



EDUCATING AND TRAINING FOR A FAIR FUTURE

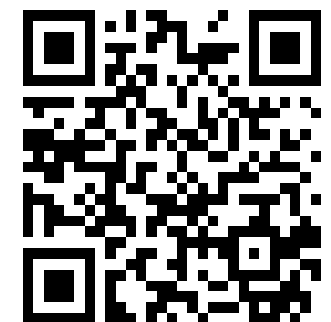
Adina Wagner

 @AdinaKrik



Psychoinformatics lab,
Institute of Neuroscience and Medicine, Brain & Behavior (INM-7)
Research Center Jülich
ReproNim/INCF fellow

Slides: [DOI 10.5281/zenodo.4541323](https://doi.org/10.5281/zenodo.4541323) (Scan the QR code)
Sources: github.com/datalad-handbook/datalad-course



TRAINING FOR A FAIR FUTURE IN NEUROSCIENCE

Neuroscience strives from interdisciplinarity
and collaboration



TRAINING FOR A FAIR FUTURE IN NEUROSCIENCE

Neuroscience strives from interdisciplinarity and collaboration



... but training groups with a diversity of skill sets is difficult



WHY ARE WE HERE? REPRONIM

- **ReproNim** is an initiative to improve the reproducibility and efficiency in neuroimaging
- It's goal is "to improve the reproducibility of neuroimaging science and extend the value of our national investment in neuroimaging research, while making the process easier and more efficient for investigators."
- **ReproNim** develops **free training materials**, supports **tool development**, and offers **training activities**. **ReproNim fellows** teach their peers in independent courses, workshops, or Hackathons about tools or methods that increase the reproducibility of their research.



The image is a composite of two parts. On the left is a terminal window with a dark background and light-colored text. It shows a sequence of Git commands and their outputs. The commands include cloning a repository, checking out a specific commit, and creating a new branch. The outputs show the progress of these operations, including file counts and commit hashes. On the right is a photograph of a woman with long brown hair and glasses, wearing a blue and white top. She is sitting at a desk in a room with wooden walls and shelves. There is a camera on a stand in front of her, and a microphone is visible. The room appears to be a home office or a studio.

RESEARCH DATA MANAGEMENT (RDM)



... TYPICAL DIFFICULTIES

Lack of formal training

Rarely included in primary/graduate/post-graduate studies. Rather: Learning "on the job" and inheriting project management from the previous person

"Science alone is hard already".

RDM comes on top of everything else, in usually tight, competitive graduate programs or jobs

Too much to know and too little guidance

Rather than a motivational problem, young researchers face the difficulty of finding out which tools exist and are helpful

Late RDM = little benefit

If RDM is only incentivized, required, or tackled as the very last step, researchers can not benefit (fully) from RDM

How can we overcome this?



PERSONAL INSIGHTS FROM RDM TRAINING



RDM overhaul in an institute

```
--- /data/BnB1/DATA/download_data/eNKI -----  
/..  
5.2 TiB [#####] /eNKI_unzipped  
3.3 TiB [#####] /eNKI_redownload  
3.2 TiB [#####] /eNKI_BIDSdownload  
724.2 GiB [#] /eNKI_20180806  
218.8 GiB [ ] /eNKI_aus_Raw_Data
```

