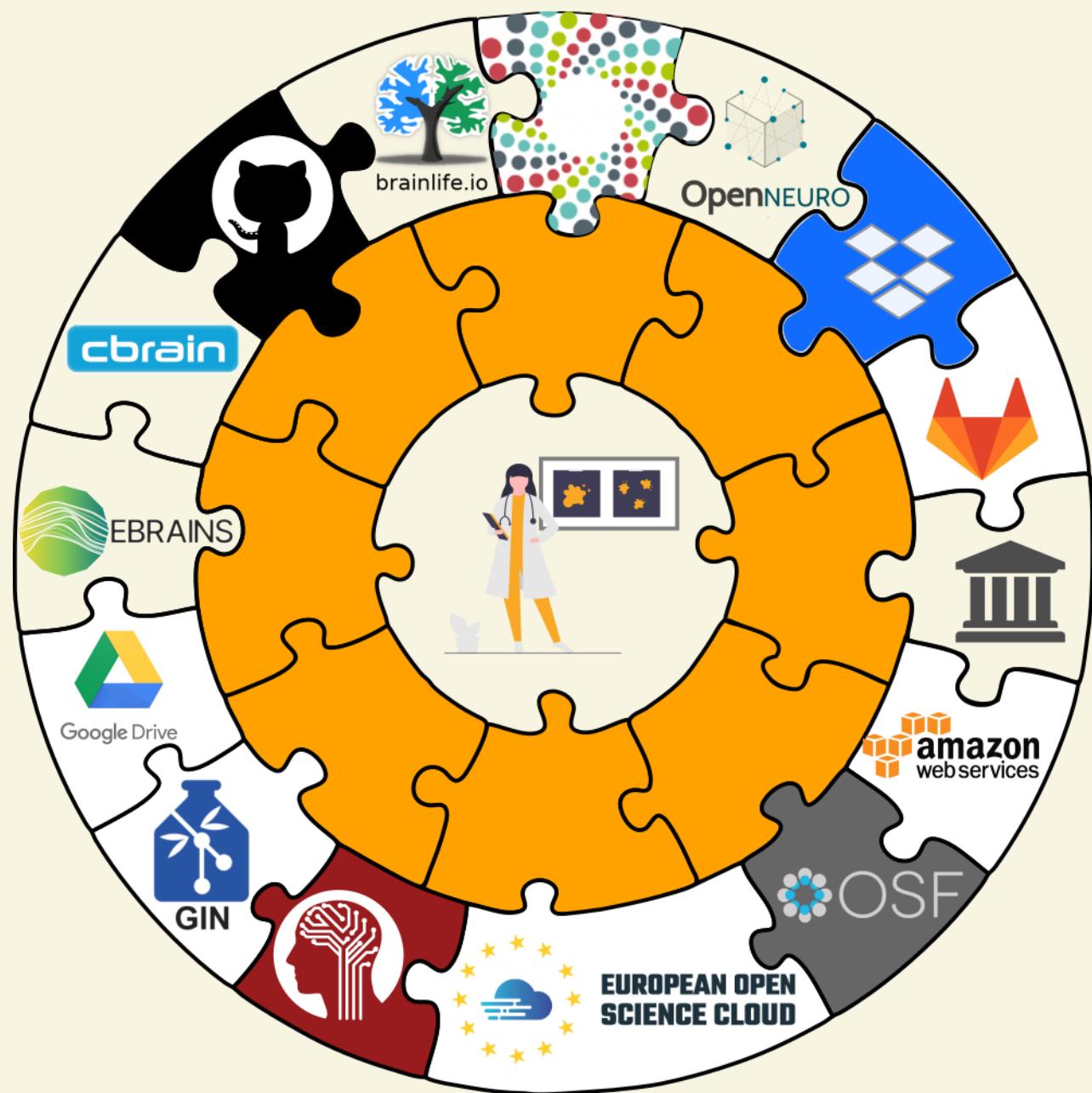


PUBLISHING DATA

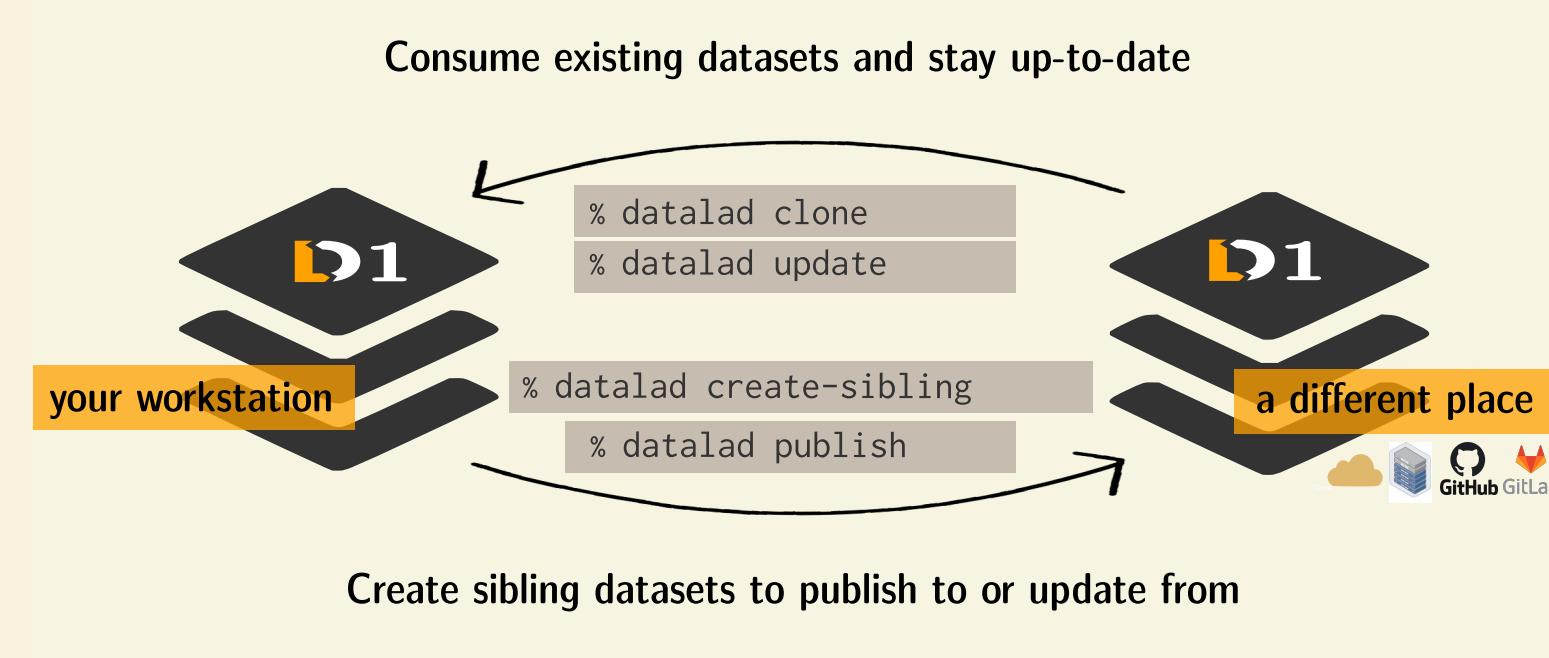


TRANSPORT LOGISTICS

- Share data like source code

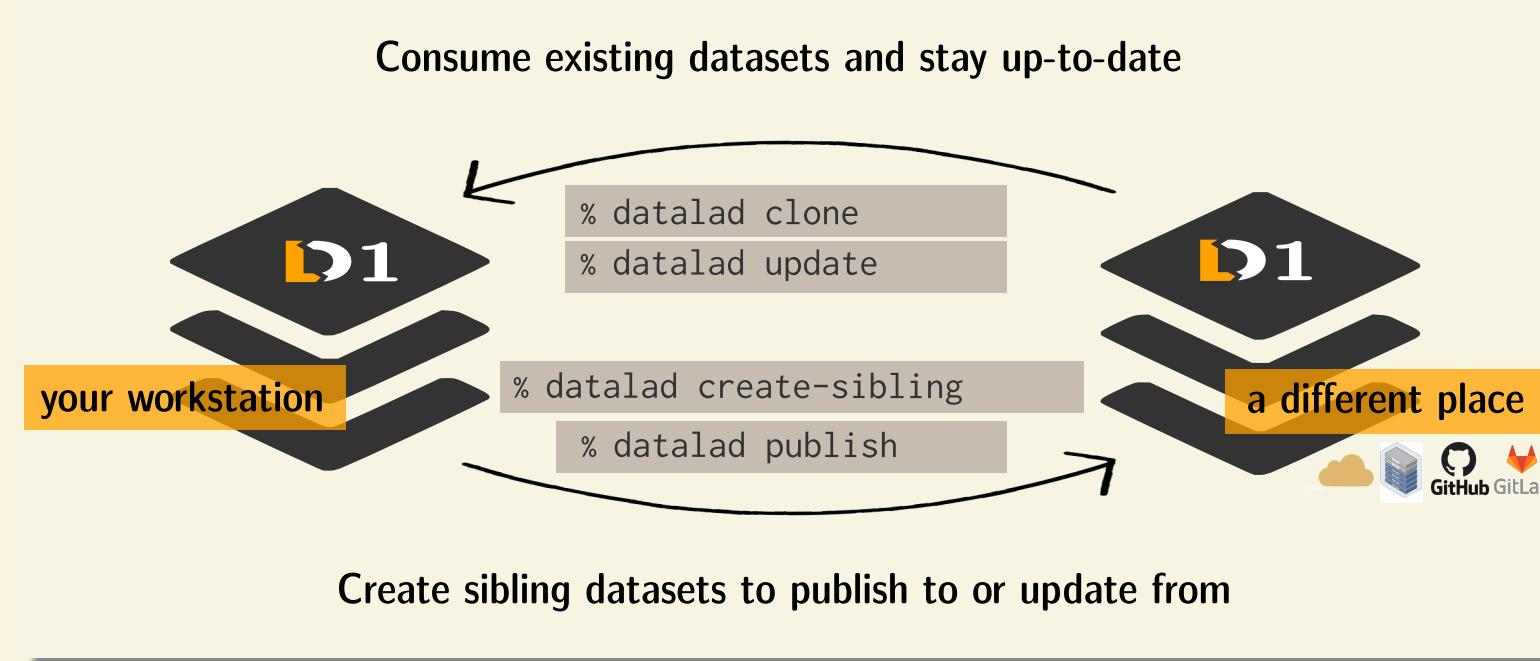
TRANSPORT LOGISTICS

- Share data like source code
- Datasets can be cloned, pushed, and updated from and to local paths, remote hosting services, external special remotes



TRANSPORT LOGISTICS

- Share data like source code
- Datasets can be cloned, pushed, and updated from and to local paths, remote hosting services, external special remotes



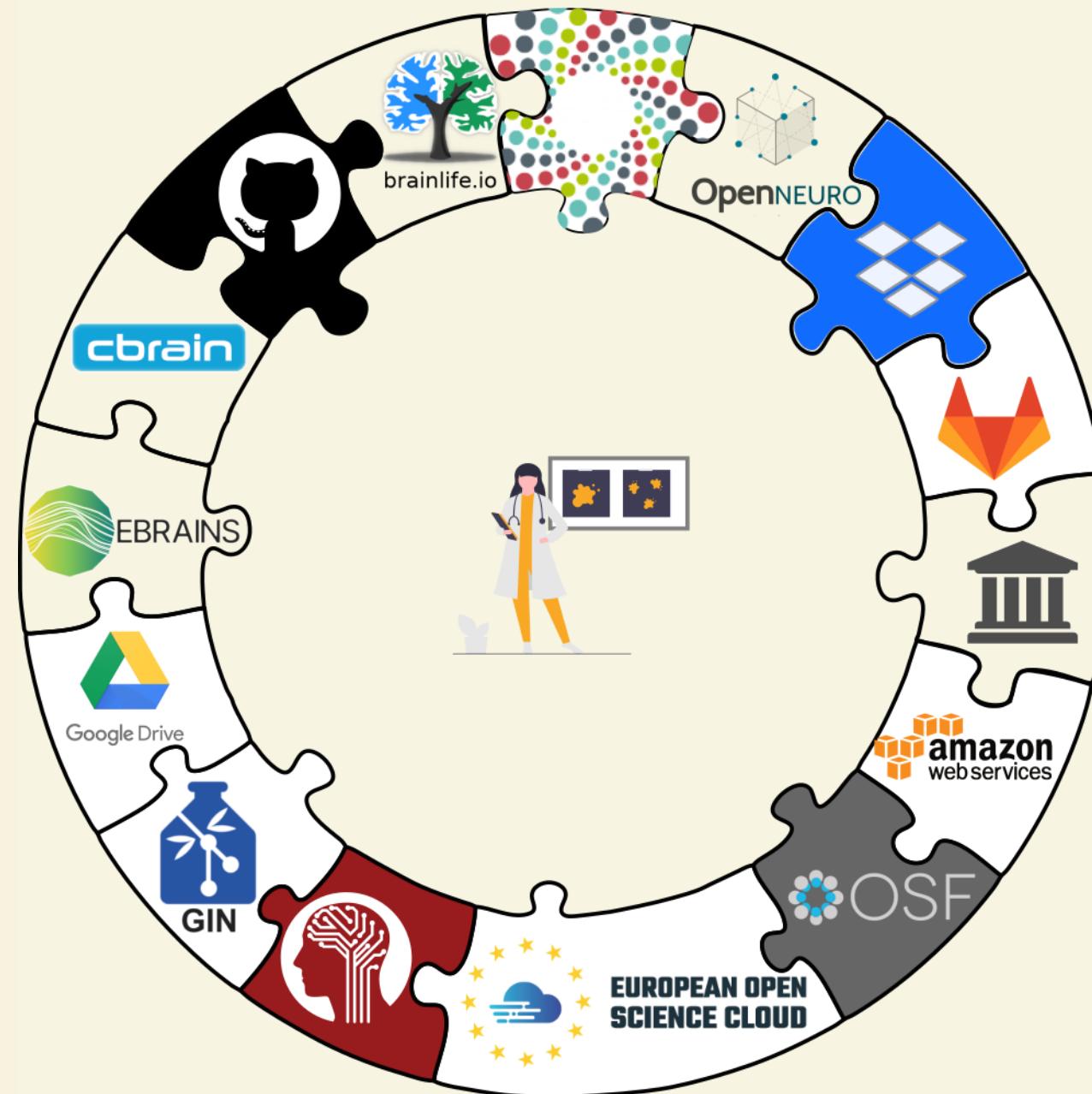
- Flexible data access management for annexed file contents based on storage location

