

# IPFS & Filecoin: A new Home for Research Data



Molly Mackinlay  
IPFS Project Lead



@momack28



# Who this talk is for

Researchers, scientists, analysts, developers, etc

- who work with large datasets and are looking for better (OSS) tools for storing & sharing that data
- who care about community-owned data, data provenance, data accessibility, and/or long-term data preservation

# 3 stories

The logo for the Environmental Data & Governance Initiative (EDGI) features the letters "EDGI" in a bold, teal, sans-serif font. Below the letters is a thick, solid black horizontal bar.

**Environmental Data  
& Governance Initiative**

**Resiliency**

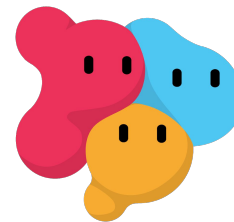
The logo for [\*starling] features the text "[\*starling]" in a black, monospaced font. To the right of the text is a small illustration of three birds in flight, arranged in a V-shape.

a framework for data integrity  
to securely capture, store and verify human history.

USC Shoah  
Foundation

Stanford | ENGINEERING  
Center for Blockchain Research  
Compression Forum

**Verifiability**

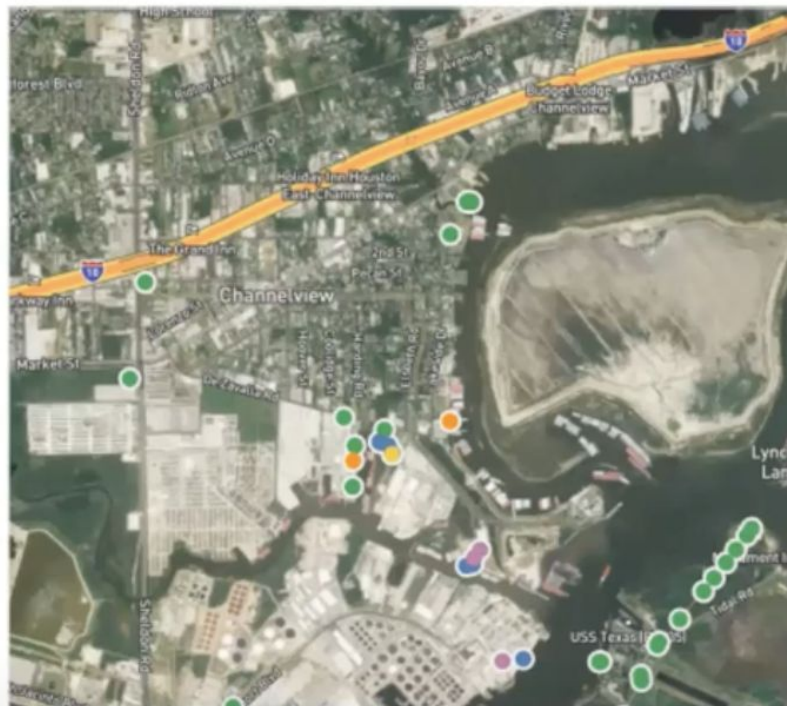


**qri.io**

**Collaboration**

☒ Show Active Monitoring Only

Sampler Type	San Sampler
<input checked="" type="checkbox"/> Automated Gas Chromatograph	TCEQ
<input checked="" type="checkbox"/> Barometric Pressure	Texas Precision Monitoring
<input checked="" type="checkbox"/> CO (High Sensitivity)	Texas Precision Monitoring
<input checked="" type="checkbox"/> Canister (Single)	Texas Precision Monitoring
<input checked="" type="checkbox"/> Carbonyl	Texas Precision Monitoring
<input checked="" type="checkbox"/> Dew Point	Texas Precision Monitoring
<input checked="" type="checkbox"/> NO <sub>2</sub> (Direct)	Texas Precision Monitoring
<input checked="" type="checkbox"/> NO <sub>y</sub> (High Sensitivity)	Texas Precision Monitoring
<input checked="" type="checkbox"/> O <sub>3</sub>	Texas Precision Monitoring
<input checked="" type="checkbox"/> PM Coarse	Texas Precision Monitoring
<input checked="" type="checkbox"/> PM <sub>2.5</sub> (FRM)	Texas Precision Monitoring
<input checked="" type="checkbox"/> PM <sub>2.5</sub> (Speciation)	Texas Precision Monitoring
<input checked="" type="checkbox"/> PM <sub>2.5</sub> (Speciation)	Texas Precision Monitoring
<input checked="" type="checkbox"/> Precipitation	Texas Precision Monitoring
<input checked="" type="checkbox"/> Relative Humidity	Texas Precision Monitoring
<input checked="" type="checkbox"/> SO <sub>2</sub> (High Sensitivity)	Texas Precision Monitoring (formerly Wooten Envir)
<input checked="" type="checkbox"/> Solar Radiation	Texas Precision Monitoring (formerly Wooten Envir)
<input checked="" type="checkbox"/> Temperature (Outdoor)	Texas Precision Monitoring (formerly Wooten Envir)
<input checked="" type="checkbox"/> UV Radiation	Texas Precision Monitoring (formerly Wooten Envir)
<input checked="" type="checkbox"/> Wind	Texas Precision Monitoring (formerly Wooten Envir)

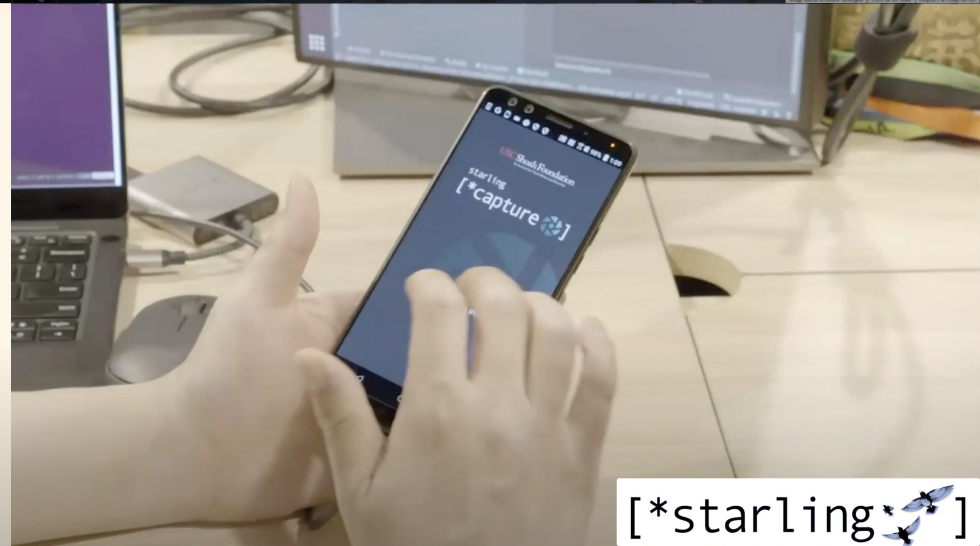
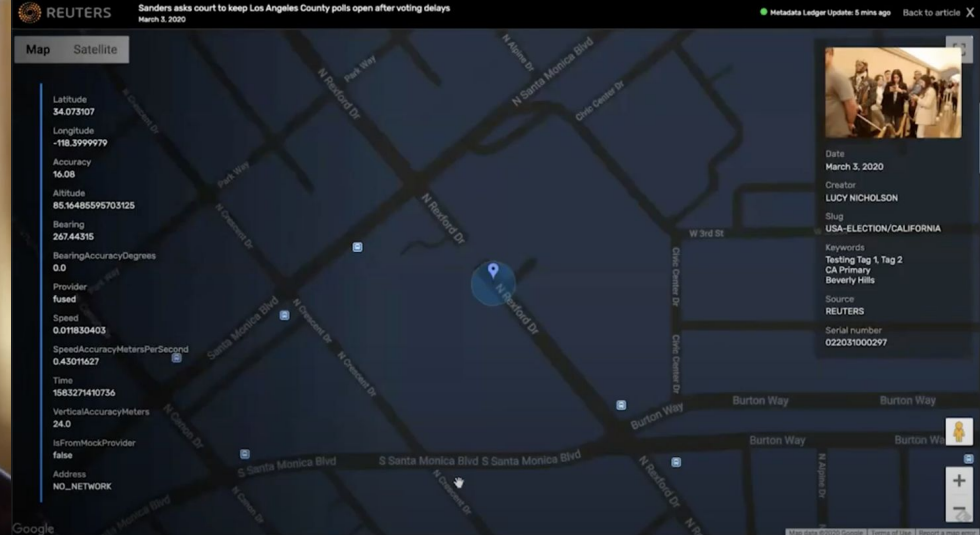


