Shareable Science:

Analysis, Archiving, and Authoring for Collaboration

Elizabeth DuPre
@emdupre_
7 August, 2019

Science today

Olah & Carter (2017). Research Debt. *Distill*.

An outline for this morning

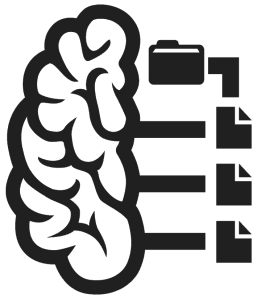
- Community-based project management
- A practical introduction to pyBIDS
- Authoring for collaboration
- A practical introduction to Jupyter Notebooks and JupyterLab

An outline for this morning

- Community-based project management
- A practical introduction to pyBIDS
- Authoring for collaboration
- A practical introduction to Jupyter Notebooks and JupyterLab

BIDS & the TIER Protocol

Community-based standards for project management



BIDS

BRAIN IMAGING DATA STRUCTURE

Brain Imaging Data Structure

Teaching Integrity in Empirical Research

Community-based project management standards

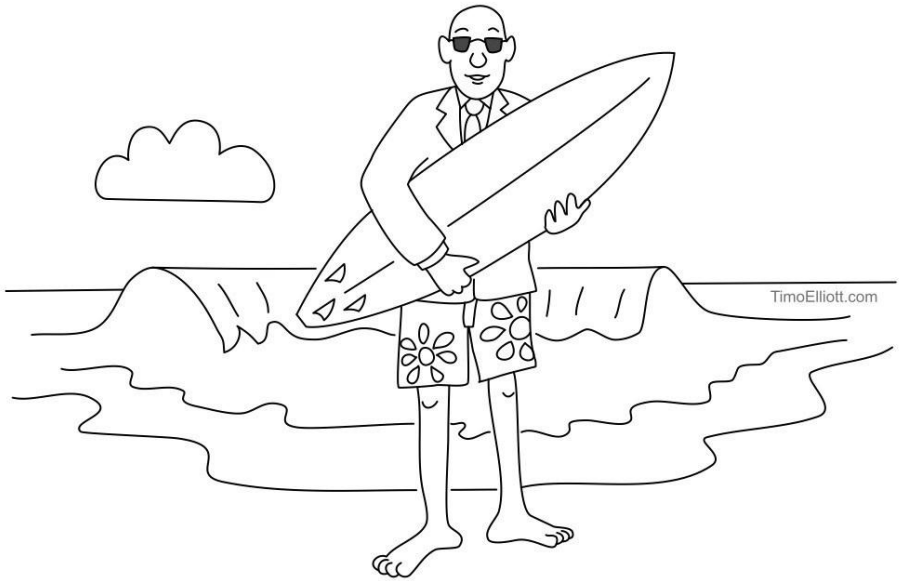
Community-based project management standards

Community-based project management standards

Meet Professor Smith

With thanks to *Chris Gorgolewski*

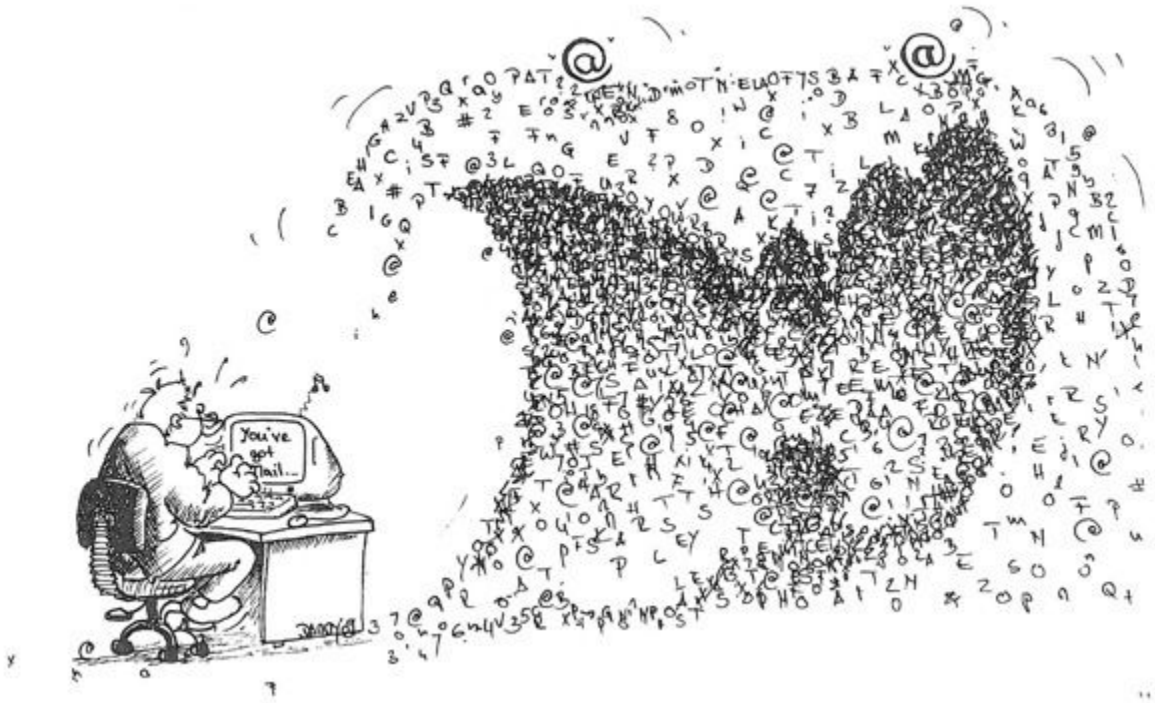
Meet Mike



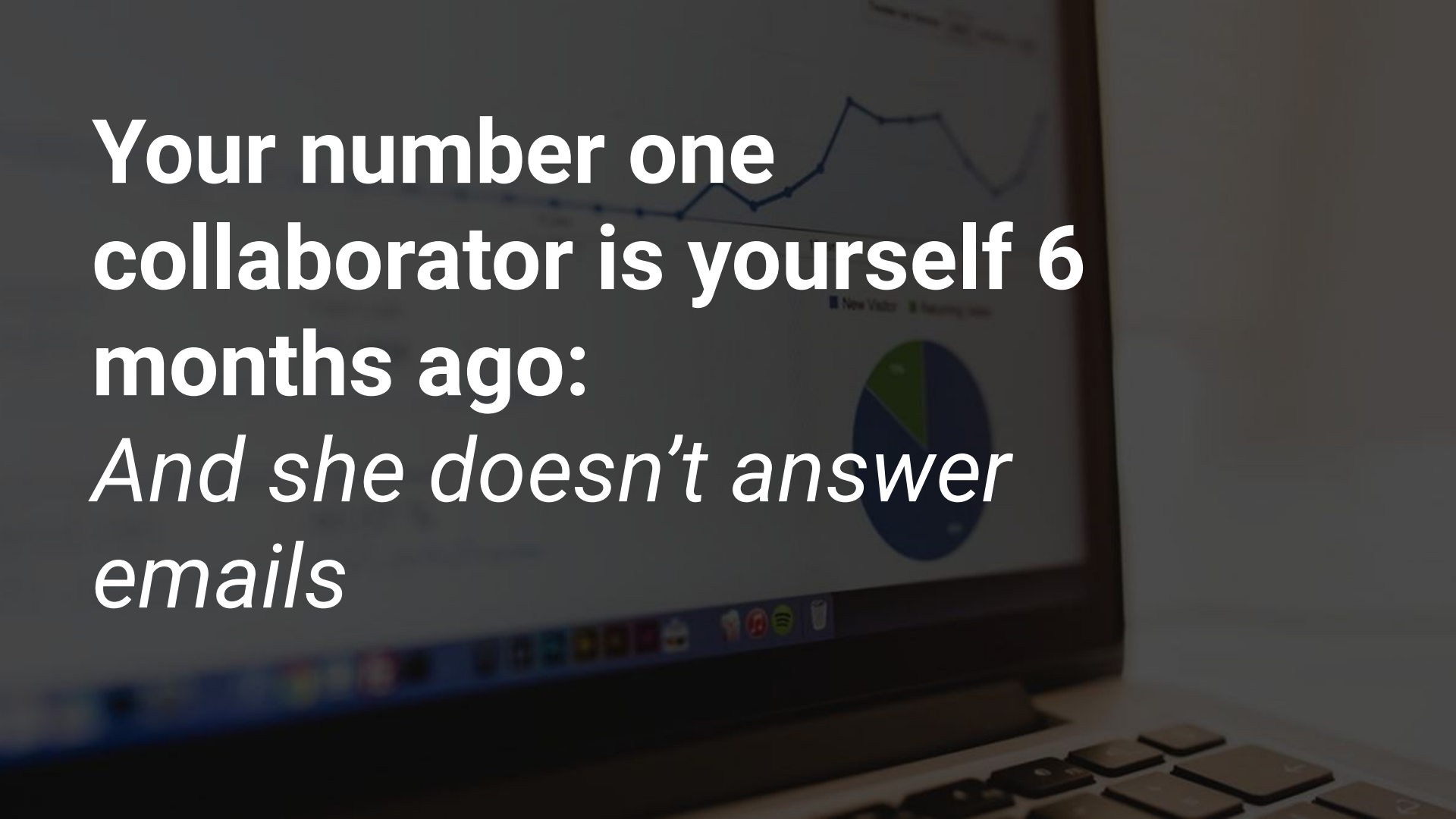
A Business Analyst that lives in California.

