

DataLad

decentralized data distribution for consumption and sharing of scientific datasets

Yaroslav O. Halchenko¹, Benjamin Poldrack², Michael Hanke²

<http://datalad.org>

1) Dept. of Psychological and Brain Sciences, Dartmouth College, NH, USA



2) Dept. of Experimental Psychology, University of Magdeburg, Germany



How could I get data?

* Data "Installation":

1. Go to website(s) and



2. Extract



* Data "upgrades"

1. Read news

2. goto 1

* Data "sharing" (I have heard about Upload button(s))

What about software?

* Linux distros

```
apt-get install anything
apt-get upgrade
```

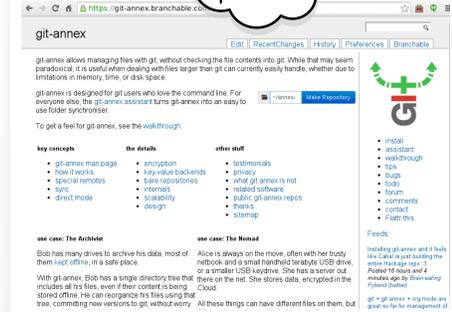
* Git-inspired folks

```
git clone git://anything
git pull
git push public
# send a pull request
```

Like that for data please?



Try git-annex

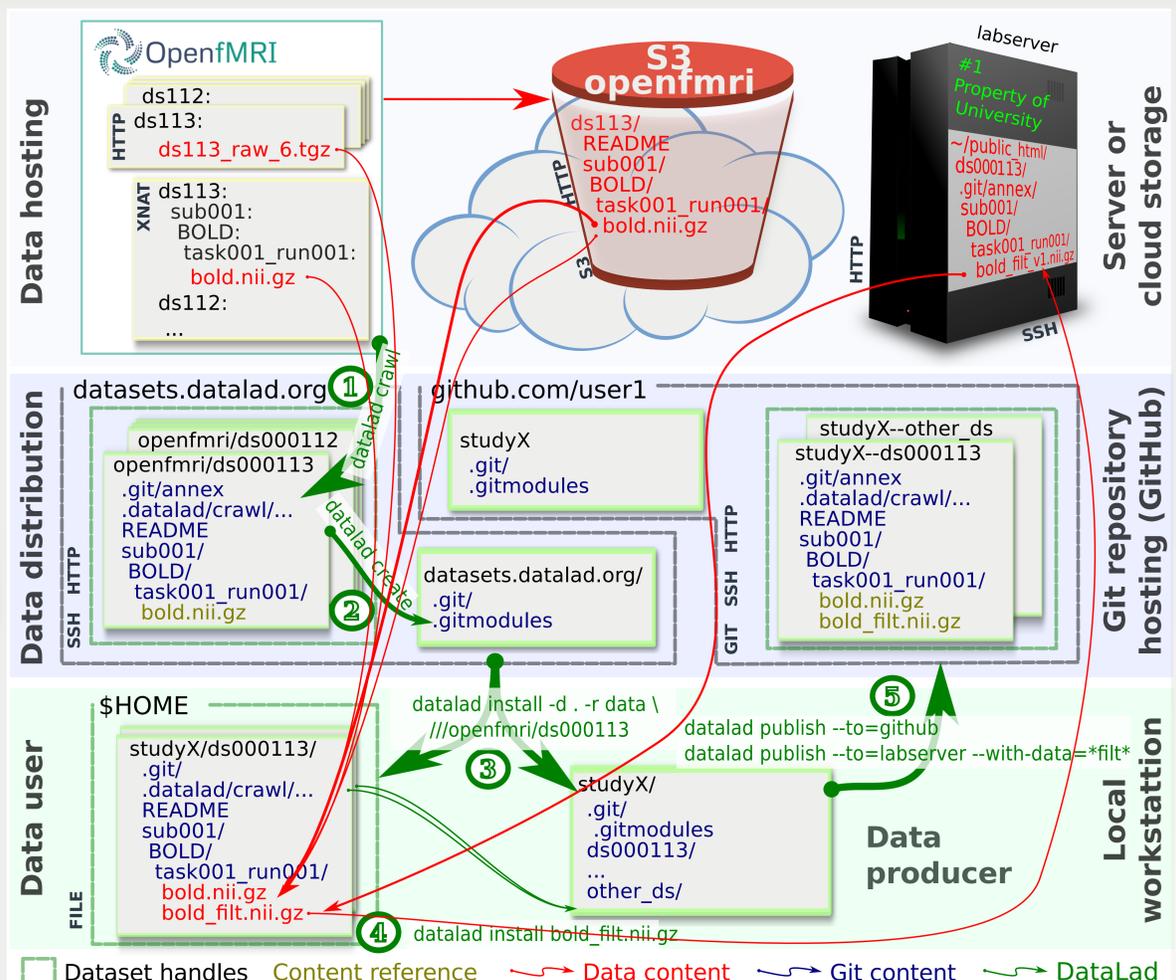


What DataLad?

1) crawl
1. Automate data access and possibly scraping of meta-information from data portals, such as
- OpenfMRI
- CRCNS.org
- HumanConnectome
- ... anything else worthwhile from XNAT, S3, COINS, ...
2. Store acquired data pointers under control Git/git-annex

2) install -d .
Collate available datasets into a super-dataset: git repository with submodules establishing versioned relationship between distribution and particular sub-datasets

3) install url
Use lean, locally installed dataset to obtain selected data files directly from original data providers (e.g. from S3 AWS storage, HTTP, XNAT, archives, etc.): same front interface, dedicated authentication and access mechanisms behind

