

dat

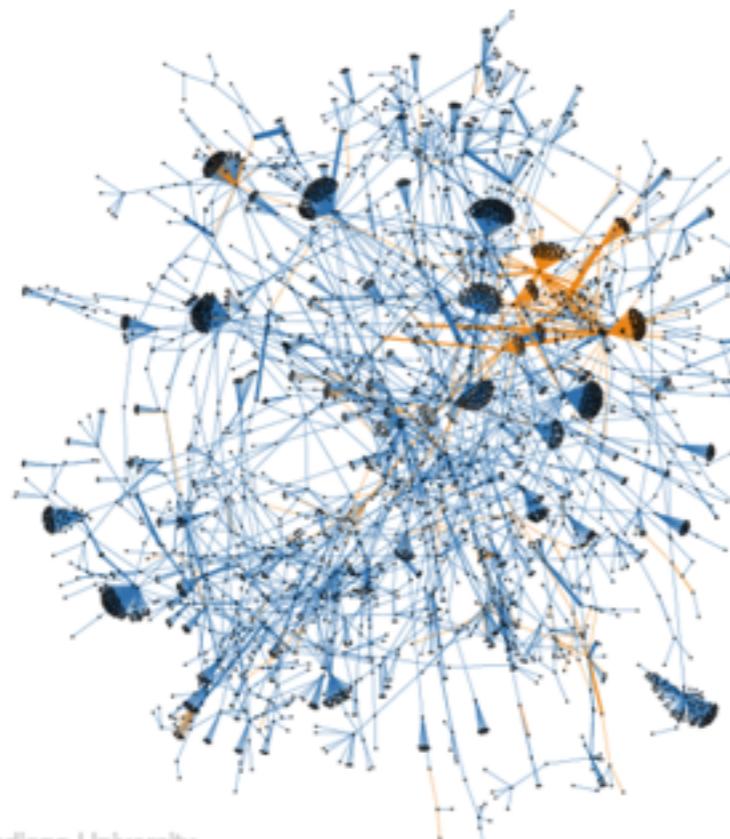
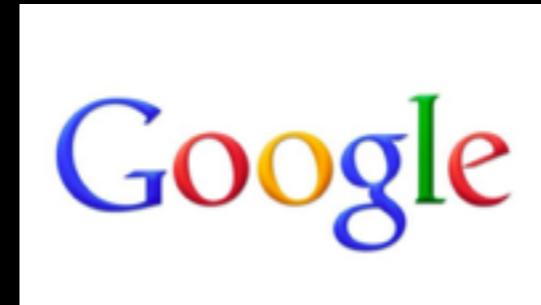
a community-driven project for
secure, real-time distributed data sync

<https://dat.foundation>

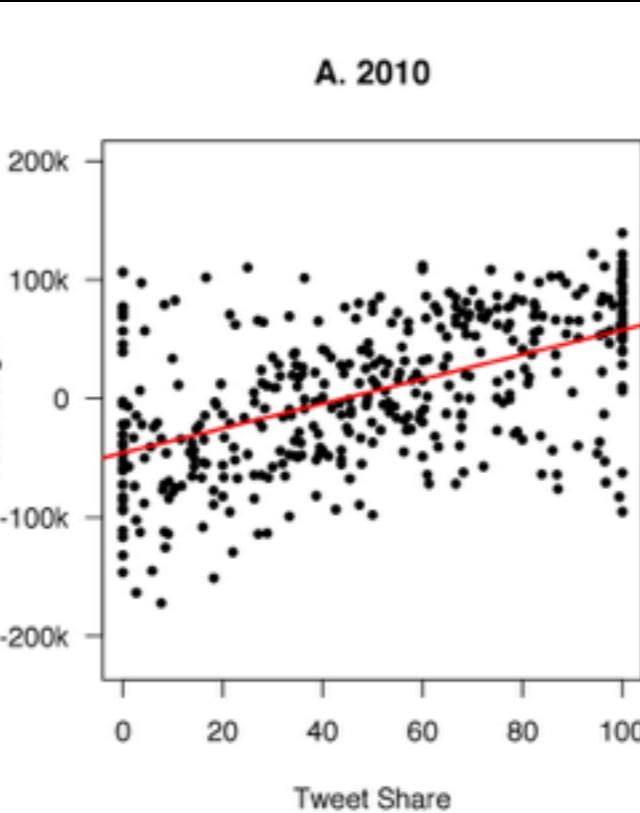
Karissa McKelvey
Founding Member
Dat Foundation
[@okdistribute](https://twitter.com/okdistribute)



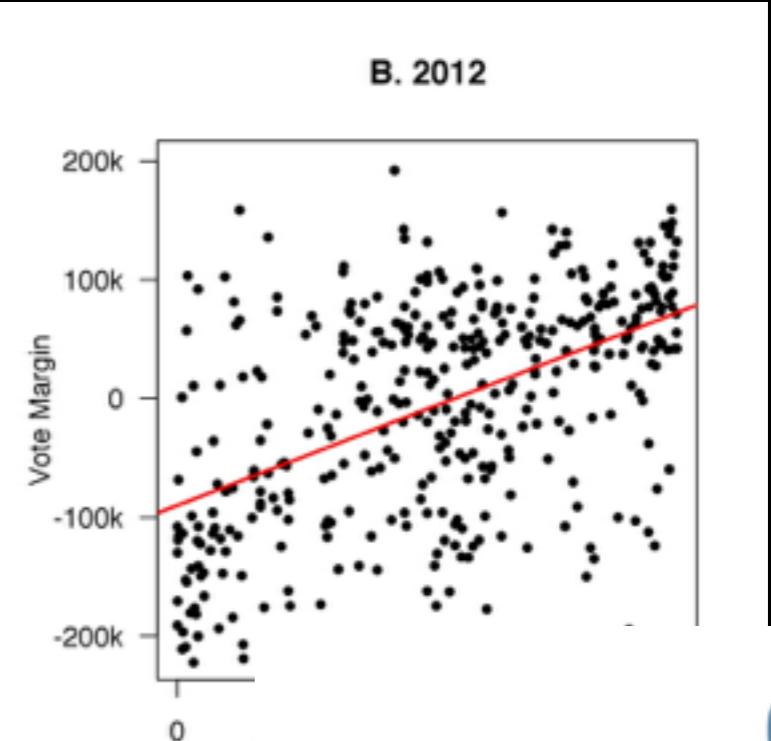
karissa \$ ~



Copyright 2010 Indiana University



[truthy.indiana.edu](http://Truthy.indiana.edu)