INTEROPERABILITY

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

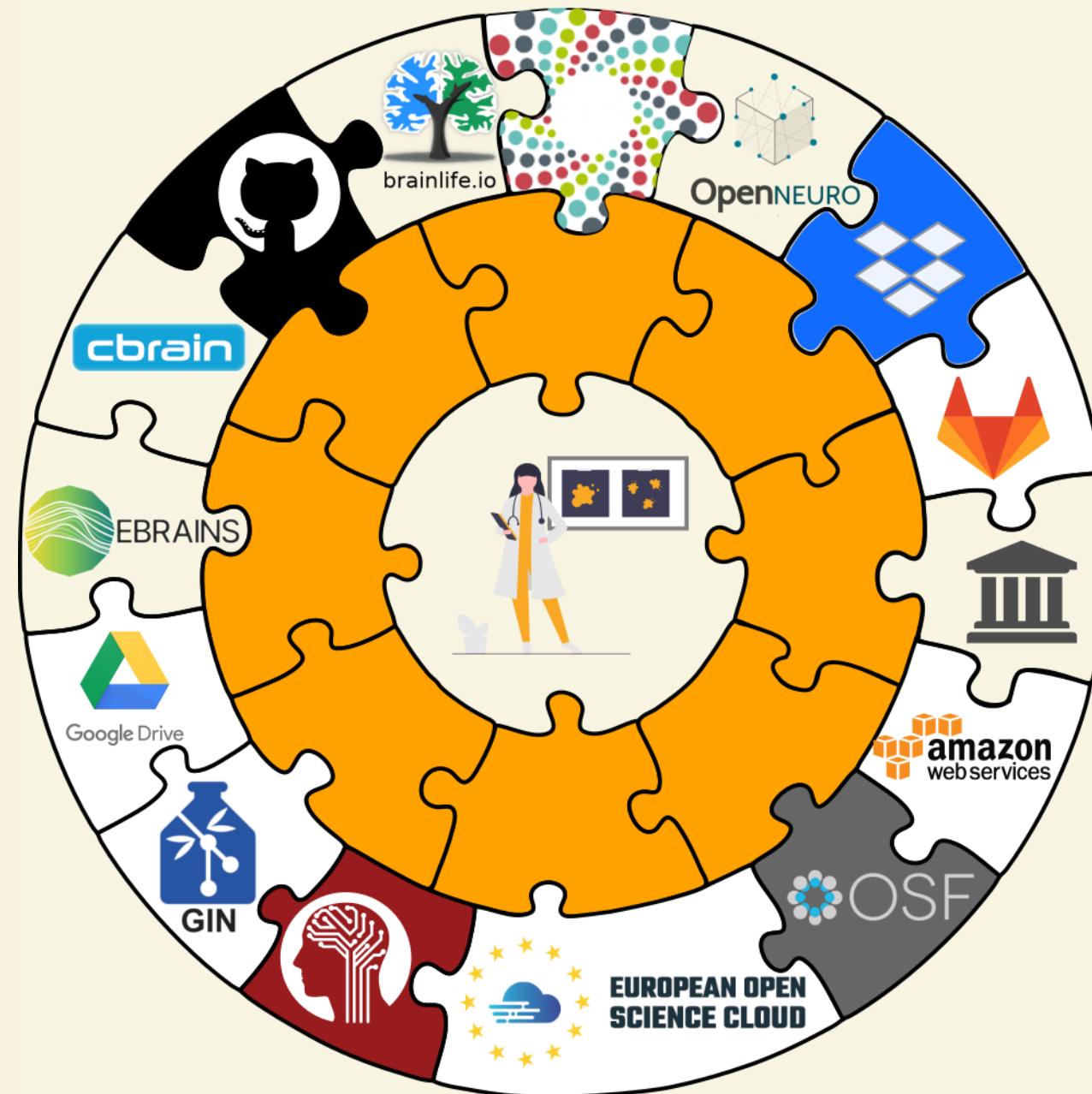
INTEROPERABILITY

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



INTEROPERABILITY

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



PUBLISHING DATASETS

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

Repository hosting services



Third party storage providers

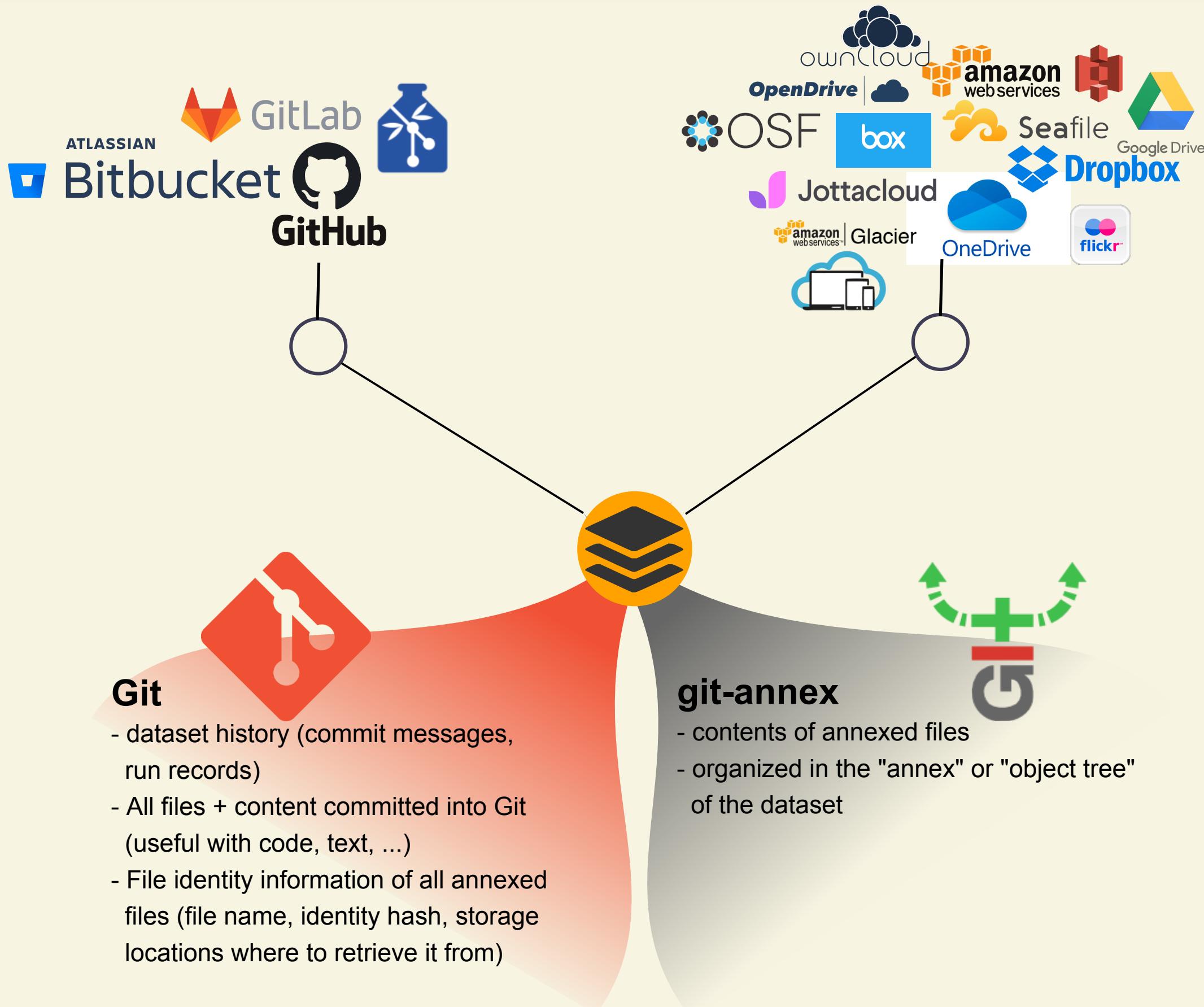
Your dataset



work stations & servers

PUBLISHING DATASETS

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



PUBLISHING DATASETS



Repository hosting

- usually no annex support & can't hold large data for free
- exposes Git history and files stored in Git
- 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 locations where to retrieve it from)

git-annex

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



PUBLISHING DATASETS



Repository hosting

- usually no annex support & can't hold large data for free
- exposes Git history and files stored in Git
- 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 locations where to retrieve it from)



Storage hosting in a special remote

- usually no Git repository hosting service
- stores the object tree/ file contents
- datasets keep track of where data is stored, `datalad get` retrieves file contents from special remote

