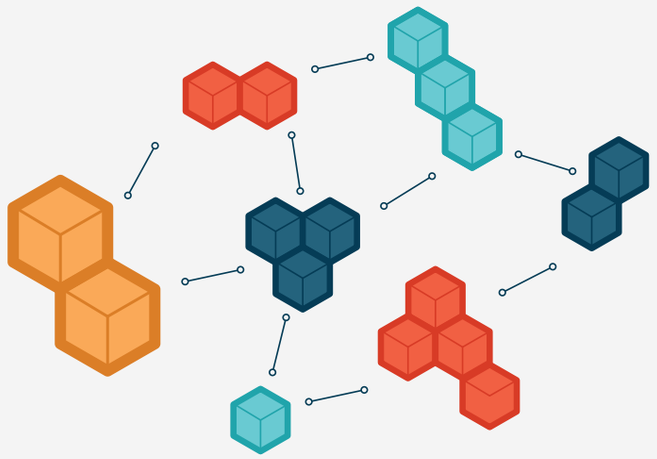
PARADIGM SHIFT:

DATA TOGETHER

COMMUNITIES USING DECENTRALIZED TECHNOLOGIES TO MAKE A BETTER WEB

MATT ZUMWALT
PROTOCOL LABS

ESIP SUMMER MEETING
25 JULY 2017



Protocol Labs

DECENTRALIZED WEB

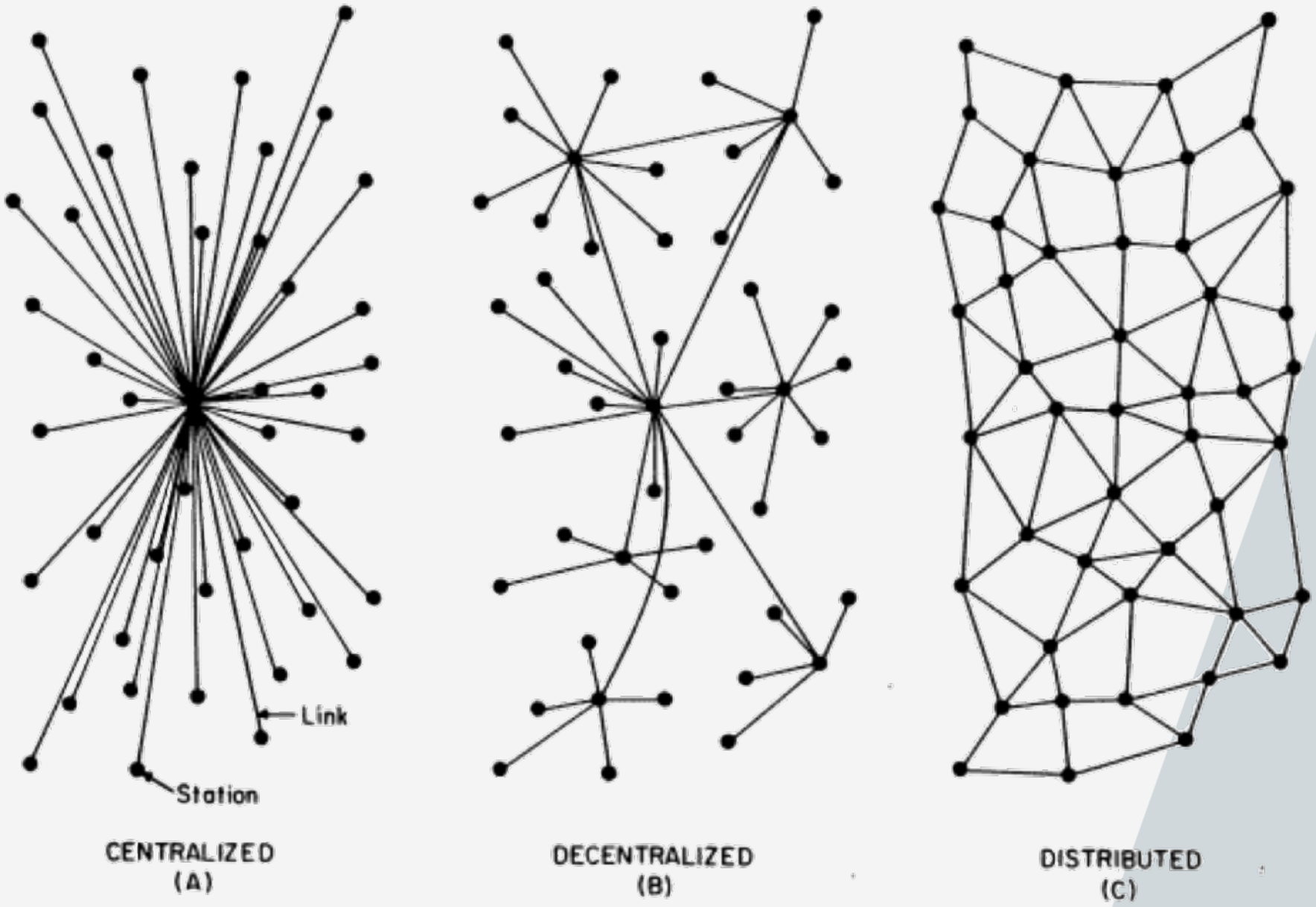


FIG. 1 — Centralized, Decentralized and Distributed Networks

Source: On Distributed Communications Networks, Paul Baran, 1962

The internet has been stolen from you. Take it back, nonviolently.

DATA TOGETHER

COMMUNITIES USING DECENTRALIZED TECHNOLOGIES TO STEWARD DATA

A MODEL FOR DECENTRALIZED ACTIVITIES
A VIEW ONTO DECENTRALIZED ACTIVITIES
TOOLS FOR DECENTRALIZED ACTIVITIES

PUBLIC RECORD
COMMUNITIES, COLLECTIONS

ACTIVITIES: **HARVESTING, MONITORING, STORING, ANALYZING, *RESCUING*** …

WHAT DO THESE HAVE IN COMMON?