## Recent Benzene Detections

Location Category	Reading Date	Concentration
Division B	3/29/2019 23:30:20	1.73 ppm
Division C	3/29/2019 23:59:41	0.53 ppm
	3/29/2019 23:56:06	0.38 ppm
	3/29/2019 23:54:30	0.03 ppm
	3/29/2019 23:49:36	0.23 ppm
	3/29/2019 23:45:15	0.13 ppm
	3/29/2019 23:44:03	0.23 ppm
	3/29/2019 23:43:08	0.20 ppm
	3/29/2019 23:24:36	3.98 ppm
	3/29/2019 23:14:50	3.08 ppm
Division E	3/29/2019 23:46:31	0.25 ppm
	3/29/2019 23:46:02	0.05 ppm
	3/29/2019 23:45:23	0.86 ppm

*Air quality data collected daily at a few of the sites in Deer Park, near Houston, Texas*





chriswhong/nyc\_capital\_project\_detail\_data\_geom

## NYC Capital Project Detail Data Geometries

Dataset published by chriswhong

preview

issues 0

history

integration

This is a lookup table for associating a geographic point geometry (latitude & longitude) with each project ID in the dataset  
chriswhong/nyc\_capital\_project\_detail\_data.

10 commits

5,207 rows

872.4KB

csv

HEAD 1susnVb Add a readme with contributor attribution

Readme

## NYC Capital Project Detail Data Geometries

This is a lookup table for associating a geographic point geometry (latitude & longitude) with each project ID in the dataset  
chriswhong/nyc\_capital\_project\_detail\_data.

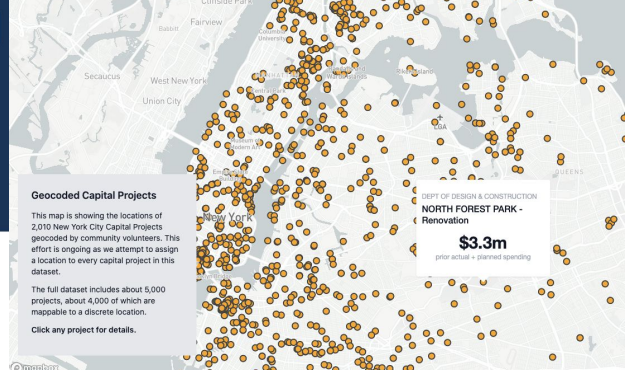
Body Preview

	project_id	project_description	managing_agency_id	managing_agency	ten_year_plan_category	community_boards_serv	borough	bud
1	LB20DUMBO	ADAMS STREET LIBRARY	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	301	BROOKLYN	LB-0101
2	LB017DRBR	DYKER INTERIOR RENOV	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	310	BROOKLYN	LB-D104
3	LB018CTRF	CORTELYOU ROOF	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	314	BROOKLYN	LB-D104
4	LB018FHS	FT HAMILTON ROOF	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	310	BROOKLYN	LB-D104
5	LB018GSMR	GRAVESEND MEETING R	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	315	BROOKLYN	LB-D104
6	LB018KBFS	KINGS BAY FIRE SAFETY	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	315	BROOKLYN	LB-D104
7	LB018KBHC	KINGS BAY BOILER	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	315	BROOKLYN	LB-D104
8	LB019CHEL	Clinton Hill Wiring Upgr	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	302	BROOKLYN	LB-D104
9	LB020CTOP	CORTELYOU OUTDOOR	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	314	BROOKLYN	LB-D104
10	LB020MPMR	MCKINLEY PARK MEETIN	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	310	BROOKLYN	LB-D104
11	LB020SAAD	STONE AVENUE ADA AC	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	316	BROOKLYN	LB-D104
12	LB020SCBM	SPRING CREEK BMS	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	305	BROOKLYN	LB-D104
13	LB021MNR	MACOM ROOF REPLACE	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	303	BROOKLYN	LB-D104
14	LB021PH2	CENTRAL RENOVATION	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	301	BROOKLYN	LB-D104
15	LB0205SWR	SYSTEM-WIDE FUNDING	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	301	BROOKLYN	LB-D104
16	LB016HLRF	HighLawn - Roof Replacer	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	311	BROOKLYN	LB-K001
17	LB016SCR	SPRING CREEK ROOF RE	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	305	BROOKLYN	LB-D104
18	LB018NLOH	NEW LOTS BRANCH OVE	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	305	BROOKLYN	LB-D104
19	LB019PLB	Brower Park Library Fit	38	BROOKLYN PUBLIC LIBR	ESSENTIAL RECONSTRU	308	BROOKLYN	LB-0107

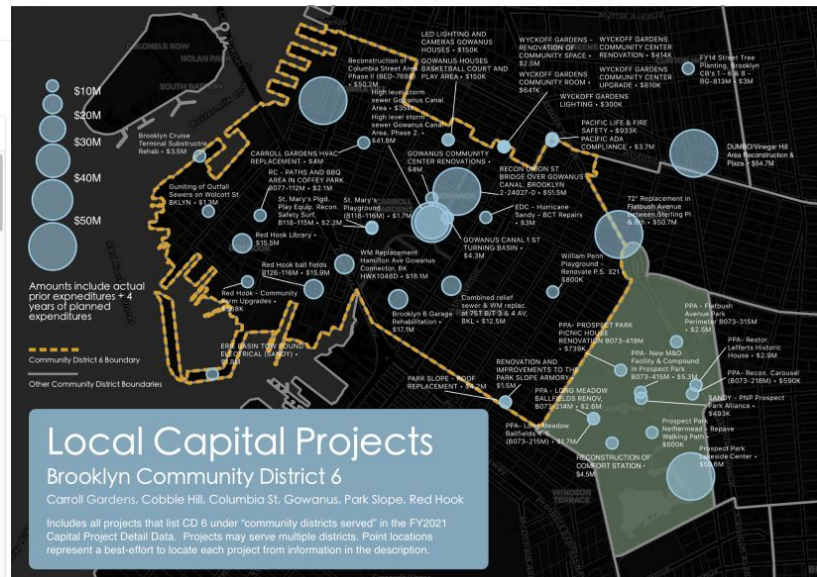
Follow 1 0

Pull

Download



We need volunteers to geocode projects in other parts of the city. If you're interested, let me know and you can claim a district! @qri\_io