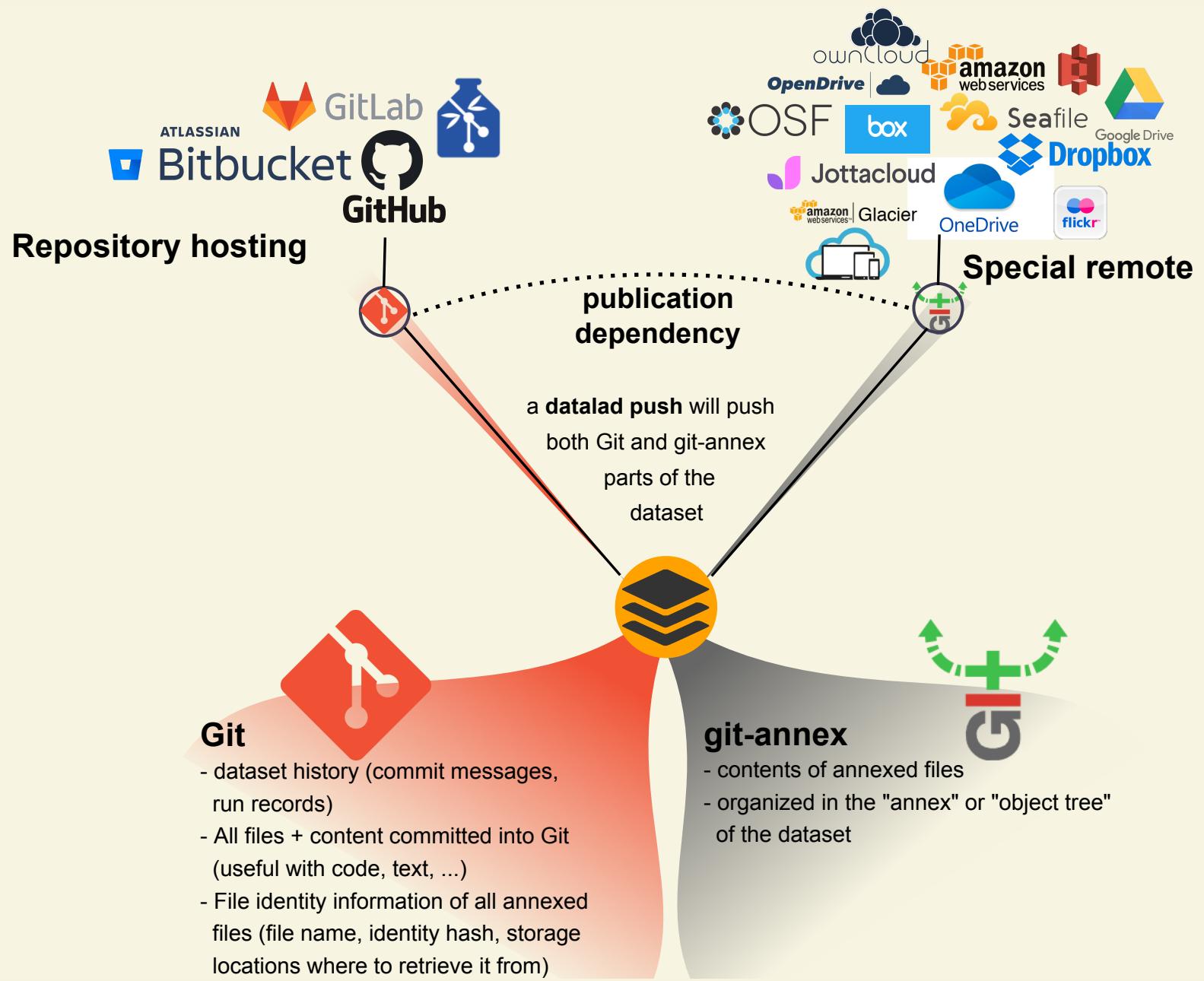
git-annex

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



PUBLISHING DATASETS

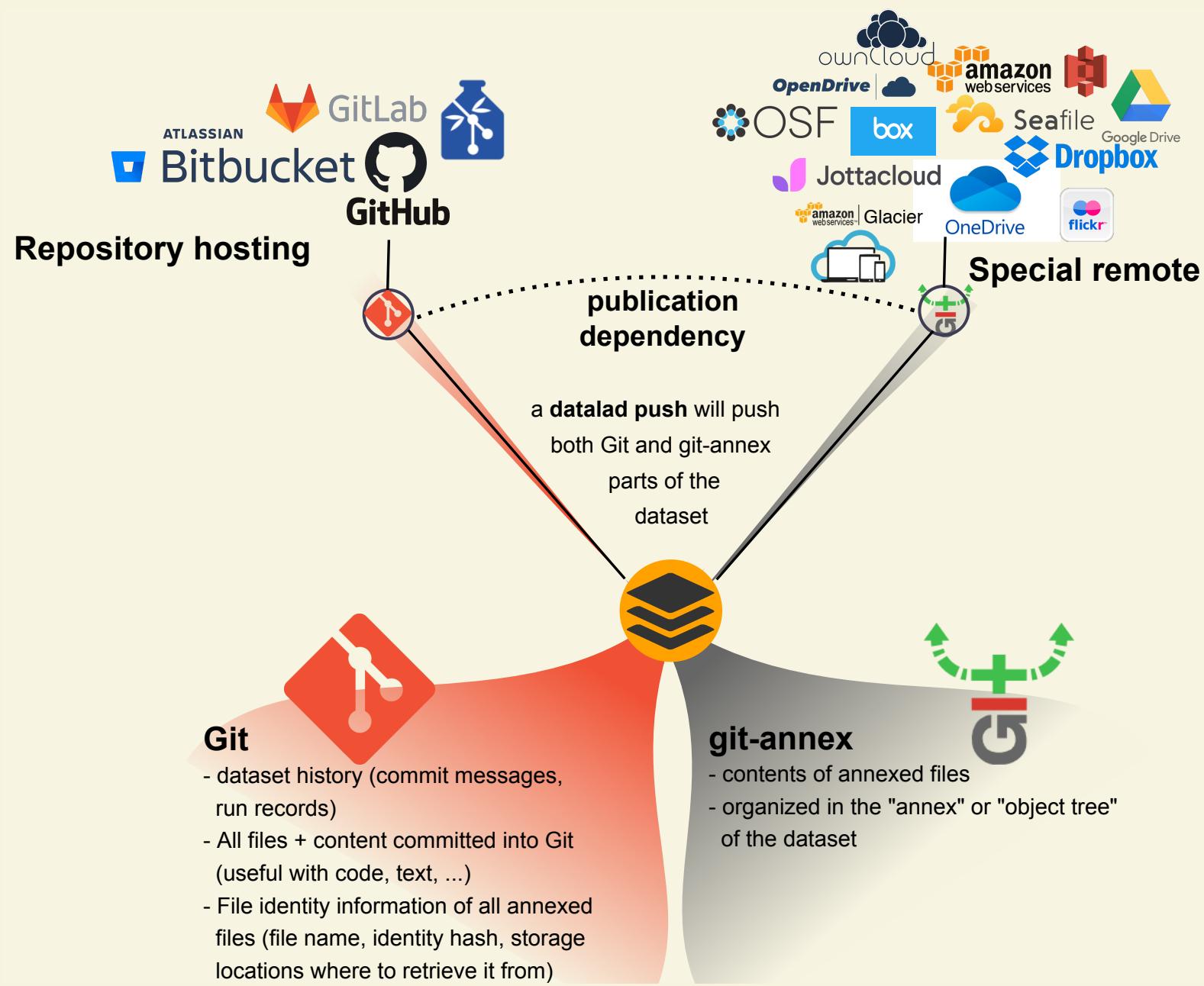
Typical case:



PUBLISHING DATASETS

Typical case:

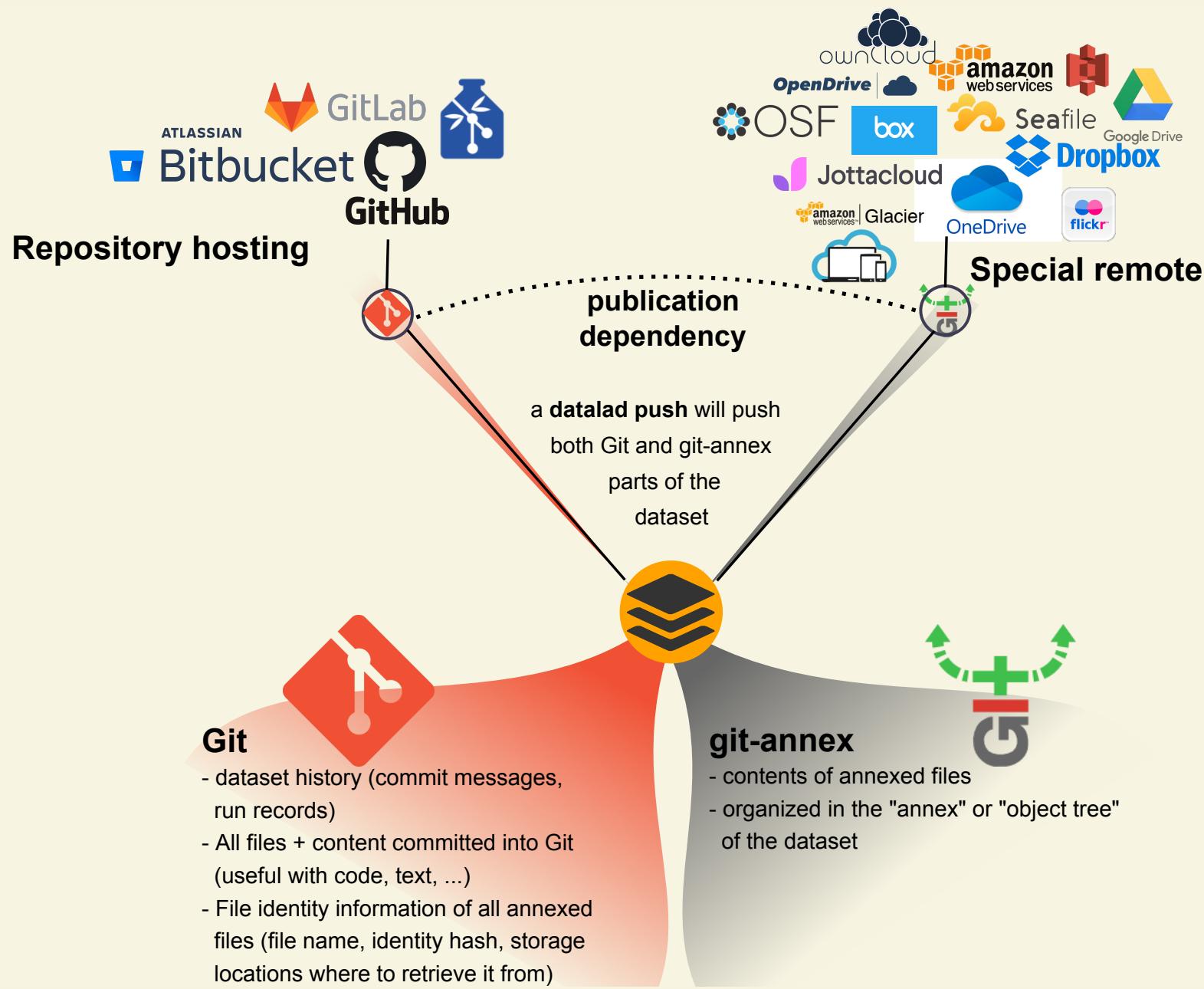
- Datasets are exposed via a private or public repository on a repository hosting service



PUBLISHING DATASETS

Typical case:

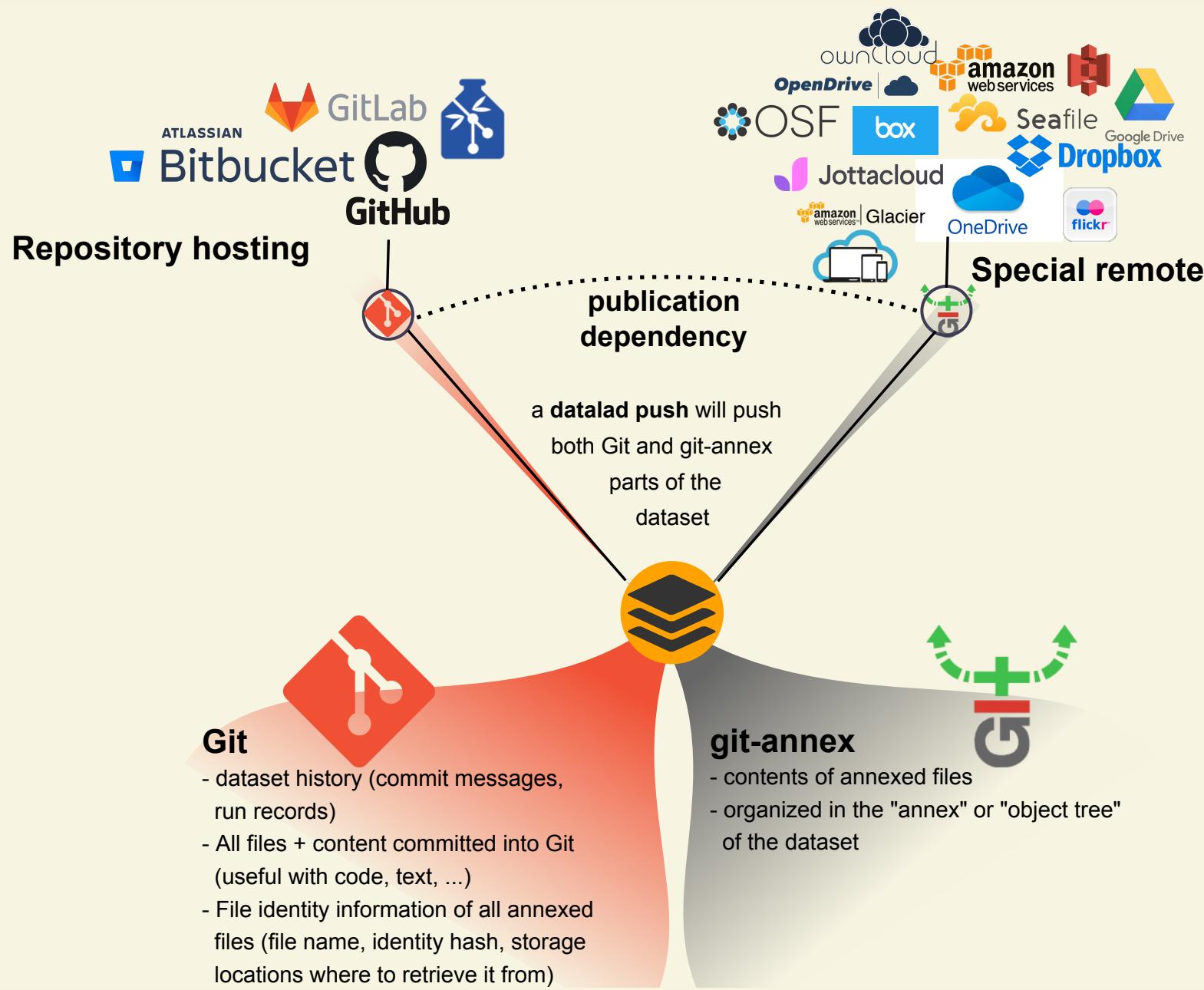
- Datasets are exposed via a private or public repository on a repository hosting service
- Data can't be stored in the repository hosting service, but can be kept in almost any third party storage



