **special remotes**



- Examples:
Local path

```
../my-projects/experiment_data
```

Remote path

```
myuser@myinstitutes.hcp.system:/home/myuser/my-projects/experiment_data
```

Hosting service

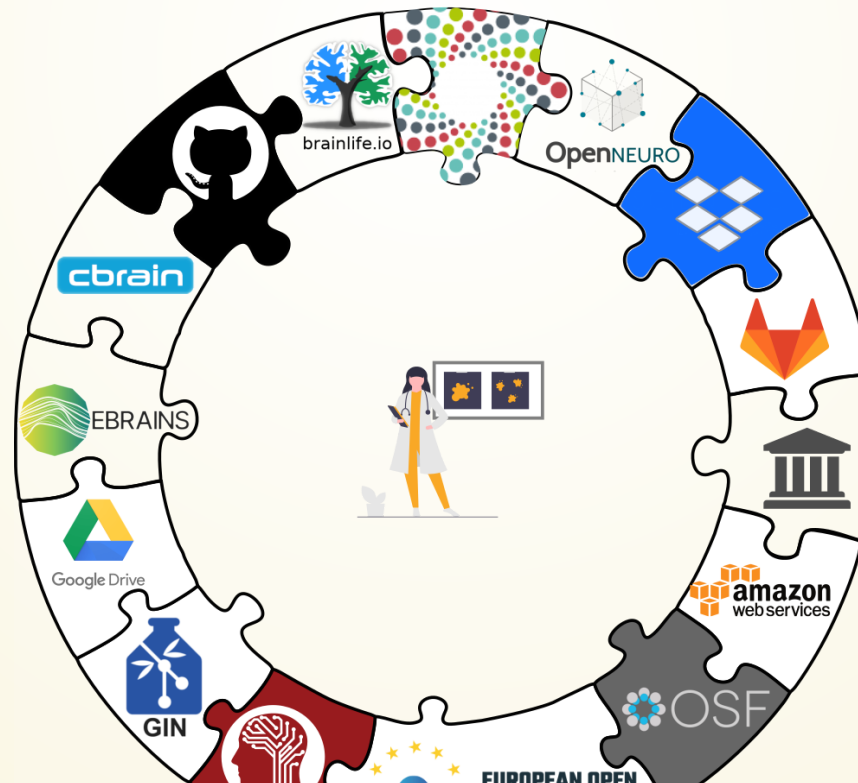
```
git.github.com:myuser/experiment_data.git
```

External special remotes

```
osf://my-osf-project-id
```

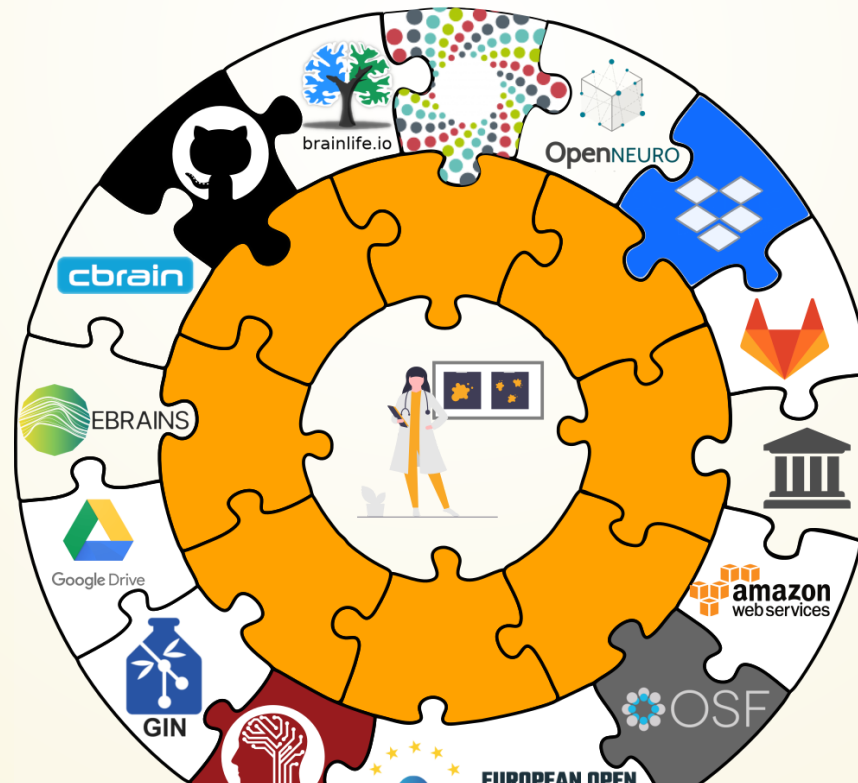
INTEROPERABILITY

- DataLad is built to maximize interoperability and use with hosting and storage technology



INTEROPERABILITY

- DataLad is built to maximize interoperability and use with hosting and storage technology



GLOSSARY

Sibling (remote)

Linked clones of a dataset. You can usually update (from) siblings to keep all your siblings in sync (e.g., ongoing data acquisition stored on experiment compute and backed up on cluster and external hard-drive)

Repository hosting service

Webservices to host Git repositories, such as GitHub, GitLab, Bitbucket, Gin, ...

Third-party storage

Infrastructure (private/commercial/free/...) that can host data. A "special remote" protocol is used to publish or pull data to and from it