Pandas



The Economist

OCTOBER 19TH-25TH 2013

Economist.com

Washington's lawyer surplus

How to do a nuclear deal with Iran

Investment tips from Nobel economists

Junk bonds are back

The meaning of Sachin Tendulkar

HOW SCIENCE GOES WRONG

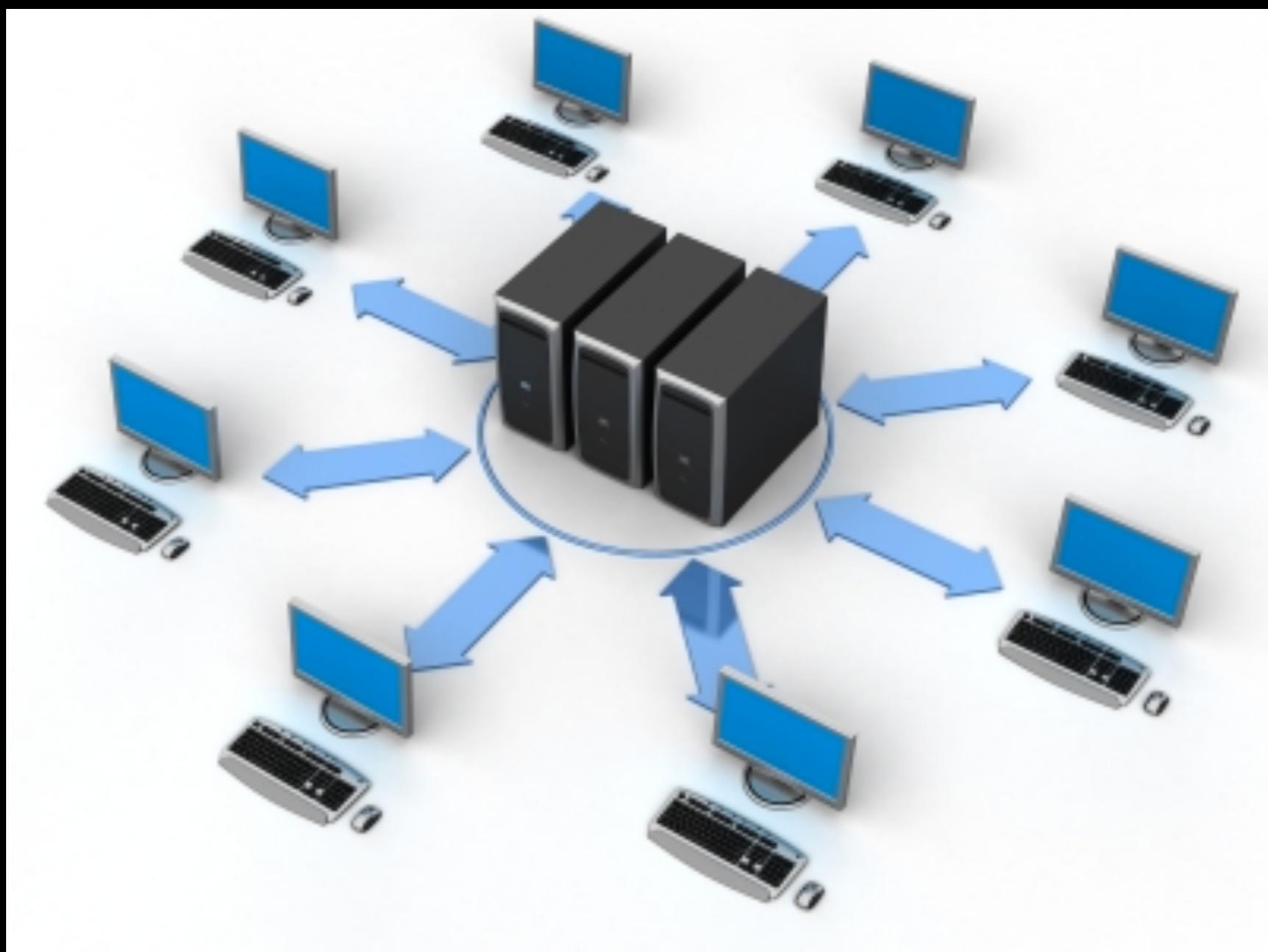
99
Einsteinium

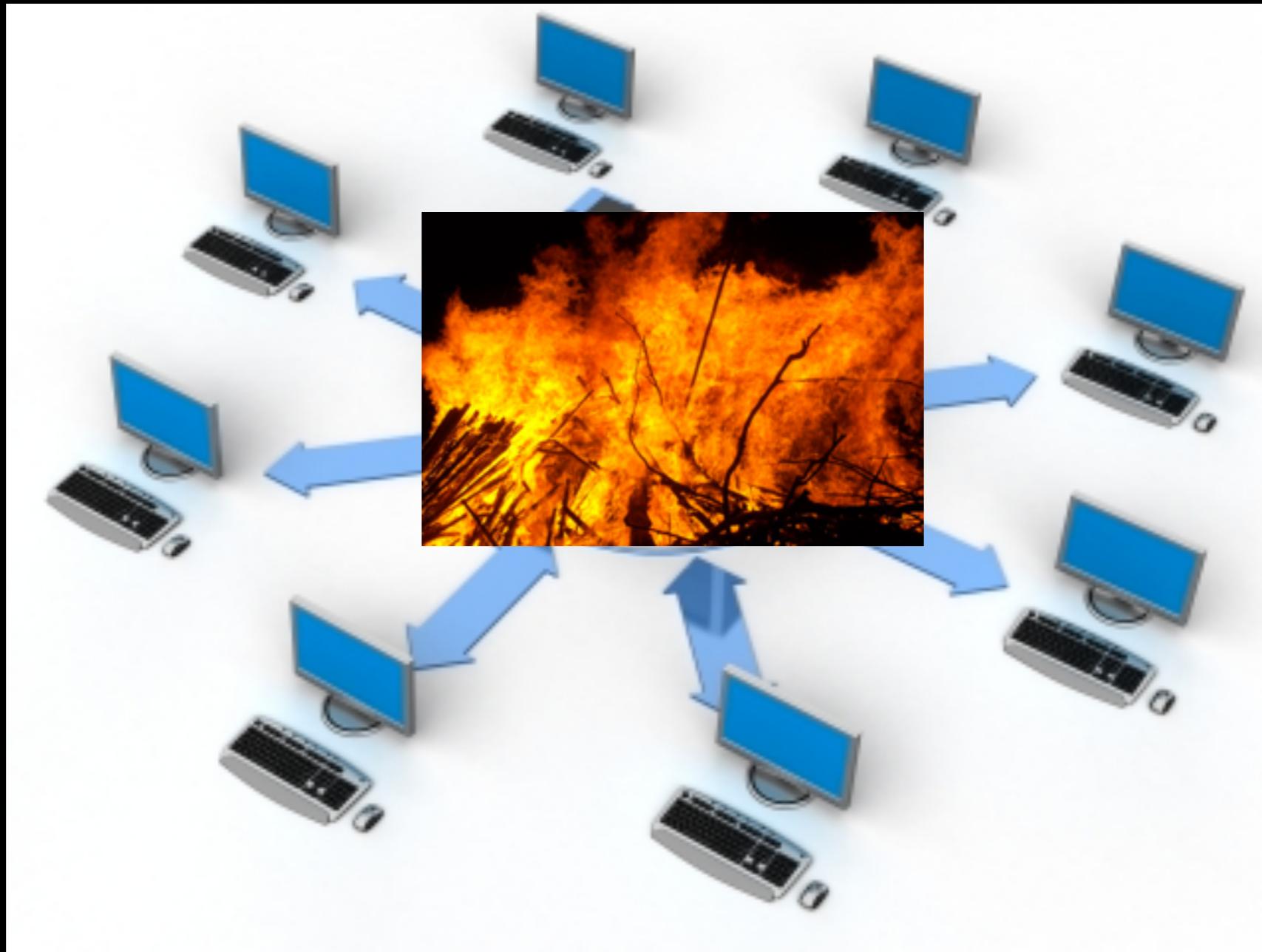
“A rule of thumb ... is that half of published research cannot be replicated”

How do we replicate research today?

collaborate on
How do we ~~replicate~~ research today?

collaborate on
How do we ~~replicate research~~ today?
data analysis





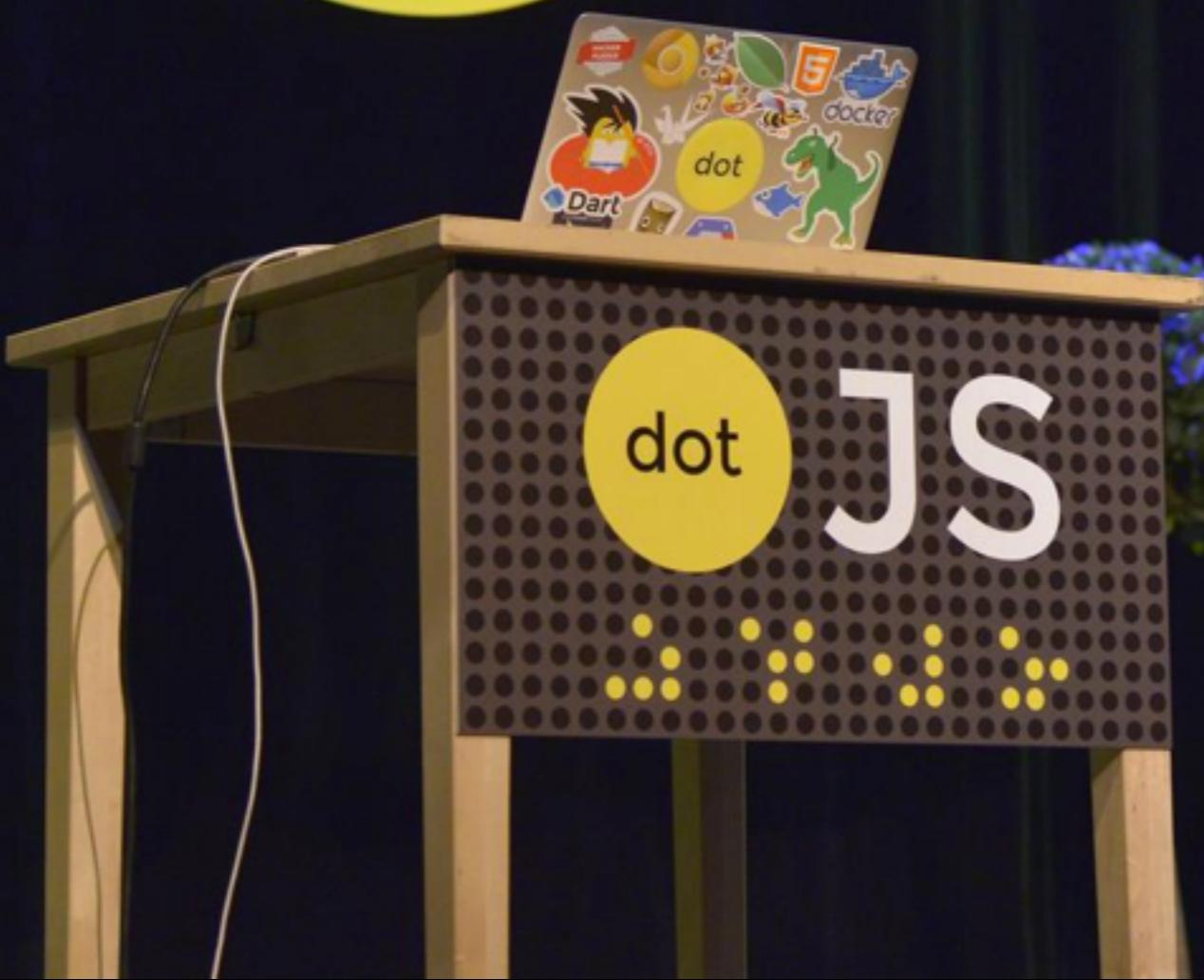
- DropBox
- S3
- rsync
- email
- git

we want to do for
data what git did
for source code





dot JS



PANELISTS

Juan Benet

IPFS
@juanbenet

Trent McConaghy

BIGCHAINDB & IPDB
@trentmc0

Karissa McKelvey

THE DAT PROJECT
@okdistribute

Denis Nazarov

MEDIACHAIN
@literature

Evan S

INTERLEDGER
@_emschwart

Moonshot Challenge: What could you do to decentralize Scientific Journal Articles?

MODERATED BY
Wendy Hanamura
INTERNET ARCHIVE
@whanamura





SCIENCE > PROFIT







- Decentralized / P2P
- Massive adopted / Simple protocol
- Only works for static files
- Scales poorly on really big data sets
- Eavesdrop & find out who is sharing what



goals

- **Easy to use** alternative to Dropbox
- Decentralized / P2P for **dynamic & real-time**
- Built for modern **scientific & activist** use cases
- Always be **faster** than other tools
- **No eavesdroppers!**

cryptographically secure dat link



<dat://8c5f5365668b13a10899d276f3b8dac998a8917bf48248b1d81f5c6d6d31af07>

cryptographically secure dat link

<dat://8c5f5365668b13a10899d276f3b8dac998a8917bf48248b1d81f5c6d6d31af07>

never changes, even as data is updated

log

Append only logs

(Append item to list) → Data item #0

Data item #0

(Append item to list)→

Data item #1

Data item #0

Data item #1

(Append item to list)→

Data item #2

How can we share append only logs?

(over a p2p network where we don't trust other people)

Merkle Trees

(a tree structure that verifies data)

Merkle Trees

(a tree structure that verifies data)
(unrelated to Angela Merkel)



Merkle Trees
(a tree structure that verifies data)
(unrelated to Angela Merkel)



git

a community-driven project for
secure, **real-time** distributed data sync

Real-time...?