With thanks to *Chris Gorgolewski*

Meet Mike's project folders



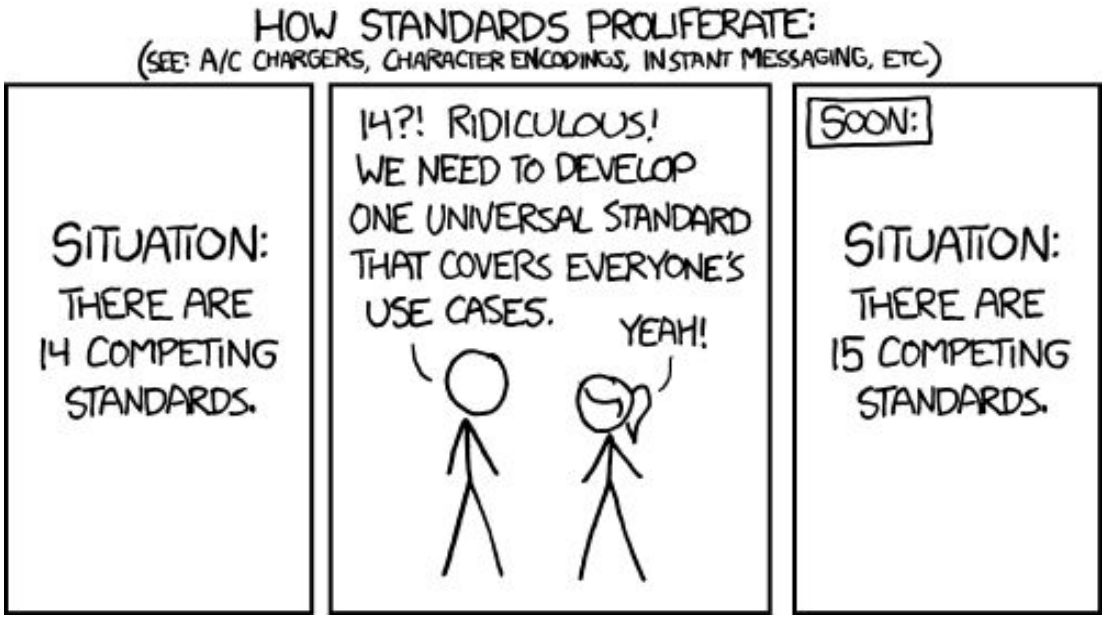
With thanks to *Chris Gorgolewski*



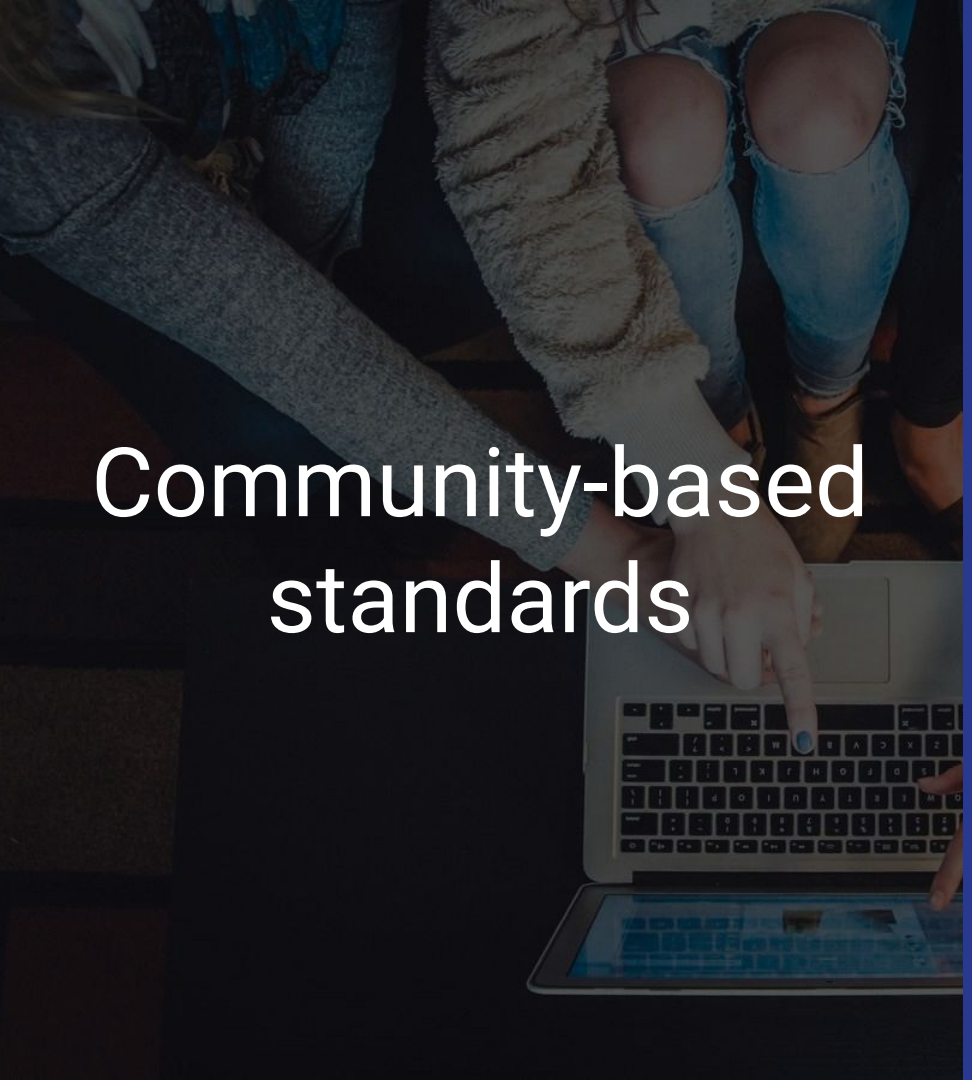
Your number one collaborator is yourself 6 months ago:
And she doesn’t answer emails

Community-based project management standards

Houston, we have a problem



XKCD



Community-based standards

- Developed **openly**
- Strive for **consensus** in decision-making
- Designed to **empower** and equip community members

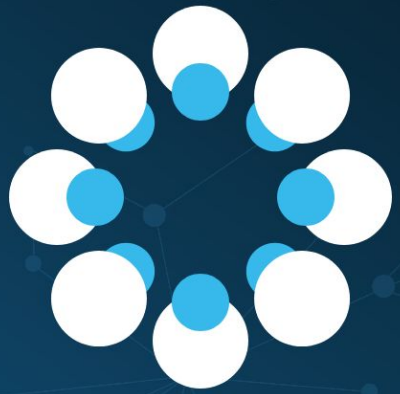
---

https://mozilla.github.io/open-leadership-training-series

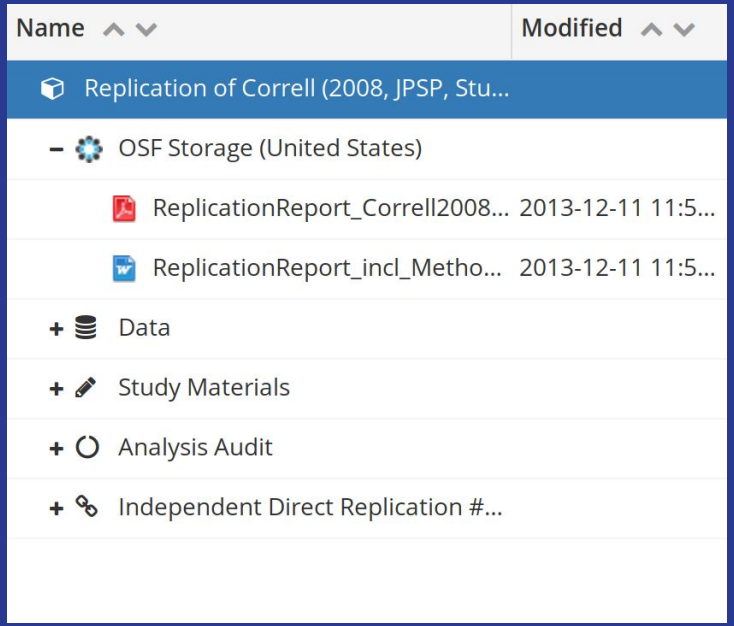
A (quick) case study

Open Science Framework

A scholarly commons to connect the entire research cycle



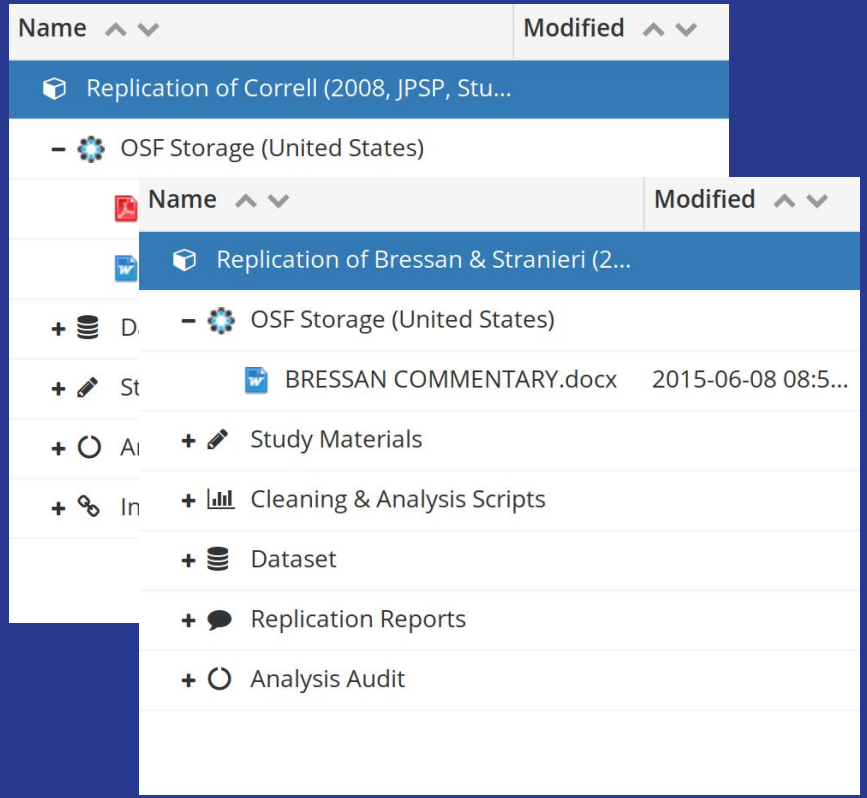
The Reproducibility Project



https://osf.io/fejxb/

The Reproducibility Project

… still did not have fully standardized project organization !



https://osf.io/fejxb/

https://osf.io/blcj6/

Teaching Integrity in Empirical Research

Project TIER

Raw Data

A copy of **every original data file** from which you extract any of data used in your study.

Your original data files **serve as a record** of the data you began the project with.