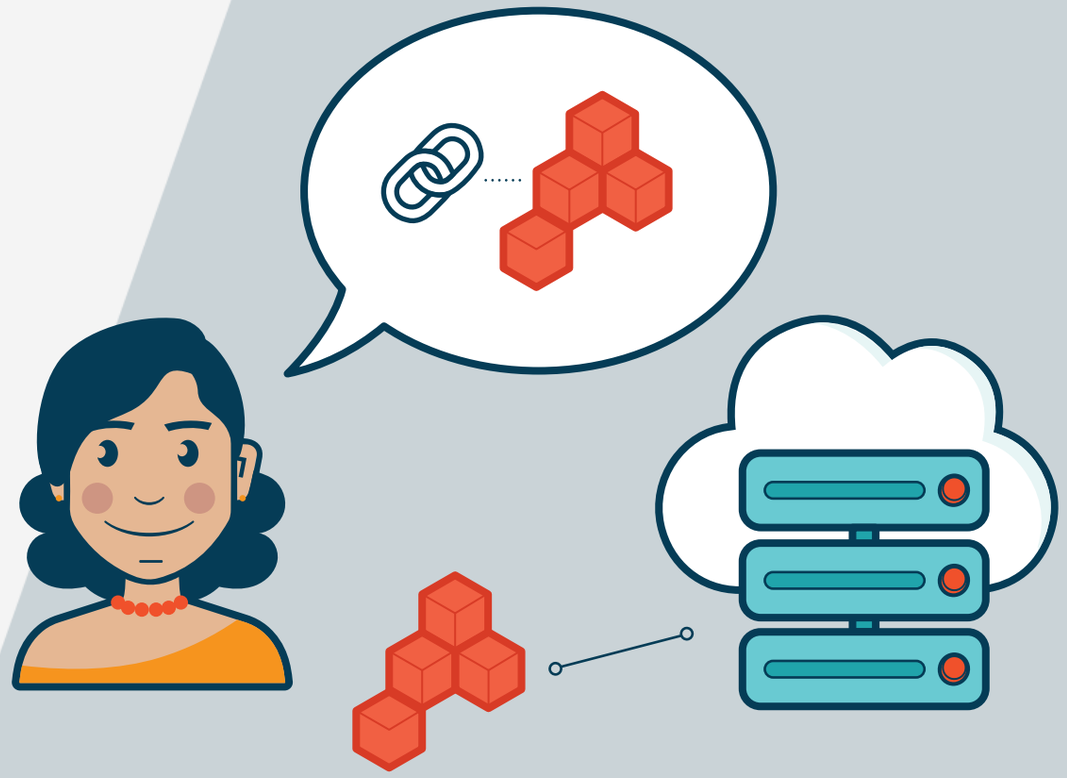
PRECARIOUS

WEB

THE LOCATION-ADDRESSED FARCE

MANY DOWNLOADS = MANY COPIES, BUT LOCATION-ADDRESSED LINKS PRETEND