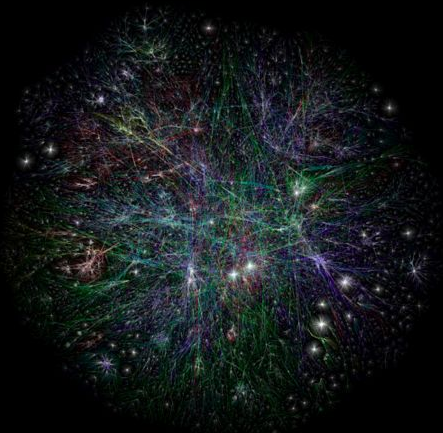


## So how did they do it?

# Web 3



# What is web3?



Internet

*wires, network*



Web 1.0

*read-only  
static*



Web 2.0

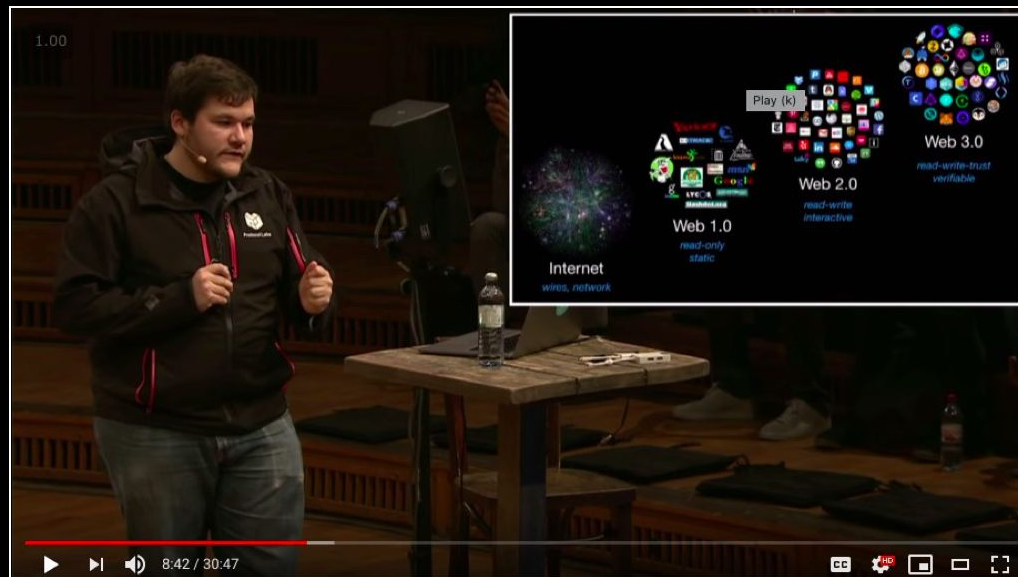
*read-write  
interactive*



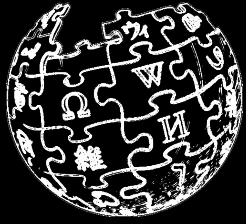
Web 3.0

*read-write-trust  
verifiable*





**What is web3?**



*Web 3.0 is a broad movement and a group of associated technologies aiming to **make the web and the internet more decentralized, verifiable, and secure.***

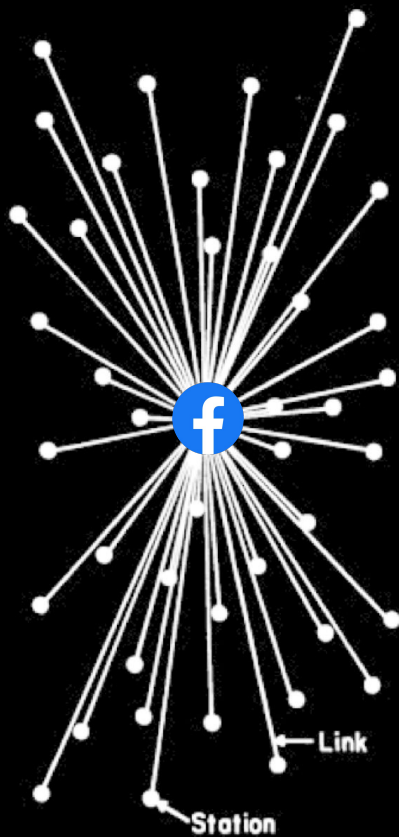
*The goals of Web 3.0 include **(a)** trustless infrastructure; **(b)** removing intermediaries; and **(c)** giving users power and ownership over their data, identity, security, and transactions.*

*The technologies add capabilities and functionality for securely linking data and programs, cryptographic verifiability, transaction processing, P2P connectivity, and trustless interoperability. They also provide decentralized computation and storage, enabling fully autonomous applications (dapps).*

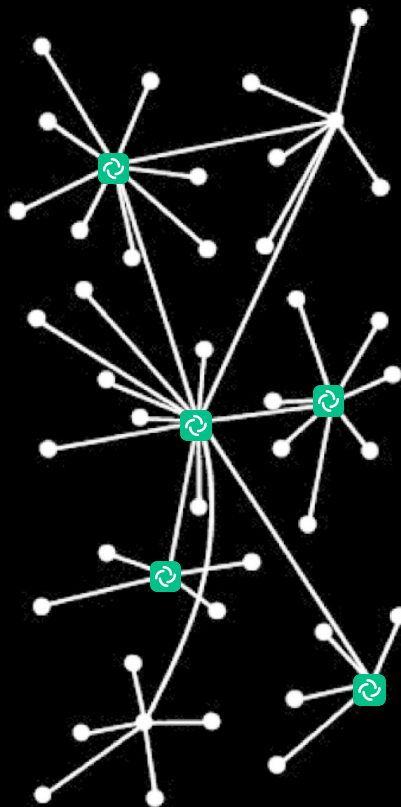
*The movement includes many blockchain and dweb projects, as well as some linked data efforts.*

# *Verifiable*

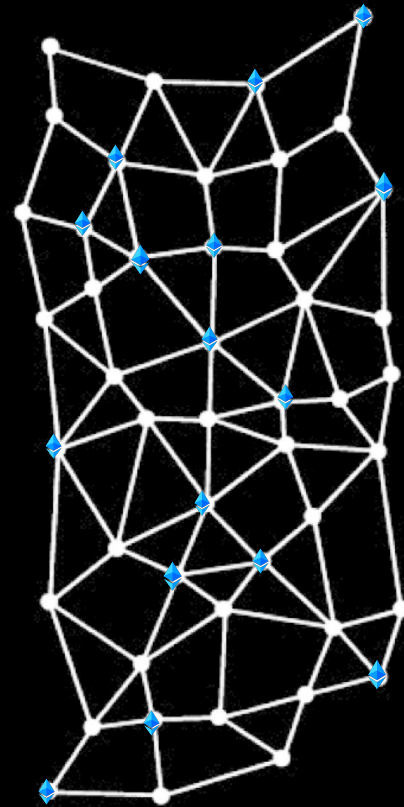
*able to be checked or demonstrated to be true*