PUBLISHING 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



PUBLISHING DATASETS

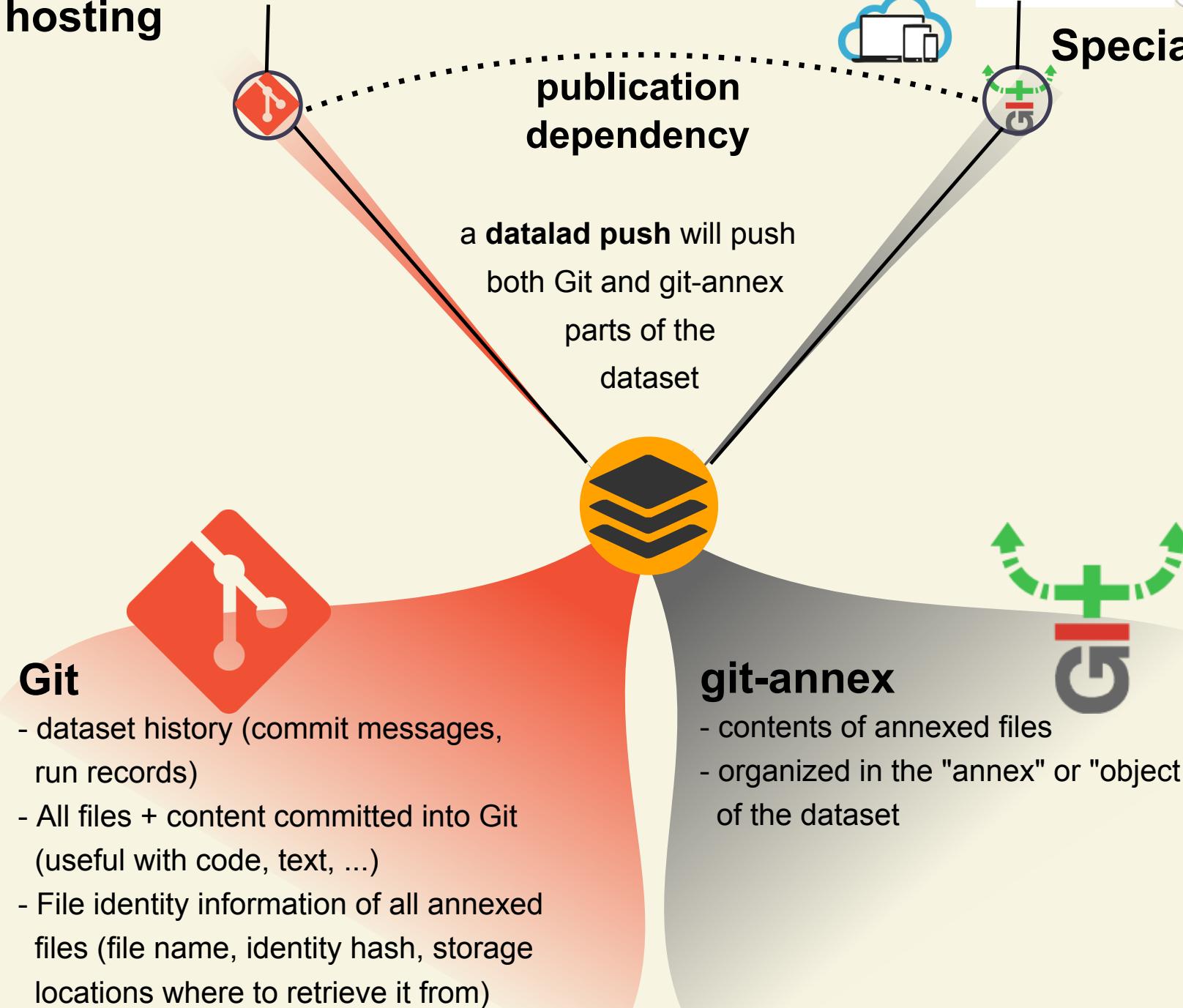
```
$ git config --local remote.github.datalad-publish-depends gdrive
```



Repository hosting



Special remote



PUBLISHING DATASETS

- Real-life example:

PUBLISHING DATASETS

- Real-life example:

The screenshot shows a GitHub repository page for 'psychoinformatics-de / studyforrest-data-phase2'. The repository has 1 pull request, 4 issues, 1 pull request, 77 commits, 2 branches, and 1 tag. The 'Code' tab is selected. The repository description states: 'studyforrest.org: Phase2 data (movie, eyetracking, remapping, visual localizers) [BIDS]'. It includes links to 'studyforrest.org', 'Readme', and 'View license'. There is one release listed: 'First public release (Latest) on Mar 26, 2016'. The 'Contributors' section lists three contributors: 'mih Michael Hanke', 'dakot Daniel Kottke', and 'adswa Adina Wagner'. The 'Languages' section shows a horizontal bar with blue and green segments.

psychoinformatics-de / **studyforrest-data-phase2**

Pull requests Issues Marketplace Explore

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

sub-19 BF: Re-import respiratory trace after bug fix in converte... 3 years ago

sub-20 BF: Re-import respiratory trace after bug fix in converte... 3 years ago

About

studyforrest.org: Phase2 data (movie, eyetracking, remapping, 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

Languages

PUBLISHING DATASETS

Special case 1: repositories with annex support

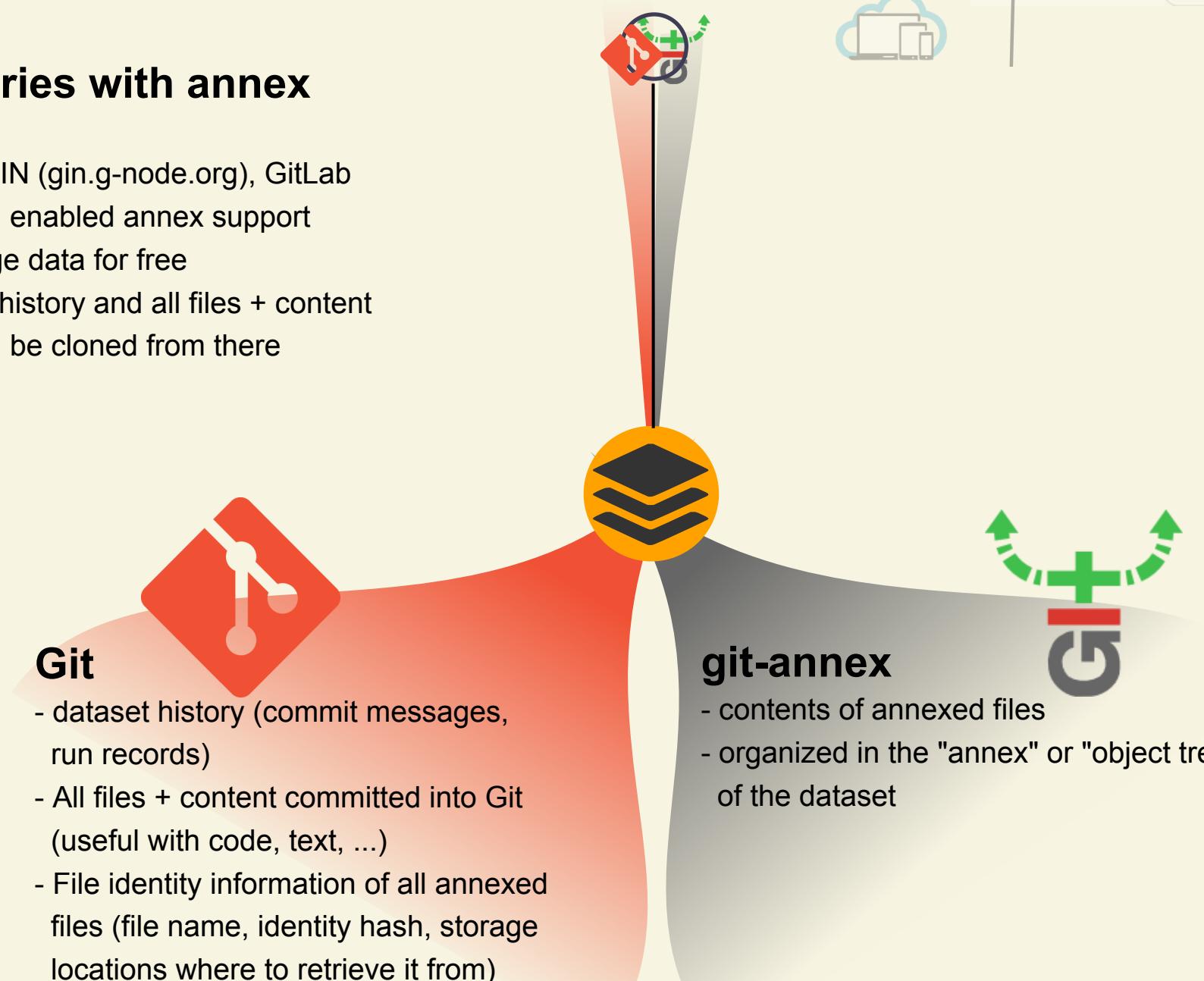
PUBLISHING DATASETS

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

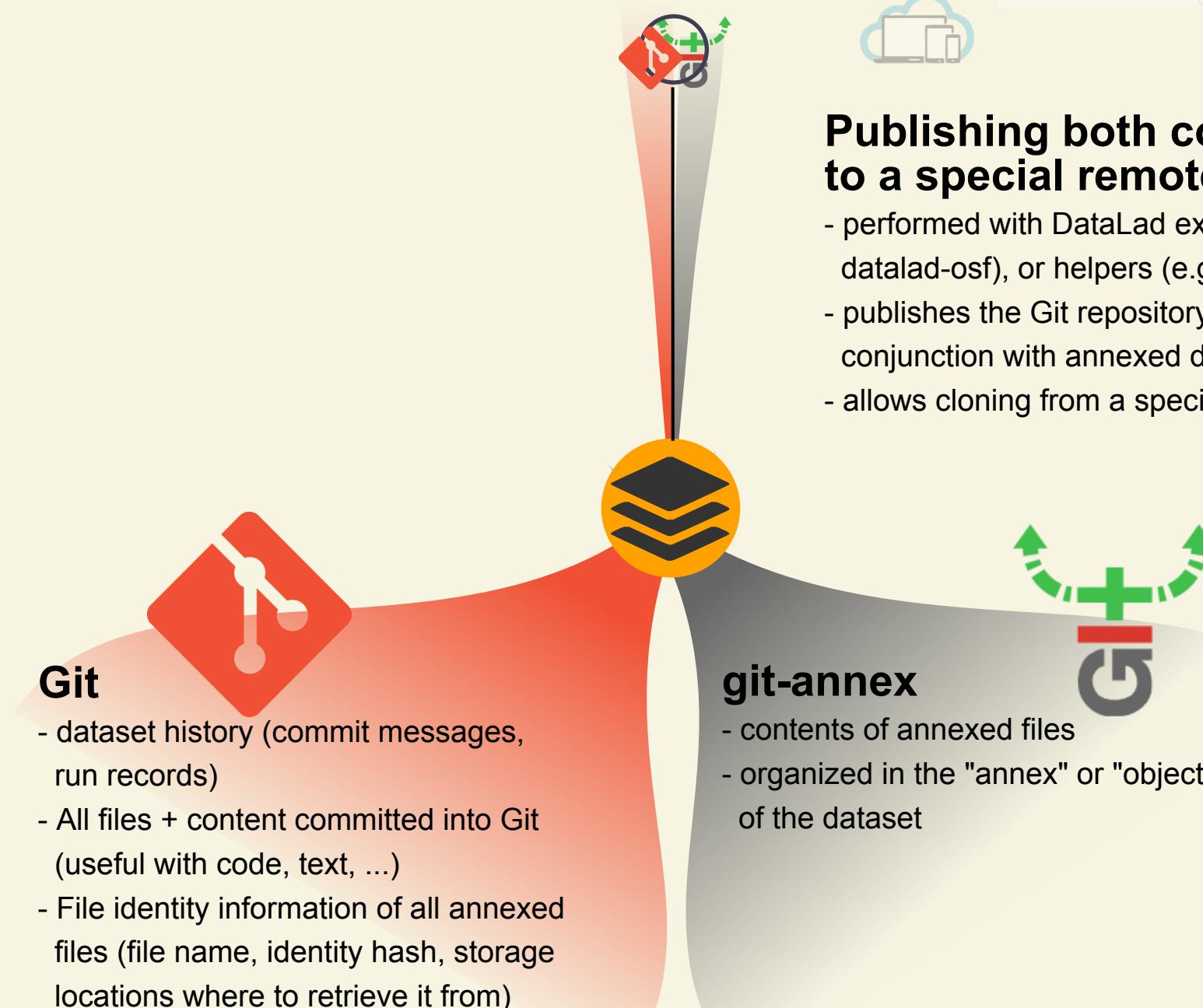
- dataset history (commit messages, run records)
- All files + content committed into Git (useful with code, text, ...)
- File identity information of all annexed files (file name, identity hash, storage locations where to retrieve it from)

git-annex

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

PUBLISHING DATASETS

Special case 2: Special remotes with repositories



Publishing both components to a special remote

- performed with DataLad extensions (e.g., `datalad-osf`), or helpers (e.g., `git-remote-rclone`)
- publishes the Git repository (in two files) in conjunction with annexed data
- allows cloning from a special remote