THERE'S ONLY ONE VALID COPY

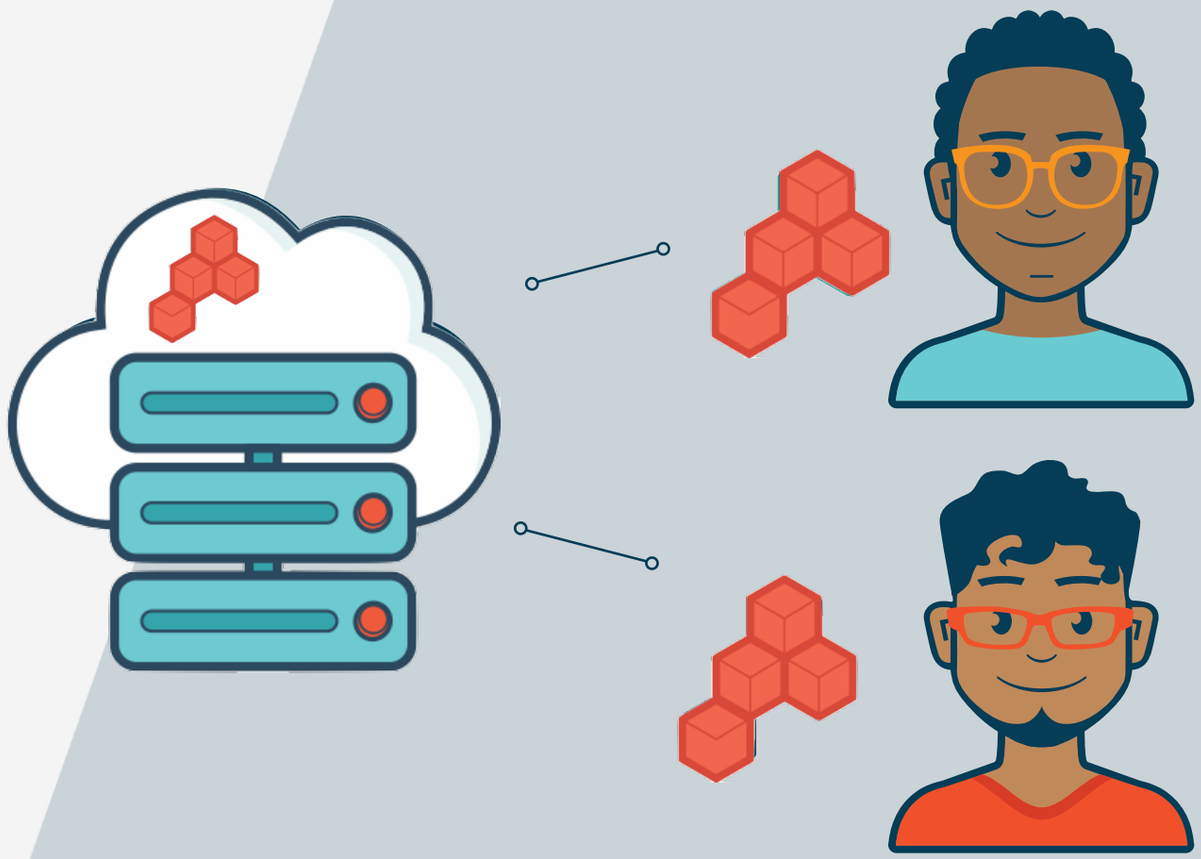
Someone puts a file on a server and passes around an http link that points to the server