Analysis Code + Analysis Data

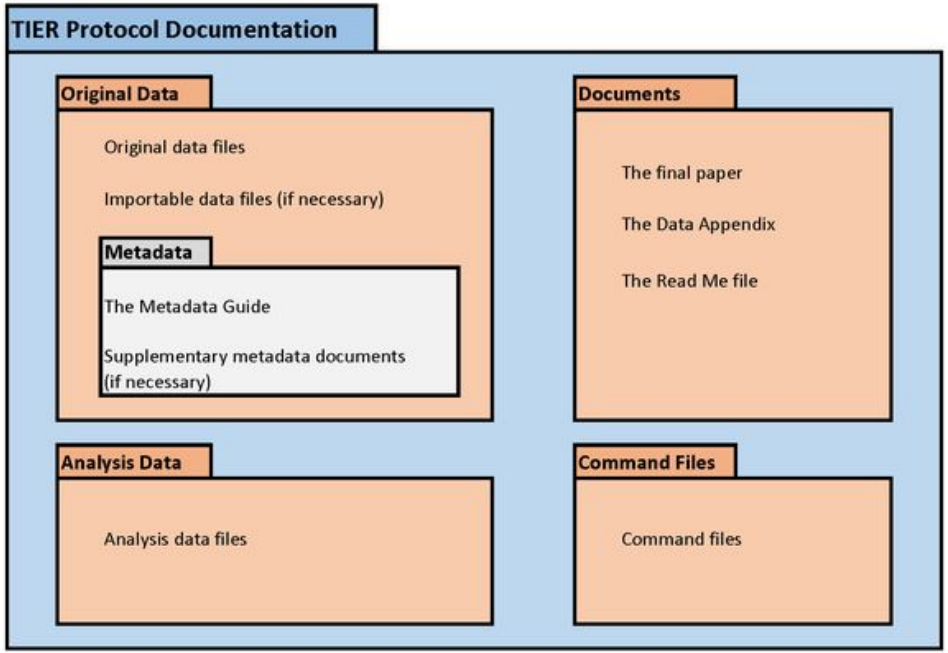
One or more files containing code used for the study... should **execute all the data processing and analysis necessary to replicate the study** and reproduce the reported results

Documents

Documentation to understand the study

- A copy of your final paper
- Your Data Appendix
- Your README file

https://www.projecttier.org/



The TIER Specification

TIER Protocol 3.0: Template

Public 92 ...

Contributors: Norm Medeiros, Richard Ball

Forked from osf.io/7g6cn on 2016-09-16 12:43 PM

Date created: 2016-09-08 11:17 PM | Last Updated: 2018-01-22 11:36 AM

Identifiers: DOI 10.17605/OSF.IO/YBZXE | ARK c7605/osf.io/ybzxe

Category: Project

Description: This project is designed to support Haverford College economics majors who produce empirical theses. The structure is based on the TIER Documentation Protocol. Additional information about Project TIER is available at http://projecttier.org

Wiki

This template was constructed for use by individuals who wish to follow the TIER Protocol for conducting and documenting an empirical research project.

Information about the TIER Protocol, and in particular about how to use this template, can be found on the Project TIER website.

Citation

Tags

Project TIER

https://osf.io/ybzxe/

Articles in this section

- Create DOIs
- Generate Citations
- Create Custom Citations
- Tag Your Project
- Tag Your Registration

Create DOIs

Sara Bowman
5 months ago

Follow

In addition to its own unique, persistent URLs, OSF offers DOIs for your public research. This help guide will walk you through how to generate DOIs for public projects and registrations.

DOIs point to the current version of the project or registration. OSF does not support DOI versioning at this time.

Articles in this section

- Create DOIs
- Generate Citations
- Create Custom Citations
- Tag Your Project
- Tag Your Registration

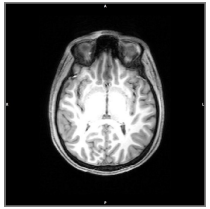
Create DOIs

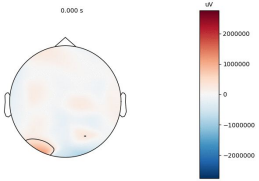
Sara Bowman
5 months ago

Follow

In addition to its own unique, persistent URLs, OSF offers DOIs for your public research. This help guide will walk you through how to generate DOIs for public projects and registrations.

DOIs point to the current version of the project or registration. OSF does not support DOI versioning at this time.





TIER Protocol Documentation

Original Data

Original data files

Importable data files (if necessary)

Metadata

The Metadata Guide

Supplementary metadata documents (if necessary)

Documents

The final paper

The Data Appendix

The Read Me file

Analysis Data

Analysis data files

Command Files

Command files

The TIER Specification

Brain Imaging Data Structure

Brain Imaging Data Structure v1.2.0

The BIDS Specification

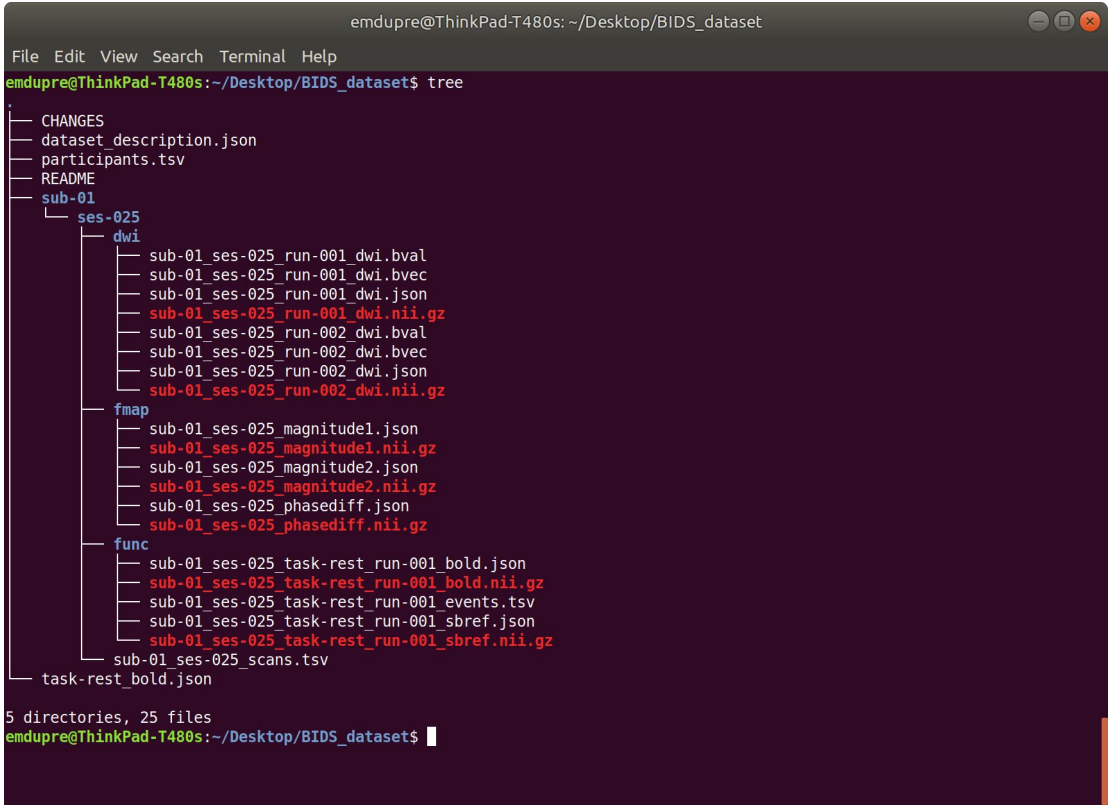
The BIDS Starter Kit

The Brain Imaging Data Structure

This site serves as an online resource to see the current state of the Brain Imaging Data Structure (BIDS) specification. It contains information about the core specification, as well as many modality-specific extensions.

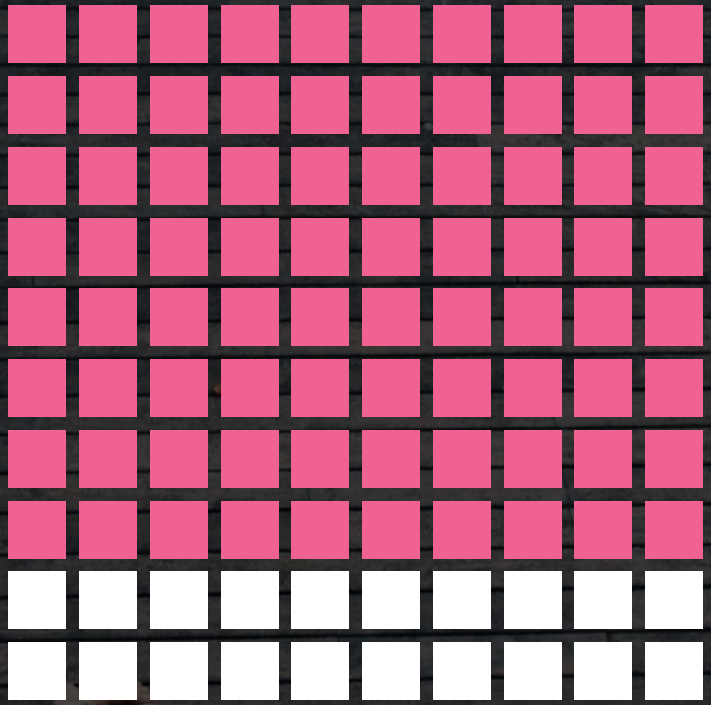
To get started, check out the introduction. If you'd like more information on how to adapt your own datasets to match the BIDS specification, we recommend exploring the bids-specification starter kit.

Example BIDS dataset layout



The 80 / 20 rule

Focus on the 80% use case to enhance clarity and collaboration



bids-standard / **bids-specification**

Unwatch releases 33 | Star 77 | Fork 47

Code | Issues 47 | Pull requests 8 | Projects 2 | Wiki | Insights

Filters | is:issue is:open | Labels 24 | Milestones 0 | New issue

47 Open ✓ 46 Closed | Author | Labels | Projects | Milestones | Assignee | Sort

- **[ENH] Update task and continuous recording metadata for EEG/MEG/iEEG** EEG MEG iEEG
  #209 opened 7 days ago by effigies
- **Proposal: BEP PRs should come from branches on the main repository** community — 4
  #203 opened 13 days ago by effigies
- **BEP Template** community — 1
  #201 opened 15 days ago by effigies
- **[FIX] Improvements to the "entity table"** good first issue help wanted
  #200 opened 18 days ago by choldgraf
- **[Infra] Add dates to releases in Spec** infrastructure
  #199 opened 19 days ago by sappelhoff
- **[ENH] Proposal for multidimensional array file format**
  #197 opened 20 days ago by tyarkoni
- **MRI: new metadata tag "ElementSpatialSize" (and may be more) to facilitate access to .nii.gz header information**
  #196 opened 21 days ago by yarikoptic

You?