Publishing datasets

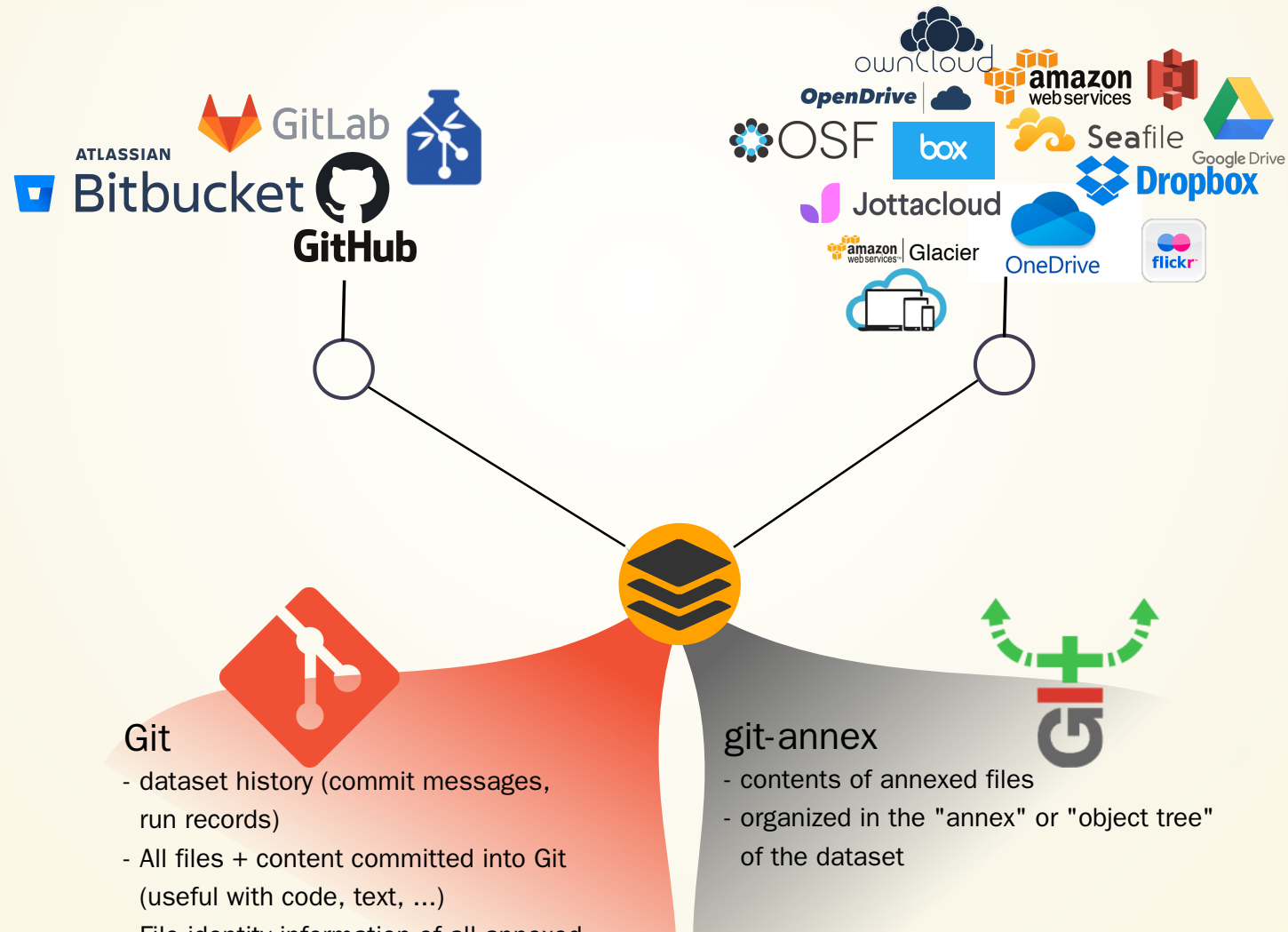
Pushing dataset contents (Git and/or annex) to a sibling using **datalad push**

Updating datasets

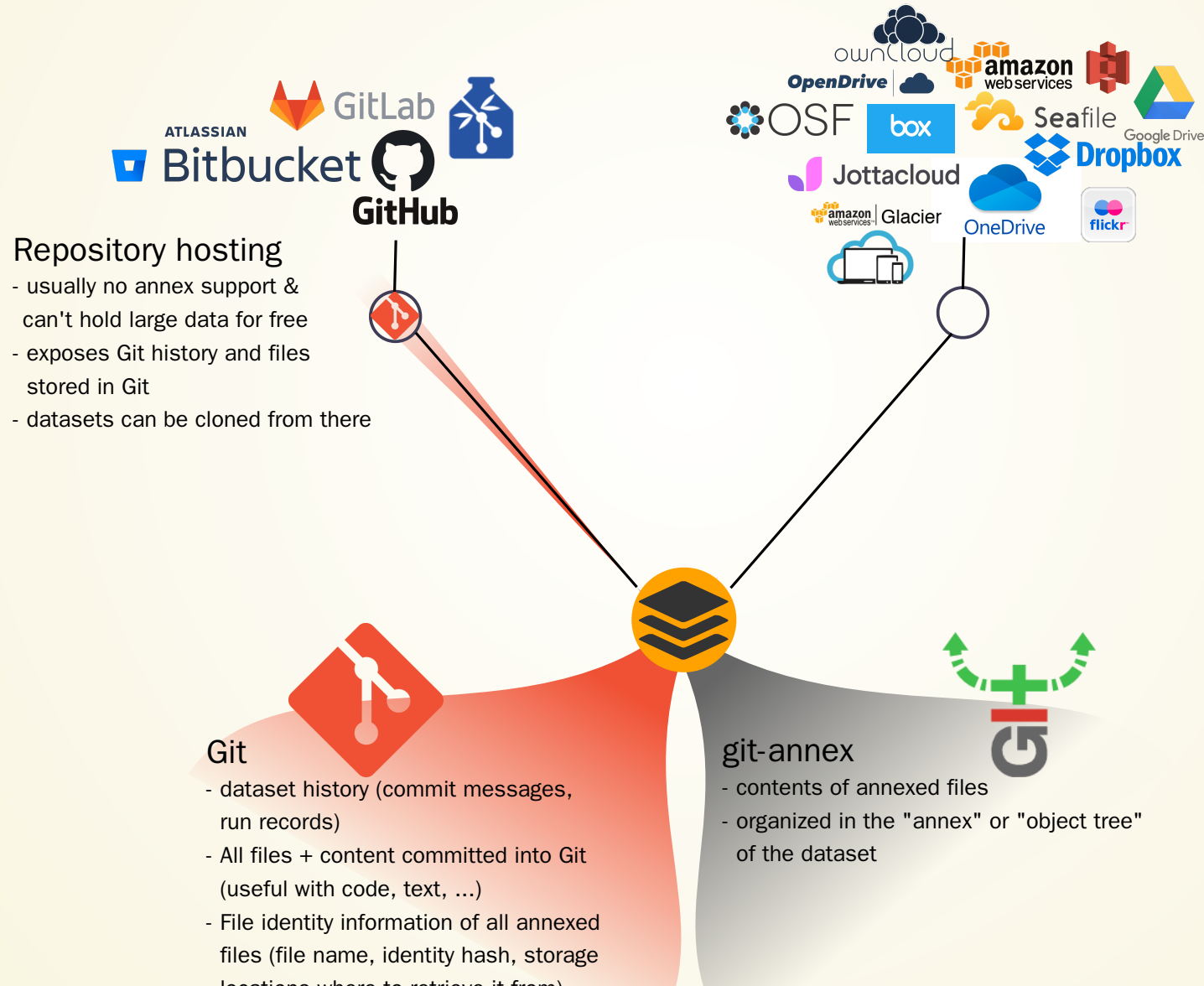
Pulling new changes from a sibling using **datalad update --merge**

