

Dat in the Lab

California Digital Library UC3
Code for Science & Society
The Dat Project

Stephen Abrams [@UC3CDL](#)
Max Ogden [@denormalize](#)
Danielle Robinson [@daniellecrobins](#)

Our common goals:

Support researchers

Improve research data management

Facilitate reproducible research

Streamline data publishing

Develop free, open solutions

Research today:

Volume and complexity of data is increasing¹

Funders, publishers moving towards open data policies ([NSF](#), [Gates](#), [NIH](#), [PLOS](#))

Many repositories to choose from²

Research today:

Data publishing and archiving via repositories as “data silos”

Inconsistent adoption of metadata standards

Research data management practices vary between groups and people

Research today:

Spotty metadata

Dead links³

Content drift⁴

... is the data backed up?



Genesis of project:

How can decentralization help
researchers to manage
share
publish
archive
data?

Genesis of project:

What should we build that would make using Dat easier?

What would we learn through close collaboration with research groups?

What's Dat?

File sharing protocol

Peer to peer file sharing, syncing,
versioning

Like git, but built for data

[Whitepaper](#)



Why Dat?

Secure data sharing within a group

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

Automatic dataset versioning

Platform agnostic



Try Dat!

For this workshop you will need:

a laptop

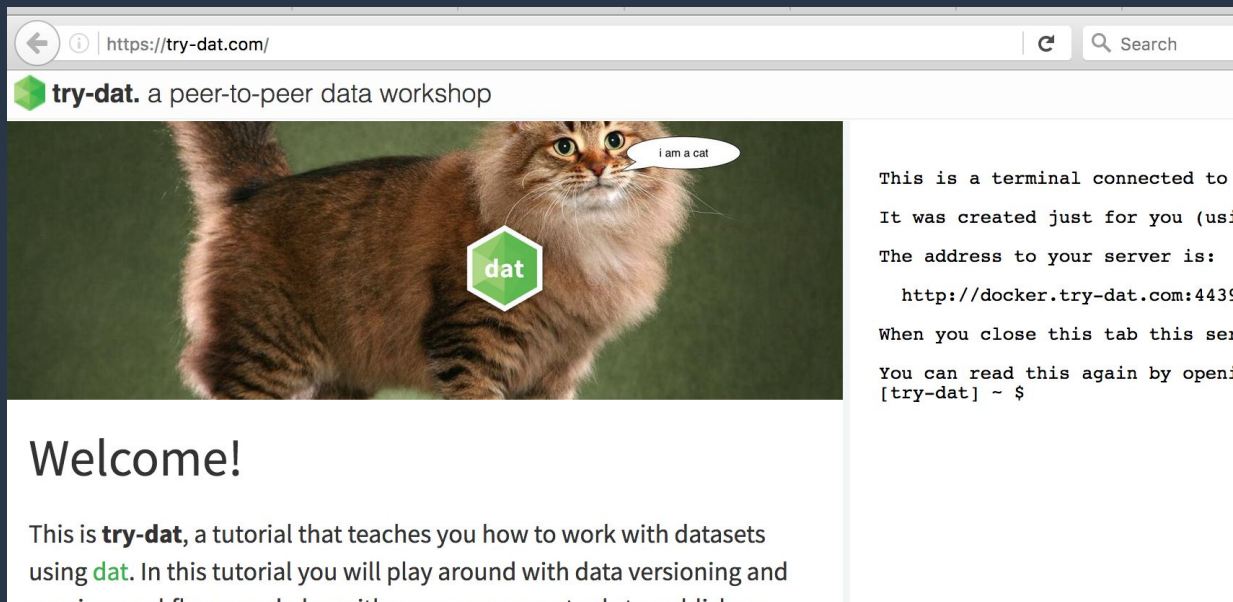
wifi

no special skills



Try Dat!

try-dat.com



The screenshot shows a web browser window with the URL `https://try-dat.com/`. The page header includes the try-dat logo and the text "a peer-to-peer data workshop". Below the header is a large image of a brown cat with a green hexagonal "dat" logo overlaid on its chest and a speech bubble saying "i am a cat". To the right of the cat image is a terminal window with the following text:

```
This is a terminal connected to  
It was created just for you (use  
The address to your server is:  
  
http://docker.try-dat.com:4439  
When you close this tab this ser  
You can read this again by open  
[try-dat] ~ $
```

Below the cat image, the text "Welcome!" is displayed. Underneath, a paragraph reads: "This is **try-dat**, a tutorial that teaches you how to work with datasets using **dat**. In this tutorial you will play around with data versioning and..."





Welcome!

This is **try-dat**, a tutorial that teaches you how to work with datasets using **dat**. In this tutorial you will play around with data versioning and syncing workflows, and play with some awesome tools to publish or share data over the peer-to-peer web.






Table of Contents

- 01-install
- 02-link
- 03-modify
- 04-live
- 05-website
- 100-beaker

If you look to your right, you'll see there is a terminal that is connected to a **docker container** -- that is a remote Linux computer, in the cloud but accessible right here in your browser!

This is a terminal connected to a virtual Ubuntu Linux server.
It was created just for you (using Docker) when you loaded the page.
The address to your server is:
`http://docker.try-dat.com:44391`
When you close this tab this server will shut down.
You can read this again by opening ``welcome.txt`` at any time.
[try-dat] ~ \$

Virtual terminal

 .bashrc
➤  .npm
 .profile
 cat.png
 welcome.txt

Files on the
virtual server

```
1
2 This is a terminal connected to a virtual Ubuntu Linux server.
3
4 It was created just for you (using Docker) when you loaded the page.
5
6 The address to your server is:
7
8   http://docker.try-dat.com:44391
9
10 When you close this tab this server will shut down.
11
12 You can read this again by opening `welcome.txt` at any time.
13
```

Contents of
`welcome.txt`

Thank you!

Hands-on portion of the workshop
starts now

UC3 [@UC3CDL](#)

Dat Project [@dat_project](#)

Stephen Abrams

Max Ogden [@denormalize](#)

Danielle Robinson [@daniellecrobins](#)

Chat with us - chat.datproject.org

Citations & links

1. Dinov ID. Volume and Value of Big Healthcare Data. *Journal of medical statistics and informatics*. 2016;4:3. doi:10.7243/2053-7662-4-3.
2. [PLOS ONE recommended data repositories list](#)
3. 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>
4. 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>