Every time we append data **Root hash** changes

Crypto to the rescue

Generate a key pair

Secret Key + Public Key



trusts Public Key

Secret Key

👩 signs the root

Root hash #2

Hash #1

Hash #0

Hash #2

Hash #4

Data #0

Data #1

Data #2

Secret Key

Root hash #3

👩 signs the new root

Hash #3

Hash #1

Hash #5

Hash #0

Hash #2

Hash #4

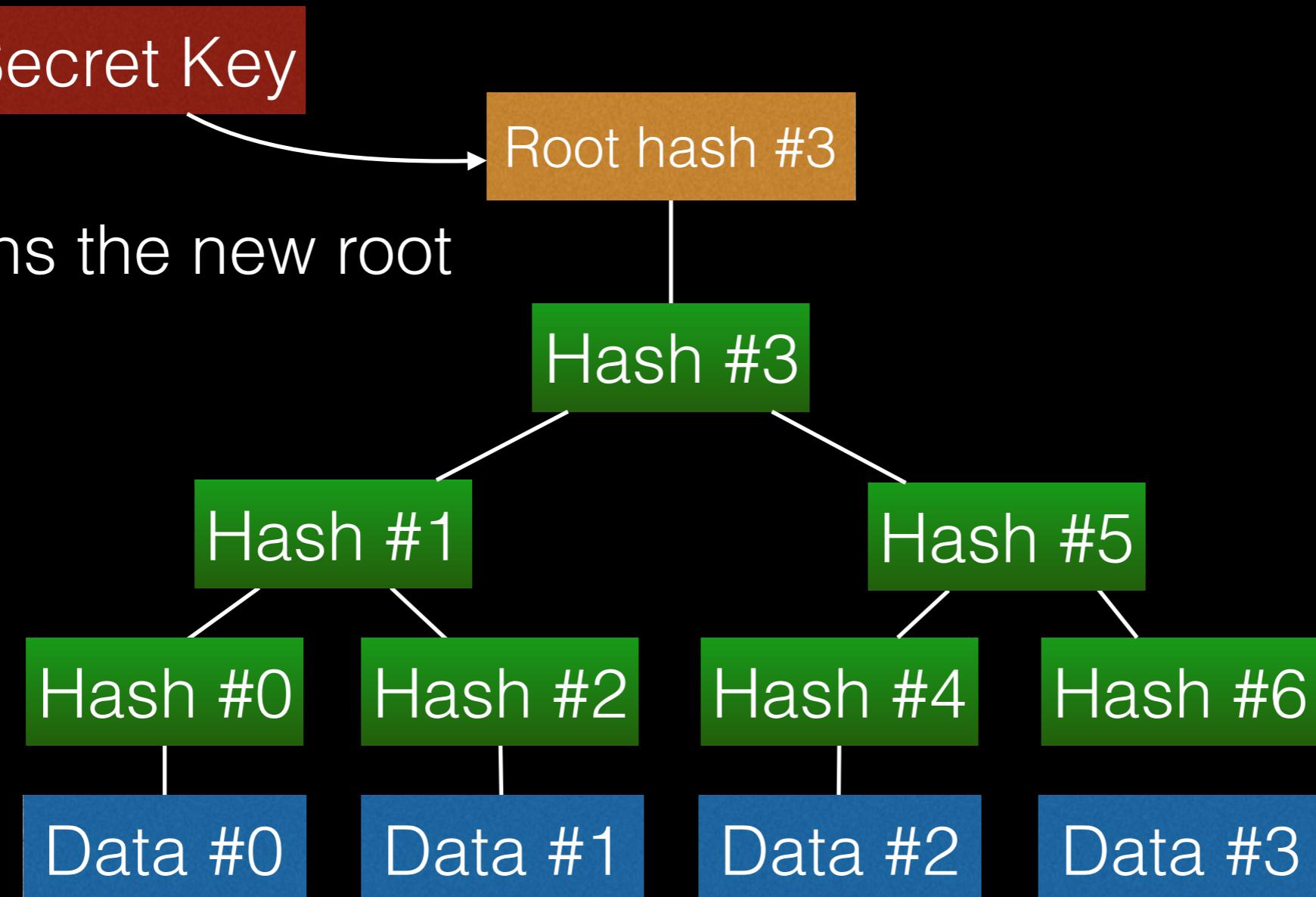
Hash #6

Data #0

Data #1

Data #2

Data #3





uses **Public Key** to verify **Root hash** signatures

npm install hypercore

log

protocol

storage

log

protocol

storage

index

log

protocol

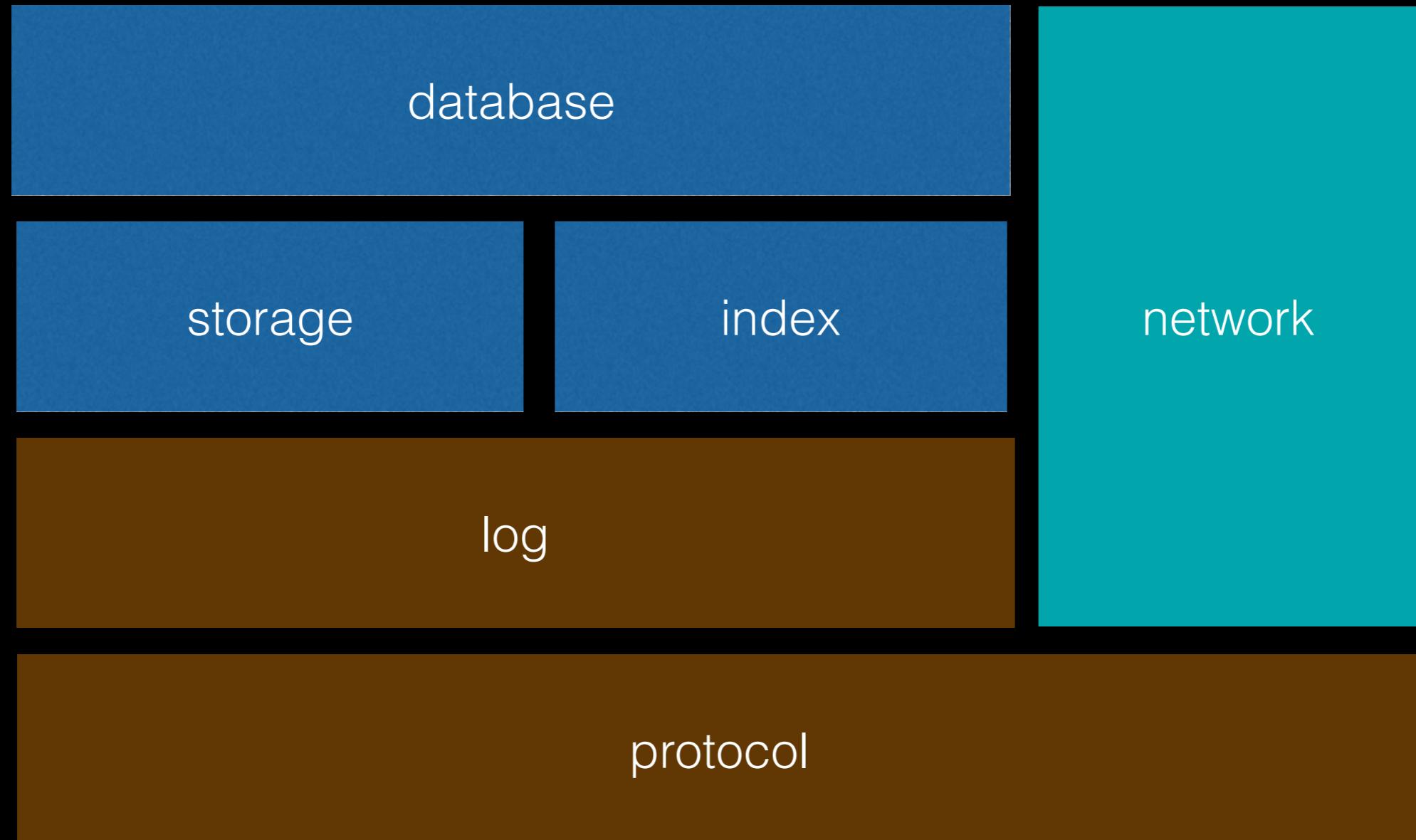
database

storage

index

log

protocol



Dat - Distributed Dataset Synchronization And Versioning

Maxwell Ogden, Karissa McKelvey, Mathias Buus Madsen

April 2017, Code for Science

Abstract

Dat is a protocol designed for syncing folders of data, even if they are large or changing constantly. A cryptographically secure register of changes to the data over time is used to ensure dataset versions are distributed as exact copies of exact versions. Any byte range of any version of any file can be efficiently streamed from a Dat repository over a network connection. Consumers can choose to fully or partially replicate the contents of a remote Dat repository, and can also subscribe to live changes. Dat uses built-in public key cryptography to encrypt network traffic, allowing it to make certain privacy and security guarantees. A group of Dat clients can connect to each other to form a public or private decentralized network to exchange data amongst each other in a swarm. A reference implementation is provided in JavaScript.

cloud storage services which fixes many issues with broken links but rely on proprietary code and services requiring users to store their data on centralized cloud infrastructure which has implications on cost, transfer speeds, vendor lock-in and user privacy.

Distributed file sharing tools can become faster as files become more popular, removing the bandwidth bottleneck and making file distribution cheaper. They also use link resolution and discovery systems which can prevent broken links meaning if the original source goes offline other backup sources can be automatically discovered. However these file sharing tools today are not supported by Web browsers, do not have good privacy guarantees, and do not provide a mechanism for updating files without redistributing a new dataset which could mean entire redownloading data you already have.