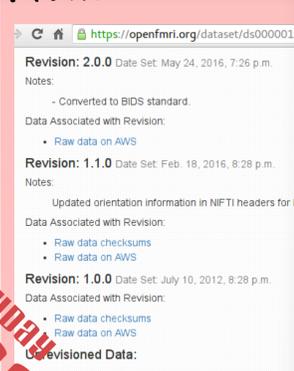


4) install file
Add new or derived data: it will be checksummed, versioned, and you could never loose track of it ever again. Since it is a git repository after all -- you can also add your scripts, documentation, etc. as well.

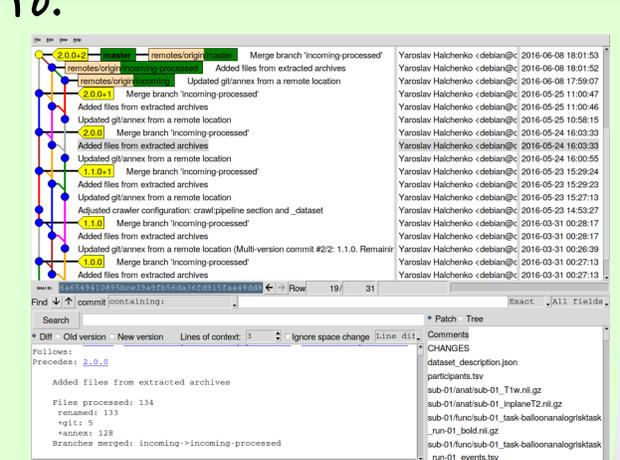
5) publish
Publish individual or collections of updated dataset handles to a service like GitHub while copying file content to a lab webserver (URL gets registered in the dataset annex) to make it publicly accessible.
Send a pull-request!

Bonus: Integrate
<http://Neuro.Debian.net>
apt-get install \ openfmri-ds11{2,3}

From:



To:



Work in Progress:

- Our growing "distribution" :
 - <http://datasets.datalad.org>
- Covered :
 - <http://openfmri.org> (S3)
 - <http://crcns.org>
 - <http://studyforrest.org>
- Coming :
 - <http://humanconnectome.org> (S3, XNAT)
 - <http://nitrc.org/ir> (INDI, FCP, etc.)
 - <http://coins.mrn.org> (COINS)
- Meta-data to facilitate search, custom views etc
- Integration NeuroDebian
 - apt-get install openfmri-ds000113
 - apt-get install openfmri

Acknowledgments

- * Linus Torvalds and Git team
- * Joey Hess (Git-annex)
- * NSF grant #1429999
- * <http://zeruch.deviantart.com> (Joey's picture)
- * OpenClipart.org contributors

EVERYBODY ONLY FOR YOU
FREE DEMO

