Dat in the Lab

Danielle Robinson, PhD
Scientific & Partnerships Director @ Code for Science & Society
@daniellecrobins

Better Science through Better Data
#scidata17
The Wellcome Collection - October 25, 2017

What's Dat?

Peer-to-peer protocol for syncing files across distributed networks

Built for sharing, versioning data

Like git, but built for data

Read the whitepaper for more[1]

@daniellecrobins

dat
DATA

Why Dat for research?

Secure data sharing, sync within a group

Fast sync/download of data from sources (equipment, people)

Automatic versioning, supporting reproducibility

@daniellecrobins

dat
DATA

Dat can automate:

Metadata creation (discoverability)

Data syncing, versioning within a group

Data versioning at point of publication[2,3]

... and post-publication data archiving[4]

dat
DATA

Genesis of project:

How can the Dat protocol help researchers to manage
share
publish
archive
find
data?

dat
DATA

What should we build that would make Dat more accessible?

What would we learn through close collaboration with research groups?

Our common goals:

Support researchers

Improve research data management
Facilitate reproducible research
Streamline data publishing
Develop free, open solutions

Learn from researchers, pilot workflows, test, iterate, return.

Dannise V. Ruiz-Ramos describes sea star genome annotation pipeline.
Read more: https://blog.datproject.org/tag/science/

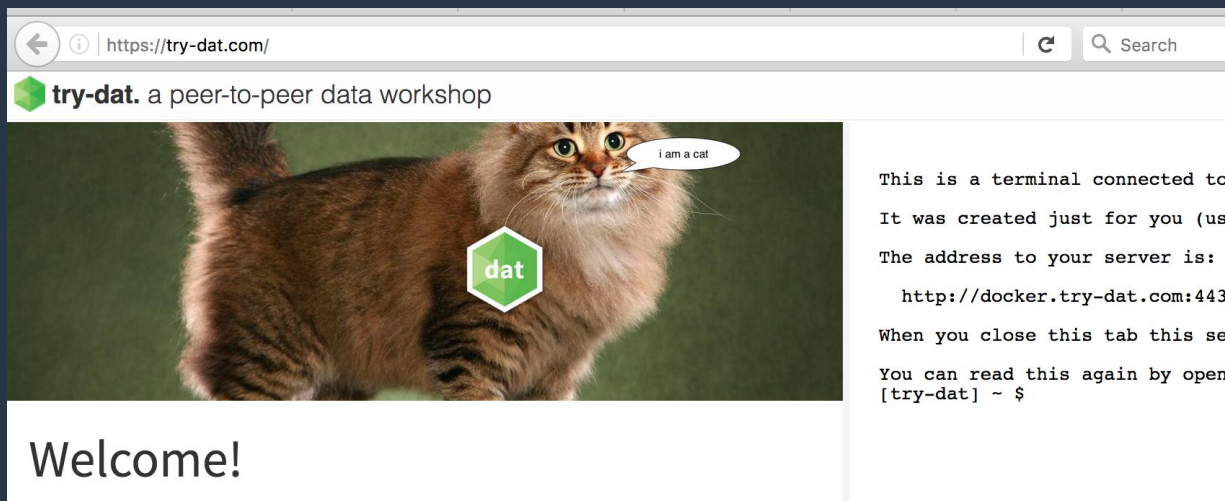Mike Dawson's Lab at UC Merced

What should we build?

- Support for versioned, reproducible research environments
- Multiwriter functionality

What are we learning?

- Institutional infrastructure
- Cross-institutional collaborations

To learn more…

[Try-dat.com](https://try-dat.com) or [join us at Mozfest](#)

Thank you!
Mozfest: Saturday 2pm, Decentralization

Web: https://codeforscience.org/
Blog: https://blog.datproject.org/tag/science/
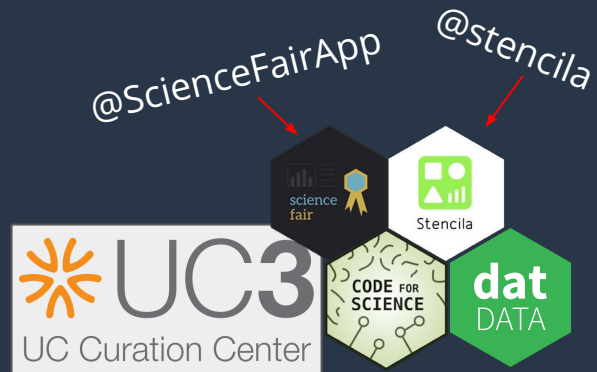Github: https://github.com/codeforscience/Dat-in-the-Lab

UC3 @UC3CDL
Dat Project @dat_project
Stephen Abrams @slabrams
Max Ogden @denormalize
Danielle Robinson @daniellecrobins

GORDON AND BETTY
MOORE
FOUNDATION

UC3
UC Curation Center

CODE FOR SCIENCE

dat DATA

Thank you!
Mozfest: Saturday 2pm, Decentralization

Web: https://codeforscience.org/
Blog: https://blog.datproject.org/tag/science/
Github: https://github.com/codeforscience/Dat-in-the-Lab

UC3 @UC3CDL
Dat Project @dat_project
Stephen Abrams @slabrams
Max Ogden @denormalize
Danielle Robinson @daniellecrobins

@ScienceFairApp

@stencila

GORDON AND BETTY
MOORE
FOUNDATION

UC3
UC Curation Center

science fair

Stencila

CODE FOR SCIENCE

dat DATA

Citations & links

1. Dat project whitepaper:
   https://github.com/datproject/docs/blob/master/papers/dat-paper.md
2. Klein M, Van de Sompel H, Sanderson R, Shankar H, Balakireva L, et al. (2014). Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLOS ONE* 9(12): e115253.
   https://doi.org/10.1371/journal.pone.0115253
3. Jones SM, Van de Sompel H, Shankar H, Klein M, Tobin R, et al. (2016). Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content. *PLOS ONE* 11(12): e0167475.
   https://doi.org/10.1371/journal.pone.0167475
4. Harmon A. Activists Rush to Save Government Science Data — If They Can Find It. The New York Times. March 6, 2017.
   https://www.nytimes.com/2017/03/06/science/donald-trump-data-rescue-science.html?_r=0

dat
DATA