~7,000 downloads / week

~7,000 downloads / week
~1,500 dependent projects (that we know of)





user interface

database

storage

index

network

log

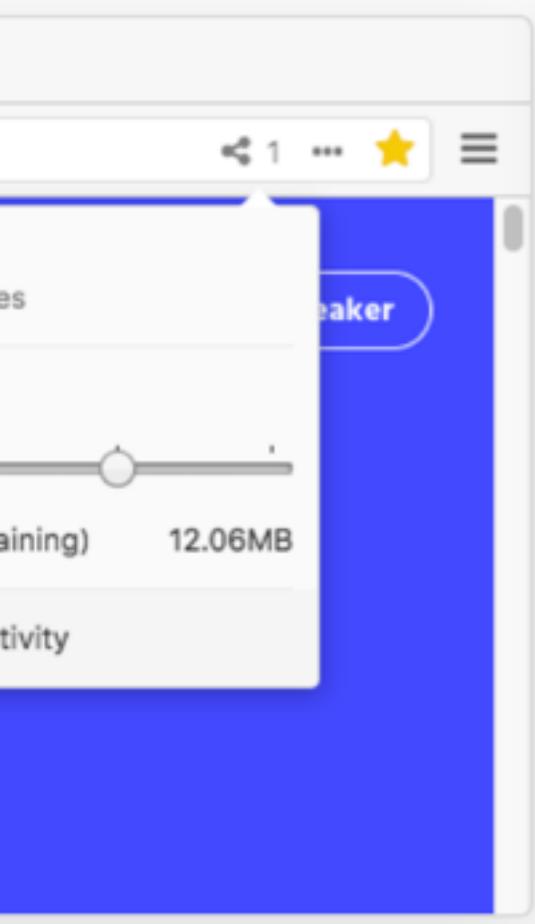
protocol



WEB BROWSER

BEAKER

<http://beakerbrowser.com>



Library - PIXEL-EYEZ

beaker://library/dat://4bcecf2846c1a60320db246ec2d84acced2806cc566c4cebcdeb87ff4d9336f6

PIXEL-EYEZ READ-ONLY

Make an editable copy

dat://4bcecf..f6/

Files Network About

.	/	60.47 kB
█	fonts	Apr 23, 2018, 2:55pm
█	images	Apr 23, 2018, 2:55pm
📄	.datignore	May 8, 2018, 12:58pm
📄	dat-links.js	Apr 23, 2018, 2:52pm
📄	dat.json	Jul 6, 2018, 4:10pm
📄	index.html	Apr 23, 2018, 3:04pm
📄	index.js	Apr 23, 2018, 10:04pm
📄	LICENSE.md	Apr 20, 2018, 9:44pm
📄	nxon.js	Apr 22, 2018, 11:26pm

PIXEL-EYEZ

Pixel size

Choose image

decentralized data preservation

community mapping

social media



COLLABORATIVE
MAP EDITOR

MAPEO

<http://mapeo.world>

Mapeo



Mac OS X interface with Mapeo application window open.

Mapeo window title bar: Mapeo File Edit Visualization Window Help

Top right status bar: Sun 23:09 5% A

Map view: Satellite imagery showing several lakes and rivers. Labeled features include "Jaijaira (laguna grande)" (highlighted in red), "Emuyai Jaira", "Sesepuaro", "Uystra", "Tul'aro Sh'a", and "Neppjo".

Toolbars: Point, Line, Area, and a zoom tool.

Syncronize USB button: Syncronize USB

Left sidebar: Edit feature panel for "Jaijaira (laguna grande)".

Field	Value
lake_type	grande
Name	Kwiña Jaira
Note	Nombre especial de la laguna por la abundante presencia de una especie de Paujil (Kwiña) que actualmente es una especie extinguida en Ecuador. En
fotos	Yes
fuente_gps	no tenemos
video	Yes
visita	

Bing en internet © 2018 Microsoft Corporation, © 2018 DigitalGlobe, ©CNES (2018) Distribution Airbus DS

1.5 km scale bar

Coordinates: -75.307040, -0.5048940

716 hidden features

2.1.0 version







MAPPING WORKSHOP, NEMONPARE



MAPPING WAORANI TERRITORY



FEI



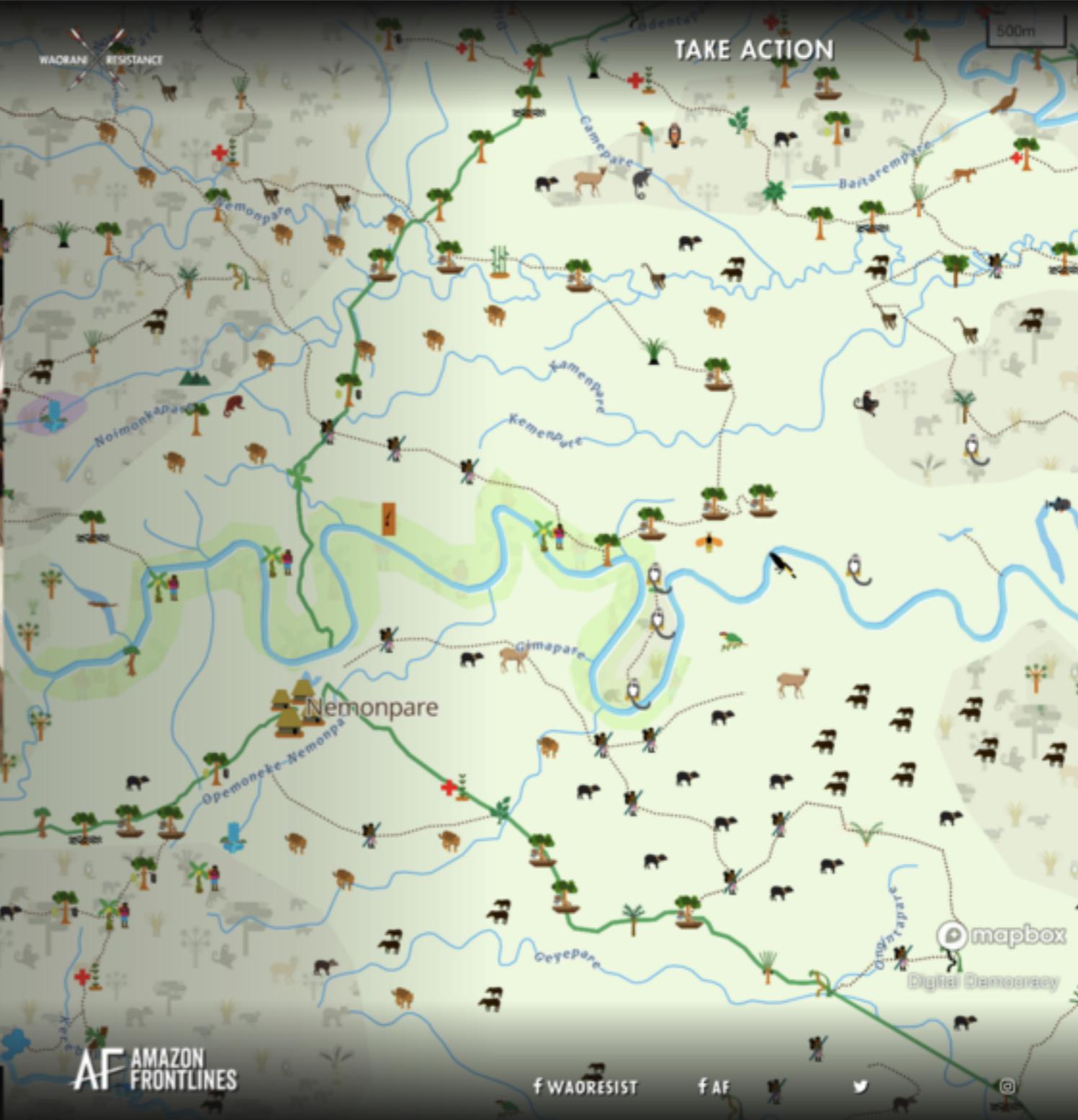


WATCH THE VIDEO

MAPS AND RESISTANCE



The Waorani communities of the Curaray river-basin (now known as Block 22 by the Government) are creating territorial maps of their lands that document the historic and actual uses of their territory, and demonstrate that their homelands are not up for grabs.



AF AMAZON
FRONTLINES

SHARE THE MOVEMENT

ESPAÑOL

f WAORESIST

f AF





WAORANI RESISTANCE IN ECUADOR





DISTRIBUTED
OPENSTREETMAP

PEERMAPS

<https://peermaps.org>

decentralized data preservation

community mapping

social media

21 7d99b45350... nikolai-desktop

58 CHANNELS +

7d # cabal-desktop

cabal-dev

cabal-mobile

default

emoji-test

farts

kappa-core

Neu

ok

test

test2

●

PEERS

- cblgh
- nikolai-desktop
- nikolaiwarner
- noffle
- RichardLitt
- seeder-nikolai-2

#cabal-desktop

8 peers connected

Share Cabal

nikolaiwarner 12:14 PM

yeah! i think it will make sense once i get it organized

Symbol 12:14 PM

Awesome :D

nikolaiwarner 12:14 PM

i think so! i think it will do hot loading ... just need to turn that on in the config

Symbol 12:16 PM

That will be super useful!

cinnamon 12:23 PM

nikolaiwarner: Ok, I won't have time for a day or two anyway! It's great how everyone is pitching in so much

noffle 12:57 PM

still blows my mind that dev chat for cabal is happening on cabal

Smokey 2:31 PM

wow, good work on the last updates!

Smokey 2:31 PM

the multi-cabal support is amazi

Smokey 2:32 PM

noffle: I would say that's a great endorsement for cabal

CHAT

CABAL

Write a message

ment Park



COLLABORATIVE WORKSPACE

Wanderer Thu, Jun 28, 2:37 PM

Looks like you'll have a lot of fun while you're there!

Thanks. Do you have

PUSHPIN

<https://inkandswitch.github.io/pushpin/>

decentralized data preservation

community mapping

social media



SCIENTIFIC DATA NETWORK

DRYAD





COOPERATIVE
HOSTING NETWORK

COBOX

<https://cobox.cloud>

peerfs

Forked from okdistribute/peerfs

Multiwriter peer-to-peer filesystem, built on kappa-core and hyperdrive

JavaScript AGPL-3.0 3 stars 19 forks 0 issues Updated yesterday



cobox-resources

A collection of public research documents, links, papers and archived material from the CoBox project.