**CENTRALIZED  
(A)**



**DECENTRALIZED  
(B)**

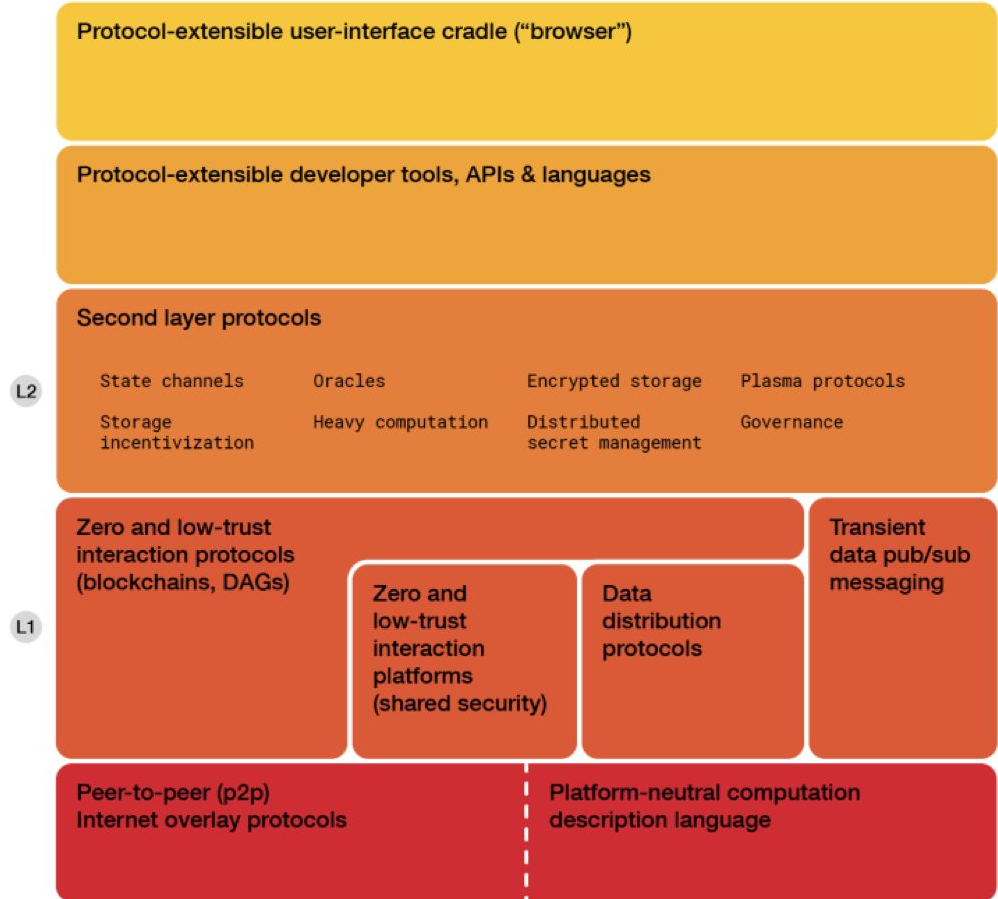


**DISTRIBUTED  
(C)**

**FIG. 1 – Centralized, Decentralized and Distributed Networks**



# Web3 Stack



# Web3 Stack



Protocol-extensible user-interface cradle ("browser")

Protocol-extensible developer tools, APIs & languages

Second layer protocols

L2

State channels

Oracles

Encrypted storage

Plasma prot

Storage  
incentivization

Heavy computation

Distributed  
secret management

Governance



Zero and low-trust  
interaction protocols  
(blockchains, DAGs)

L1



Zero and  
low-trust  
interaction  
platforms  
(shared security)

Data  
distribution  
protocols



Transient  
data pub/sub  
messaging



Peer-to-peer (p2p)  
Internet overlay protocols

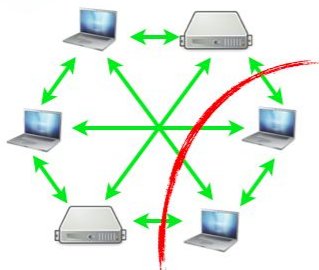


Platform-neutral computation  
description language





***content addressing***



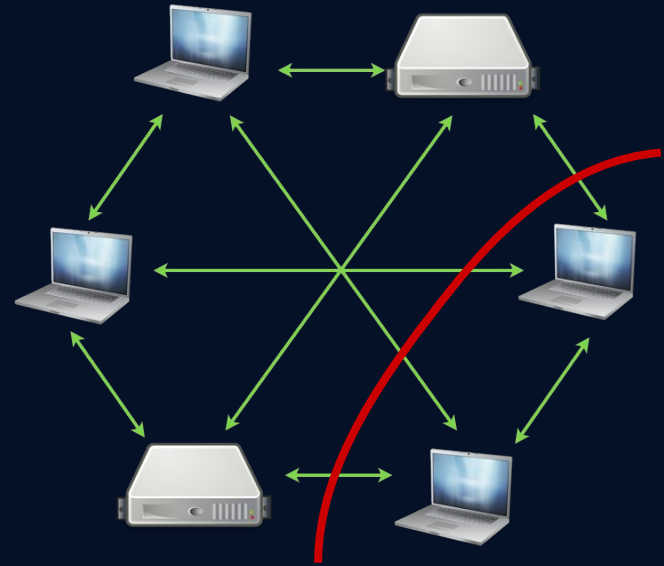
***peer-to-peer file sharing***

*IPFS makes the web work peer-to-peer*

**HTTP**



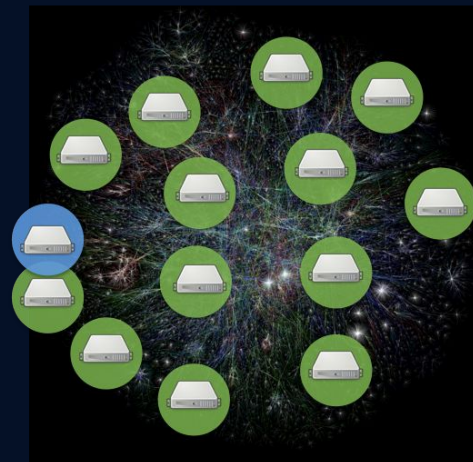
**IPFS**





domain name

/dns/example.com/foo/bar/baz.png



content address

/ipfs/QmW98pJrc6FZ6/foo/bar/baz.png

[About](#)[Install](#)[Docs](#)[Team](#)[Blog](#)[Help](#)

# IPFS powers the Distributed Web

A peer-to-peer hypermedia protocol  
designed to make the web faster, safer, and more open.

[Get started](#)[How it works](#)