SHARING DATASETS

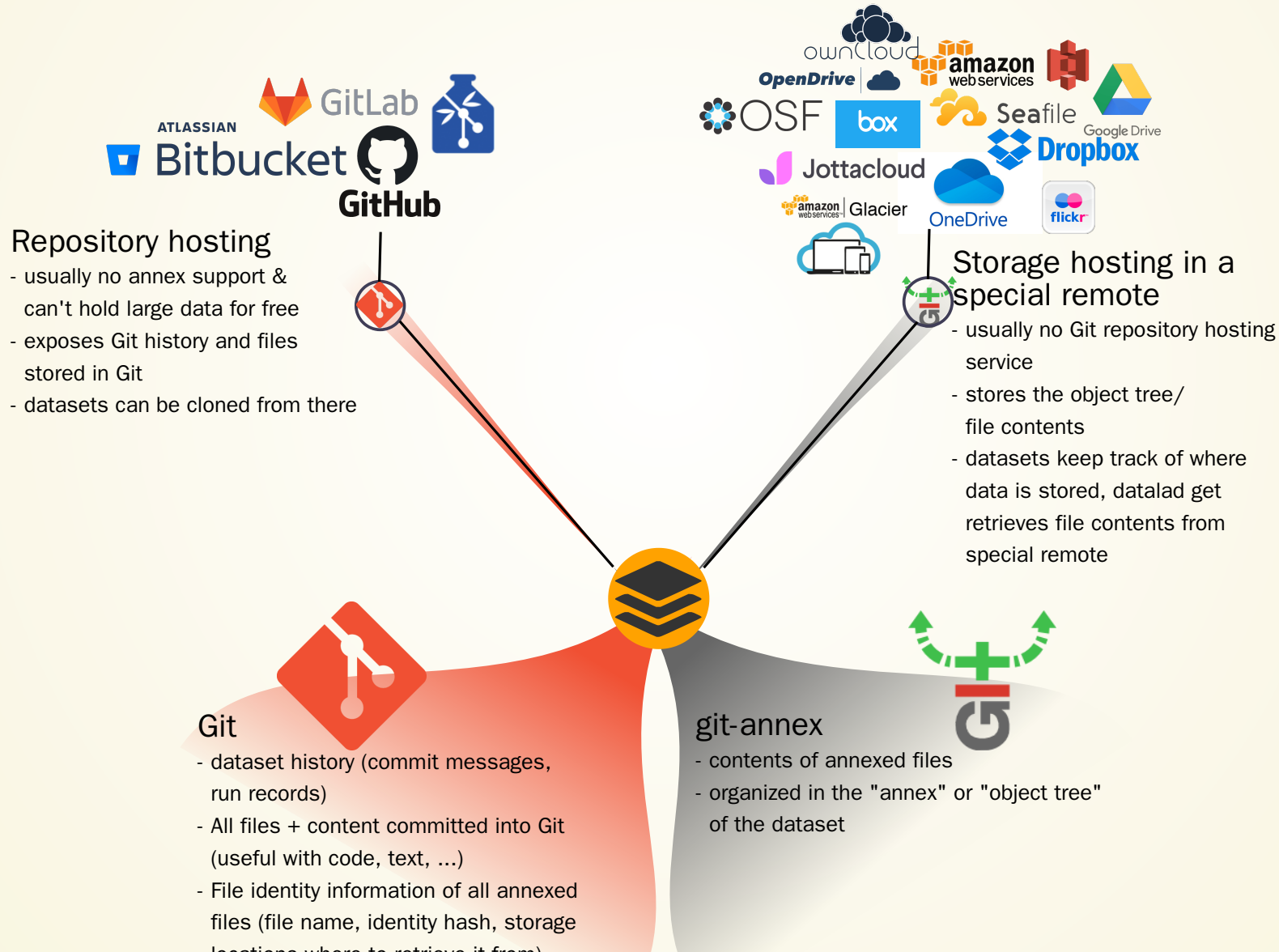
- Most public datasets separate content in Git versus git-annex behind the scenes



SHARING DATASETS



SHARING DATASETS



SHARING DATASETS

Typical case:

- Datasets are exposed via a private or public repository on a repository hosting service
- Data can't be stored in the repository hosting service, but can be kept in almost any third party storage
- Publication dependencies automate pushing to the correct place, e.g.,

```
$ git config --local remote.github.datalad-publish-depends gdrive  
# or  
$ datalad siblings add --name origin --url git@git.jugit.fzj.de:adswa/experiment-data.git --publish-depends s3
```

SHARING DATASETS

- Real-life example 1:
GitHub for repository hosting, data hosting via datapub.fz-juelich.de + GNODE

The screenshot shows the GitHub repository page for `psychoinformatics-de/studyforrest-data-phase2`. The repository is owned by `psychoinformatics-de` and has 1 pull request, 4 issues, 6 stars, and 8 forks. The main branch is `master` with 2 branches and 1 tag. The repository contains a list of files and folders, including `.datalad`, `code`, `src`, `stimuli`, and a series of subfolders (`sub-01` to `sub-18`). The `code` folder has a commit message "Fix type in physio log converter (fixes gh-11)" from 3 years ago. The `src` folder has a commit message "Recover lost segment from eyetracker (closes gh-3)" from 5 years ago. The `stimuli` folder has a commit message "Add BIDS-compatible stimuli/ directory (with symlinks)" from 4 years ago. The subfolders (`sub-01` to `sub-18`) all have commit messages "BF: Re-import respiratory trace after bug fix in converte..." from 3 years ago. The repository also has a "About" section with a description: "studyforrest.org: Phase2 data (movie, eyetracking, retmapping, visual localizers) [BIDS]" and a link to `studyforrest.org`. There is a "Releases" section with one release: "First public release" on Mar 26, 2016. There is also a "Packages" section with the message "No packages published". Finally, there is a "Contributors" section with three contributors: `mih` Michael Hanke, `dakot` Daniel Kottke, and `adswa` Adina Wagner.

Search or jump to... Pull requests Issues Marketplace Explore

psychoinformatics-de / studyforrest-data-phase2

Unwatch 1 Unstar 6 Fork 8

Code Issues 4 Pull requests 1 Actions Projects Security Insights

master 2 branches 1 tag Go to file Add file Code

mih Merge pull request #15 from adswa/ENH/README b5306e2 on May 7 77 commits

.datalad	[DATALAD] dataset aggregate metadata update	2 years ago
code	Fix type in physio log converter (fixes gh-11)	3 years ago
src	Recover lost segment from eyetracker (closes gh-3)	5 years ago
stimuli	Add BIDS-compatible stimuli/ directory (with symlinks)	4 years ago
sub-01	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-02	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-03	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-04	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-05	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-06	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-09	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-10	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-14	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-15	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-16	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-17	BF: Re-import respiratory trace after bug fix in converte...	3 years ago
sub-18	BF: Re-import respiratory trace after bug fix in converte...	3 years ago