0 stars 6 forks 0 issues 1 pull request Updated 3 days ago



kappa-private

Use private-box for sending encrypted messages between kappa-core feeds

JavaScript AGPL-3.0 0 stars 9 forks 0 issues 0 pull requests Updated 7 days ago



cobox-group

a single p2p private group with a mountable drive and personalised feeds

JavaScript AGPL-3.0 0 stars 5 forks 0 issues 0 pull requests Updated 20 days ago

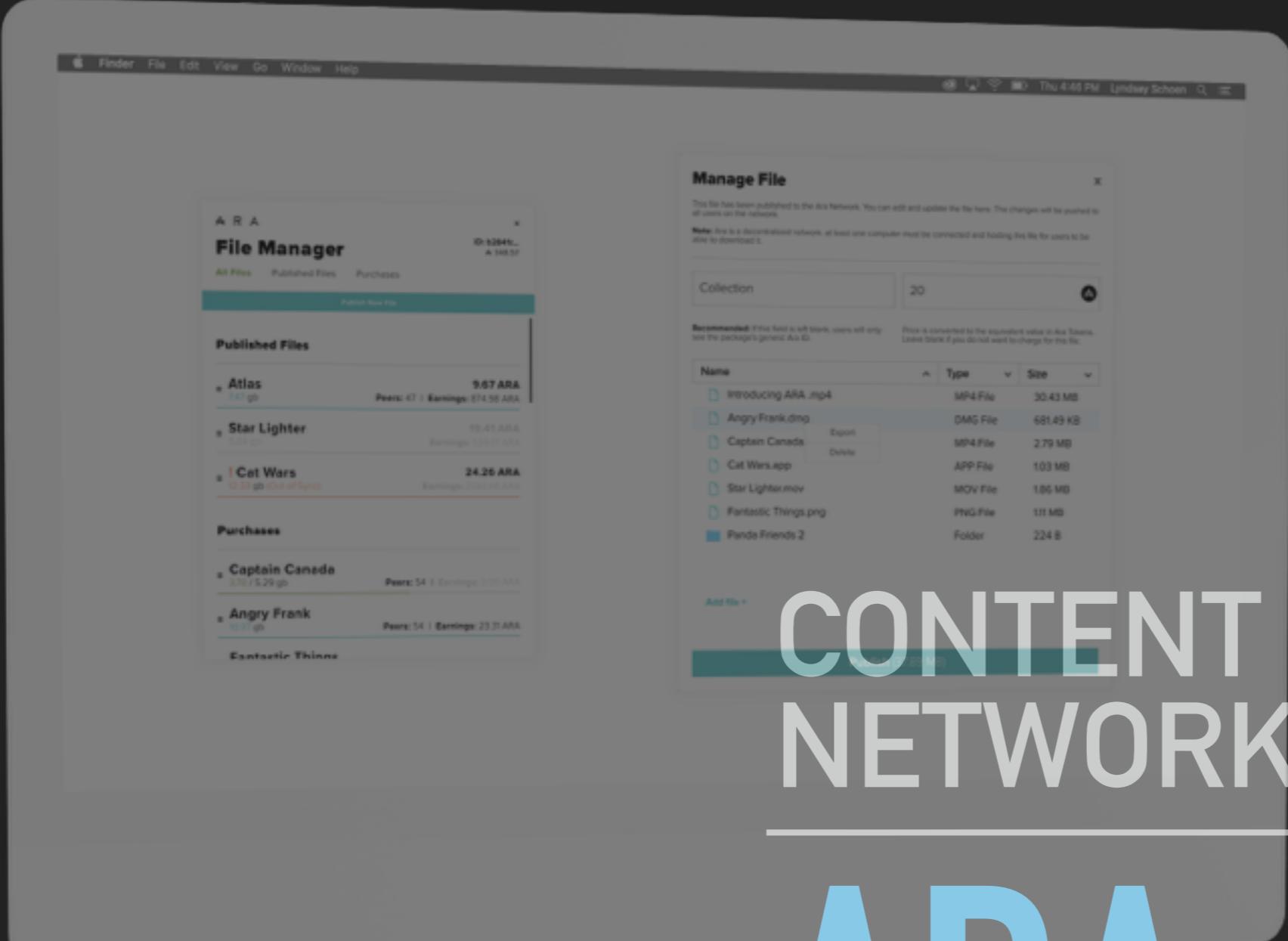


cobox-config

Read and write a local configuration file storing cobox identities and groups

JavaScript AGPL-3.0 0 stars 0 forks 0 issues 0 pull requests Updated 20 days ago





CONTENT REWARD NETWORK

ARA

<http://ara.one>

ara-contracts

Blockchain interactions in Ara modules.



[identity](#) [crypto](#) [filesystem](#) [blockchain](#) [smart](#) [ara](#) [contracts](#)



JavaScript



LGPL-3.0



0



5



7



1

Updated 4 days ago

ara-util

Utility functions for common Ara tasks



[JavaScript](#) [GPL-3.0](#) [0 stars](#) [1 fork](#) [1 open issue](#) Updated 4 days ago

ara-identity-server

A http server to create Ara Identities



identity

manager

ara

ckms



JavaScript



LGPL-3.0



0



0

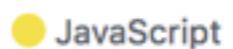


0

Updated 6 days ago

ara-reward-dcdn

Ara-Reward-DCDN



JavaScript



LGPL-3.0



0



1



14



0

Updated 17 days ago

ara-identity

Create and resolve decentralized identity based Ara identifiers.



identity

decentralized

ara

aid

did



WIRELINE

APP TOOLKIT

WIRELINE

wireline.io

user interface

database

storage

index

network

log

protocol

user interface

database

storage

index

network

log

protocol

user interface

database

storage

index

network

log

protocol

user interface

software development toolkit

database

storage

index

network

log

protocol

user interface

software development toolkit

database 2.2

storage

index

network 2.0

log v8

protocol v7

user interface

dat-sdk

hyperdrive v10

random-access-*

hypertrie

hyperswarm

hypercore v8

hypercore-protocol v7

dat 2.0

\$\$\$....?

\$1.6 million since 2013



foundation organization

- U.S. based non-profit, started in 2013, formalized 2017
- \$1.6 million over the past 6 years
- Private foundations: Sloan, Knight