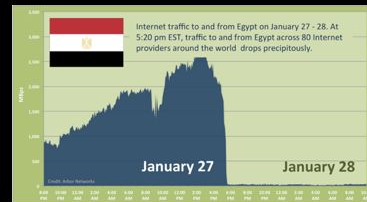


emerging networks

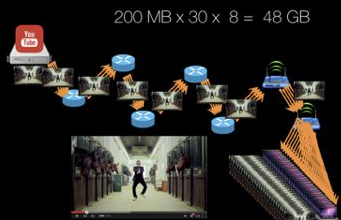
# Problems



# Addresses



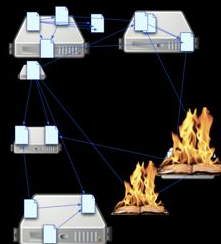
censorship



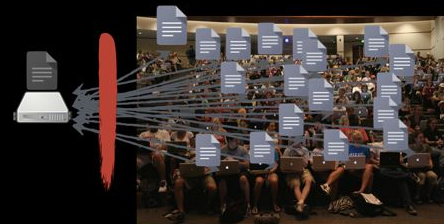
huge inefficiency



bad security model



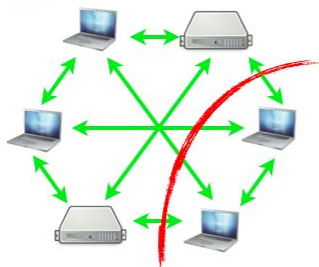
links break



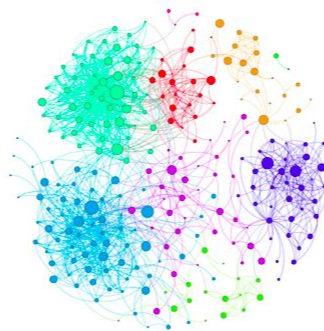
no offline use



***content addressing***



***peer-to-peer file sharing***



***peer-to-peer networking  
DHTs***

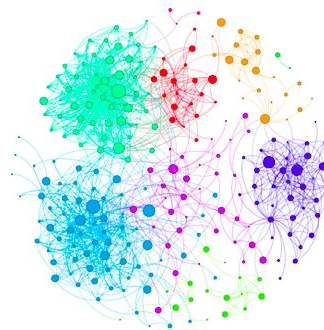


# *libp2p is a modular network stack for p2p protocols*





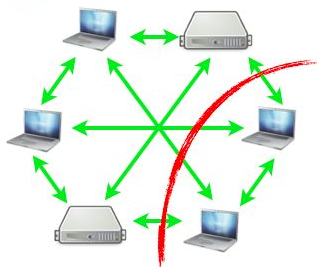
***content addressing***



***peer-to-peer networking  
DHTs***

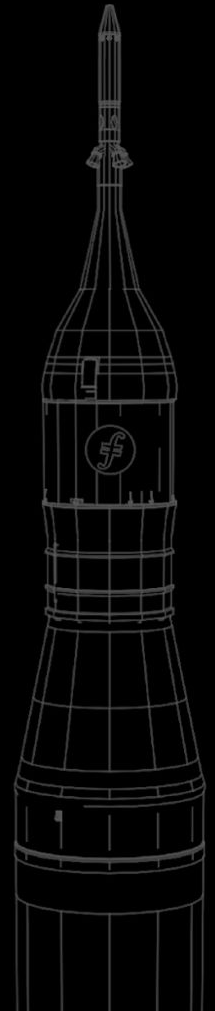


***incentives for data persistence***



***peer-to-peer file sharing***

Filecoin is the  
decentralized storage network  
for web3 and beyond.



# The Mission of Filecoin

to create a **decentralized**, **efficient**, and **robust** foundation for humanity's information.

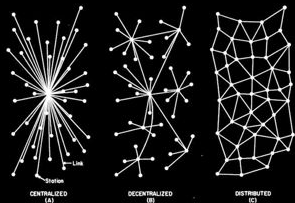
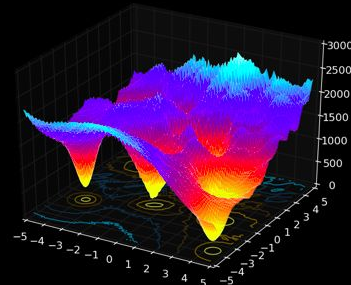
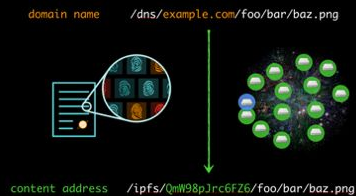


FIG. 1 - Centralized, Decentralized and Distributed Networks



## Filecoin is a...

- Decentralized storage network
- Decentralized storage protocols
- Payment mechanism
- Thriving ecosystem