PUBLISHING DATASETS

Special case 1: repositories with annex support

PUBLISHING DATASETS

Special case 1: repositories with annex support

(handbook) adina@muninn in /tmp

adswa/comics - G-Node - Mozilla Firefox

adswa/comics - G-Node x +
https://gin.g-node.org/adswa/comics
adswa / comics
Files Issues 0 Wiki Settings
Quick Guide
Get this repository with the gin client
gin get adswa/comics
Or Clone this repository Need help cloning? Visit Help!
HTTPS SSH git@gin.g-node.org:/adswa/comics.git
Create a new repository on the command line
touch README.md
git init
git add README.md

The screenshot shows a terminal window at the top with the command `(handbook) adina@muninn in /tmp` and a file transfer progress bar. Below it is a Firefox browser window displaying a G-Node repository page for `adswa/comics`. The browser title bar says "adswa/comics - G-Node - Mozilla Firefox". The page content includes instructions for cloning the repository using the `gin` client or SSH, and a section for creating a new repository on the command line.

PUBLISHING DATASETS

Special case 2: Special remotes with repositories

HTTP request sent, awaiting response... 200 OK
Length: 27246 (27K) [image/png]
Saving to: 'interdisciplinary.png'

interdisciplinary.png 100%[=====] 26.61K ---KB/s in 0.005s

2020-11-12 13:19:58 (4.94 MB/s) - 'interdisciplinary.png' saved [27246/27246]

(handbook) adina@muninn in /tmp/comics on git:master
❯ datalad save -m "Add a comic"
add(ok): interdisciplinary.png (file)
save(ok): . (dataset)
action summary:
 add (ok: 1)
 save (ok: 1)
(handbook) adina@muninn in /tmp/comics on git:master
❯ datalad siblings add -s gin --url git@gin.g-node.org:/adswa/comics.git
[INFO] Could not enable annex remote gin. This is expected if gin is a pure Git remote, or happens if it is not accessible.
[WARNING] Could not detect whether gin carries an annex. If gin is a pure Git remote, this is expected.
. : gin(-) [git@gin.g-node.org:/adswa/comics.git (git)]
(handbook) adina@muninn in /tmp/comics on git:master
❯ datalad push --to gin
copy(ok): interdisciplinary.png (file) [to gin...]
publish(ok): . (dataset) [refs/heads/git-annex->gin:refs/heads/git-annex c28dd74..2fa0974]
publish(ok): . (dataset) [refs/heads/master->gin:refs/heads/master [new branch]]
(handbook) adina@muninn in /tmp/comics on git:master
❯

adswa/comics - G-Node GIN - Mozilla Firefox

adswa/comics - G-Node X +

https://gin.g-node.org/adswa/comics/

We use cookies to ensure you get the best experience on our website.

Dashboard Issues Pull Requests Explore Help News

adswa / comics

How to publish Add DataCite file

Unwatch 1 Star 0 Fork 0

Files Issues 0 Pull Requests 0 Wiki Settings

No Description

2 Commits 2 Branches 0 Releases

Branch: master comics New file Upload file GIN HTTPS SSH git@gin.g-node.org:/adswa/ci

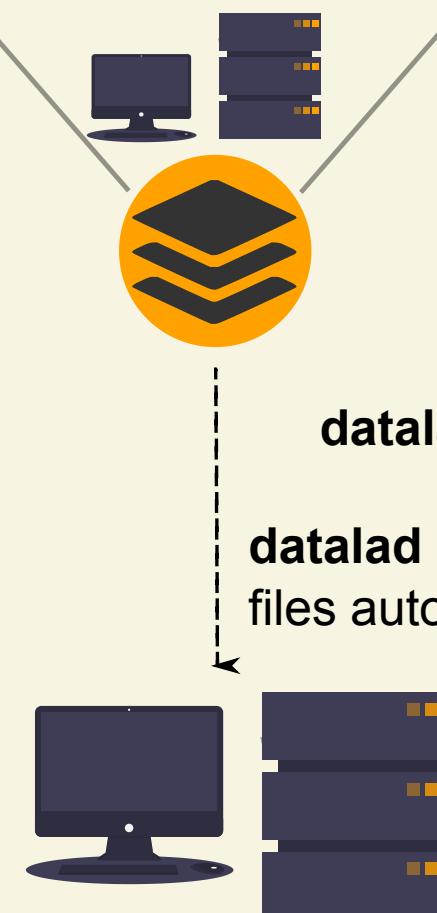
Adina Wagner 8e03e7803b Add a comic 8 minutes ago
.datalad 4e0bfa3b9c [DATALAD] new dataset 8 minutes ago
.gitattributes 4e0bfa3b9c [DATALAD] new dataset 8 minutes ago
interdisciplinary.png 8e03e7803b Add a comic 8 minutes ago

CLONING DATALAD DATASETS

How does cloning dataset feel like for a consumer?



Clone a local path (on the same machine)

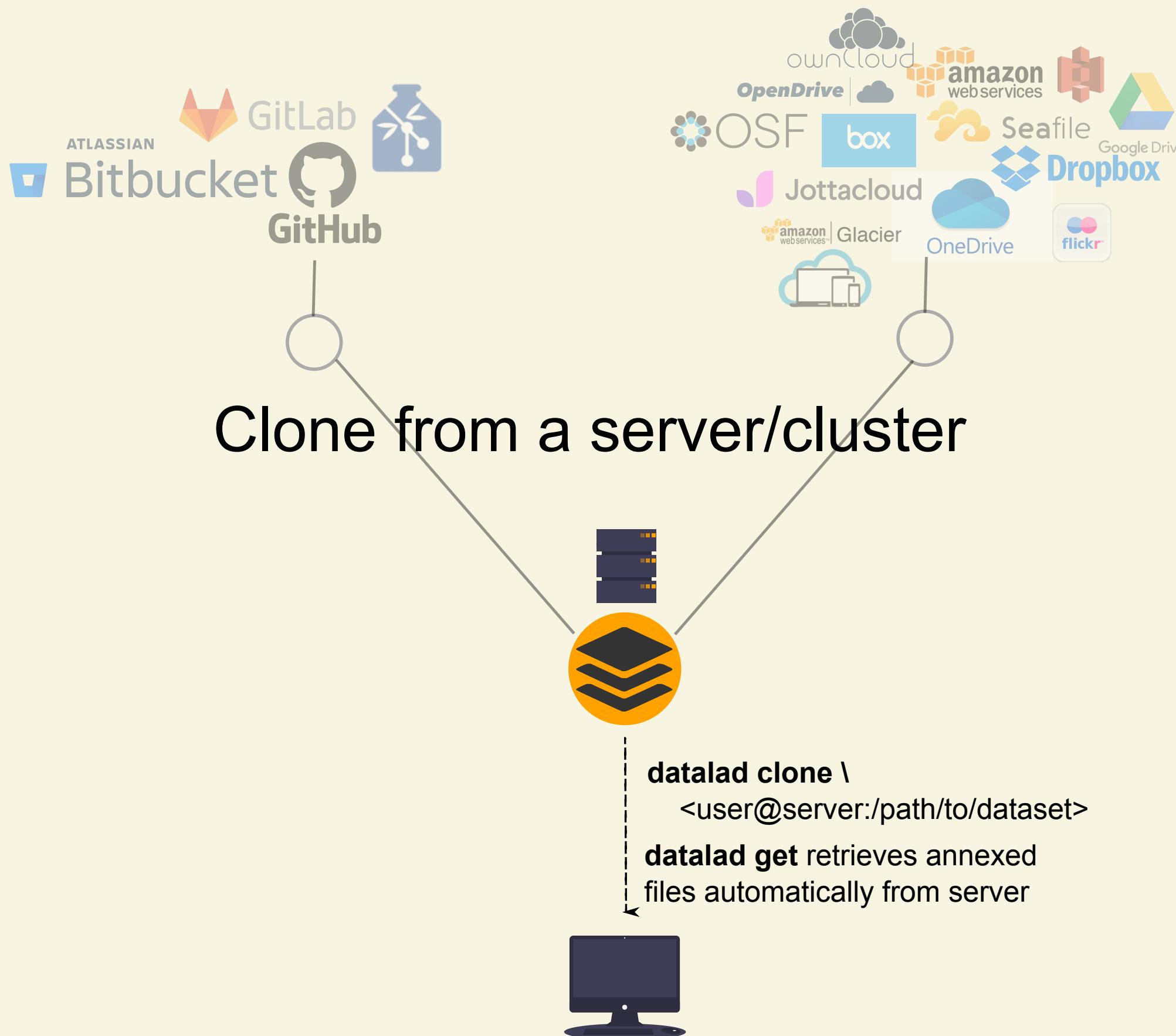


datalad clone <path/to/dataset>

datalad get retrieves annexed
files automatically from local path

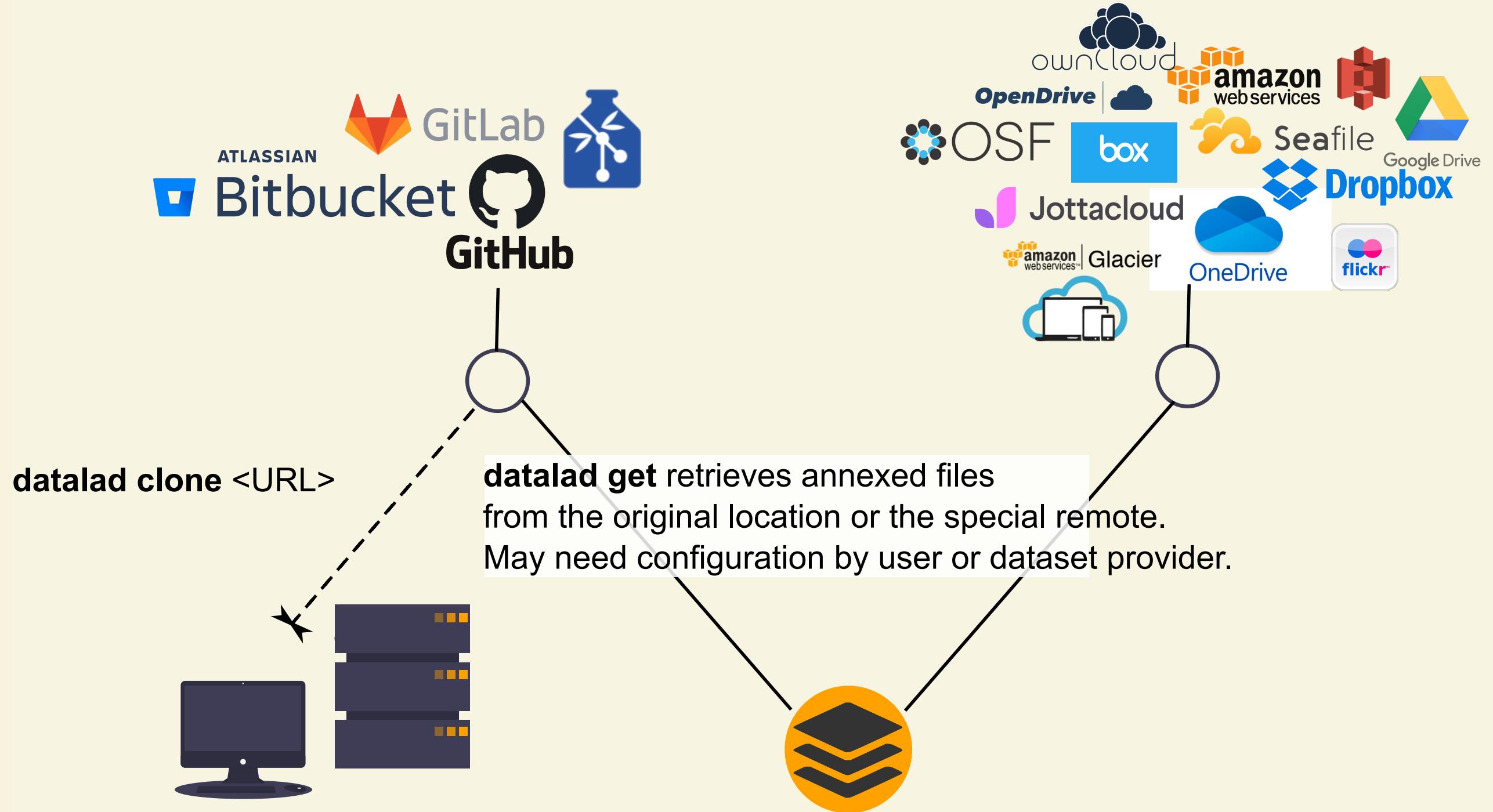
CLONING DATALAD DATASETS

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:

```
(handbook) adina@muninn in /tmp/mycloneofcomics on git:master
>
```