Knight Foundation



Current operating budget: \$295,000

Money left in the bank: \$40,000





dat:// foundation values

- Community first
- Transparently funded
- Open governance



SCIENCE > PROFIT





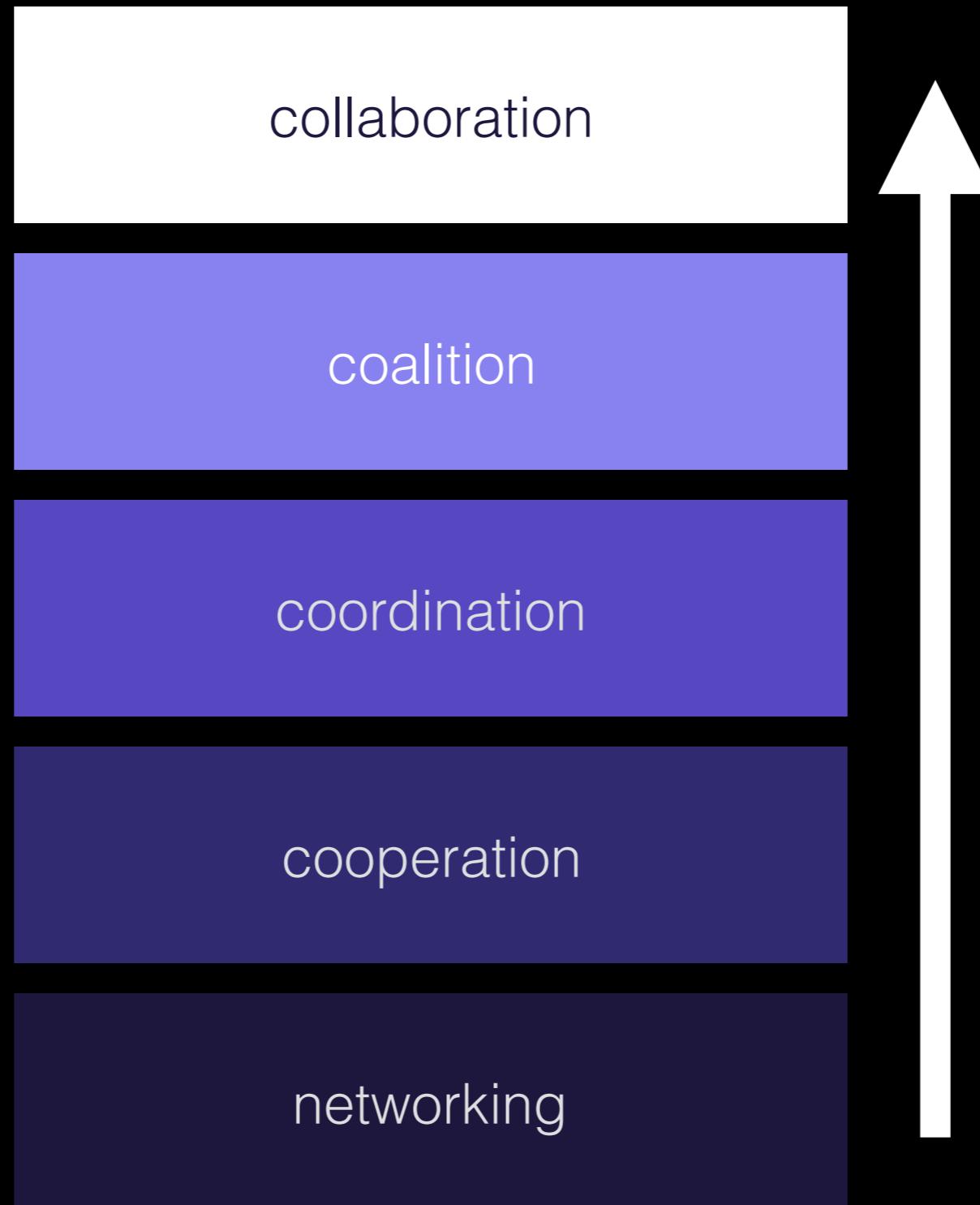
collaboration

coalition

coordination

cooperation

networking





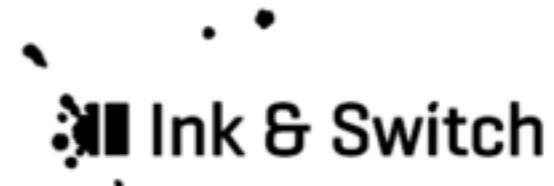
foundation plans

- Membership model



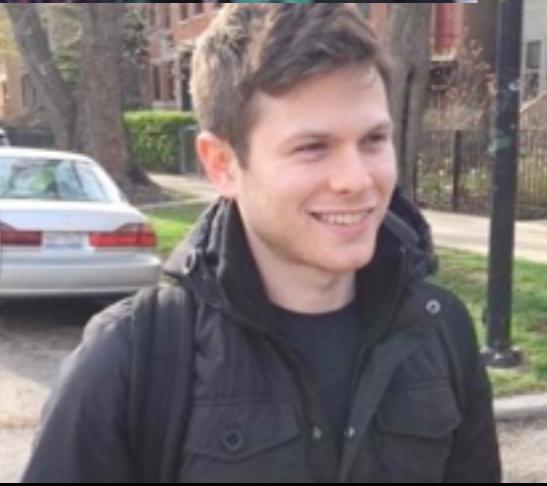
foundation plans

- Membership model
- Summer 2020 community event
- Community grants program



cabal





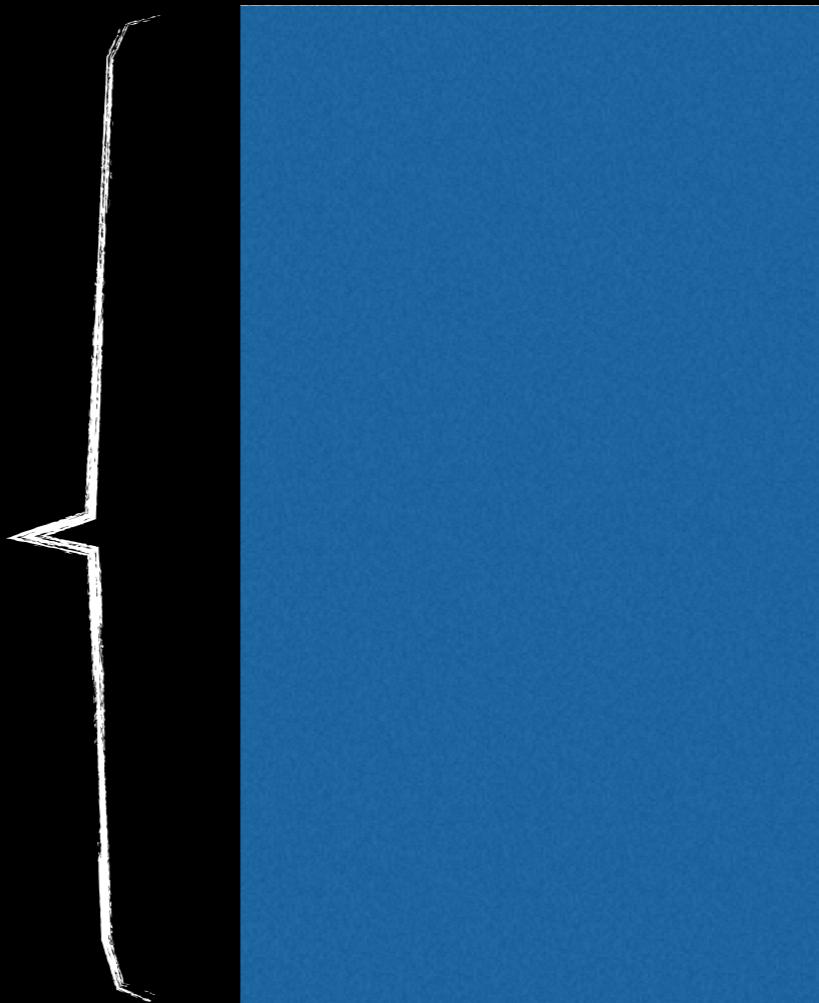
THANK YOU!
@OKDISTRIBUTE



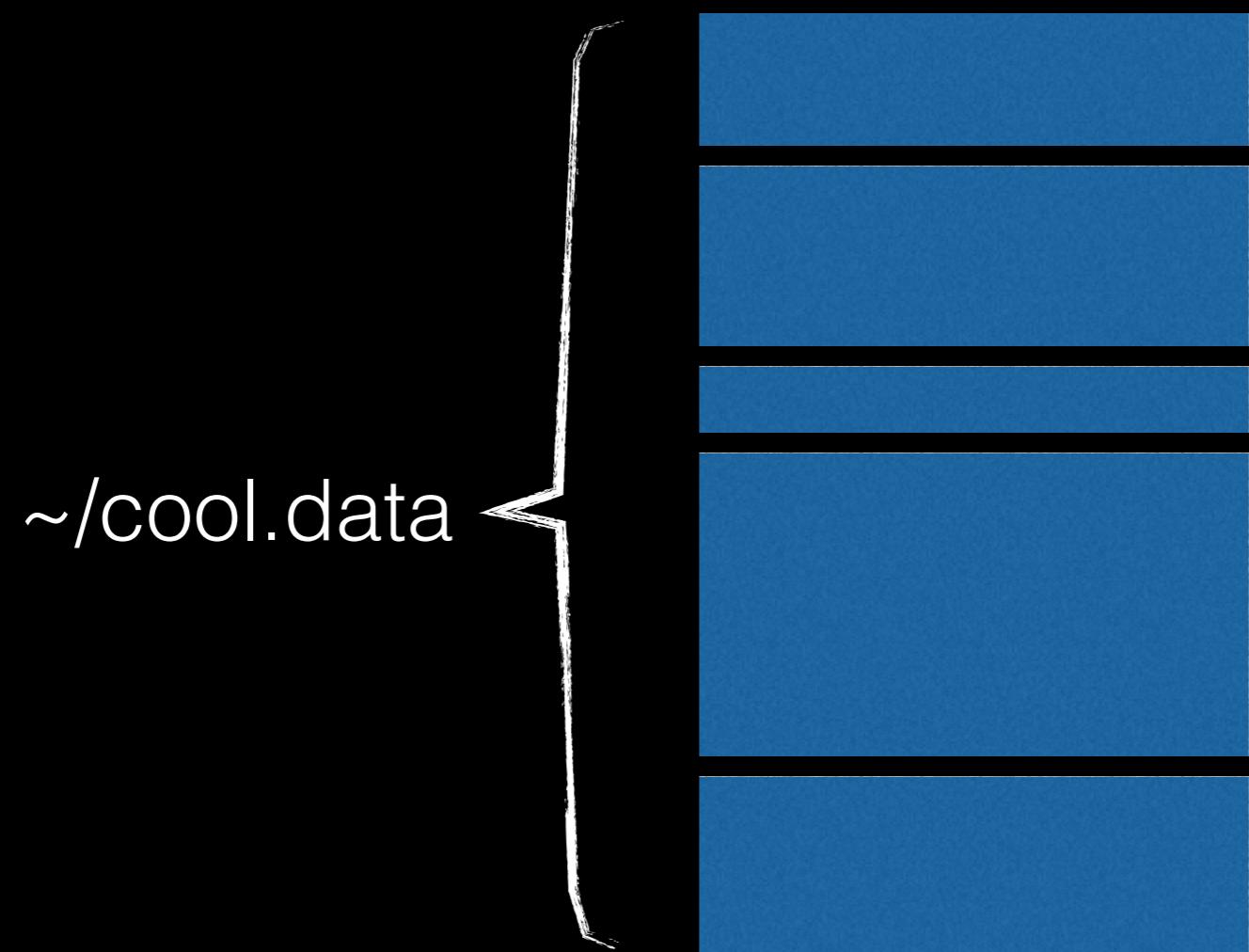
How do we turn append only logs into a file sharing tool?