THE LOCATION-ADDRESSED FARCE

MANY DOWNLOADS = MANY COPIES, BUT LOCATION-ADDRESSED LINKS PRETEND THERE'S ONLY ONE VALID COPY

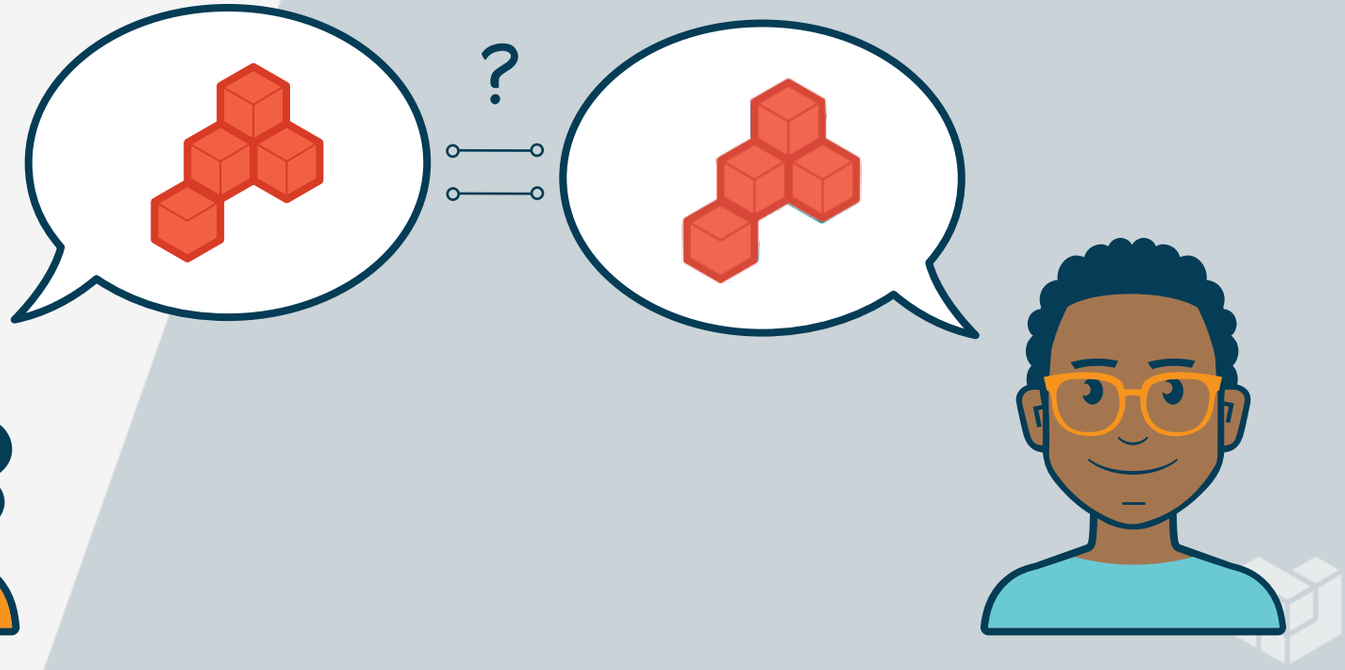
Many people download the file. Now there are many copies of the file — all of the downloaded copies



THE LOCATION-ADDRESSED FARCE

MANY DOWNLOADS = MANY COPIES, BUT LOCATION-ADDRESSED LINKS PRETEND THERE'S ONLY ONE VALID COPY

… But we pretend that there's **still only one copy**



Endangered Data
Woven into a Precarious Web

precarious adj. Dangerously lacking in security or stability: *a precarious posture; precarious footing on the ladder*.adj. Subject to chance or unknown conditions.