We need your expertise !
Join the BIDS community on GitHub

Welcome to the BIDS Starter Kit

How to get started with the Brain Imaging Data Structure

A community-curated collection of tutorials, wikis, and templates to get you started with creating BIDS compliant datasets.

BIDS Homepage | Wiki | Standard | Tutorials | Chat | Forum

Click to view the intro video!

BIDS Validator

Select a BIDS dataset to validate

Browse… No directory selected.

Note: Selecting a dataset only performs validation. Files are never uploaded.

BIDS Validator

life

Summary

- 192 Files, 1GB
- 19 - Subjects
- 1 - Session

Available Tasks

- life

Available Modalities

- bold
- events
- T1w

Your dataset is not a valid BIDS dataset.

view 4 errors in 164 files

Error 1: [Code 55] JSON_SCHEMA_VALIDATION_ERROR 77 files

Click here for more information about this issue

Invalid JSON file. The file is not formatted according the schema.

NeuroStars search topics, posts, users, or categories

BIDS validator

+ New Topic

50+ results for **BIDS validator** Sort by Relevance

BIDS Validator error: TR mismatch between NIFTI header and JSON file, and **BIDS Validator** is somehow finding TR=0. Best solution?

fmri bids

Jun '18 - ...a from ~700 participants. Importantly, you get NIFTIs, but not the raw DICOMs . I had to rework their file directory structure a bit to make it fully **BIDS**-compatible, but this is what I did (notably, each subject has JSON files in the same directory as the corresponding NIFTI files, but I didn't see a n...

Inconsistent errors with **BIDS validator**

bids bids-validator

Sep '18 - Hi All, I am trying to get a dataset to pass **BIDS** validation, but am running into a couple of errors regarding the stimuli folder that I can't seem to figure out. I am specifically understand why I d...

Advanced Search

Posted by

Categorized

All categories

Tagged

All the above tags

Only return topics/posts...

Matching in title only

I liked

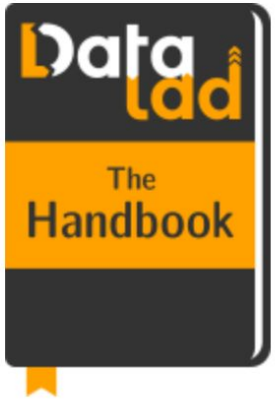
In my messages

I read

any

Where topics

NeuroStars.org



Star

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

The DataLad Handbook

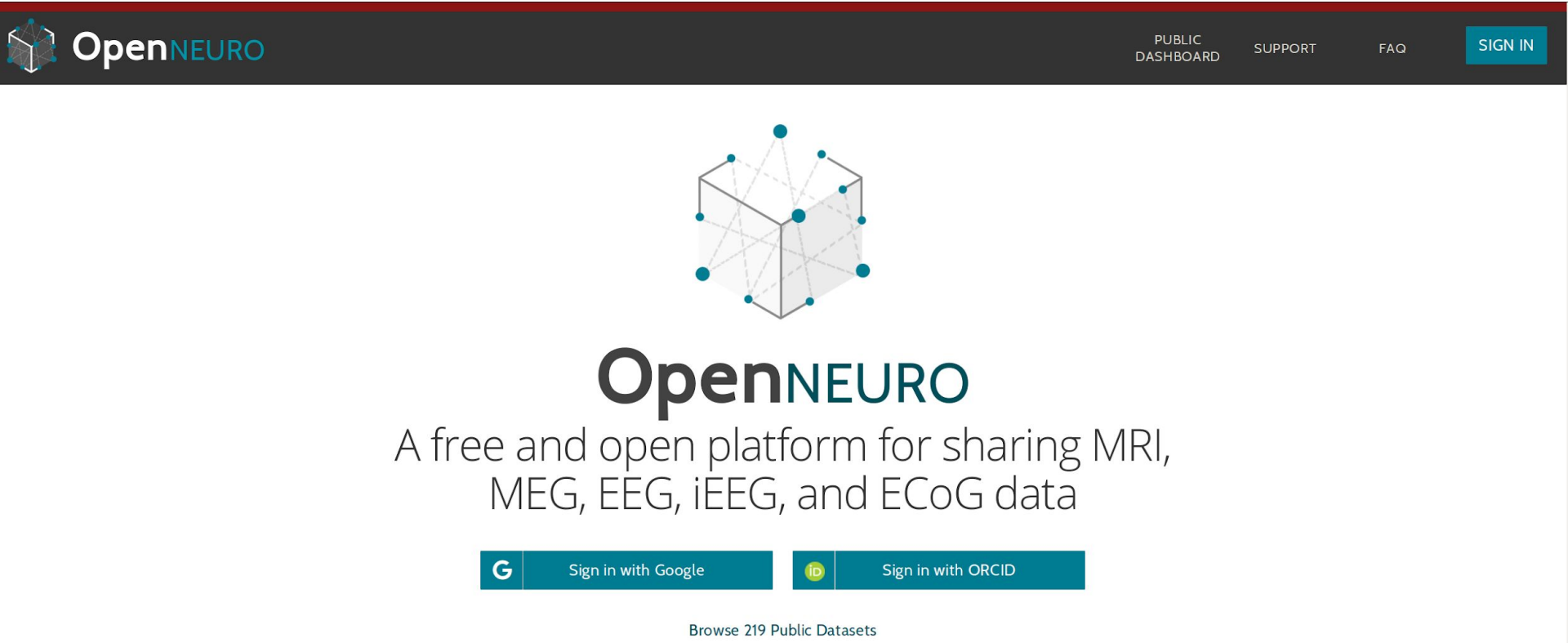
Important:

Welcome to the DataLad handbook!

handbook.datalad.org

Checking in on Professor Smith

With thanks to *Chris Gorgolewski*



OpenNeuro.org

Recent Activity

Most Active

20,329	UCLA Consortium for Neuropsychiatric Phenomics LA5c Study
19,376	The Midnight Scan Club (MSC) dataset
10,341	Visual object recognition
9,449	Forrest Gump
8,599	Classification learning

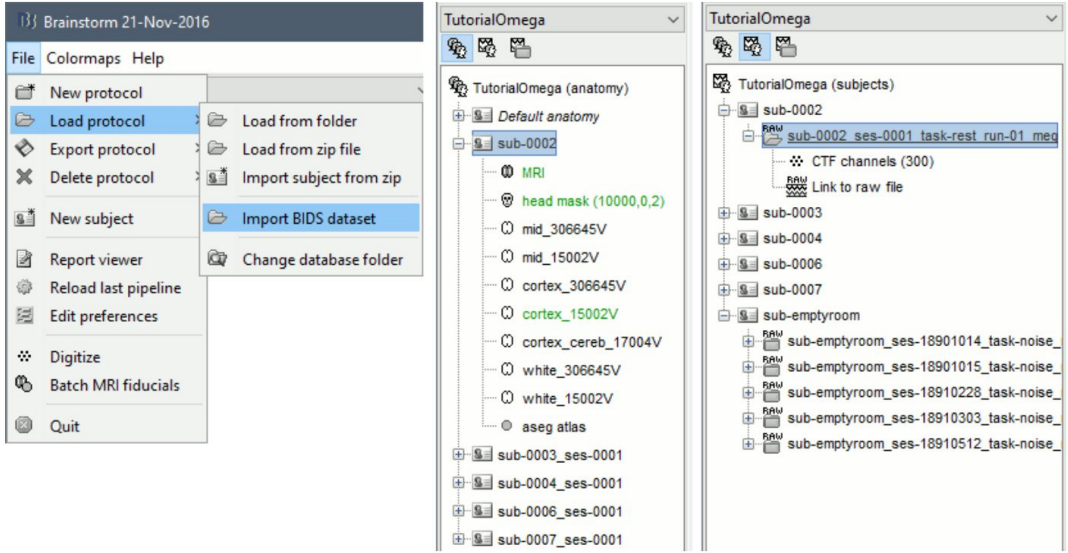
View More...

Recently Published

The Reading Brain Project L1 Adults	4 DAYS AGO
Parallel Adaptation of Symbols, Quantities, and Physical Size	10 DAYS AGO
A Probability Distribution over Latent Causes, in the Orbitofrontal Cortex	11 DAYS AGO
Spinal stimulation stepping and standing dataset	17 DAYS AGO
Handedness and Symbolic Number Representation	17 DAYS AGO

Recent activity on OpenNeuro.org

- Select the menu File > Load protocol > **Import BIDS dataset** > Select folder **OMEGA_RestingState_sample**.
- Keep the default values for all the questions that may be asked during the import process (eg. number of vertices in the cortex surfaces). Once done, you should be able to access the data for the 5 subjects in your database explorer: anatomy, and subject and noise recordings.



https://neuroimage.usc.edu/brainstorm/Tutorials/RestingOmega#BIDS_specifications

BIDS Apps

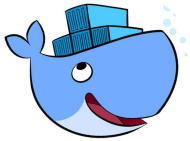
portable neuroimaging pipelines that understand BIDS datasets

About Tutorials Apps

How it works

Containerization

Package your software to use in any computing environment.