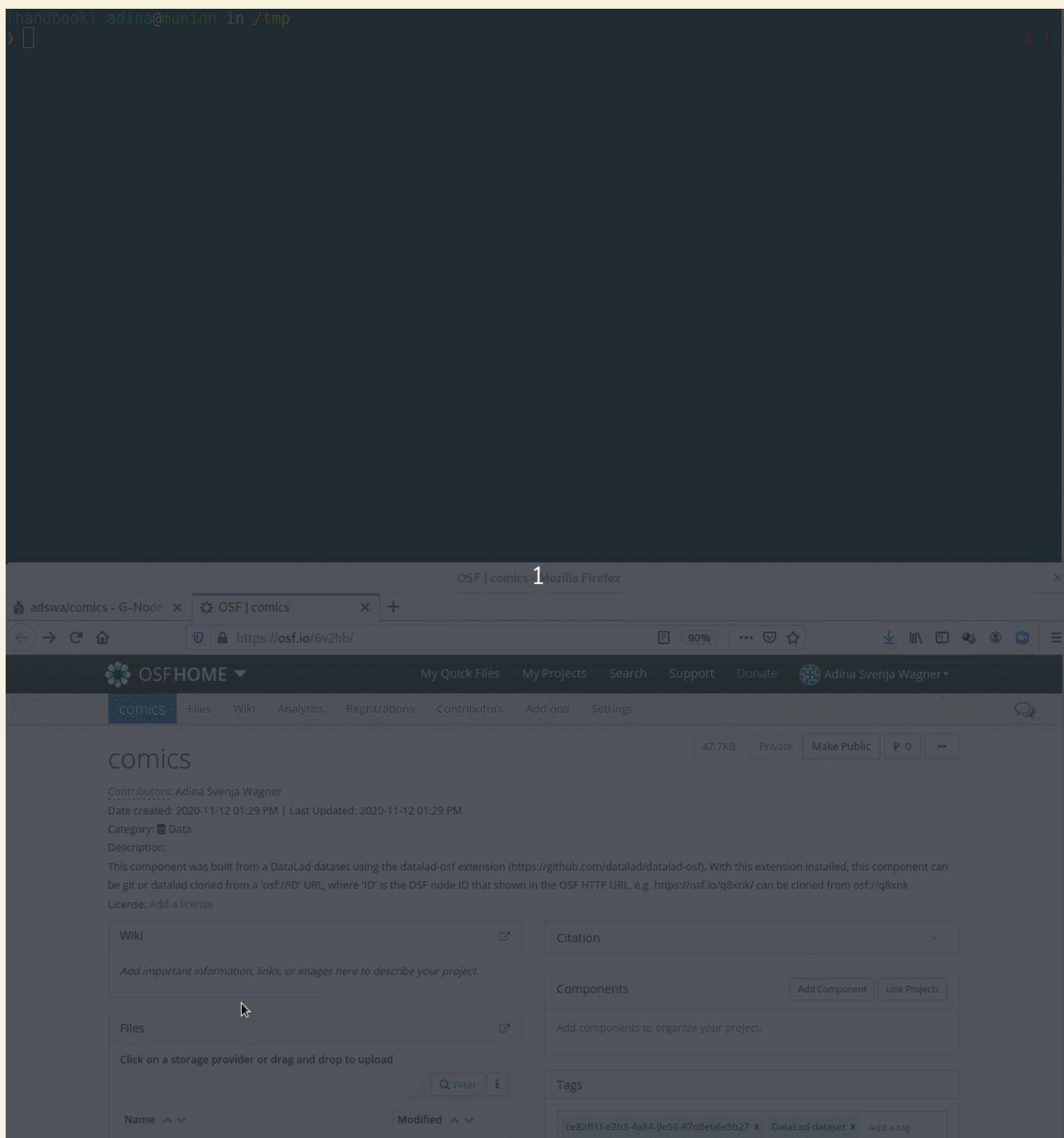
The screenshot shows a terminal window at the top with the command `(handbook) adina@muninn in /tmp/mycloneofcomics on git:master`. Below it is a Firefox browser window titled "adswa/comics - G-Node GIN - Mozilla Firefox". The browser address bar shows `https://gin.g-node.org/adswa/comics`. The page content includes a header with "Dashboard", "Issues", "Pull Requests", "Explore", "Help", and "News" buttons. It displays a "No Description" message and a commit history with four entries:

- Adina Wagner 8e03e7803b Add a comic 8 minutes ago
- .datalad 4e0bfa3b9c [DATALAD] new dataset 8 minutes ago
- .gitattributes 4e0bfa3b9c [DATALAD] new dataset 8 minutes ago
- interdisciplinary.png 8e03e7803b Add a comic 8 minutes ago

At the bottom of the browser window, there are buttons for "New file", "Upload file", "GIN", "HTTPS", "SSH", and download links.

CLONING DATALAD DATASETS

Let's take a look at the special cases:



DATA SHARING USING SEAFILE / KEEPER



More info: DataLad Handbook chapter on **Third party infrastructure**

CREATE A DATASET

CREATE A DATASET

Create a new DataLad dataset for your cartoon collection:

```
$ datalad create cartoon-collection
[INFO] Creating a new annex repo at /Users/wittkuhn/Desktop/cartoon-collection
[INFO] Scanning for unlocked files (this may take some time)
create(ok): /Users/wittkuhn/Desktop/cartoon-collection (dataset)
```

CREATE A DATASET

Create a new DataLad dataset for your cartoon collection:

```
$ datalad create cartoon-collection
[INFO] Creating a new annex repo at /Users/wittkuhn/Desktop/cartoon-collection
[INFO] Scanning for unlocked files (this may take some time)
create(ok): /Users/wittkuhn/Desktop/cartoon-collection (dataset)
```

```
$ cd cartoon-collection
$ wget https://imgs.xkcd.com/comics/compiling.png
--2020-11-11 14:04:27--  https://imgs.xkcd.com/comics/compiling.png
Resolving imgs.xkcd.com (imgs.xkcd.com)... 2a04:4e42:3::67, 151.101.12.67
Connecting to imgs.xkcd.com (imgs.xkcd.com)|2a04:4e42:3::67|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 28315 (28K) [image/png]
Saving to: 'compiling.png'

compiling.png
100%[=====] 27.65K --.-KB/s
in 0.02s
2020-11-11 14:04:28 (1.50 MB/s) - 'compiling.png' saved [28315/28315]
```

CREATE A DATASET

Create a new DataLad dataset for your cartoon collection:

```
$ datalad create cartoon-collection
[INFO] Creating a new annex repo at /Users/wittkuhn/Desktop/cartoon-collection
[INFO] Scanning for unlocked files (this may take some time)
create(ok): /Users/wittkuhn/Desktop/cartoon-collection (dataset)
```

```
$ cd cartoon-collection
$ wget https://imgs.xkcd.com/comics/compiling.png
--2020-11-11 14:04:27--  https://imgs.xkcd.com/comics/compiling.png
Resolving imgs.xkcd.com (imgs.xkcd.com)... 2a04:4e42:3::67, 151.101.12.67
Connecting to imgs.xkcd.com (imgs.xkcd.com)|2a04:4e42:3::67|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 28315 (28K) [image/png]
Saving to: 'compiling.png'

compiling.png
100%[=====] 27.65K --.-KB/s
in 0.02s
2020-11-11 14:04:28 (1.50 MB/s) - 'compiling.png' saved [28315/28315]
```

```
$ datalad save -m "add funny xkcd comic"
add(ok): compiling.png (file)
save(ok): . (dataset)
action summary:
  add (ok: 1)
  save (ok: 1)
```

CREATE A DATASET

Create a new DataLad dataset for your cartoon collection:

```
$ datalad create cartoon-collection
[INFO] Creating a new annex repo at /Users/wittkuhn/Desktop/cartoon-collection
[INFO] Scanning for unlocked files (this may take some time)
create(ok): /Users/wittkuhn/Desktop/cartoon-collection (dataset)
```

```
$ cd cartoon-collection
$ wget https://imgs.xkcd.com/comics/compiling.png
--2020-11-11 14:04:27--  https://imgs.xkcd.com/comics/compiling.png
Resolving imgs.xkcd.com (imgs.xkcd.com)... 2a04:4e42:3::67, 151.101.12.67
Connecting to imgs.xkcd.com (imgs.xkcd.com)|2a04:4e42:3::67|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 28315 (28K) [image/png]
Saving to: 'compiling.png'

compiling.png
100%[=====] 27.65K --.-KB/s
in 0.02s
2020-11-11 14:04:28 (1.50 MB/s) - 'compiling.png' saved [28315/28315]
```

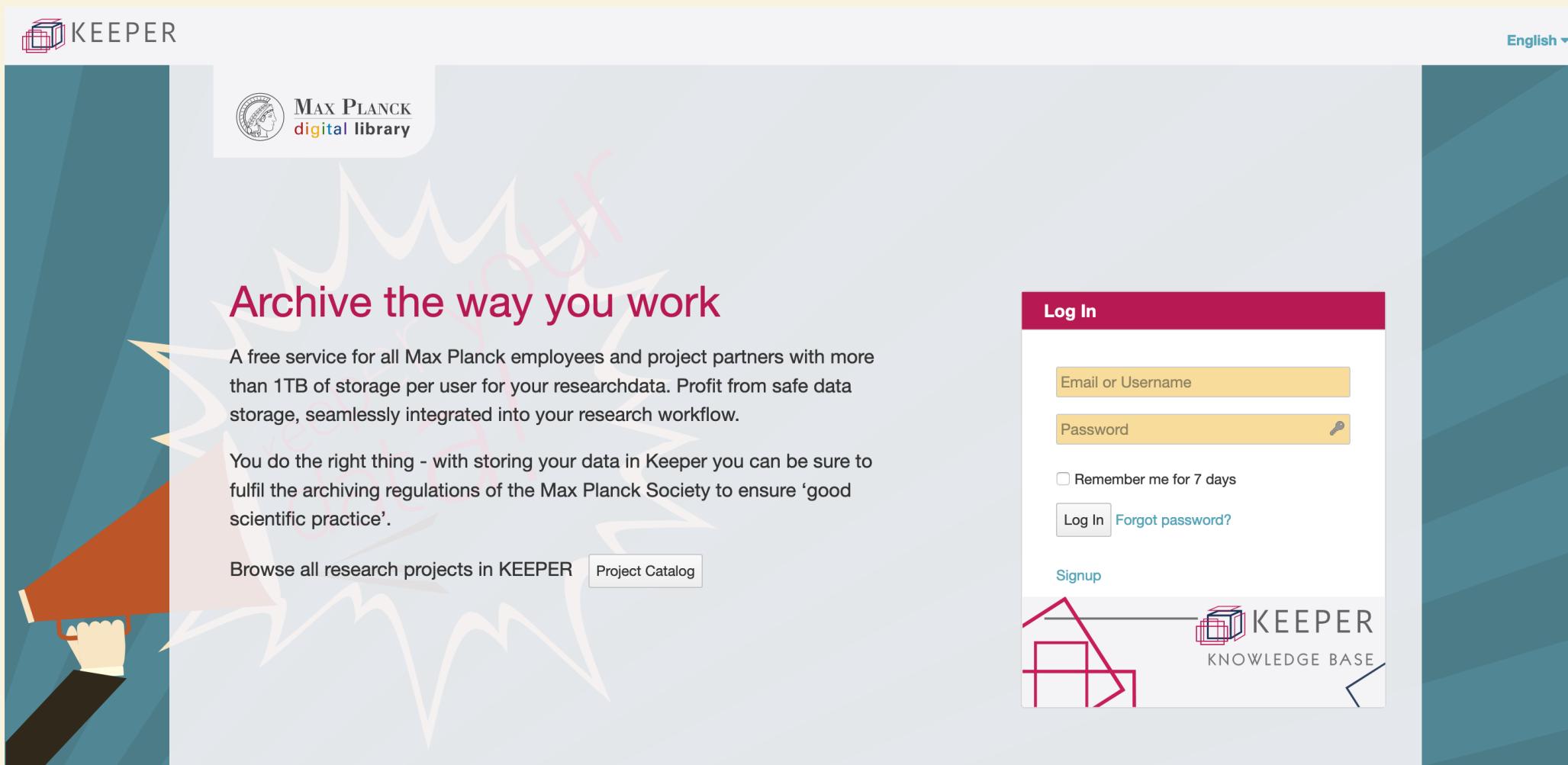
```
$ datalad save -m "add funny xkcd comic"
add(ok): compiling.png (file)
save(ok): . (dataset)
action summary:
  add (ok: 1)
  save (ok: 1)
```

"How can I share my cartoon-collection with my friend Adina?" 🤔

KEEPER (SEAFILE): A FLEXIBLE DATALAD SIBLING

<https://keeper.mpdl.mpg.de/>

Keeper offers **1TB** flexible storage to all Max Planck employees *



* Even if you end up not using DataLad, you might want to check Keeper out and get those 1TB of storage space!