Centralization is a disease that the web is suffering from. It makes the web unstable, insecure and vulnerable to exploitation. We can address that disease by changing the way we link to information.

Instructions for Saving Endangered Data. Zumwalt. Dec 2016.

CORE CONCEPT

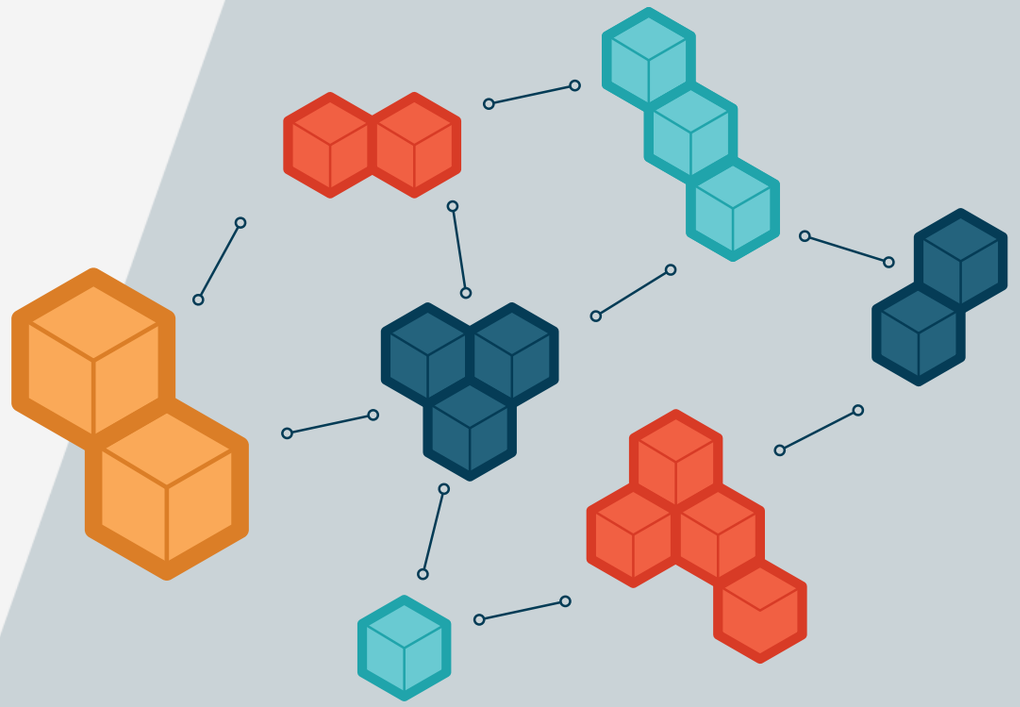
CONTENT ADDRESSING

Benefit: If *anyone on the network* has a copy of the *content you will be able to find and retrieve it*

CORE CONCEPT

CONTENT ADDRESSING

Key idea: It doesn't matter where the content is stored. What matters is being able to know that you're getting exactly the content you requested*