Structured queries

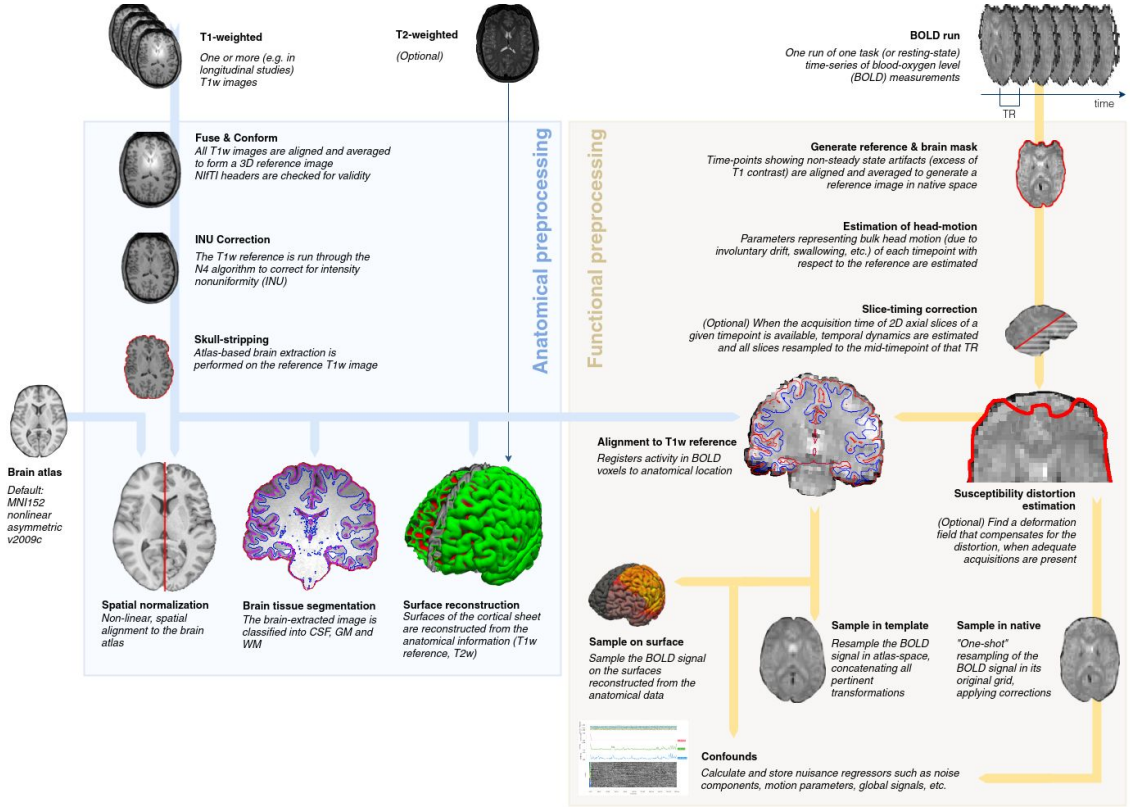
Know that the provided data meets certain attributes.

```
  name": "subject",
 "pattern": "(sub-\\d+)",
 "directory": "{{root}}/{subject}",
},
{
 "name": "session",
 "pattern": "(ses-\\d)",
 "directory": "{{root}}/{subject}/{session}",
},
{
 "name": "run",
 "pattern": "(run-\\d+)"
},
{
 "name": "type",
 "pattern": ".*_(.*?)\\."
```

Analysis code

Currently 26 supported BIDS Apps. Wrap your own software with the example app.

One example: fMRIPrep

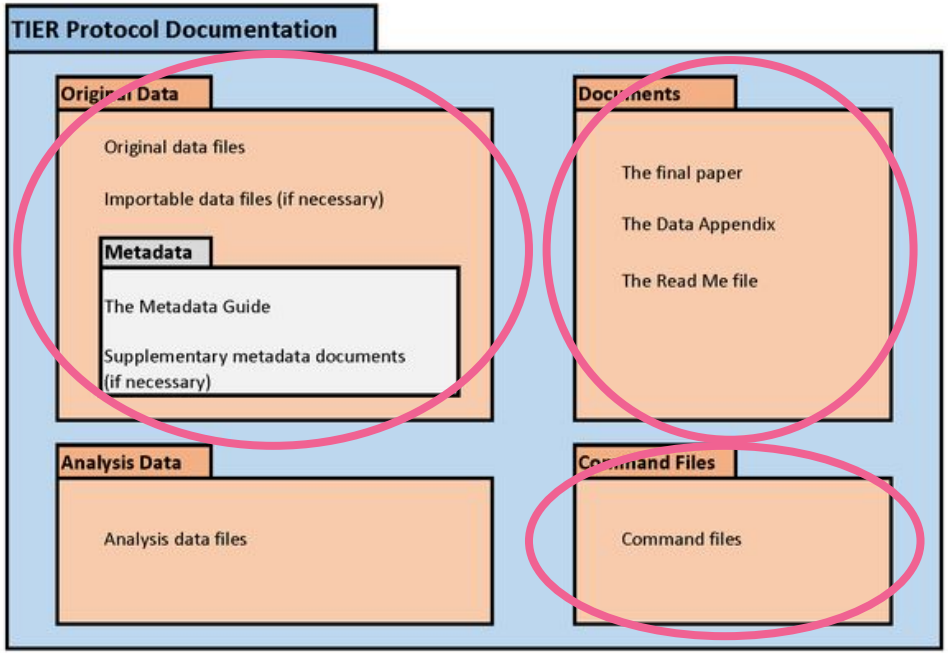



```
>> fmriprep     \
    data/ out/   \
    participant \
    -w work/
```

Professor Smith (circa 2030)

With thanks to *Chris Gorgolewski*

Reproln



RMarkdown, Jupyter

Cookiecutter

The TIER Specification

Making Your Code Citable

10 minute read

Digital Object Identifiers (DOI) are the backbone of the academic reference and metrics system. If you're a researcher writing software, this guide will show you how to make the work you share on GitHub citable by archiving one of your GitHub repositories and assigning a DOI with the data archiving tool Zenodo.

Intro

Choosing Your Repo

Login to Zenodo

Check Repo Settings

Zenodo GitHub integration

Acknowledgements

- Stefan Appelhoff 📖💬🤔🐛💡💻👀⚠️📢
- Tibor Auer 💬📖💡🔧📢
- Sylvain Baillet 📖🔍
- Elizabeth Bock 📖💡
- Eric Bridgeford 📖🔧
- Teon L. Brooks 📖💻⚠️💬👀🤔🔧🐛📢
- Suyash Bhogawar 📖💡⚠️🔧💬
- Vince D. Calhoun 📖
- Alexander L. Cohen 🐛💻📖💬
- R. Cameron Craddock 📖📢
- Samir Das 📖
- Alejandro de la Vega 🐛💻⚠️
- Arnaud Delorme 📖💡
- Eugene P. Duff 📖
- Elizabeth DuPre 📖💡🔍🤔
- Eric A. Earl 🤔
- Anders Eklund 📖📢💻
- Oscar Esteban 📖🔧🤔💬💻
- Franklin W. Feingold 📋📝✔️
- Guillaume Flandin 📖💻
- Remi Gau 📖💻
- Satrajit S. Ghosh 📖💻
- Tristan Glatard 📖💻
- Mathias Goncalves 💻🔧📢
- Krzysztof J. Gorgolewski 📖💻💬🤔🔍📢📝💡➰
- Alexandre Gramfort 📖💡
- Jeffrey S. Grethe 💬🐛✔️📢
- Yaroslav O. Halchenko 📖📢🔧💬🐛
- Daniel A. Handwerker 📖
- Michael Hanke 📖🤔🔧🐛📢
- Michael P. Harms 📖⚠️🔧
- Richard N. Henson 📖
- Dora Hermes 📖💻✔️🔍🤔
- Katja Heuer 🔧
- Chris Holdgraf 📖🤔
- Jean-Christophe Houde 📖
- International Neuroinformatics Coordinating Facility 💵📋
- Mainak Jas 📖💻
- David Keator 📖
- James Kent 💬💻
- Gregory Kiar 📖💻🎨🔧
- Pamela LaMontagne 📖💡
- Kevin Larcher 💬
- Laura and John Arnold Foundation 💵
- Xiangrui Li 📖💻

Acknowledgements

- Dan Lurie 🤔📖🔧🔌💻💬
- Camille Maumet 📖
- Christopher J. Markiewicz 💬📖💻
- Jeremy Moreau 📖💡
- Zachary Michael 📖
- Ezequiel Mikulan 📖💻
- Michael P. Milham 💡🔍
- Henk Mutsaerts 📖
- National Institute of Mental Health 💵
- B. Nolan Nichols 📖
- Thomas E. Nichols 📖📢🔧
- Dylan Nielson 📖💻🔧
- Guiomar Niso 📖💡📢
- Robert Oostenveld 📖🔧📢💡✔️⚠️🤔
- Dimitri Papadopoulos Orfanos 📖💡🤔💬
- Patrick Park 📖💡💬
- Dianne Patterson 📖
- John Pellman 📖
- Cyril Pernet 💬📖💡📋
- Franco Pestilli 📖
- Dmitry Petrov 📖💻
- Christophe Philips 📖
- Russell A. Poldrack 📖🔍📢
- Jean-Baptiste Poline 📖📢🤔🎨
- Vasudev Raguram 💻🎨📖🔧
- Ariel Rokem 📖
- Matt Sanderson 📖💻
- Gunnar Schaefer 📖
- Jan-Mathijs Schoffelen 📖
- Vanessa Sochat 📖
- Francois Tadel 📖🔌💡
- Roberto Toro 🔧
- William Triplett 📖
- Jessica A. Turner 📖
- Joseph Wexler 📖💡
- Kirstie Whitaker 📖💡🔍🤔📢💬
- Gaël Varoquaux 📖
- Tal Yarkoni 💻📖🤔🔍🔌👀📢🐛🎨

Take-home ideas

- **Project management** is for everyone
- **Community-driven standards** enable new kinds of science

@emdupre

An outline for this morning

- Community-based project management
- A practical introduction to pyBIDS
- Authoring for collaboration
- A practical introduction to Jupyter Notebooks and JupyterLab

An outline for this morning

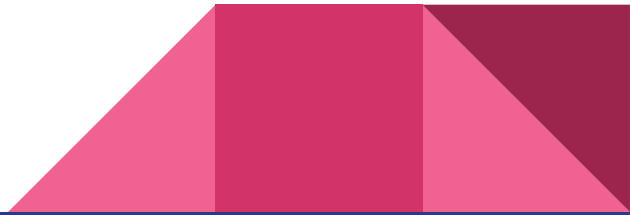
- Community-based project management
- A practical introduction to pyBIDS
- Authoring for collaboration
- A practical introduction to Jupyter Notebooks and JupyterLab