About

studyforrest.org: Phase2 data (movie, eyetracking, retmapping, visual localizers) [BIDS]

studyforrest.org

Readme

View license

Releases 1

First public release Latest on Mar 26, 2016

Packages

No packages published

Contributors 3

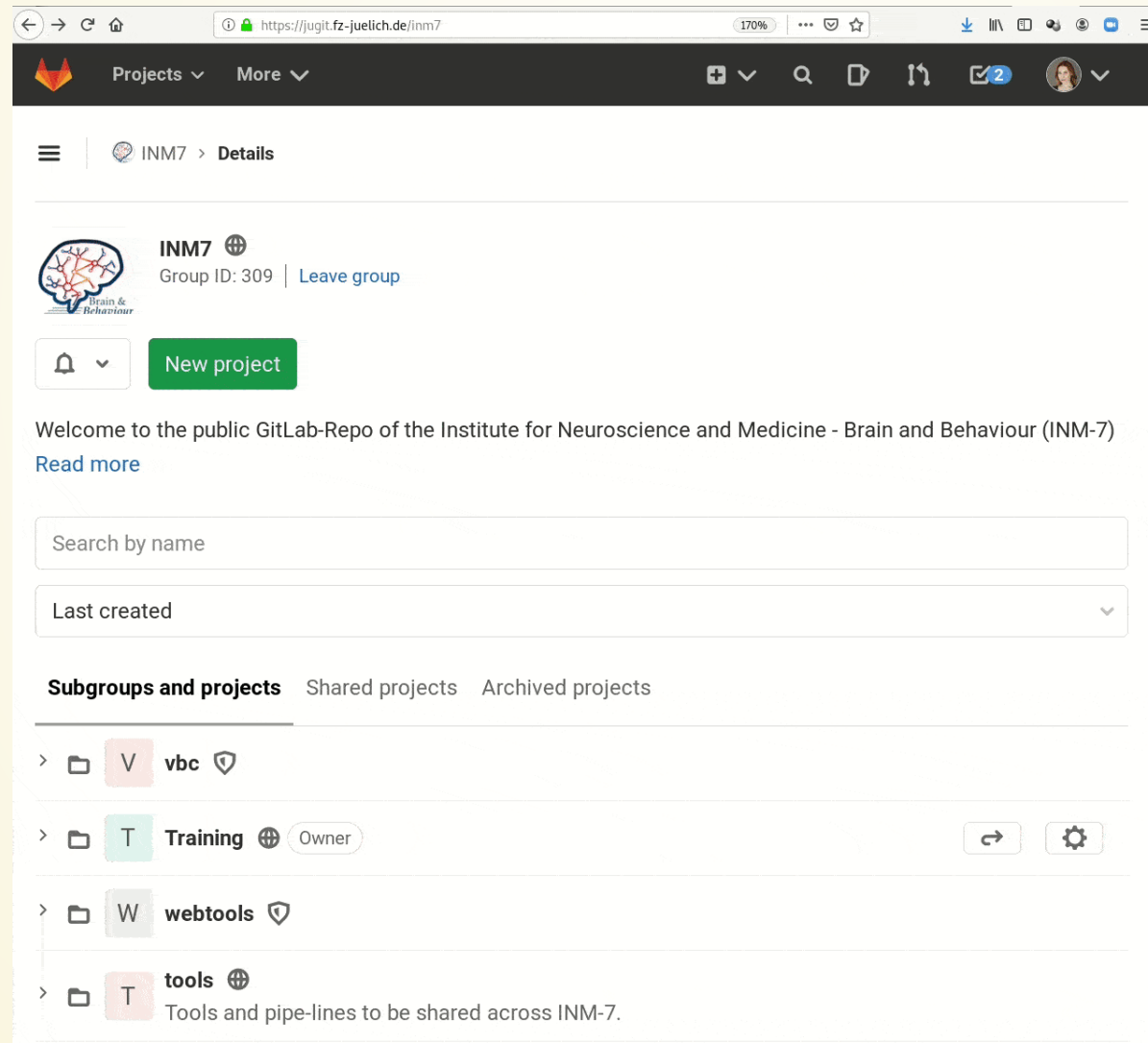
mih Michael Hanke

dakot Daniel Kottke

adswa Adina Wagner

SHARING DATASETS

- Real-life example 2:
GitLab for repository hosting, data hosting via internal webserver (access restricted)



SHARING DATASETS

- Real-life example 3:
GitHub for repository hosting, data hosting via Amazon S3 (requires DUA)

Search or jump to...

Pull requestsIssuesMarketplaceExplore

datalad-datasets / human-connectome-project-openaccess

Unwatch7Unstar11Fork1

<> Code1 Issues5⬆️ Pull requests1🔗 Actions📁 Projects📖 Wiki🔒 Security🔍 Insights...

master2 branches0 tagsGo to fileAdd fileCode

adswa Merge pull request #18 from datalad-datasets/RES...c71d3b628 days ago22 commits

.datalad	Let DataLad find subdatasets in the main dataset store ...	14 months ago
HCP1200	add a range of MNINonLinear/Results/ files for subjects...	28 days ago
.gitmodules	Replace participants broken participant sub(sub)datasets	14 months ago
.noannex	Turn into an actual DataLad dataset (no annex, though)	14 months ago
DATA_USE_AGREEMENT...	add copy of Data Usage Agreement from the HCP	14 months ago
README.md	Adjust README with note on data from ConnectomeDB	7 months ago

README.md

Get data from the Human Connectome Project Open Access dataset with DataLad

Made with

This dataset enables data retrieval with DataLad (0.12.2 or later) from the [HCP Open Access dataset](#) for users that accepted the WU-Minn HCP Consortium Open Access Data Use Terms and obtained valid AWS credentials via [db.humanconnectome.org](#).

Human Connectome Project

The Human [Connectome Project \(HCP\)](#) aims to construct a map of the complete structural and functional neural connections in vivo within and across individuals.

Its 'WU-Minn HCP Open Access Data' data release includes high-resolution 3T MR scans from young healthy adult twins and non-twin siblings (ages 22-35) using four imaging modalities:

About

WU-Minn HCP1200 Data:
3T/7T MR scans from young healthy adults twins and non-twin siblings (ages 22-35)
[T1w, T2w, resting-state and task fMRI, high angular resolution dMRI]

[db.humanconnectome.org/...](#)

Readme

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Contributors 3

adswa Adina Wagner

mih Michael Hanke

yarikoptic Yaroslav Halche...

SHARING DATASETS

Special case 1: repositories with annex support



Repositories with annex support

- examples: GIN (gin.g-node.org), GitLab instances with enabled annex support
- can hold large data for free
- exposes Git history and all files + content
- datasets can be cloned from there



Git

- dataset history (commit messages, run records)
- All files + content committed into Git (useful with code, text, ...)
- File identity information of all annexed files (file name, identity hash, storage

git-annex

- contents of annexed files
- organized in the "annex" or "object tree" of the dataset

SHARING DATASETS

Special case 2: Special remotes with repositories



SHARING DATASETS

Special case 1: repositories with annex support

[LIVE DEMO GIN]

SHARING DATASETS

Special case 2: Special remotes with repositories

Requires the DataLad extension `datalad-osf`

[LIVE DEMO OSF]

SHARING DATASETS

Special case 2: Special remotes with repositories

Requires the DataLad extension `datalad-next`

[DEMO WEBDAV]

The screenshot shows a web browser window with the URL `b2drop.eudat.eu/apps/user_saml/saml/selectUserBackend?redirectUri=`. The page title is "Deposit and retrieve DataLad datasets with WebDAV services" and it includes a "GO TO EUDAT WEBSITE" link. The header features the B2DROP and EUDAT logos, along with navigation links for "FAQS", "WHAT IS B2DROP", "USER GUIDE", and "CONTACT". A "Share" button is visible in the top right. The main content area displays "Login options:" and a prominent blue "Login" button. Below the button is a video player icon. On the left side, there is a large, stylized graphic of concentric circles with an arrow pointing upwards. The footer contains text about collaborative data infrastructure and copyright information for 2019 EUDAT.