***Clients want storage***



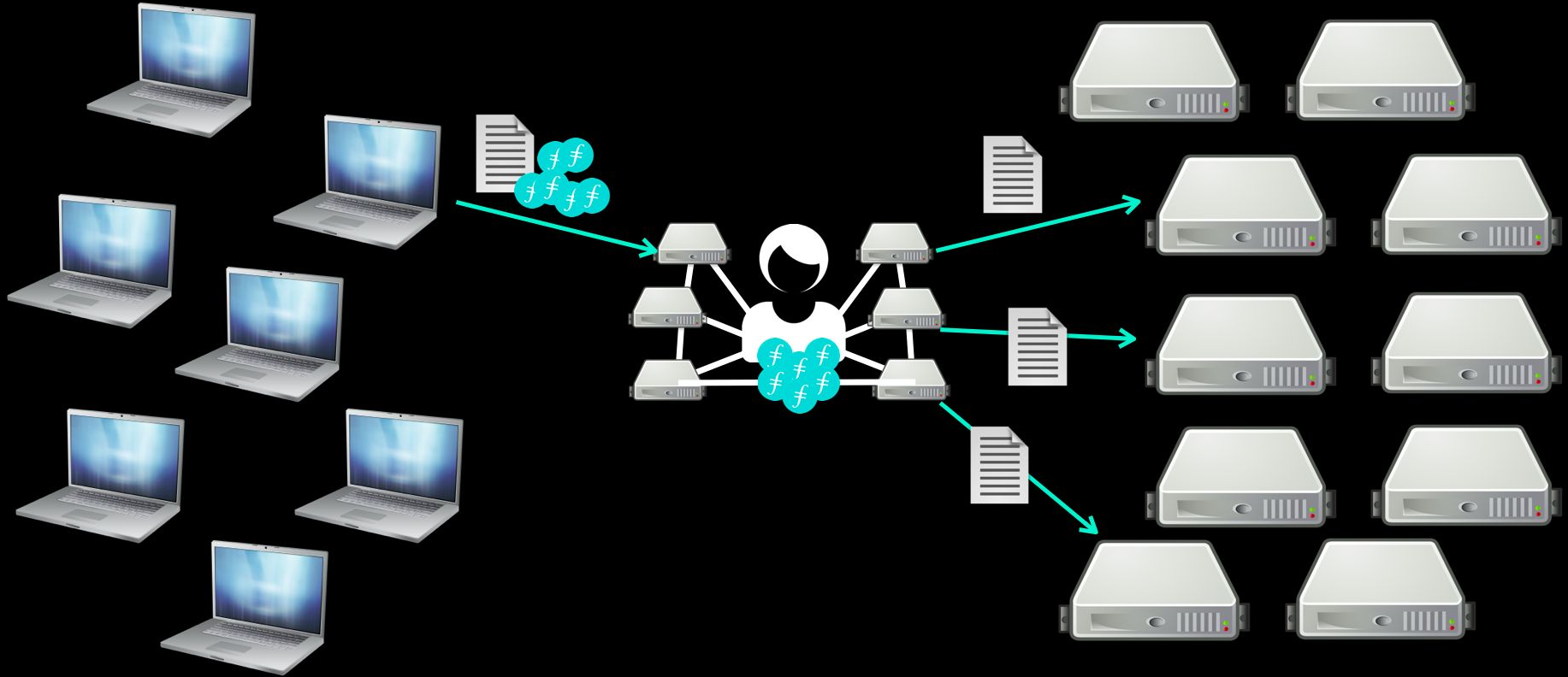
***Network manages***



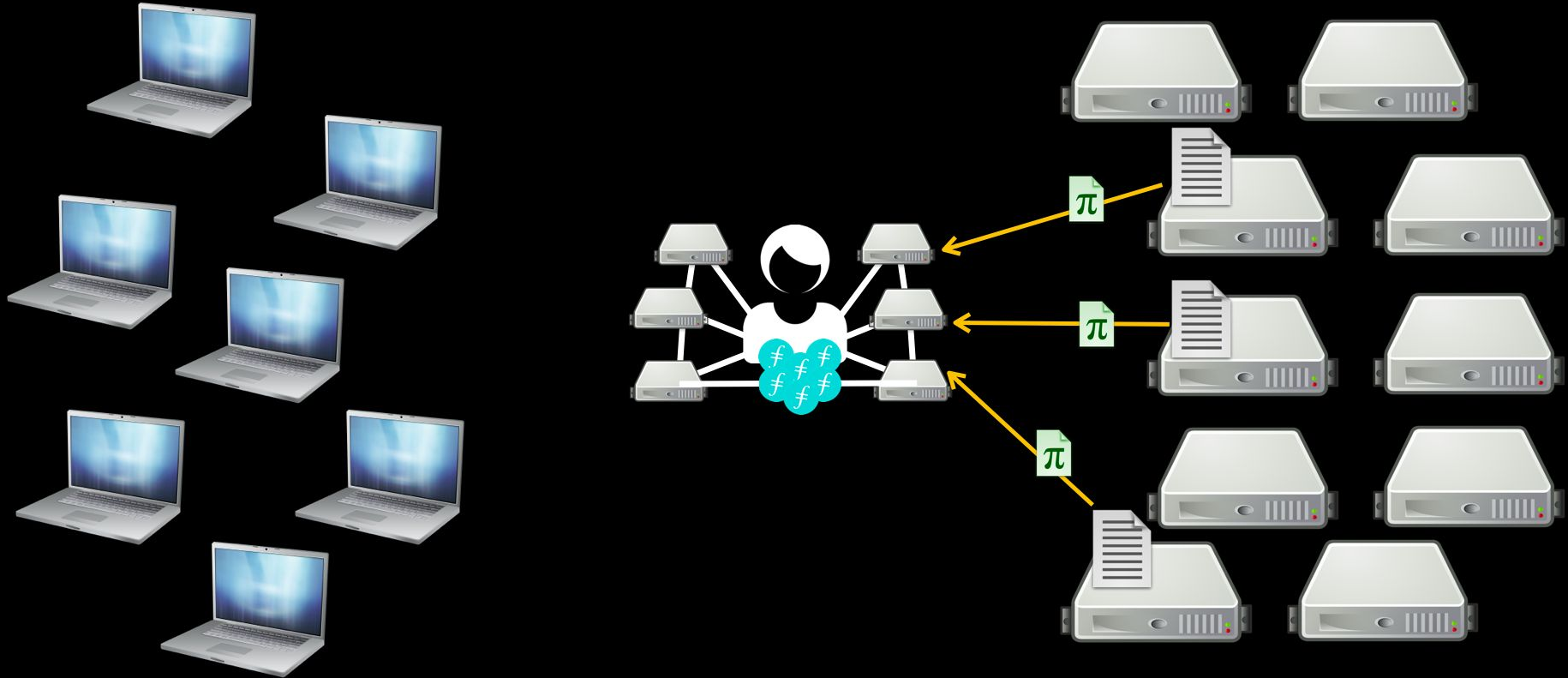
***Miners provide storage***



The **Network** acts as an intermediary between **Clients** and **Miners**

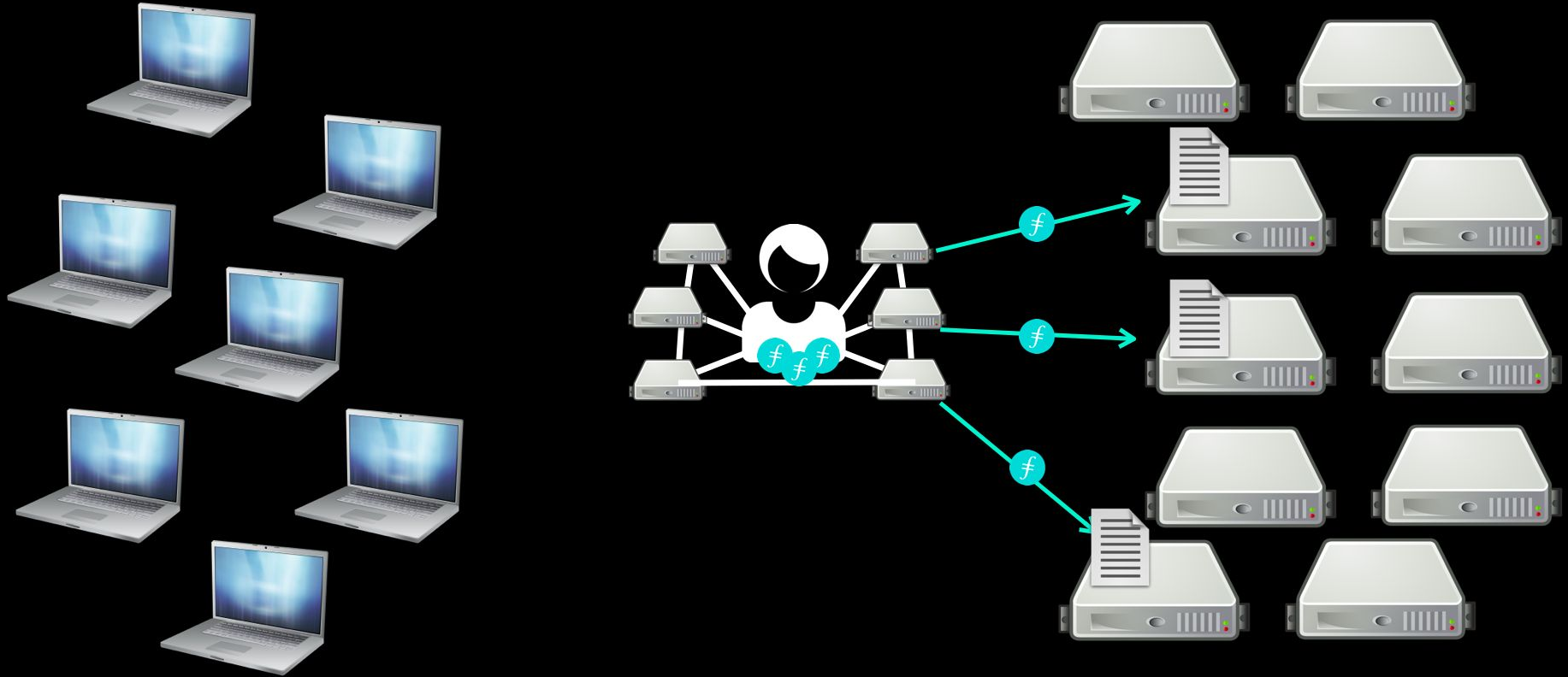


The **Network** checks miners are storing data over time

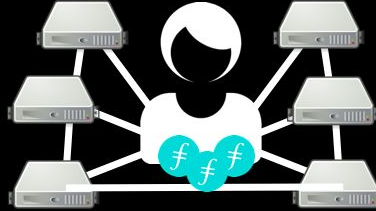


$\pi$  = *Proof-of-Spacetime*

The **Network** pays out to **Miners** over time



**Clients & Miners** must be able to trust the **Network** to preserve data







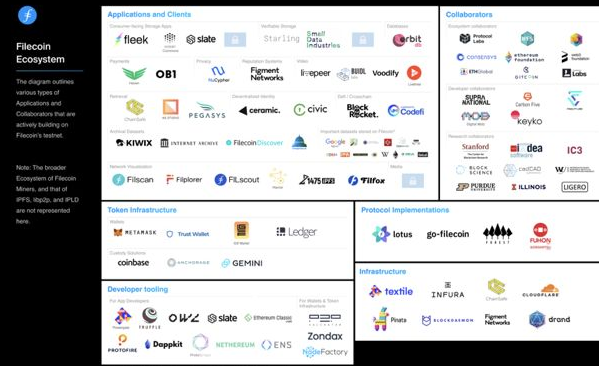
# Filecoin Network Roles



## Clients want storage



## Miners provide storage



## Core Developers & Ecosystem Partners



**Filecoin:** Decentralized Storage Network



**IPFS:** Distributed Web Protocol



**IPLD:** authenticated data model & formats



**libp2p:** modular p2p networking library



**Multiformats:** future-proofing & upgradability

# IPFS and Filecoin Ecosystems:

## Applications and Clients

**Consumer-facing Storage Apps**

- fleek
- ocean community
- slate
- CoinSummer
- Block Rocket
- Starling
- Small Data Industries
- orbit db

**Payments**

- Haven
- OBI
- NuCypher

**Privacy**

- Figment Networks

**Reputation Systems**

- livepeer
- BUIDL labs
- Voodify
- Livetree

**Video**

**Retrieval**

- ChainSafe
- AE STUDIO
- PEGASYS

**Decentralized Identity**

- ceramic.
- civic

**Defi / Crosschain**

- Huobi
- Codefi

**Archival Datasets**

- KIWIX
- INTERNET ARCHIVE
- FilecoinDiscover
- Important datasets stored on Filecoin\*
- W
- YEM
- IPMA
- OpenStreetMap
- OpenSea
- MEWA
- Bitfury