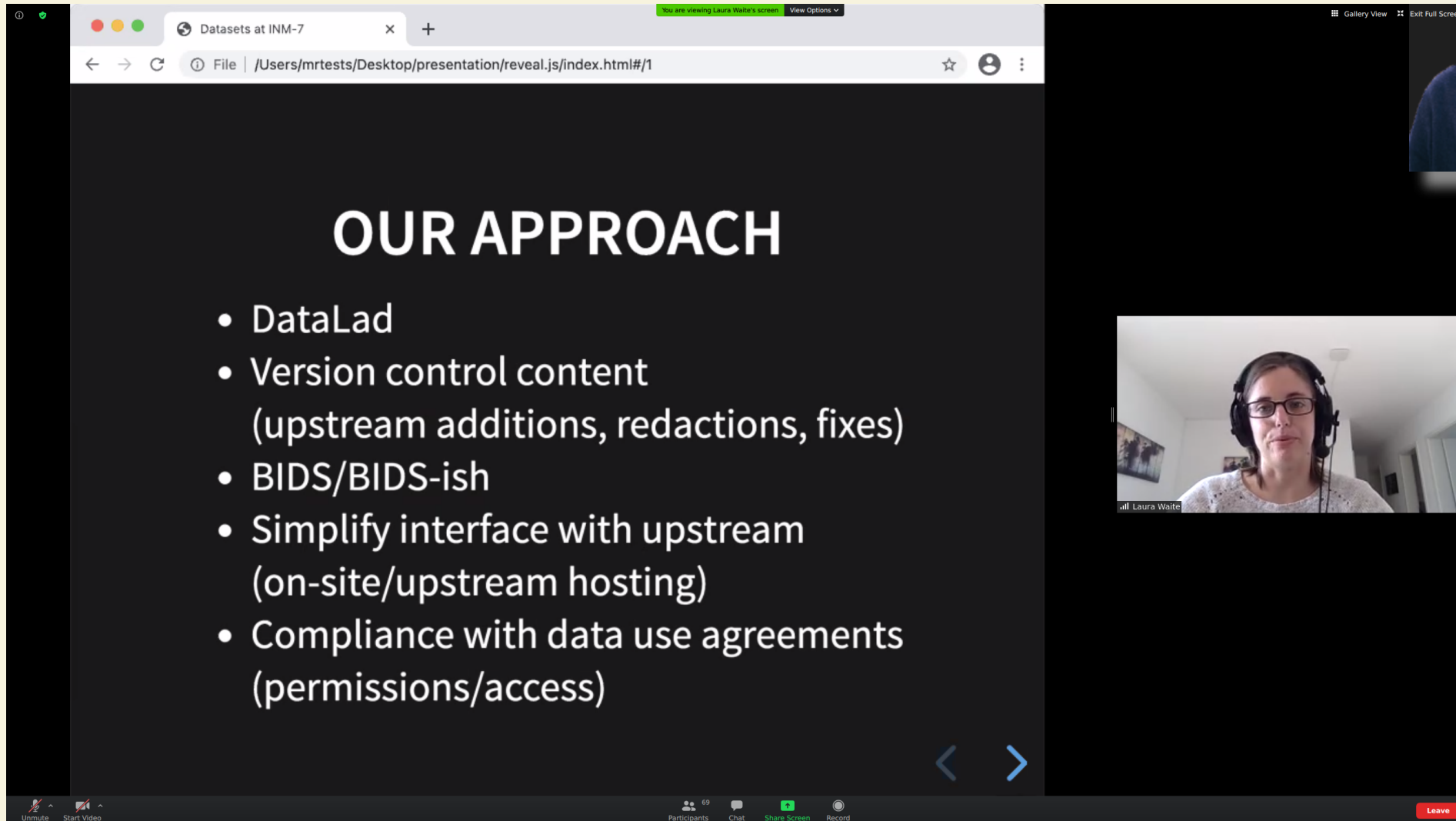


User-focused software documentation

```
In [1]: from datalad.api import create  
  
In [2]: create?  
Signature:  
create(  
    path=None,  
    initopts=None,  
    force=False,  
    description=None,  
    dataset=None,  
    no_annex=<class 'datalad.core.local.create._NoAnnexDefault'>,  
    annex=True,  
    fake_dates=False,  
    cfg_proc=None,  
)  
Docstring:  
Create a new dataset from scratch.  
  
This command initializes a new dataset at a given location, or the  
current directory. The new dataset can optionally be registered in an  
existing superdataset (the new dataset's path needs to be located  
within the superdataset for that, and the superdataset needs to be given
```

Aim: Researchers turn to RDM because it improves their science and work, not because they are forced to

RDM OVERHAUL IN AN INSTITUTE



The screenshot shows a Zoom meeting interface. The main window displays a presentation slide titled "OUR APPROACH" with a bulleted list of points. The browser address bar at the top shows the file path: `/Users/mrtests/Desktop/presentation/reveal.js/index.html#/1`. In the top right corner, there is a small portrait of a woman with short brown hair and glasses. In the bottom right corner, there is a video feed of a woman wearing a headset, identified as Laura Waite. The Zoom control bar at the bottom includes buttons for Unmute, Start Video, Participants (69), Chat, Share Screen, Record, and a red Leave button.

OUR APPROACH

- DataLad
- Version control content
(upstream additions, redactions, fixes)
- BIDS/BIDS-ish
- Simplify interface with upstream
(on-site/upstream hosting)
- Compliance with data use agreements
(permissions/access)

- System-wide DataLad deployment
- All large datasets are managed and accessed via DataLad (to avoid duplication, for provenance-tracked, reproducible transformations)
- Effort led by Laura Waite

TRAINING AN INSTITUTE



- In-person and virtual training
- On-demand user documentation at handbook.datalad.org

Search with Amazon.com

Search with Google

Mail - a.wagner@fz-juelich.de — [webmail.fz-juelich.de/owa/](mailto:a.wagner@fz-juelich.de)

The Handbook — The DataLad Handbook — [Switch to Tab](#)

Linguee | Deutsch-Englisch Wörterbuch (und weitere Spr — linguee.de

Google Docs — docs.google.com/document/u/0/?usp=direct_url

twitter — twitter.com

Google Scholar — scholar.google.de

The DataLad handbook will supply you with everything you need to get started and break new grounds with DataLad.

Get the gist!