Deposit and retrieve DataLad datasets with WebDAV services
GO TO EUDAT WEBSITE

B2DROP EUDAT

FAQS WHAT IS B2DROP USER GUIDE CONTACT

Login options:

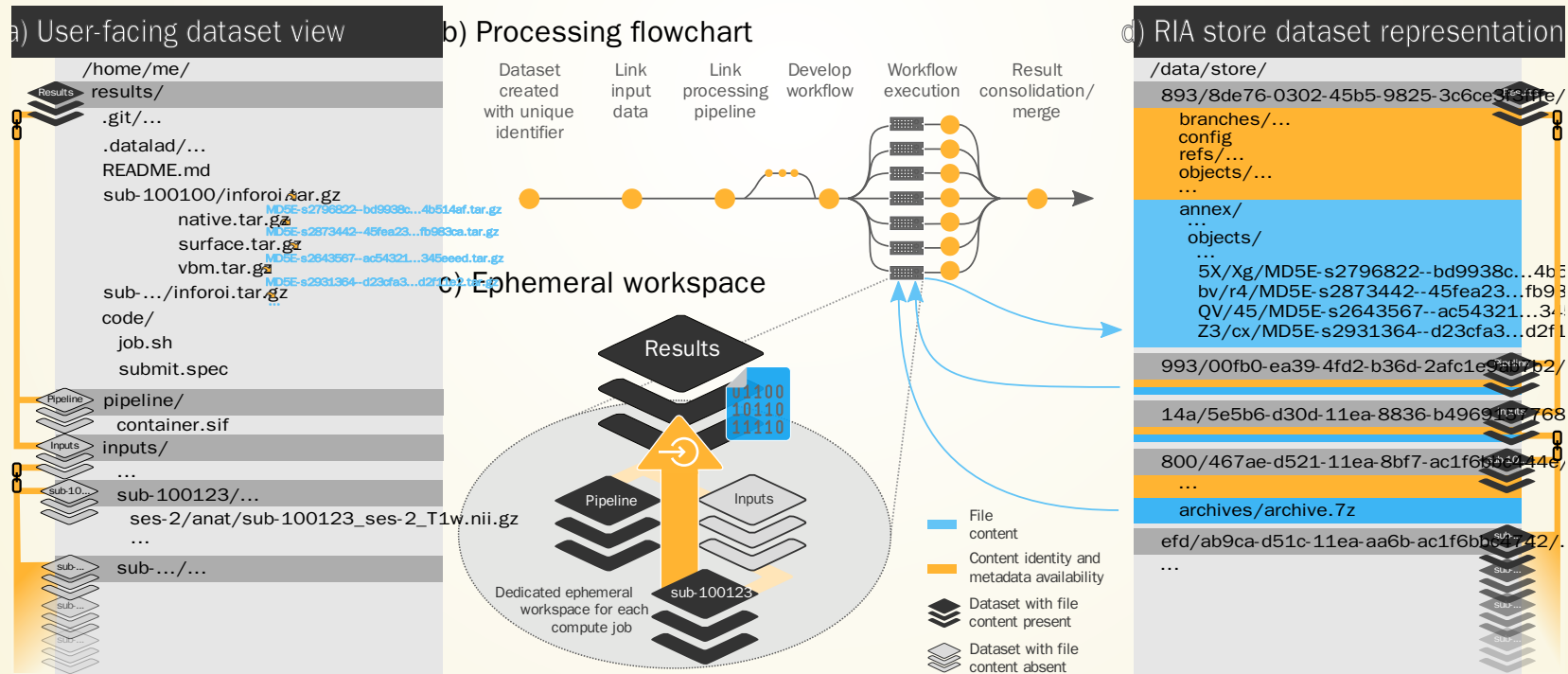
Login

How to collaborate with DataLad datasets using any WebDAV-enabled cloud storage (example: European Open Science Cloud's B2DROP)

SHARING DATASETS

Special case 3: RIA stores for dataset hosting/backup

Tutorial for large scale, reproducible computation: github.com/psychoinformatics-de/fairly-big-processing-workflow



SHARING DATASETS

DataLad can create siblings from the command line for the following services:

GitHub

```
datalad create-sibling-github
```

GitLab

```
datalad create-sibling-gitlab
```

Gin

```
datalad create-sibling-gin
```

Gogs

```
datalad create-sibling-gogs
```

local or remote paths

```
datalad create-sibling
```

RIA stores

```
datalad create-sibling-ria
```

Open Science Framework (needs datalad-osf)

```
datalad create-sibling-osf
```

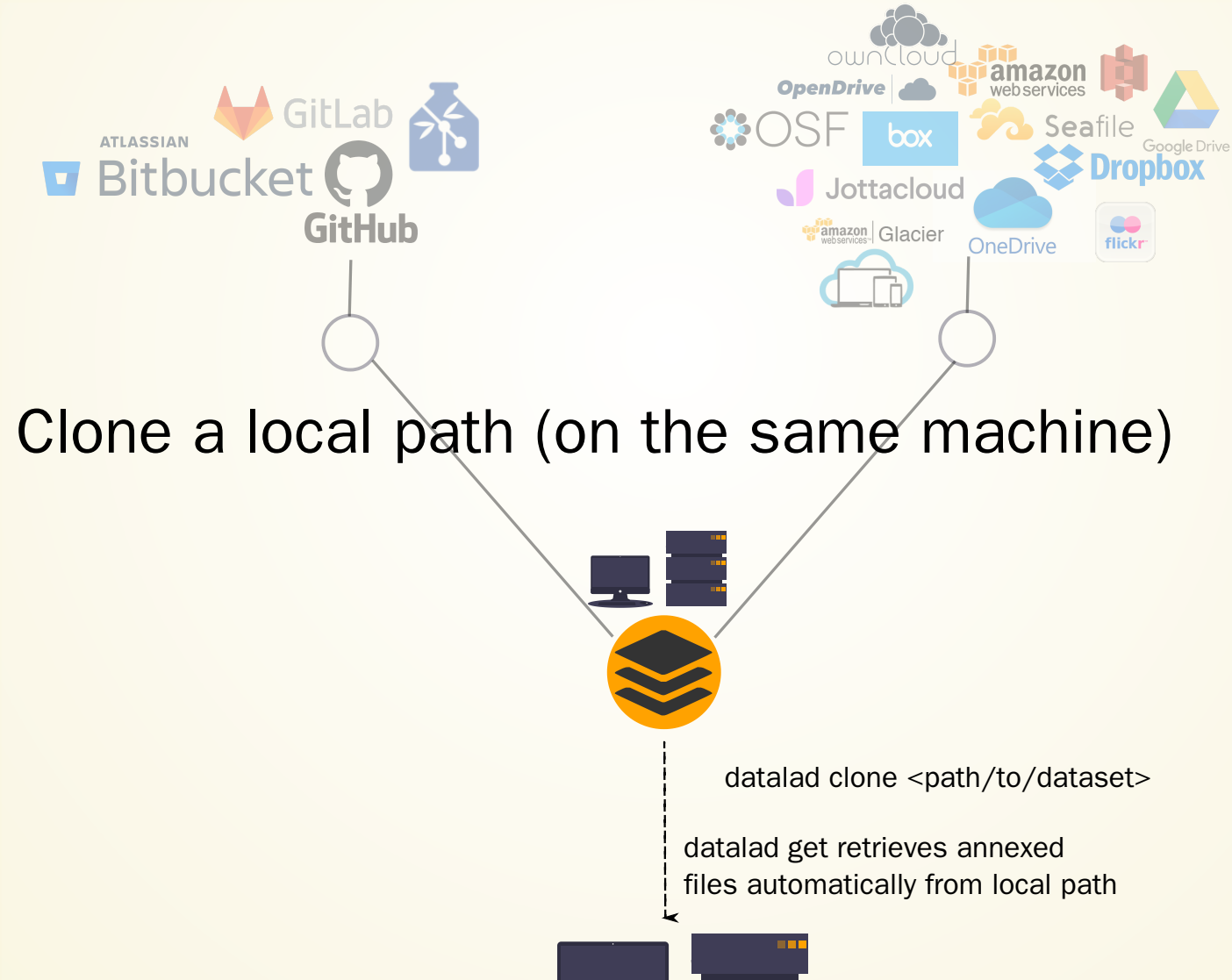
WebDAV-based hosting (e.g., Sciebo, EOSC; needs datalad-next)

```
datalad create-sibling-webdav
```