* Data are stored on servers of the Max Planck Society - your data protection officer will be pleased!

OK, SHOW ME HOW IT'S DONE! 💪

(see the DataLad Handbook chapter "Beyond shared infrastructure"):

1. Sign-up to Keeper / Seafile
2. Create a GitHub / GitLab repository to host meta data of your dataset and all content that is not stored in git-annex
3. Install `rclone` and the `git-annex-remote-rclone` wrapper to make `rclone` usable with `git-annex`

Disclaimer: Yes, things can go wrong here, but given that installation is platform-dependent we don't get into details

CREATING A REMOTE REPOSITORY

On GitLab (here, of the MPIB) ...

New project

A project is where you house your files (repository), plan your work (issues), and publish your documentation (wiki), [among other things](#).

All features are enabled for blank projects, from templates, or when importing, but you can disable them afterward in the project settings.

Information about additional Pages templates and how to install them can be found in our [Pages getting started guide](#).

Tip: You can also create a project from the command line. [Show command](#)

Blank project	Create from template	Import project
Project name comic-collection		
Project URL <input type="text" value="https://git.mpib-berlin.mpg.de/"/> wittkuhn	Project slug comic-collection	
Want to house several dependent projects under the same namespace? Create a group .		
Project description (optional) Description format		
Visibility Level ? <input type="radio"/> Private Project access must be granted explicitly to each user. If this project is part of a group, access will be granted to members of the group. <input type="radio"/> Internal The project can be accessed by any logged in user. <input checked="" type="radio"/> Public The project can be accessed without any authentication. <input type="checkbox"/> Initialize repository with a README Allows you to immediately clone this project's repository. Skip this if you plan to push up an existing repository.		
Create project		Cancel

CREATING A REMOTE REPOSITORY

... or on GitHub:

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Repository template
Start your repository with a template repository's contents.

No template ▾

Owner * **Repository name ***

 Innrtwttkhn / comic-collection ✓

Great repository names are short and memorable. Need inspiration? How about [studious-chainsaw](#)?

Description (optional)

 **Public**
Anyone on the internet can see this repository. You choose who can commit.

 **Private**
You choose who can see and commit to this repository.

Initialize this repository with:
Skip this step if you're importing an existing repository.

Add a README file
This is where you can write a long description for your project. [Learn more](#).

Add .gitignore
Choose which files not to track from a list of templates. [Learn more](#).

Choose a license
A license tells others what they can and can't do with your code. [Learn more](#).

Create repository

CONFIGURE AN RCLONE REMOTE FOR KEEPER / SEAFILE

CONFIGURE AN RCLONE REMOTE FOR KEEPER / SEAFILE

Run rclone config and create a new remote

```
$ rclone config
e) Edit existing remote
n) New remote
d) Delete remote
r) Rename remote
c) Copy remote
s) Set configuration password
q) Quit config
e/n/d/r/c/s/q> n # enter n and press enter
```

CONFIGURE AN RCLONE REMOTE FOR KEEPER / SEAFILE

Run rclone config and create a new remote

```
$ rclone config
e) Edit existing remote
n) New remote
d) Delete remote
r) Rename remote
c) Copy remote
s) Set configuration password
q) Quit config
e/n/d/r/c/s/q> n # enter n and press enter
```

Enter the name of the rclone configuration
("seafile" or "keeper" could be a good idea)

```
name> seafile # enter name and press enter
```

Select "seafile" as the type of storage to configure

```
1 Type of storage to configure.
2 Enter a string value. Press Enter for the default ("").
3 Choose a number from below, or type in your own value
4 1 / 1Fichier
5 \ "fichier"
6 2 / Alias for an existing remote
7 \ "alias"
8 3 / Amazon Drive
9 \ "amazon cloud drive"
10 4 / Amazon S3 Compliant Storage Provider (AWS, Alibaba, Ceph, Digital Ocean, Dreamhost, IBM COS, Minio
11 \ "s3"
12 5 / Backblaze B2
13 \ "b2"
14 6 / Box
15 \ "box"
16 7 / Cache a remote
17 \ "cache"
18 8 / Citrix Sharefile
19 \ "sharefile"
20 9 / Dropbox
21 \ "dropbox"
22 38 / seafile
23 \ "seafile"
24 Storage> seafile # enter 'seafile' or '38' and press enter
25 ** See help for seafile backend at: https://rclone.org/seafile/ **
```

Notice the many storage types you can configure with rclone! 🐱
You can also configure Dropbox, Google Drive, and more as DataLad remotes.
Quite a few options had to be removed because the list is so long ...

Enter the URL of the Seafile host (here, Keeper)

```
URL of seafile host to connect to
Enter a string value. Press Enter for the default ("").
Choose a number from below, or type in your own value
 1 / Connect to cloud.seafile.com
\ "https://cloud.seafile.com/"
url> https://keeper.mpdl.mpg.de/
```

Enter the URL of the Seafile host (here, Keeper)

```
URL of seafile host to connect to
Enter a string value. Press Enter for the default ("").
Choose a number from below, or type in your own value
 1 / Connect to cloud.seafile.com
\ "https://cloud.seafile.com/"
url> https://keeper.mpdl.mpg.de/
```

Enter the email address of your Seafile / Keeper account

```
User name (usually email address)
Enter a string value. Press Enter for the default ("").
user> wittkuhn@mpib-berlin.mpg.de # enter YOUR email address!
```

Enter the URL of the Seafile host (here, Keeper)

```
URL of seafile host to connect to
Enter a string value. Press Enter for the default ("").
Choose a number from below, or type in your own value
 1 / Connect to cloud.seafile.com
\ "https://cloud.seafile.com/"
url> https://keeper.mpdl.mpg.de/
```

Enter the email address of your Seafile / Keeper account

```
User name (usually email address)
Enter a string value. Press Enter for the default ("").
user> wittkuhn@mpib-berlin.mpg.de # enter YOUR email address!
```

Enter your Keeper password

```
Password
y) Yes type in my own password
g) Generate random password
n) No leave this optional password blank (default)
y/g/n> y # type 'y' and press enter
Enter the password:
password: # enter password here
Confirm the password:
password: # enter password again
```


Enable two-factor authentication if used for Keeper

```
Two-factor authentication ('true' if the account has 2FA enabled)
Enter a boolean value (true or false). Press Enter for the default ("false").
2fa> # just press enter if you don't use 2FA
```

Enable two-factor authentication if used for Keeper

```
Two-factor authentication ('true' if the account has 2FA enabled)
Enter a boolean value (true or false). Press Enter for the default ("false").
2fa> # just press enter if you don't use 2FA
```

Don't enter a Keeper library name! Here, we will configure one Seafile rclone remote that
you can re-use across your projects 😎

```
Name of the library. Leave blank to access all non-encrypted libraries.
Enter a string value. Press Enter for the default ("").
library> # leave blank and press enter
```

Enable two-factor authentication if used for Keeper

```
Two-factor authentication ('true' if the account has 2FA enabled)
Enter a boolean value (true or false). Press Enter for the default ("false").
2fa> # just press enter if you don't use 2FA
```

Don't enter a Keeper library name! Here, we will configure one Seafile rclone remote that you can re-use across your projects 😎

```
Name of the library. Leave blank to access all non-encrypted libraries.
Enter a string value. Press Enter for the default ("").
library> # leave blank and press enter
```

Enable library encryption if desired

```
Library password (for encrypted libraries only). Leave blank if you pass it through the command line.
y) Yes type in my own password
g) Generate random password
n) No leave this optional password blank (default)
y/g/n> # leave blank (or type 'n') and press enter to continue without encryption
```


Edit advanced configuration

```
Edit advanced config? (y/n)
  y) Yes
  n) No (default)
y/n> y # type 'y' and press enter
```

Edit advanced configuration

```
Edit advanced config? (y/n)
  y) Yes
  n) No (default)
y/n> y # type 'y' and press enter
```

Allow rclone to create non-existing libraries on Keeper

```
Should rclone create a library if it doesn't exist?
Enter a boolean value (true or false). Press Enter for the default ("false").
create_library> true # type 'true' and press enter
```

Edit advanced configuration

```
Edit advanced config? (y/n)
  y) Yes
  n) No (default)
y/n> y # type 'y' and press enter
```

Allow rclone to create non-existing libraries on Keeper

```
Should rclone create a library if it doesn't exist?
Enter a boolean value (true or false). Press Enter for the default ("false").
create_library> true # type 'true' and press enter
```

Keep the encoding at the default

```
This sets the encoding for the backend.
See: the [encoding section in the overview] (/overview/#encoding) for more info.
Enter a encoder.MultiEncoder value. Press Enter for the default ("Slash,DoubleQuote,BackSlash,Ctl,InvalidU
encoding> # press enter to keep the default
```

You successfully configured your Seafile / Keeper rclone remote! 🙌🍾🏆

```
Remote config
Two-factor authentication is not enabled on this account.
-----
[seafile]
type = seafile
url = https://keeper.mpdl.mpg.de/
user = wittkuhn@mpib-berlin.mpg.de
pass = *** ENCRYPTED ***
create_library = true
-----
y) Yes this is OK (default)
e) Edit this remote
d) Delete this remote
y/e/d> y # leave blank (or enter 'y') and press enter to confirm
```

List of your configured rclone remotes:

```
Current remotes:
Name          Type
====          ====
seafile       seafile
```

You successfully configured your Seafile / Keeper rclone remote! 🙌🍾🏆

```
Remote config
Two-factor authentication is not enabled on this account.
-----
[seafile]
type = seafile
url = https://keeper.mpdl.mpg.de/
user = wittkuhn@mpib-berlin.mpg.de
pass = *** ENCRYPTED ***
create_library = true
-----
y) Yes this is OK (default)
e) Edit this remote
d) Delete this remote
y/e/d> y # leave blank (or enter 'y') and press enter to confirm
```

List of your configured rclone remotes:

```
Current remotes:
Name          Type
====          ====
seafile       seafile
```

Impatient? Configure rclone in one line of code:

```
rclone config create seafile seafile url https://keeper.mpdl.mpg.de/ \
user wittkuhn@mpib-berlin.mpg.de pass Maga2020! create-library true
```

INITIALIZE THE RCLONE SPECIAL REMOTE

In our dataset, we run the `git annex initremote` command:

```
$ git annex initremote seafile \
type=external externaltype=rclone chunk=50MiB encryption=none \
target=seafile prefix=cartoon-collection
```

- it's helpful if the name of the remote and the target are the same
(we name both 'seafile' in this example)
- the prefix will be the name of the library on Keeper and should be the same as our dataset

```
initremote seafile ok
(recording state in git...)
lip-osx-003854:cartoon-collection wittkuhn$ datalad siblings
.: here(+) [git]
.: seafile(+) [rclone]
```

ADD THE GITLAB / GITHUB SIBLING

ADD THE GITLAB / GITHUB SIBLING

Let's add our GitLab / GitHub sibling and configure the Seafile / Keeper siblings as a publication dependency:

```
$ datalad siblings add --name origin \
--url git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git \
--publish-depends seafile
[INFO] Configure additional publication dependency on "seafile"
[INFO] Could not enable annex remote origin. This is expected if origin is a pure Git remote, or happens i
[WARNING] Could not detect whether origin carries an annex. If origin is a pure Git remote, this is expecte
.: origin(-) [git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git (git)]
```

ADD THE GITLAB / GITHUB SIBLING

Let's add our GitLab / GitHub sibling and configure the Seafile / Keeper siblings as a publication dependency:

```
$ datalad siblings add --name origin \
--url git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git \
--publish-depends seafile
[INFO] Configure additional publication dependency on "seafile"
[INFO] Could not enable annex remote origin. This is expected if origin is a pure Git remote, or happens i
[WARNING] Could not detect whether origin carries an annex. If origin is a pure Git remote, this is expecte
..: origin(-) [git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git (git)]
```

```
$ datalad siblings
..: here(+) [git]
..: seafile(+) [rclone]
..: origin(-) [git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git (git)]
```