**Network Visualization**

- Filfox
- Filscan
- FILScout
- Filplorer
- 1475 IPFS
- FilStats
- Hactar

## Token Infrastructure

**Wallets**

- METAMASK
- Trust Wallet
- Gif Wallet
- Ledger
- For Wallets & Token Infrastructure
- Zondax
- NodeFactory

**Custody Solutions**

- coinbase
- ANCHORAGE
- GEMINI

## Developer tooling

**For App Developers**

- Powergate
- TRUFFLE
- OWZ
- Ethereum Classic LABS
- fission
- ceramic.
- PROTOFIRE
- Dappkit
- NETHEREUM
- ENS
- ProtoSchool
- slate

## Collaborators

**Ecosystem collaborators**

- Protocol Labs
- CONSENSYS
- ETHGlobal
- IPFS
- ethereum foundation
- GITCOIN
- web3 foundation
- Labs

**Developer collaborators**

- SUPRA NATIONAL
- Carbon Five
- keyko
- Digital Mob

**Research collaborators**

- Stanford
- The Center for Blockchain Research
- cadCAD
- ILLINOIS
- LIGERO
- IC3
- idea software

**Data**

- orbit
- qri.io
- arbore
- filelek
- Catena
- BLOCKCHAIN EXPLORER

**Identity**

- civic
- ION
- UNSTOPPABLE DOMAINS
- NOVIOX
- ZINC
- uport
- handshake

**Persistence**

- INFURA
- eternum
- textile
- TEMPORAL

**Marketplace**

- OpenBazaar
- ORIGIN
- Bounty0x
- Ethlance

**NFT**

- dlux
- DIGITAL ART CHAIN
- Glossy
- mokens
- Decentraland

**Content**

- D.tube
- EVERIPEDIA
- ALEXANDRIA
- Matters
- PEERGOS
- dlive
- PeerPad
- Partyshare

**Other**

- ipwb
- Dappkit
- kauri
- Simple As Water
- MOIB
- SprtsHub
- adXchain
- MONITOR CHAIN
- IPSE
- IKU
- stake.fish
- W IPFS
- Cyber
- FILESTO
- tallylab
- edCha
- ROBONOMICS
- fission
- IPFS DATA
- actyx
- EDGI
- Environmental Data & Governance Initiative
- DAppN

**Social Media**

- 3 BOX
- Identifi
- berty
- KARMA
- Orbit
- AKASHA
- Indorse
- Peepeth

**Integrations & Collaborations**

- NETFLIX
- NixOS
- Microsoft Azure
- CLOUDFLARE
- KIVIX
- Guix

**Prediction**

- AUGUR
- CryptoBets
- MÖBIUS 2D
- VIRTUE POKER

**Governance**

- GovBlocks
- Democracy Earth
- ARAGON ONE

**Exchange**

- Dether
- Swap .online

**Finance**

- coinomi
- REQUEST NETWORK
- RAVENCOIN
- Bloom
- colu.
- SETTLE
- MyEtherWallet
- kyber network
- Uniswap
- MARKETPROTOCOL

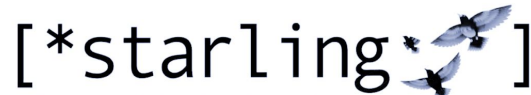


# How these stories become possible on web3

The logo for the Environmental Data & Governance Initiative (EDGI) features the letters "EDGI" in a bold, teal, sans-serif font. A thick black horizontal bar is positioned directly beneath the text.

**Environmental Data  
& Governance Initiative**

**Resiliency**

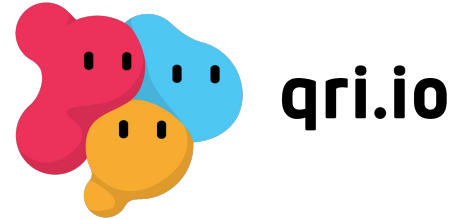
The logo for [\*starling] features the text "[\*starling]" in a black, monospaced font. To the right of the closing bracket is a small illustration of three birds in flight, arranged in a V-shape.

a framework for data integrity  
to securely capture, store and verify human history.

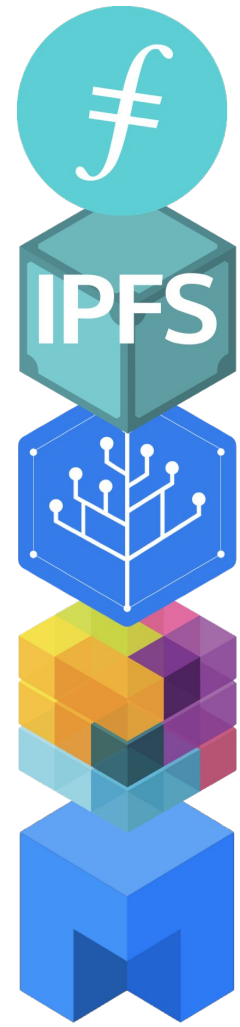
USC Shoah  
Foundation

Stanford ENGINEERING  
Center for Blockchain Research  
Compression Forum

**Verifiability**

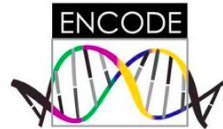


**Collaboration**



# Archiving and Distributing Precious Data

from these organizations and many more





# 3 ways it's getting easier in 2021

## Filecoin Plus

*Incentivizing Useful Storage on Filecoin*



filecoin

Powergate v2.0.0 is here!