Take a file

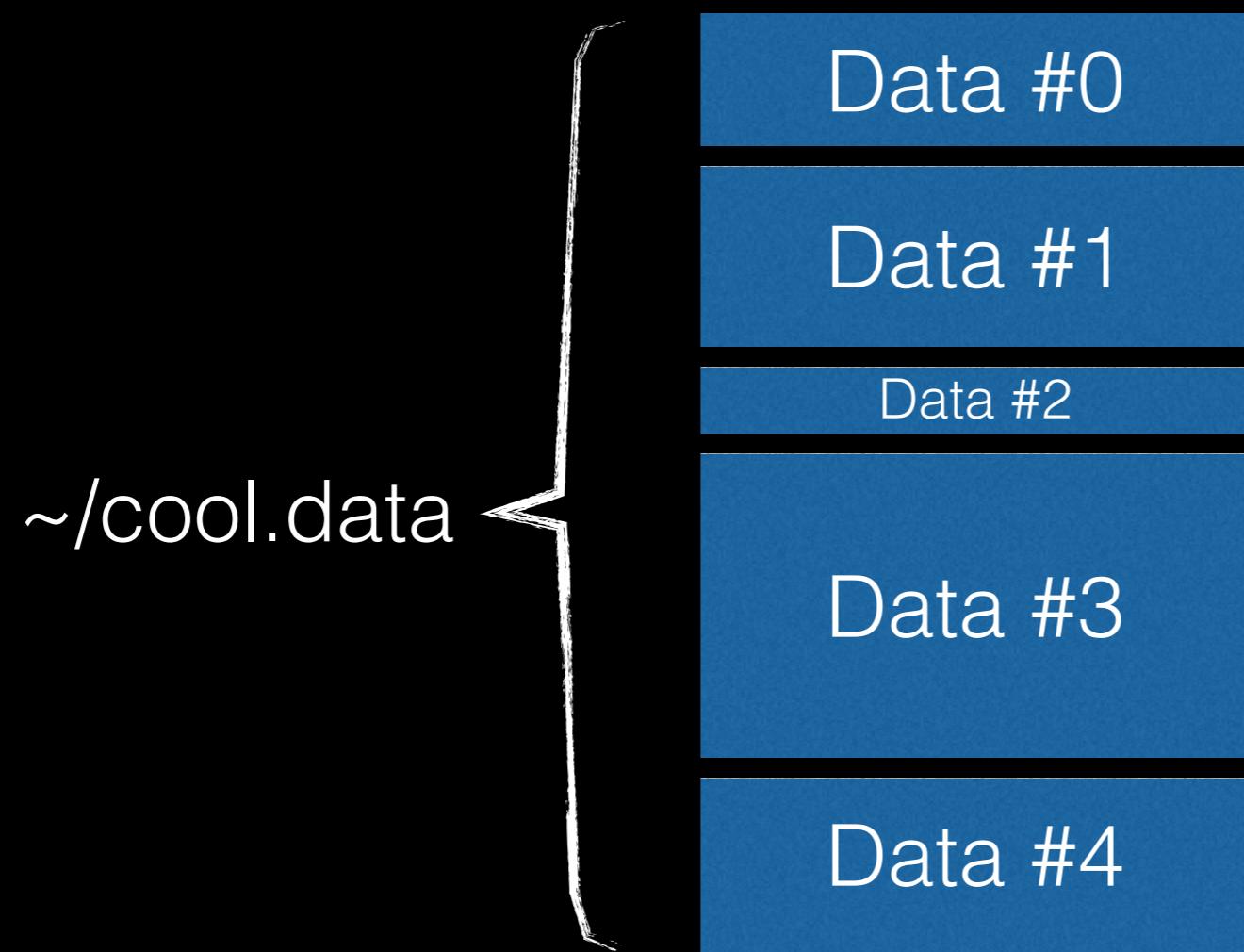
~/cool.data



Cut it into pieces



Insert each piece into the log



Insert filename + file metadata in another log

npm install hyperdrive

software development toolkit

hyperdrive v10

random-access-*

hypertrie

hyperswarm

hypercore v8

hypercore-protocol v7

dat 2.0

software development toolkit

hyperdrive v10

random-access-*

hypertrie

hyperswarm

hypercore v8

hypercore-protocol v7

Data #0

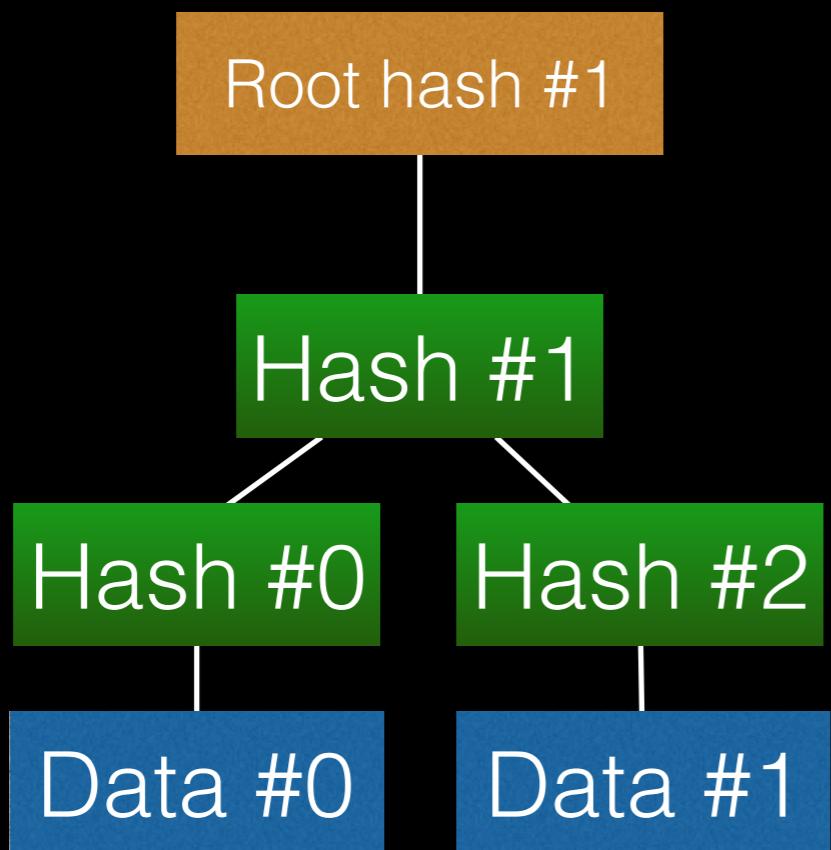
Root hash #0

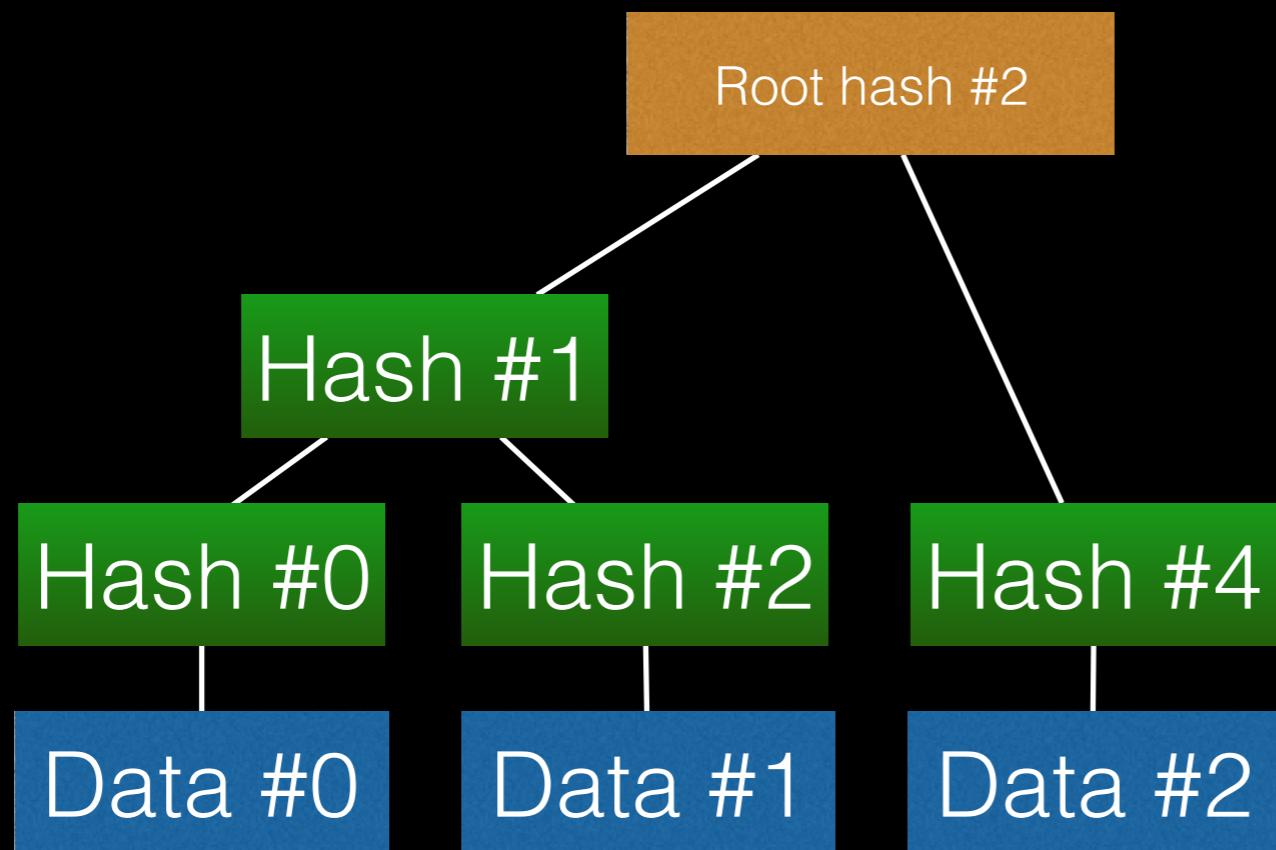