ADD THE GITLAB / GITHUB SIBLING

Let's add our GitLab / GitHub sibling and configure the Seafile / Keeper siblings as a publication dependency:

```
$ datalad siblings add --name origin \
--url git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git \
--publish-depends seafile
[INFO] Configure additional publication dependency on "seafile"
[INFO] Could not enable annex remote origin. This is expected if origin is a pure Git remote, or happens i
[WARNING] Could not detect whether origin carries an annex. If origin is a pure Git remote, this is expecte
..: origin(-) [git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git (git)]
```

```
$ datalad siblings
..: here(+) [git]
..: seafile(+) [rclone]
..: origin(-) [git@git.mpib-berlin.mpg.de:wittkuhn/cartoon-collection.git (git)]
```

Quiz question: "Why did Lennart name the GitLab sibling 'origin'?" 🤔

THE MOMENT OF TRUTH: DATALAD PUSH

THE MOMENT OF TRUTH: DATALAD PUSH

Let's add our GitLab / GitHub sibling, configure the Seafile / Keeper siblings as a publication dependency and publish our dataset:

```
$ datalad push --to origin
copy(ok): compiling.png (file) [to seafile...]
publish(ok): . (dataset) [refs/heads/master->origin:refs/heads/master [new branch] ]
publish(ok): . (dataset) [refs/heads/git-annex->origin:refs/heads/git-annex [new branch] ]
```

THE MOMENT OF TRUTH: DATALAD PUSH

Let's add our GitLab / GitHub sibling, configure the Seafile / Keeper siblings as a publication dependency and publish our dataset:

```
$ datalad push --to origin
copy(ok): compiling.png (file) [to seafile...]
publish(ok): . (dataset) [refs/heads/master->origin:refs/heads/master [new branch] ]
publish(ok): . (dataset) [refs/heads/git-annex->origin:refs/heads/git-annex [new branch] ]
```

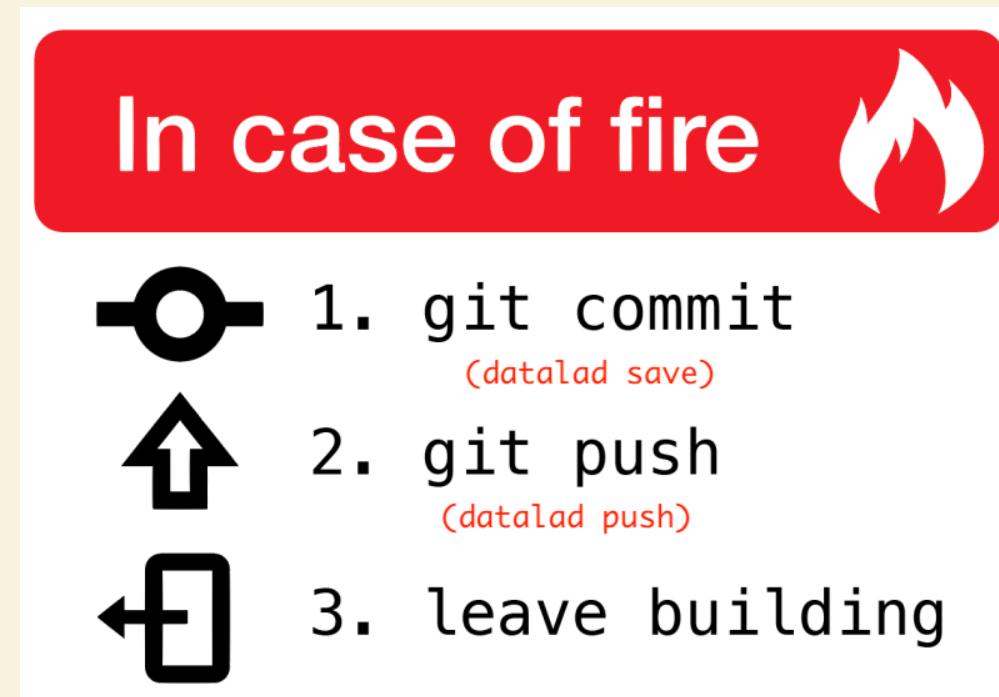
- DataLad successfully published our dataset to GitLab and pushed the annexed contents to Seafile / Keeper 🎉🥳🥳

THE MOMENT OF TRUTH: DATALAD PUSH

Let's add our GitLab / GitHub sibling, configure the Seafile / Keeper siblings as a publication dependency and publish our dataset:

```
$ datalad push --to origin
copy(ok): compiling.png (file) [to seafile...]
publish(ok): . (dataset) [refs/heads/master->origin:refs/heads/master [new branch]]
publish(ok): . (dataset) [refs/heads/git-annex->origin:refs/heads/git-annex [new branch]]
```

- DataLad successfully published our dataset to GitLab and pushed the annexed contents to Seafile / Keeper 🎉🥳🥳



THE MOMENT OF TRUTH: DATALAD PUSH

Let's add our GitLab / GitHub sibling, configure the Seafile / Keeper siblings as a publication dependency and publish our dataset:

```
$ datalad push --to origin  
copy(ok): compiling.png (file) [to seafile...]  
publish(ok): . (dataset) [refs/heads/master->origin:refs/heads/master [new branch]]  
publish(ok): . (dataset) [refs/heads/git-annex->origin:refs/heads/git-annex [new branch]]
```

- DataLad successfully published our dataset to GitLab and pushed the annexed contents to Seafile / Keeper 🎉🥳🥳

Image credit: <https://co.pinterest.com/pin/551128073121451139//imgcredit>



⬆️ That's another backup of your precious data right there! ⬆️

OUR COMIC COLLECTION IS PUBLISHED!

A screenshot of a web-based file library interface. At the top, there are buttons for "New Library" and "More", a search bar labeled "Search Files", and a user profile icon. Below this is a section titled "My Libraries" with a table. The table has columns for "Name", "Size", and "Last Update". There is one item listed: "cartoon-collection" (Size: 215.8 KB, Last Update: 12 minutes ago).

A screenshot of a library details page. The top navigation includes "Upload", "New", "Share", "Archive", and other icons. A search bar says "Search files in this library". Below is a table of files in the "cartoon-collection" library. The files listed are "510" (Size: 14 minutes ago), "archive-metadata.md" (Size: a year ago), and "Cared-Data-Certificate-HowTo.pdf" (Size: 4 years ago).

A screenshot of a Git annex project details page for "cartoon-collection". The top shows a message: "You pushed to [git-annex](#) at [Lennart Wittkuhn / cartoon-collection](#) 16 minutes ago" and a "Create merge request" button. Below is a project summary: "cartoon-collection" (Project ID: 470), 2 Commits, 2 Branches, 0 Tags, 348 KB Files, 348 KB Storage. It shows a commit by Lennart Wittkuhn: "add funny xkcd comic" (9c0588c5). At the bottom, there are buttons for "Add README", "Add LICENSE", "Add CHANGELOG", "Add CONTRIBUTING", "Add Kubernetes cluster", and "Set up CI/CD". A table shows recent commits:

Name	Last commit	Last update
.datalad	[DATALAD] new dataset	1 day ago
.gitattributes	[DATALAD] new dataset	1 day ago
compiling.png	add funny xkcd comic	1 day ago

FOR THE PERSON WHOM WE SHARE THE DATASET WITH

FOR THE PERSON WHOM WE SHARE THE DATASET WITH

If the repository is publicly available, a datalad clone with the URL will install the dataset:

```
$ datalad clone https://git.mpib-berlin.mpg.de/wittkuhn/cartoon-collection.git
[INFO] Scanning for unlocked files (this may take some time)
[INFO] Unable to parse git config from https://git.mpib-berlin.mpg.de/wittkuhn/cartoon-collection.git/conf
[INFO] Remote origin not usable by git-annex; setting annex-ignore
[INFO] access to 1 dataset sibling seafile not auto-enabled, enable with:
| datalad siblings -d "/Users/wittkuhn/Desktop/cartoon-collection" enable -s seafile
install(ok): /Users/wittkuhn/Desktop/cartoon-collection (dataset)
```

FOR THE PERSON WHOM WE SHARE THE DATASET WITH

If the repository is publicly available, a datalad clone with the URL will install the dataset:

```
$ datalad clone https://git.mpib-berlin.mpg.de/wittkuhn/cartoon-collection.git
[INFO] Scanning for unlocked files (this may take some time)
[INFO] Unable to parse git config from https://git.mpib-berlin.mpg.de/wittkuhn/cartoon-collection.git/conf
[INFO] Remote origin not usable by git-annex; setting annex-ignore
[INFO] access to 1 dataset sibling seafile not auto-enabled, enable with:
| datalad siblings -d "/Users/wittkuhn/Desktop/cartoon-collection" enable -s seafile
install(ok): /Users/wittkuhn/Desktop/cartoon-collection (dataset)
```

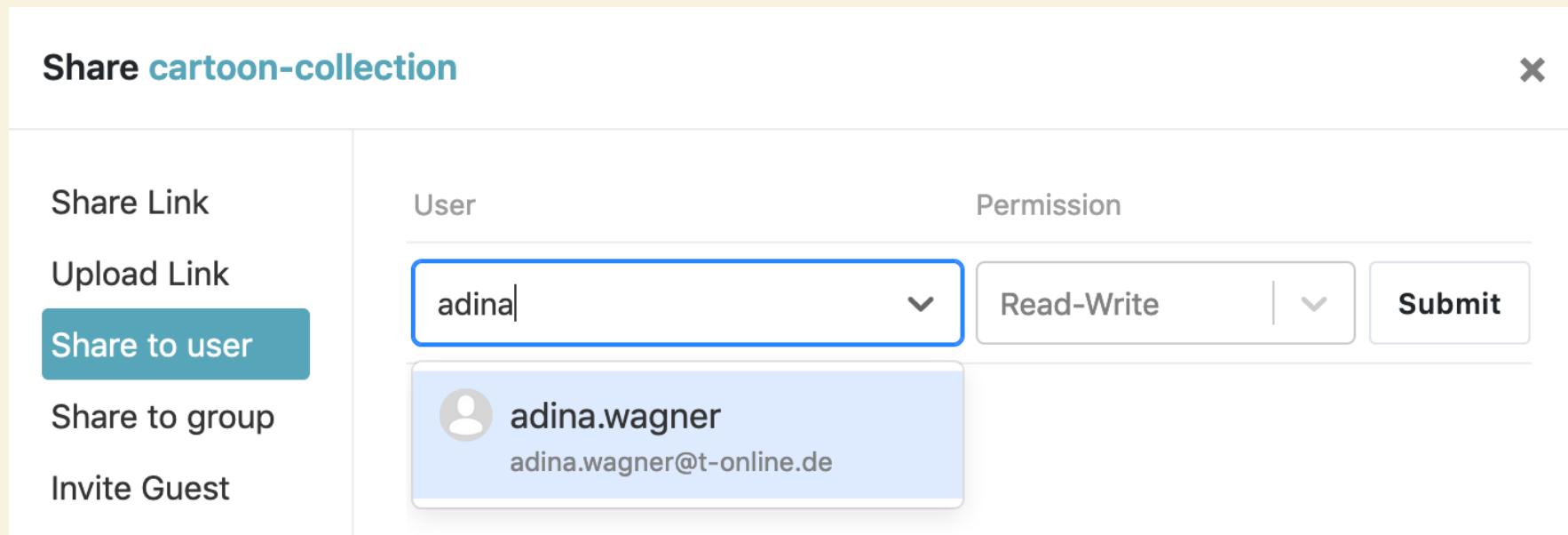
⚠ Pay attention to one crucial information in this output: ⚠

```
[INFO] access to 1 dataset sibling seafile not auto-enabled, enable with:
| datalad siblings -d "/Users/wittkuhn/Desktop/cartoon-collection" enable -s seafile
```

Someone who wants to access the data from Keeper / Seafile needs to enable the special rclone remote (using the same configuration)!

And you need to give them access to the Keeper library (see next slide)

Sharing the Keeper library with your collaborators ...



... will allow them to access the annexed contents of your dataset!
(after they successfully configured the seafolder rclone special remote)

```
$ datalad siblings -d "/Users/wittkuhn/Desktop/cartoon-collection" enable -s seafolder
.: seafolder(?) [git]
```

```
$ datalad siblings # just checking if all siblings are configured
.: here(+) [git]
.: seafolder(+) [rclone]
.: origin(-) [https://git.mpib-berlin.mpg.de/wittkuhn/cartoon-collection.git (git)]
```

```
$ datalad get compiling.png
get(ok): compiling.png (file) [from seafolder...]
```