a day ago • 9 min read

★ Featured



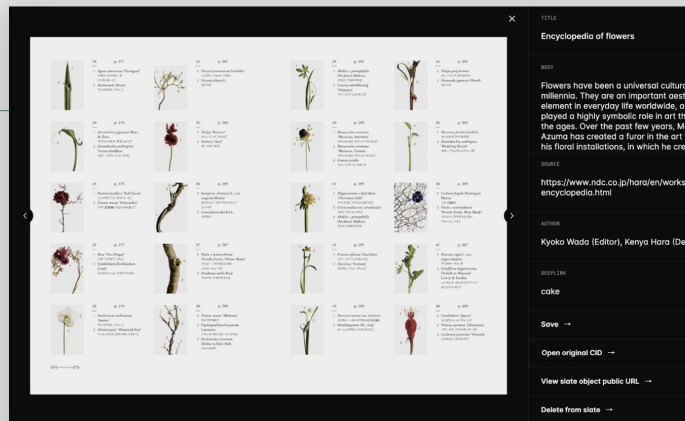
slate

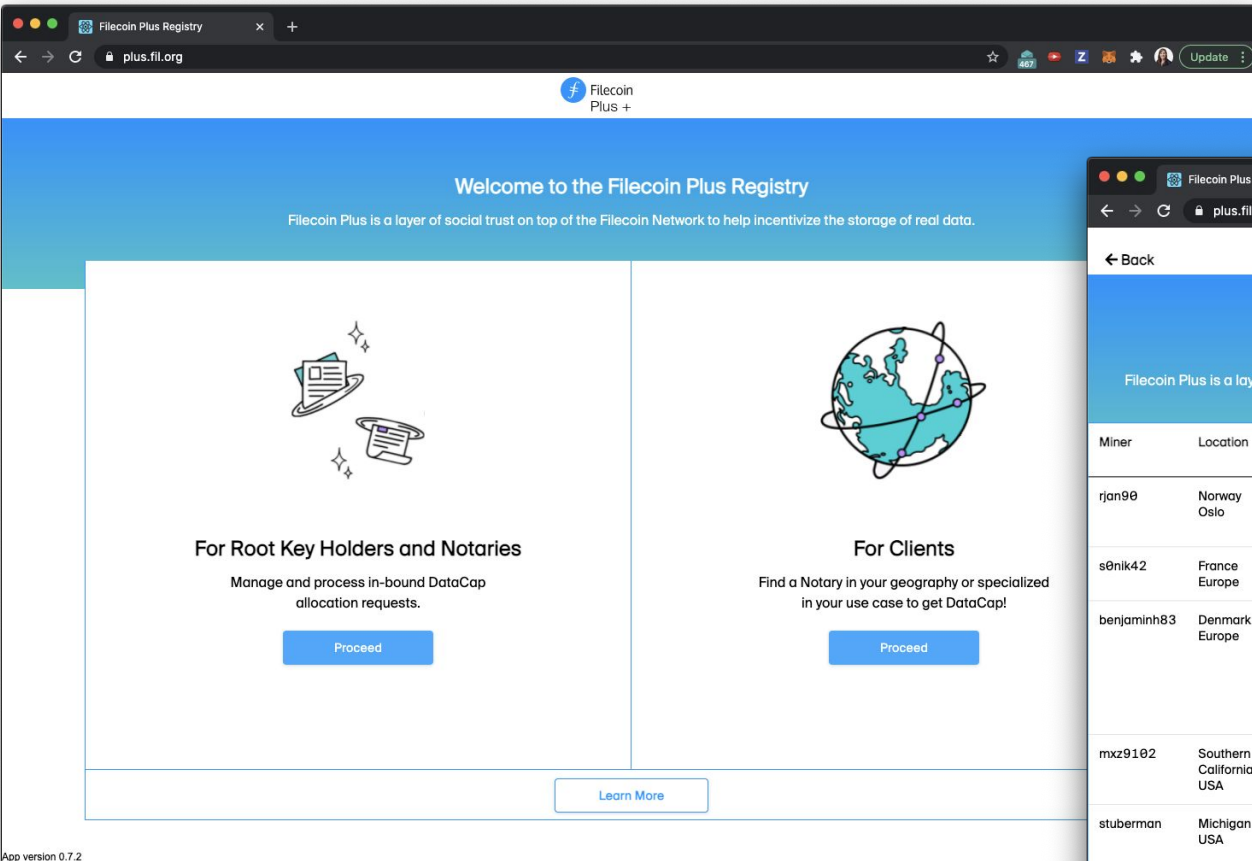
## Store, annotate, cite, and link your files

Slate is a new home for information  
that matters to you.

### Books

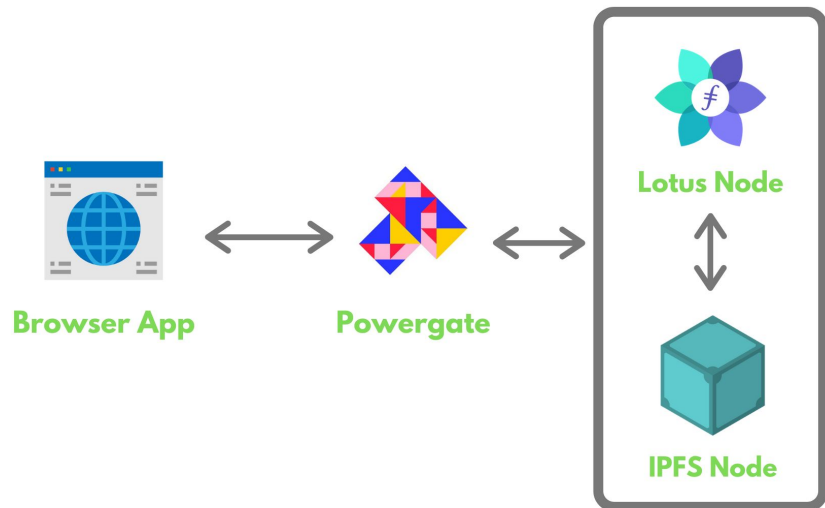
- White papers
- Writing notes
- Films
- Videos
- Photographs
- Illustrations
- Data graphs
- Maps
- Designs
- Music albums
- Podcasts
- Games
- Code





The screenshot shows the Filecoin Plus Registry miners page. The browser address bar displays 'plus.fil.org/miners'. The page has a blue header with the Filecoin Plus logo. Below the header, a blue banner contains the text 'Welcome to the Filecoin Plus Registry' and 'Filecoin Plus is a layer of social trust on top of the Filecoin Network to help incentivize the storage of real data.' The main content area is a table listing miners. The table has six columns: Miner, Location, Miner ID, Contact Info, Special Offering, and Features. The 'Special Offering' column is highlighted with a green border. The table lists seven miners: rjan90, s0nik42, benjaminh83, mxz9102, stuberman, and Shekelme. The 'Special Offering' column contains text such as '0 FIL for verified data', '0 FIL for early FIL+ Clients', and '0.000000000001 FIL GiB/epoch'. The 'Features' column contains detailed information about each miner's setup, including location, network, and storage capabilities. At the bottom of the table, there is a pagination bar showing '1' and '2'.

Miner	Location	Miner ID	Contact Info	Special Offering	Features
rjan90	Norway Oslo	f023467	Slack <a href="#">Website</a>	0 FIL for verified data	Servers housed in a tier 3 datacenter in Oslo, Norway. Sub millisecond latency to the most densely populated areas of Norway.
s0nik42	France Europe	f010479	Slack <a href="#">Email</a>	0 FIL for early FIL+ Clients	
benjaminh83	Denmark Europe	f02576	Slack	0 FIL for verified data	- 1 Gbit Fiber, Static IP behind NextGen Firewall, No VPN. - 1.5TiB daily sealing power - Running on 75-95% renewable energy - Used energy is recirculated for heating - Contributor to Calibration/SR1/SR2 and multiple slack community champion
mxz9102	Southern California USA	f09848	Slack	0 FIL for verified data	- 1 Gbit Fiber, Static IP - 2TiB daily sealing power
stuberman	Michigan USA	f09848	Slack	0.000000000001 FIL GiB/epoch	- 1 Gbit Fiber - UPS protected system - Min size 2 GiB - Max size 32 GiB - 160GiB daily sealing power
Shekelme	Novosibirsk Russia	f08240, f0791817, f080480	Slack <a href="#">Email</a>	0.000000002 FIL GiB/epoch (price may change subject to basefee)	- 100 Mbit PON (can change to 1 Gbit if necessary) - UPS protected operation - Min size 100 MiB - Max size 32 GiB - Raidz2 array to store your data.



```
> pow --help
A client for storage and retrieval of filecoin data

Usage:
  filecoin [command]

Available Commands:
  asks      Provides commands to view asks data
  config    Provides commands to manage storage configuration
  deal      Interactive storage deal
  deals     Provides commands to manage storage deals
  ffs       Provides commands to manage ffs
  health    Display the node health status
  help      Help about any command
  init      Initializes a config file with the provided values or defaults
  miners    Provides commands to view miners data
  net       Provides commands related to peers and network
  reputation Provides commands to view miner reputation data
  slashing  Provides commands to view slashing data
  wallet    Provides commands about filecoin wallets

Flags:
  --config string      config file (default is $HOME/.powergate.yaml)
  -h, --help           help for filecoin
  --serverAddress string address of the filecoin service api (default "/ip4/127.0.0.1/tcp/5002")

Use "filecoin [command] --help" for more information about a command.
~/Textile/filecoin/powergate/docker master
>
```

Powergate CLI





# We'd love to hear from you!

- Tell us about your top pain-points!
- Suggest additional tools & capabilities
- Get involved in pilots and hackathons applying these new tools to your research/datasets/work
- Help build new applications on this stack!





# Thanks! ▲

*Get in touch:*  *@momack28*

# Today

- ♦ Who this talk is for
- ♦ What is web3 or “the distributed web”
- ♦ What are IPFS & Filecoin
- ♦ 3 reasons IPFS & Filecoin may be a great fit for you
- ♦ Tools to get started