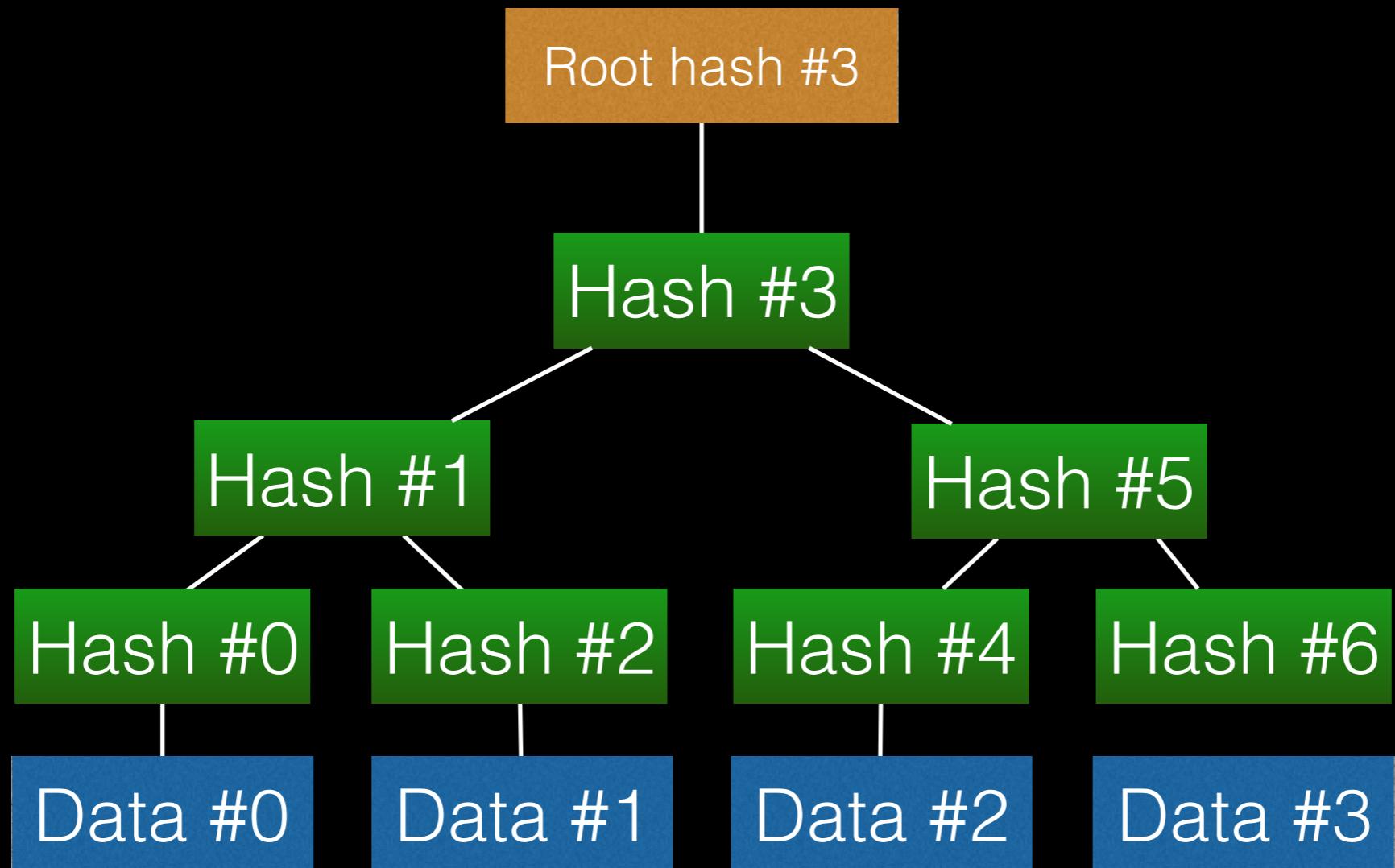
Hash #0



Data #0

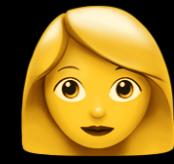






Root hash #3

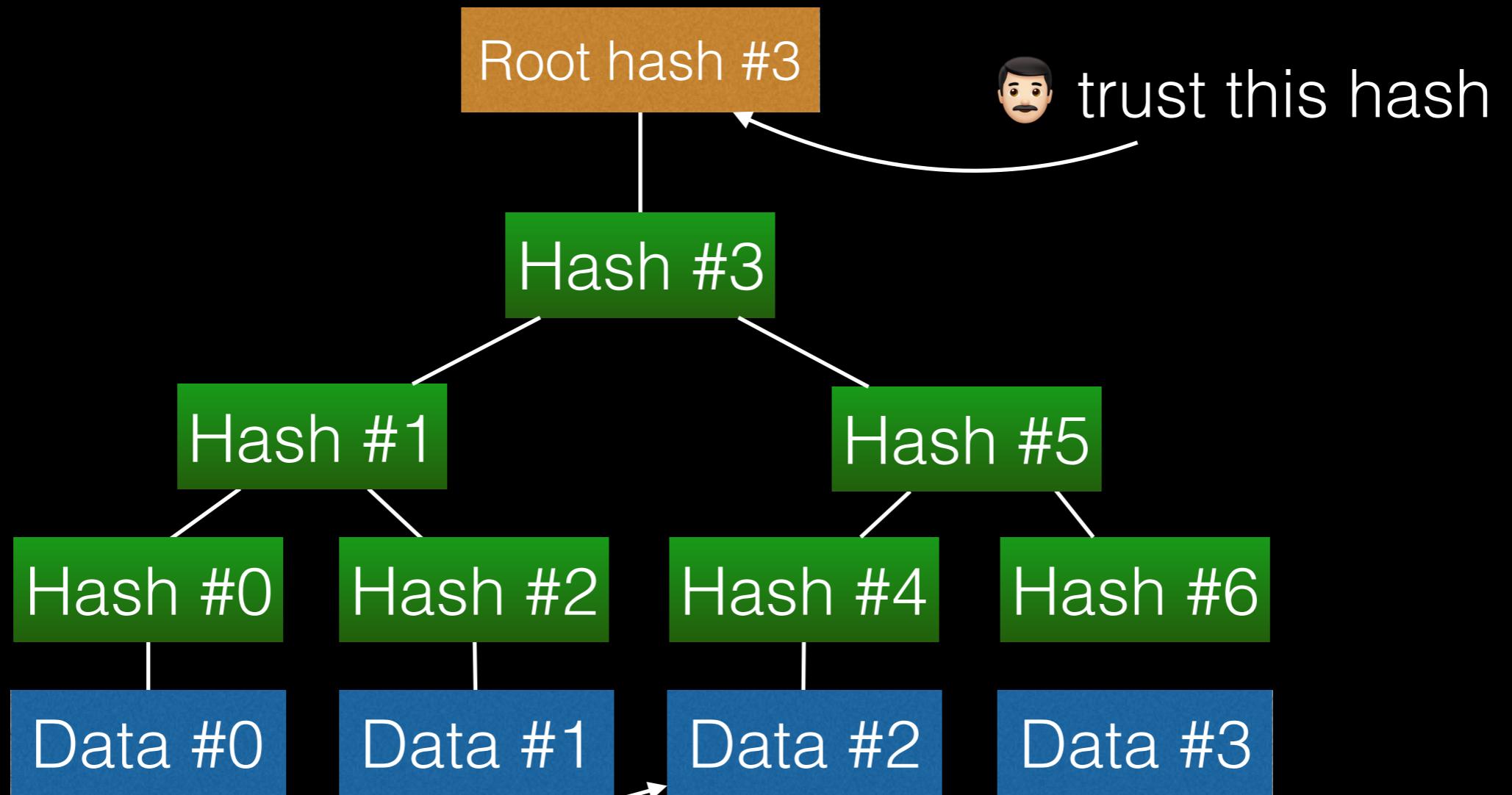
verifies all the data



wants to share

Data #2





👱‍♀️ wants to share this

Root hash #3



trust this hash

Hash #1

Hash #6



needs to share these

Data #2

Root hash #3

Hash #1

Hash #4

Hash #6

Data #2

Root hash #3

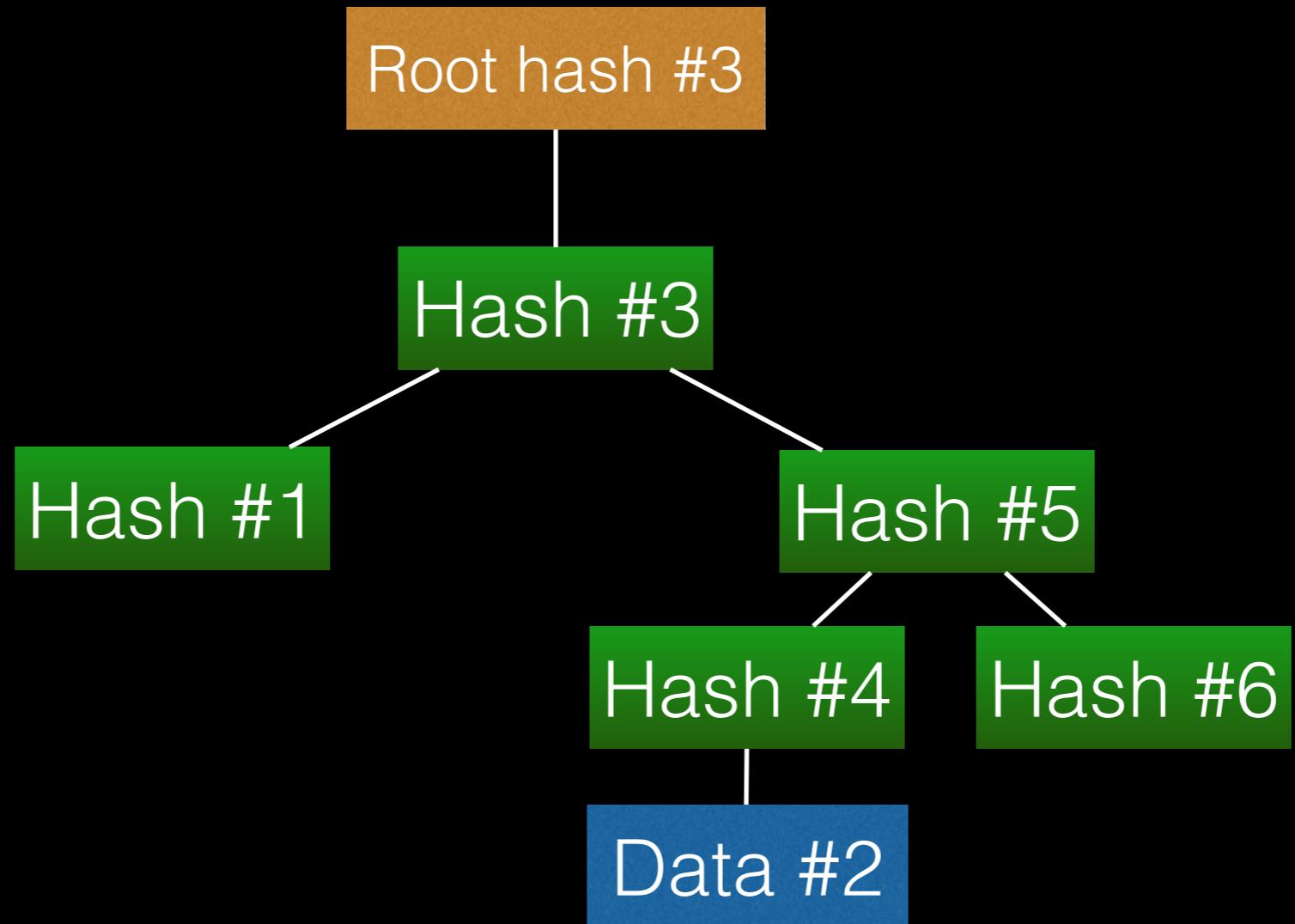
Hash #1

Hash #5

Hash #4

Hash #6

Data #2





checks that Hash #3 match Root hash #3

 only needs to send $O(\log(n))$ hashes to 