(Additional services being worked on at this moment: Dataverse, ebrains;
Get in touch with additional service support requests)

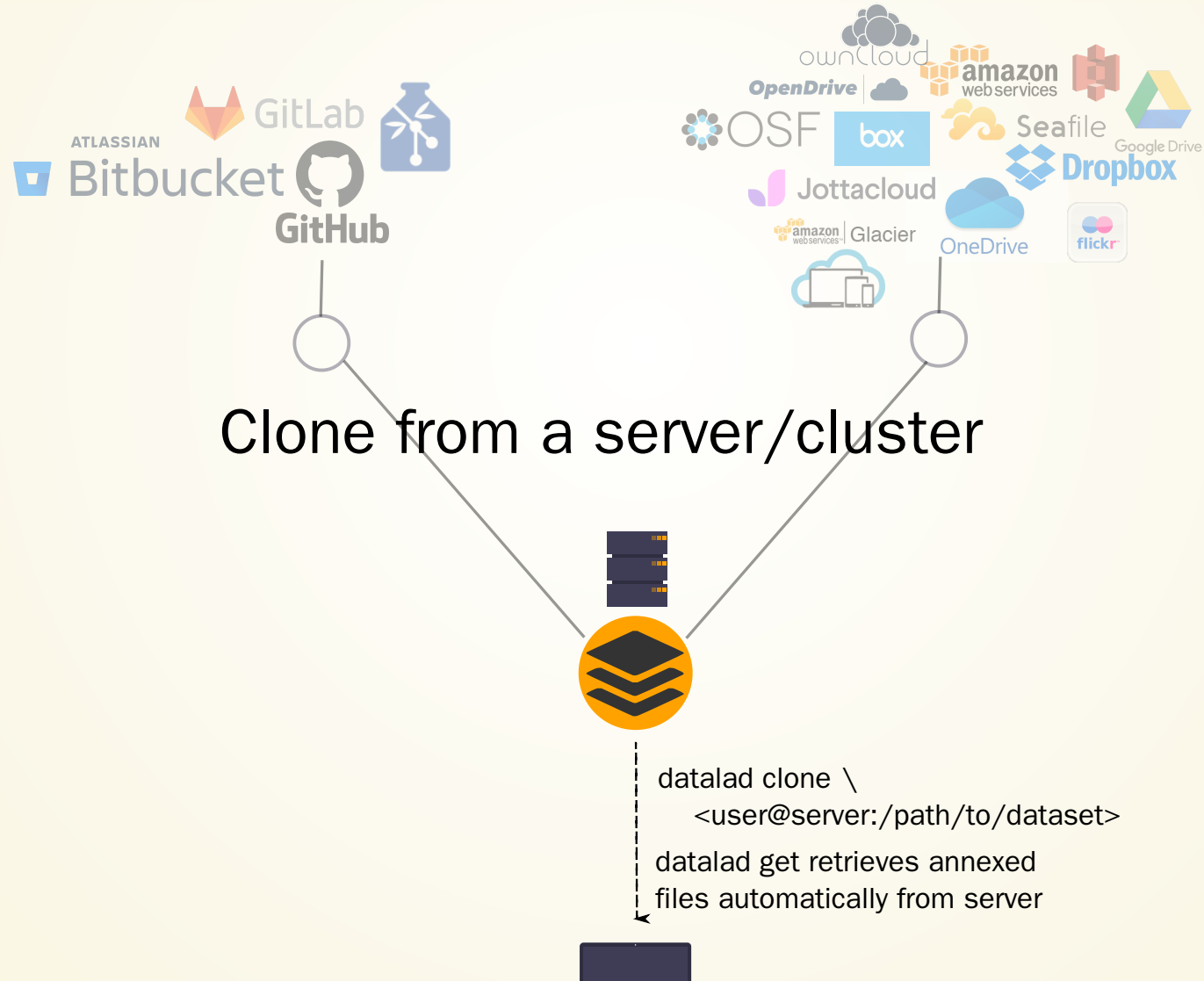
CLONING DATALAD DATASETS

How does cloning dataset feel like for a consumer?