Jupyter

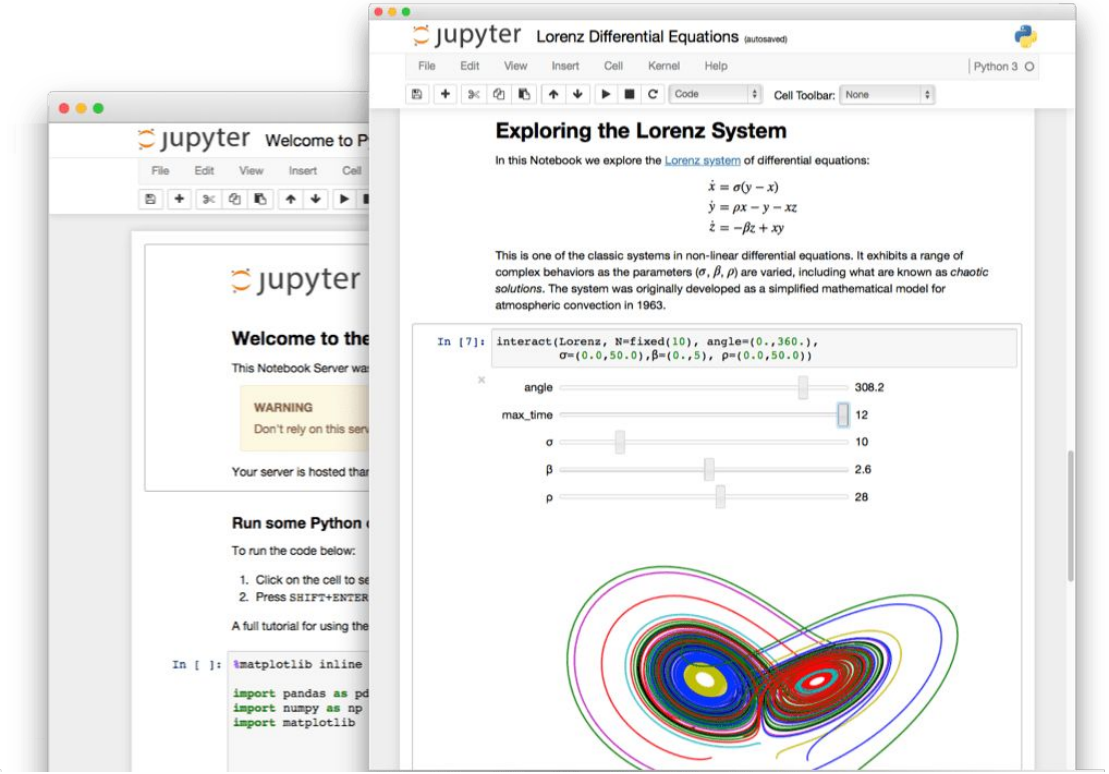
Authoring the next generation article

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.

-- Buckheit and Donoho
WaveLab and Reproducible Research, 1995



For example: the Jupyter Notebook



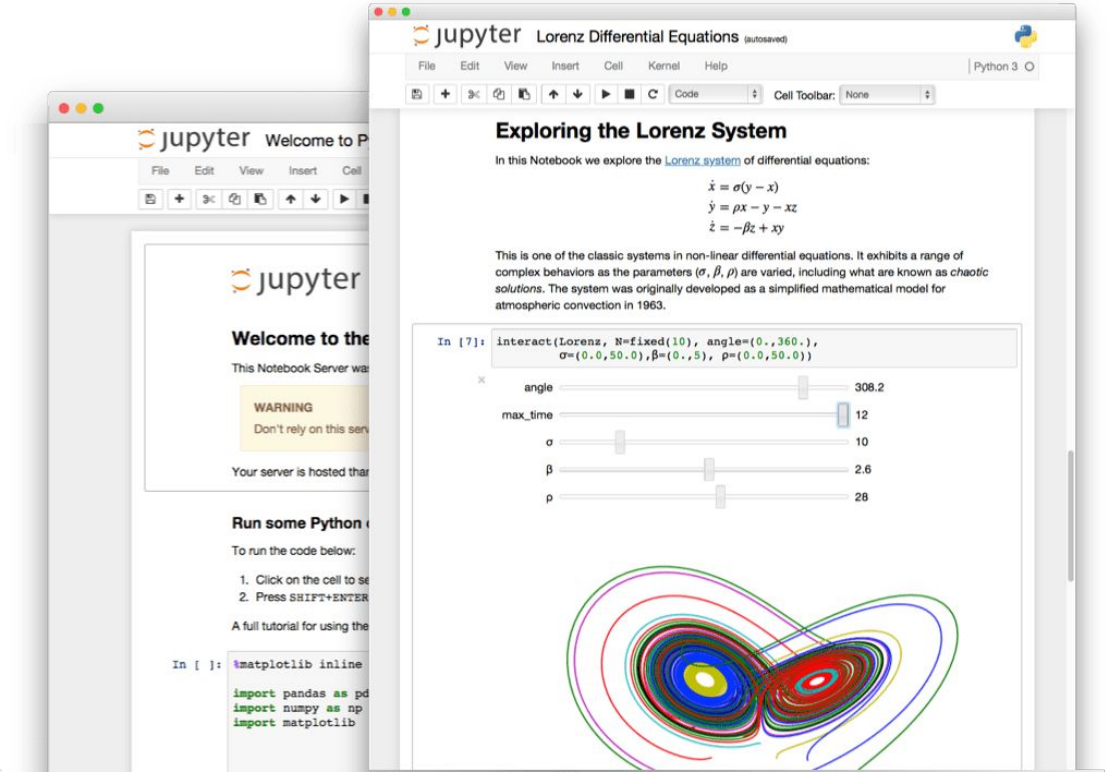
a *community* of people and an *ecosystem* of open tools and standards for interactive computing

With thanks to *Chris Holdgraf*

create things that are **language-agnostic** and **modular**.
Empower people to use **other open tools**.

With thanks to *Chris Holdgraf*

For example: the Jupyter Notebook



complete set of instructions

open-source languages are the raw material

complete set of instructions

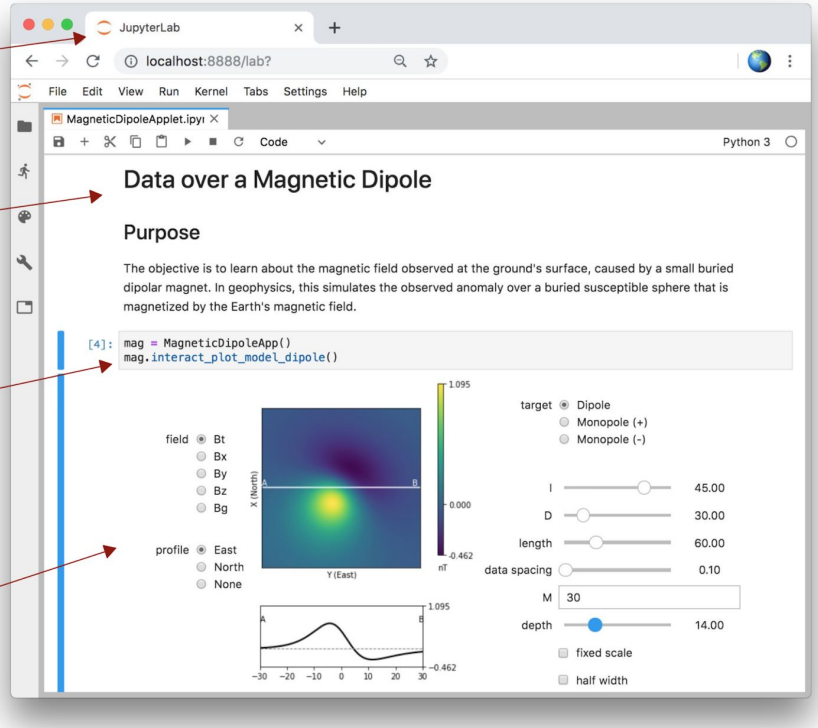
capture the steps

Runs in web browser

Text, comments, equations

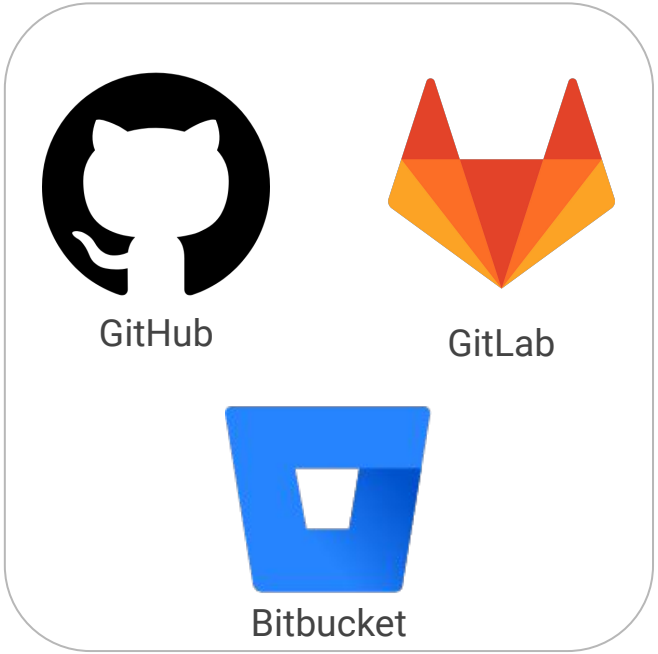
Lines of code
(Python, Julia, R, Matlab, …)

Output: values, images, plots, tables, interactive widgets



complete set of instructions

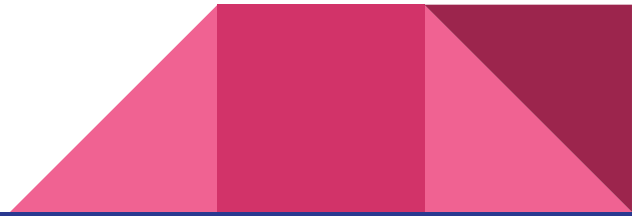
maintenance and sharing



- version control
- issue tracking
- licensing
- integrations with
  - testing services
  - documentation hosting
  - ...

What can we do with it ?

- Interleave explanatory text, code, and generated figures
- Add in additional kernels
- Create outputs in a variety of formats (HTML, PDF, Markdown)



Magic !