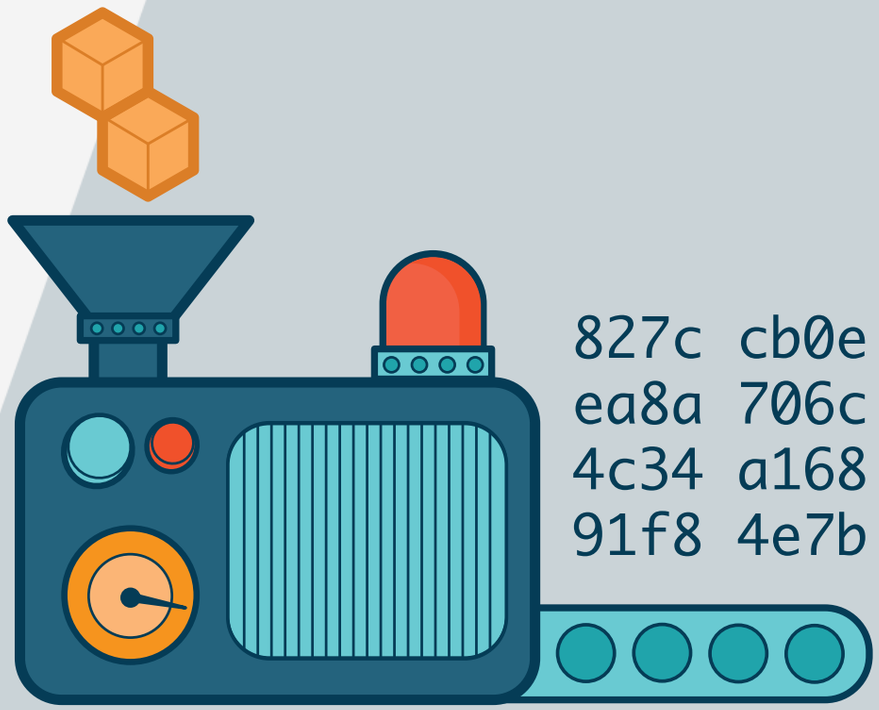


*also important to consider who you got the link/address from.

CORE CONCEPT

CONTENT ADDRESSING

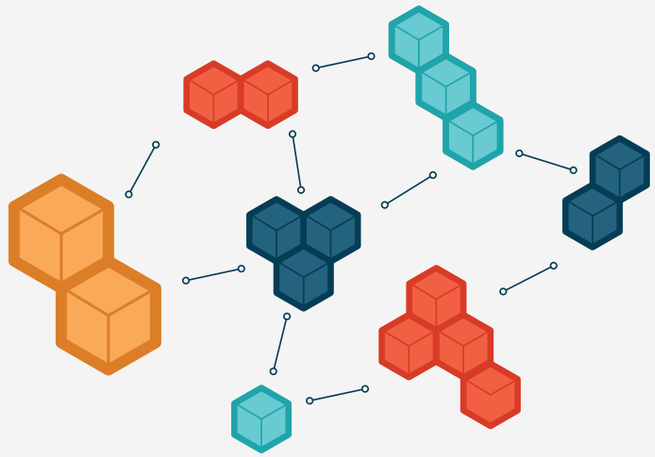
How: identify content by its cryptographic hash



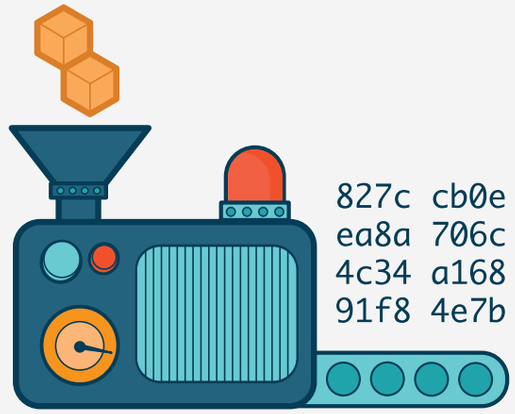
CORE CONCEPT

CONTENT ADDRESSING

Benefit: If *anyone on the network* has a copy of the *content you will be able to find and retrieve it*



Key idea: It doesn't matter where the content is stored. What matters is being able to know that you're getting exactly the content you requested*

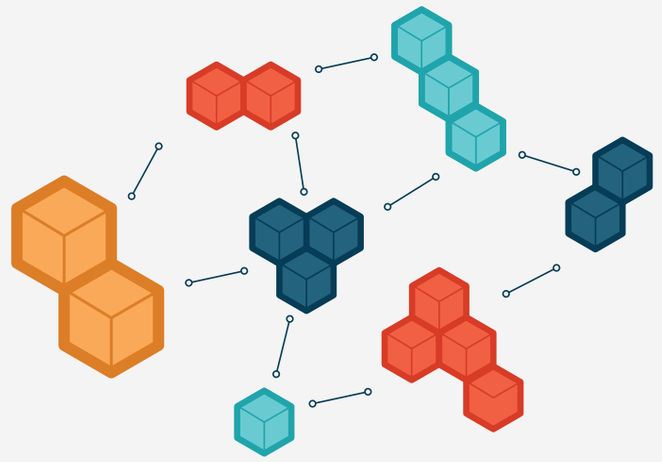


How: Identify content by its cryptographic hash

*also important to consider who you got the link/address from.