Contributors

This guide is the result of the collaboration of [many people](#), and your contributions are welcome!

Useful Links

- [DataLad Website](#)
- [Developer Docs](#)
- [DataLad@GitHub](#)
- [Handbook@GitHub](#)
- [Frequently Asked Questions](#)
- [Handbook Index](#)
- [DataLad cheat sheet](#)

DataLad Cheat Sheet

Feedback

We highly appreciate your feedback on the [handbook](#) and on DataLad

Quick search

Welcome!

This handbook is a living resource about why and – more importantly – *how* to use DataLad. It aims to provide novices and advanced users of all backgrounds with both the basics of DataLad and start-to-end use cases of specific applications. If you want to get hands-on experience and learn DataLad, the *Basics* part of this book will teach you. If you want to know what is possible, the *use cases* will show you. And if you want to help others to get started with DataLad, the [companion repository](#) provides [free and open source teaching material](#) tailored to the handbook.

Before you read on, please note that the handbook is based on **DataLad version 0.12**, but the section [Installation and configuration](#) will set you up with what you need if you currently do not have DataLad 0.12 or higher installed.

If you're new here, please start the handbook [here](#). Alternatively, try to identify with one of several user-types in this [user specific guide to the handbook](#).

Important:

The handbook is currently in beta stage. If you would be willing to provide feedback on its contents, please [get in touch](#).

Introduction

What this is all about

- [Introduction](#)

v: latest

THE DATALAD HANDBOOK

Its structure reflects different needs of different stakeholders in science:



Intro-
duction

Starting must be easy

- High-level function/command overviews, Installation, Configuration, Cheatsheet



Basics

Tutorials for everyone

- Narrative-based code-along course
- Independent on background/skill level, suitable for data management novices



Advanced

Tutorials for experts

- Complex functionality and workflows
- Examples: Computationally reproducible analysis on big data



Use
cases

Overviews for decision makers

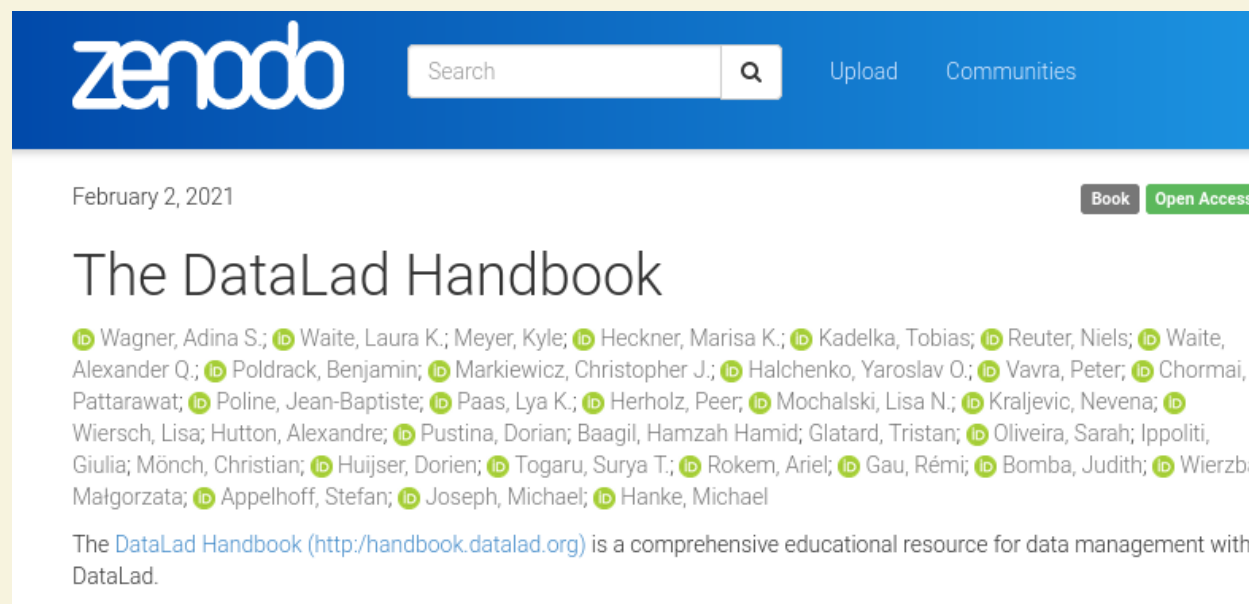
- Step-by-step solutions to common data management problems, like how to make a reproducible paper

TAKE HOME MESSAGES

- Complex software tools need accessible documentation
- Way to adoption of tools or principles via requirements/force/external incentives, or through immediate, personal benefits of good RDM (the latter keeps trainers sane)

A few Handbook metrics:

- Three Handbook releases
- 36 coauthors



- 500 pages of content
- dozens of openly shared slides, code lists, and other training materials
- about 250 unique views per day