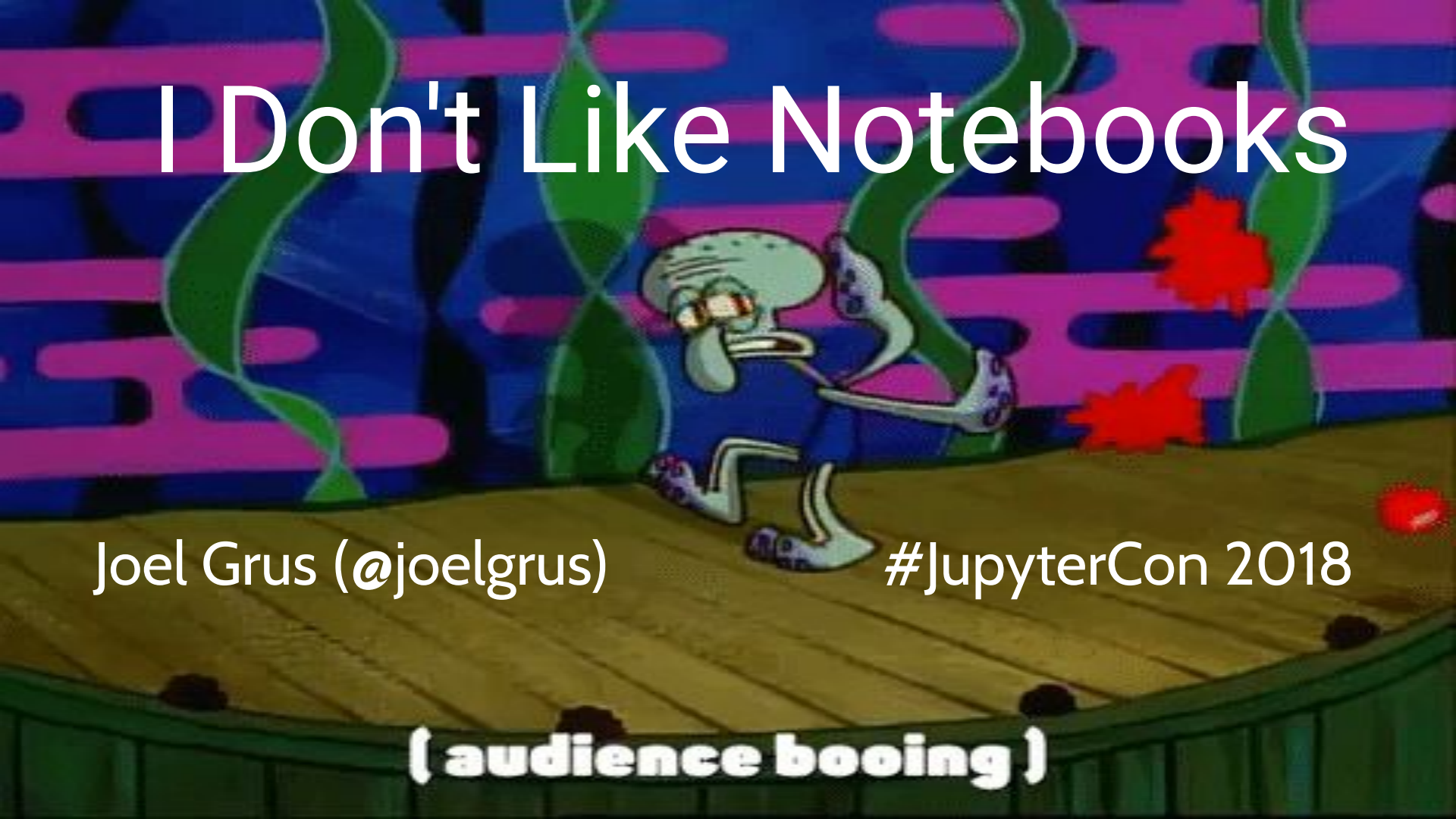
- `%matplotlib`
- `%timeit`
- `%load`
- `%history`

Here’s one workflow for collaborative authoring

With thanks to [Chris Holdgraf]

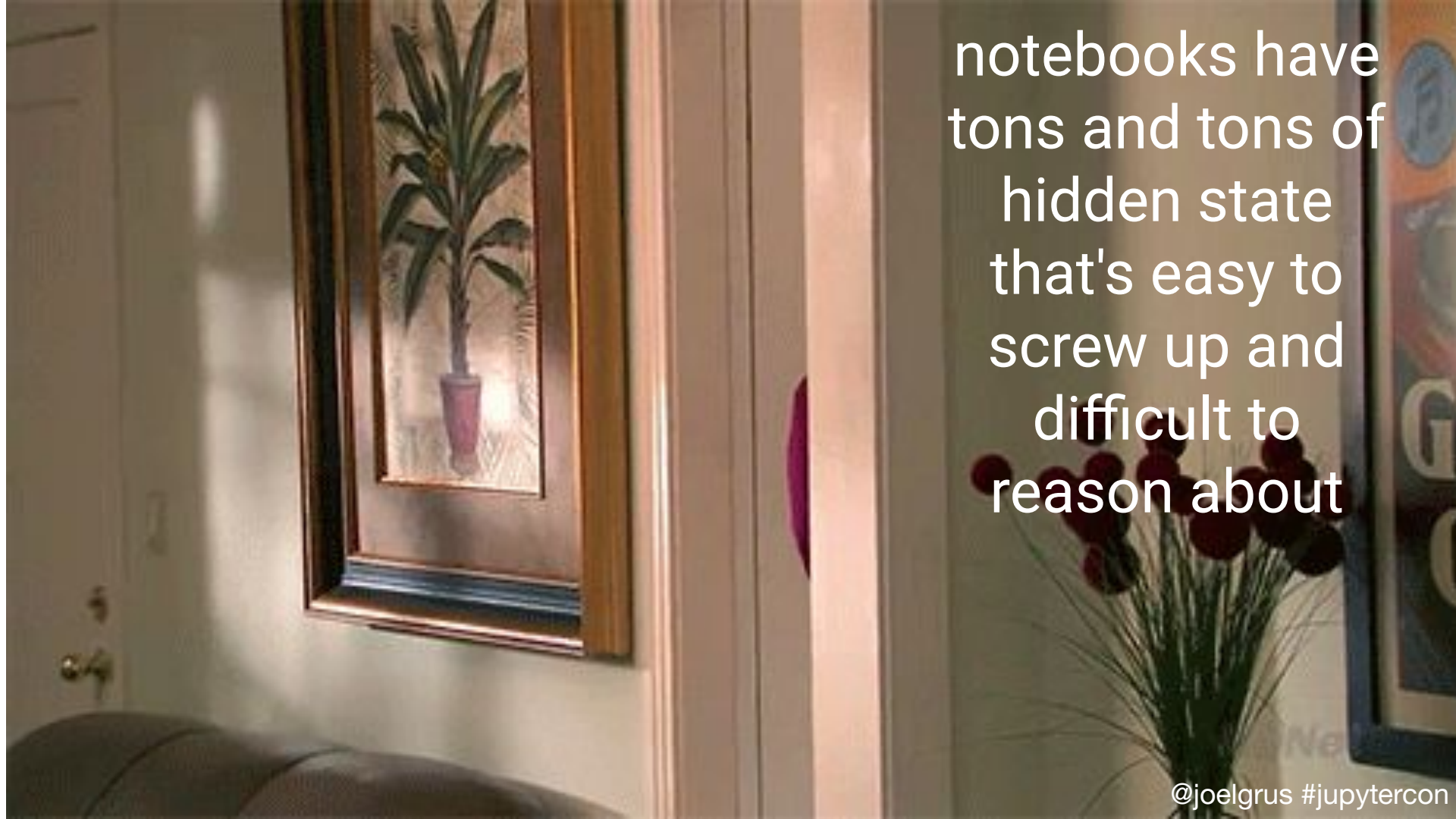
With thanks to Chris Holdgraf

I Don't Like Notebooks



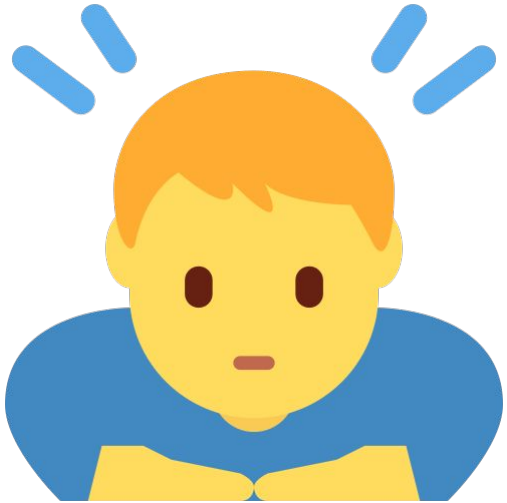
Joel Grus (@joelgrus)

#JupyterCon 2018


notebooks have tons and tons of hidden state that's easy to screw up and difficult to reason about
@joelgrus #jupytercon

We can respect statefulness

- `Kernel restart and run-all`
- Think about using tools beyond just the Notebook