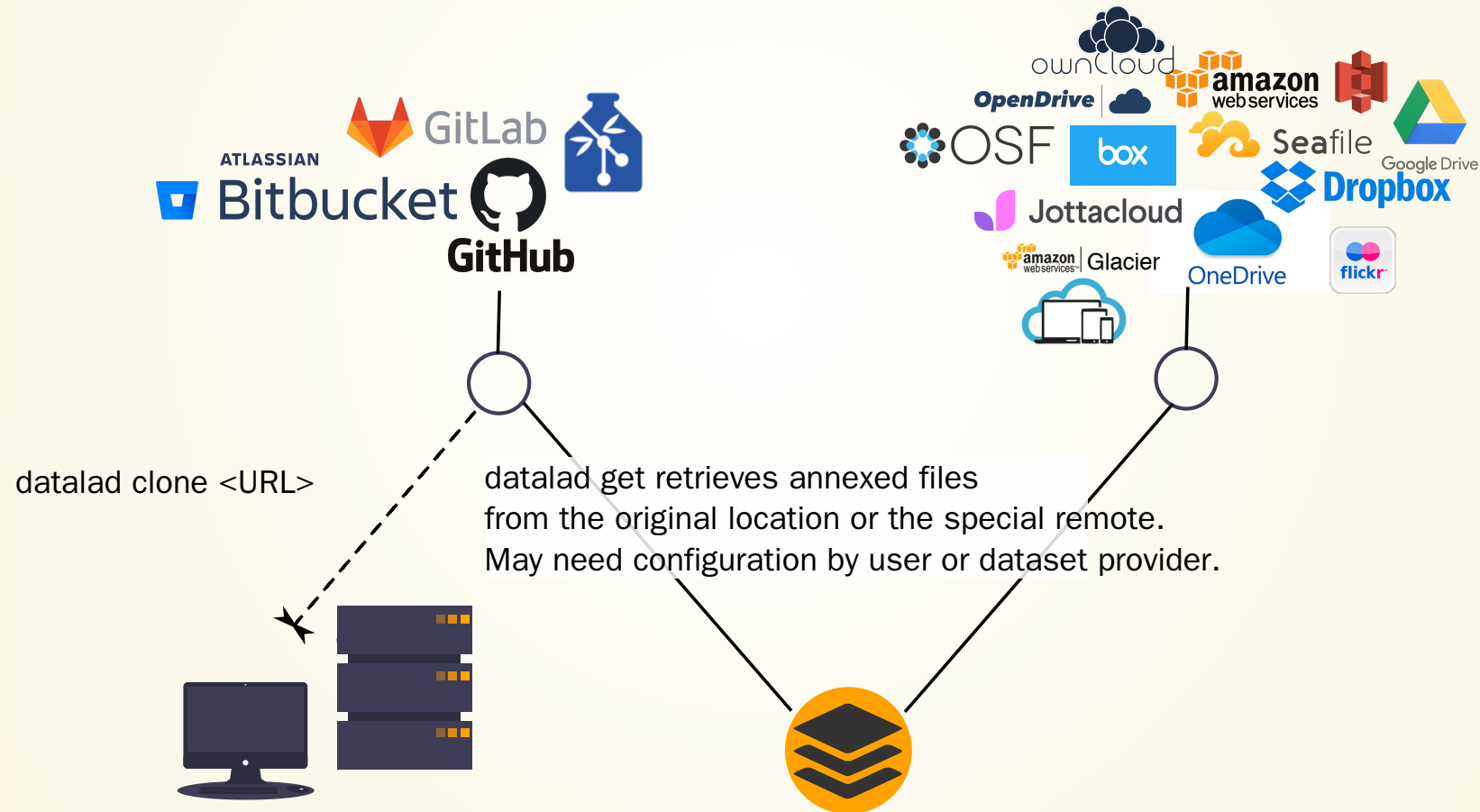
CLONING DATALAD DATASETS

How does cloning dataset feel like for a consumer?



CLONING DATALAD DATASETS

How does cloning dataset feel like for a consumer?



Clone from a remote URL

CLONING DATALAD DATASETS

Let's take a look at the special cases:
[LIVE DEMO CLONING GIN]

CLONING DATALAD DATASETS

Let's take a look at the special cases:

Requires the DataLad extension `dataad-osf`

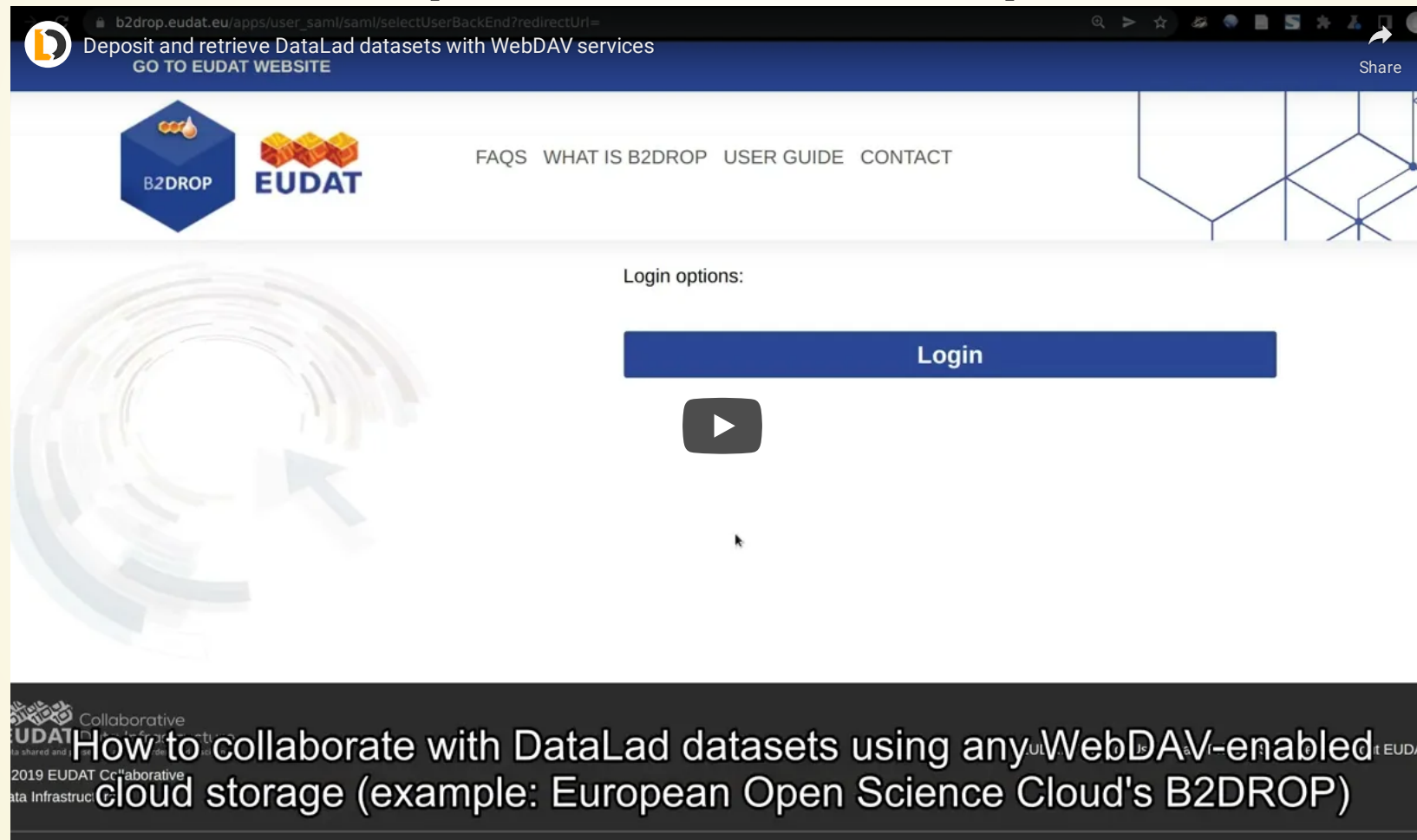
[LIVE DEMO CLONING OSF]

CLONING DATALAD DATASETS

Let's take a look at the special cases:

Requires the DataLad extension `datalad-next`

[DEMO CLONING WebDAV]