THE KEY:

HASH-LINKED DATA STRUCTURES



SOME TOOLS THAT USE

HASH-LINKED DATA STRUCTURES

A BENEFIT OF HASH-LINKING:

CRYPTOGRAPHIC INTEGRITY CHECKING

WHEN RESOLVING A LINK **YOU CAN USE THE LINK VALUE (A HASH) TO VALIDATE THE RESULT**. THIS ALLOWS WIDE, SECURE, EXCHANGES OF DATA (E.G. GIT OR BITTORRENT)

A BENEFIT OF HASH-LINKING:

IMMUTABLE DATA STRUCTURES

DATA STRUCTURES WITH HASH LINKS **CANNOT MUTATE**. THIS IS USEFUL FOR VERSIONING, FOR REPRESENTING DISTRIBUTED MUTABLE STATE, AND FOR LONG TERM ARCHIVING.

A BENEFIT OF HASH-LINKING:

DECOUPLE CONTENT FROM LOCATION

DECOUPLE *WHERE THE CONTENT IS STORED* FROM THE IDENTITY OF THAT CONTENT, SO THAT **DATA CAN EXIST IN MANY PLACES AND PASS THROUGH MANY HANDS** WITHOUT LOSING INTEGRITY

PUTTING IT TOGETHER

CONTENT-ADDRESSED PROTOCOL

examples of links to a hash "`QmZt34a`":

dweb:/ipfs/QmZt34a

dweb address space: general address space for decentralized protocols

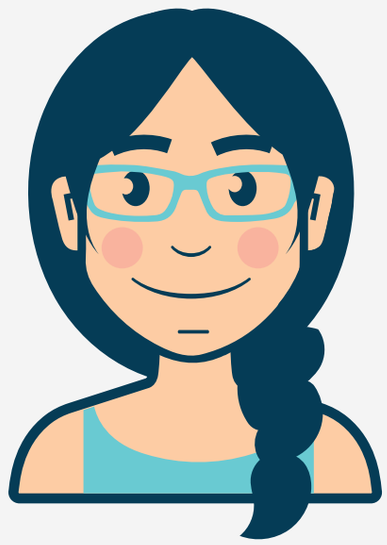
http://ipfs.io/ipfs/QmZt34a

http gateway: IPFS nodes expose an http interface that is a gateway to the p2p IPFS network

*also important to consider who you got the link/address from.

STORING DATA TOGETHER

PEERS COORDINATING



“This information is important to me. Please hold a copy at the library.”