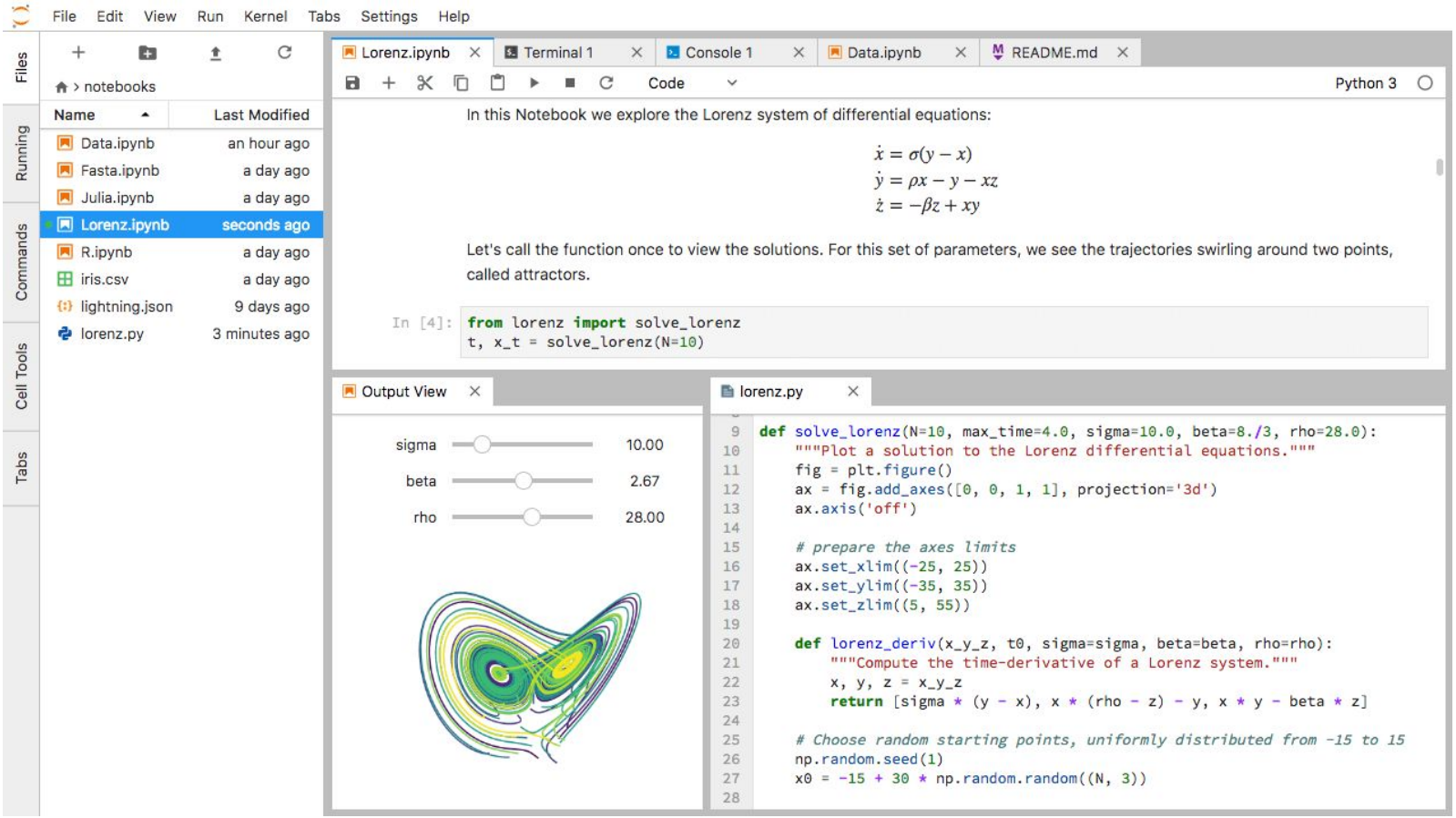


JupyterLab

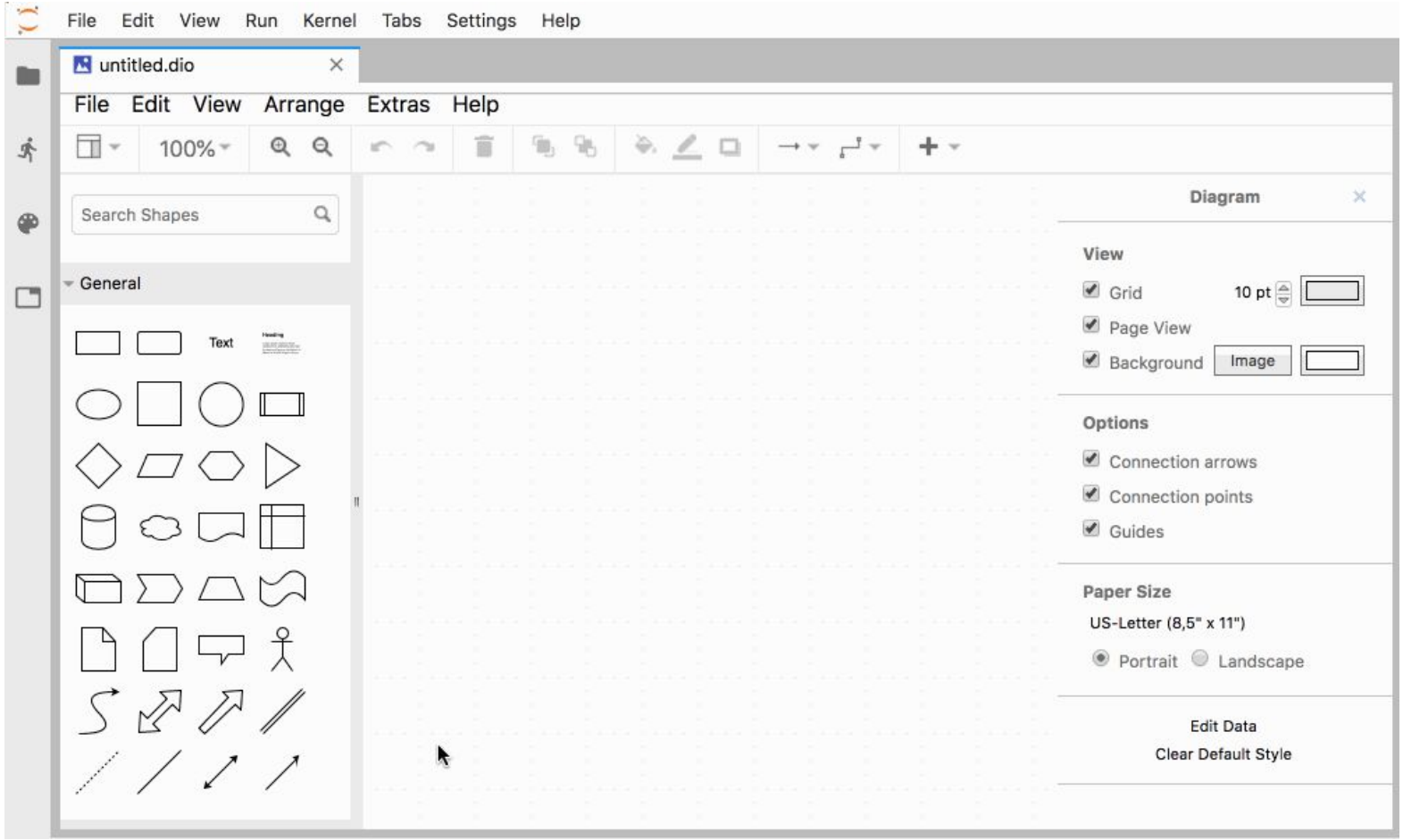
The next generation Jupyter interface

jupterlab.readthedocs.io

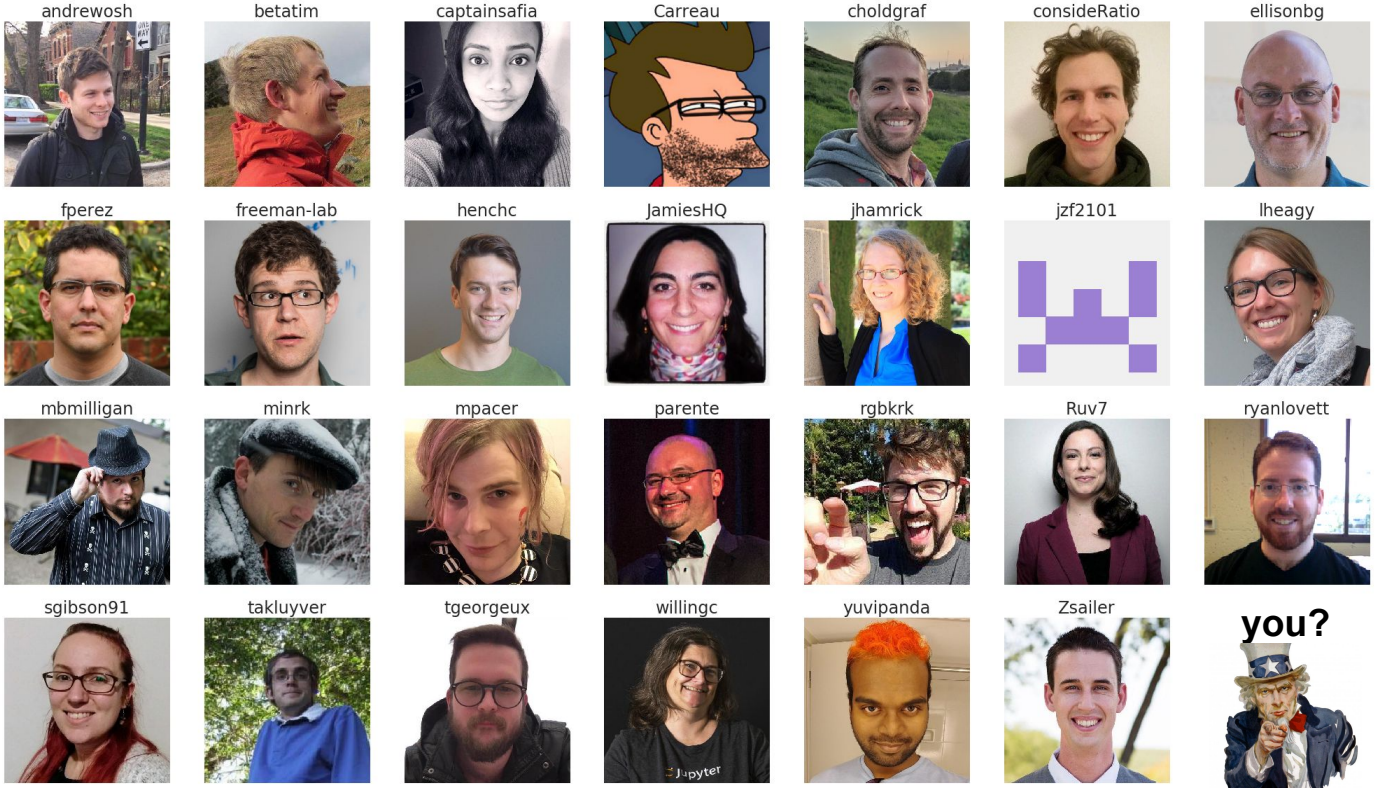
A complete software development environment



JupyterLab Extensions


File Edit View Run Kernel Tabs Settings Help
untitled.dio
File Edit View Arrange Extras Help
100%
Search Shapes
General
Text
Diagram
View
Grid
10 pt
Page View
Background
Image
Options
Connection arrows
Connection points
Guides
Paper Size
US-Letter (8,5" x 11")
Portrait
Landscape
Edit Data
Clear Default Style

Acknowledgements



With thanks to *Chris Holdgraf*

Take-home ideas

- There’s a whole galaxy **beyond the PDF,** and Jupyter can help us get there

@emdupre_

An outline for this morning

- Community-based project management
- A practical introduction to pyBIDS
- Authoring for collaboration
- A practical introduction to Jupyter Notebooks and JupyterLab