The screenshot shows the B2DROP web interface. The browser address bar displays `b2drop.eudat.eu/apps/user_saml/saml/selectUserBackEnd?redirectUri=`. The page header includes the B2DROP and EUDAT logos, a navigation menu with links for FAQs, WHAT IS B2DROP, USER GUIDE, and CONTACT, and a 'Share' button. The main content area features a large circular graphic on the left and a 'Login options:' section with a prominent blue 'Login' button and a play button icon below it. The footer contains the text 'How to collaborate with DataLad datasets using any WebDAV-enabled cloud storage (example: European Open Science Cloud's B2DROP)' and the EUDAT logo.

b2drop.eudat.eu/apps/user_saml/saml/selectUserBackEnd?redirectUri=

Deposit and retrieve DataLad datasets with WebDAV services

GO TO EUDAT WEBSITE

Share

B2DROP EUDAT

FAQS WHAT IS B2DROP USER GUIDE CONTACT

Login options:

Login

How to collaborate with DataLad datasets using any WebDAV-enabled cloud storage (example: European Open Science Cloud's B2DROP)

EUDAT

SUMMARY: DATA PUBLICATION

In case of fire 

 1. git commit
(datalad save)

 2. git push
(datalad push)

 3. leave building

datasets can have "siblings", linked clones in other places

Those can be local or remote, on commercial, free, or personal infrastructure

Typical repository hosting services do not host annexed contents

A notable exception is Gin

Typical storage providers do not host Git repositories

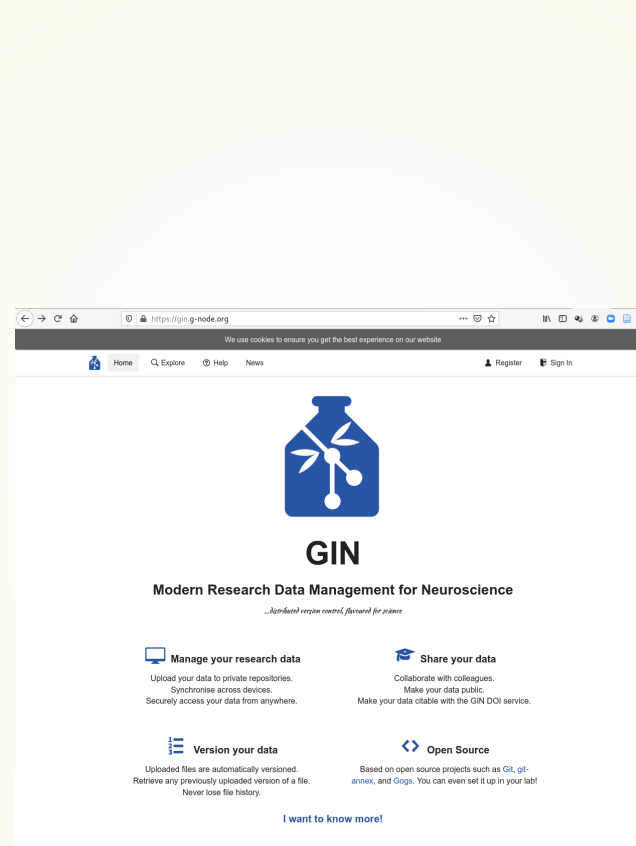
but datalad extensions can make it possible for certain services, such as the OSF

Despite the different possible services, operations are streamlined

PUBLISH YOUR OWN DATASET

Code: psychoinformatics-de.github.io/rdm-course/03-remote-collaboration/index.html#publishing-datasets-to-gin

USING GIN FOR DATA PUBLICATION



Gin has a few advantages for publishing data

- DataLad Integration: Convenience commands to create siblings
- Annex support: Easiest possible publication, preview and individual download of annexed contents in the webinterface
- Open Science support: Archive datasets to obtain a DOI; ensures minimal metadata and a licence

USING GIN FOR DATA PUBLICATION

The screenshot displays the GIN web interface for the repository 'Innrwttkhn / highspeed-bids'. The browser address bar shows the URL 'https://gin.g-node.org/Innrwttkhn/highspeed-bids'. The repository page includes a navigation bar with links to Dashboard, Issues, Pull Requests, Explore, Help, and News. Below the repository name, there is a DOI '10.12751/g-node.4ivuv8' and buttons for Watch (1), Star (1), and Fork (1). The main content area shows the repository description: 'DataLad dataset containing BIDS-converted MRI data used in Wittkuhn & Schuck, 2020, Nature Communications. For details, see: https://wittkuhn.mpib.berlin/highspeed/'. It also displays statistics: 60 Commits, 2 Branches, and 0 Releases. A section for cloning the repository is visible, with options for GIN, HTTPS, and SSH. The commit history is shown as a table with columns for the commit author, commit hash, commit message, and time ago.

Author	Commit Hash	Commit Message	Time Ago
Lennart Wittkuhn	0f424f45da	Merge remote-tracking branch 'origin/master'	5 months ago
	a4652dfd2b	Configure metadata type(s)	1 year ago
	a2b905b439	add defaced BIDS-converted MRI data after running heudiconv ...	1 year ago
	36bb8a3e68	update documentation	1 year ago
	f4b832b93c	[DATALAD] Recorded changes	1 year ago
	4223e4692d	add empty logs directory	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago
	5dd8eb86d8	[DATALAD RUNCMD] create IntendedFor fields in fieldmap .json...	1 year ago

USING GIN FOR DATA PUBLICATION

- Step 1: Create a Gin account (requires an email address)
- Step 2: Generate and upload an SSH key
- Step 3: Create and register a sibling repository
- Step 4: Publish your dataset
- Step 5: Update your dataset

SUMMARY: PUBLISHING AND UPDATING DATA (GIN)

Gin is a free repository hosting service

To publish datasets to Gin, you need an account and an SSH key

DataLad has built-in integration with `datalad create-sibling-gin`

This requires generating an access token

Gin has annex support

`datalad push` published all dataset contents and the Git history

The dataset can be cloned from Gin by others

If the dataset is public, this does not even require a Gin account

You can still publish your dataset to (your lab's) GitHub/GitLab/other places

and use Gin only for data hosting. Walkthrough: handbook.datalad.org/basics/101-139-gin.html#ginbts

Next: Let's collaborate!

PUBLICATION AND COLLABORATION EXERCISE

	Local	Published	Collaborator	
	Local	Published	Local	Published
Clone dataset				
Save dataset modification	 + 		 + 	
Create dataset sibling on GIN	 + 		 + 	
Push dataset to GIN	 + 	 + 	 + 	 + 
Safely remove local dataset		 + 		 + 
Clone own dataset from GIN again	 + 	 + 	 + 	 + 
Save another dataset modification	 + 	 + 	 + 	 + 
Push dataset update to GIN	 + 	 + 	 + 	 + 
Add a collaborator's GIN project as a sibling to local dataset	 + 	 + 	 + 	 + 
Merge collaborator's work	 +  + 	 + 	 +  + 	 +  + 
Push dataset update to GIN	 +  + 	 +  + 	 +  + 	 +  + 