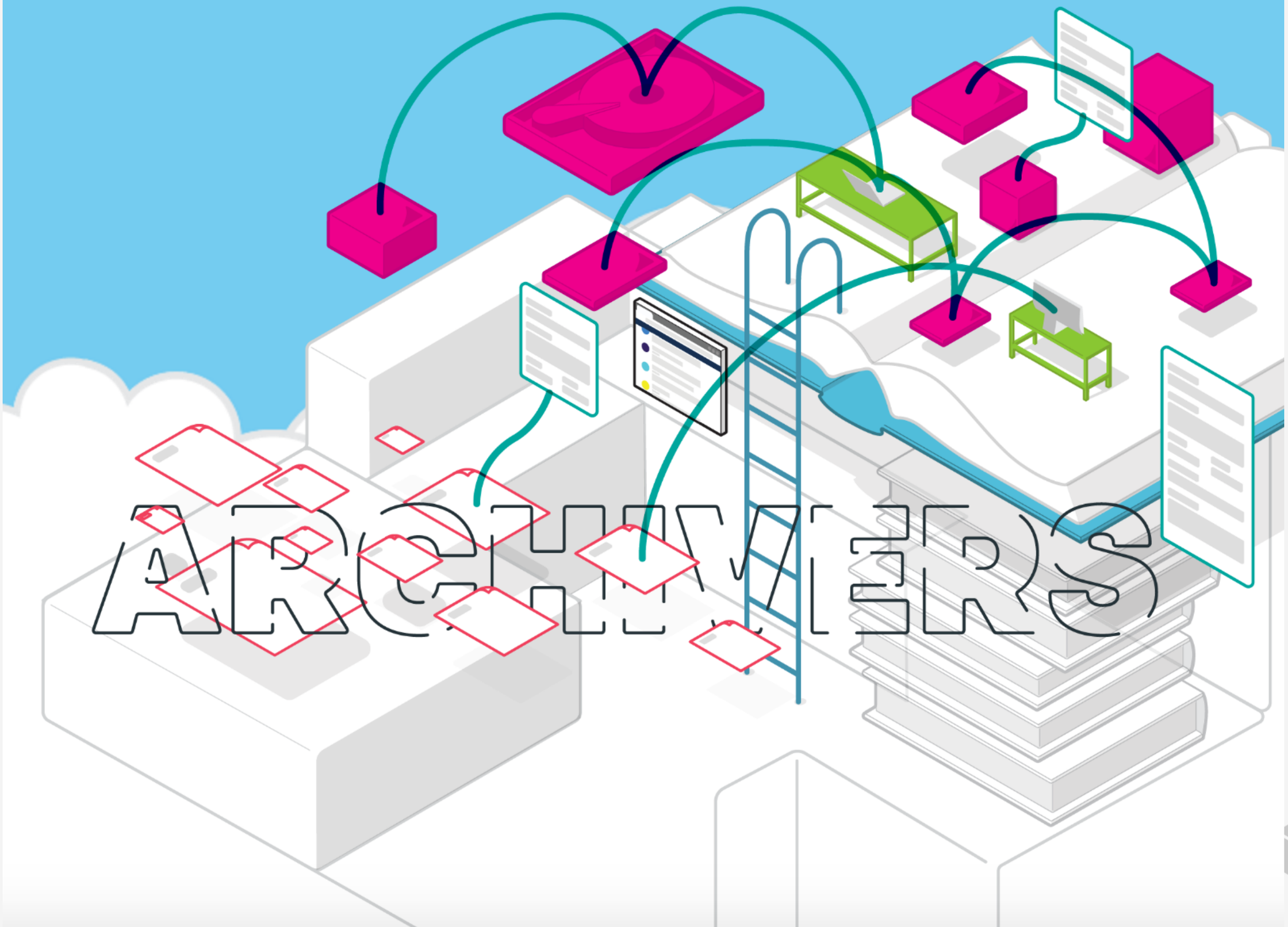
“This information is important to me. I have a copy, in case anyone needs it.”

“I’ve got some spare storage. I’ll help preserve a little piece of the stuff we all care about.”

Secure https://alpha.archivers.space

ARCHIVERS 2.0 ALPHA

LOGIN MENU



Primers:

In order to make our archiving efforts as thorough and systematic as possible, we use Agency Archiving Primers to identify key programs, datasets, and documents that are vulnerable to change and loss. Primers are composed of *sources*, which each specify a url as a starting point for archiving.

HUD

US Department of Housing & Urban Development
17014 urls
914 / 3115 completed

NASA

National Aeronautics and Space Administration
0 urls 0 / 0 completed

NOAA

National Oceanic and Atmospheric Administration
0 urls 0 / 0 completed

PRIMER

Environmental Protection Agency

The mission of the Environmental Protection Agency is to protect human health and the environment through the development and enforcement of regulations. The EPA is responsible for administering a number of laws that span various sectors, such as agriculture, transportation, utilities, construction, and oil and gas. In the budget for FY 2017, the agency lays out goals to better support communities and address climate change following the President's Climate Action Plan. Additionally, the agency aims to improve community water infrastructure, chemical plant safety, and collaborative partnerships among federal, state, and tribal levels.

urls	content	documented
3205779	**45353**	**2276**

Sources:

EPA Open Data

0/1061

Hazardous Air Pollutants

0/0

epa.gov

2272/43273

Environmental Dataset Gateway

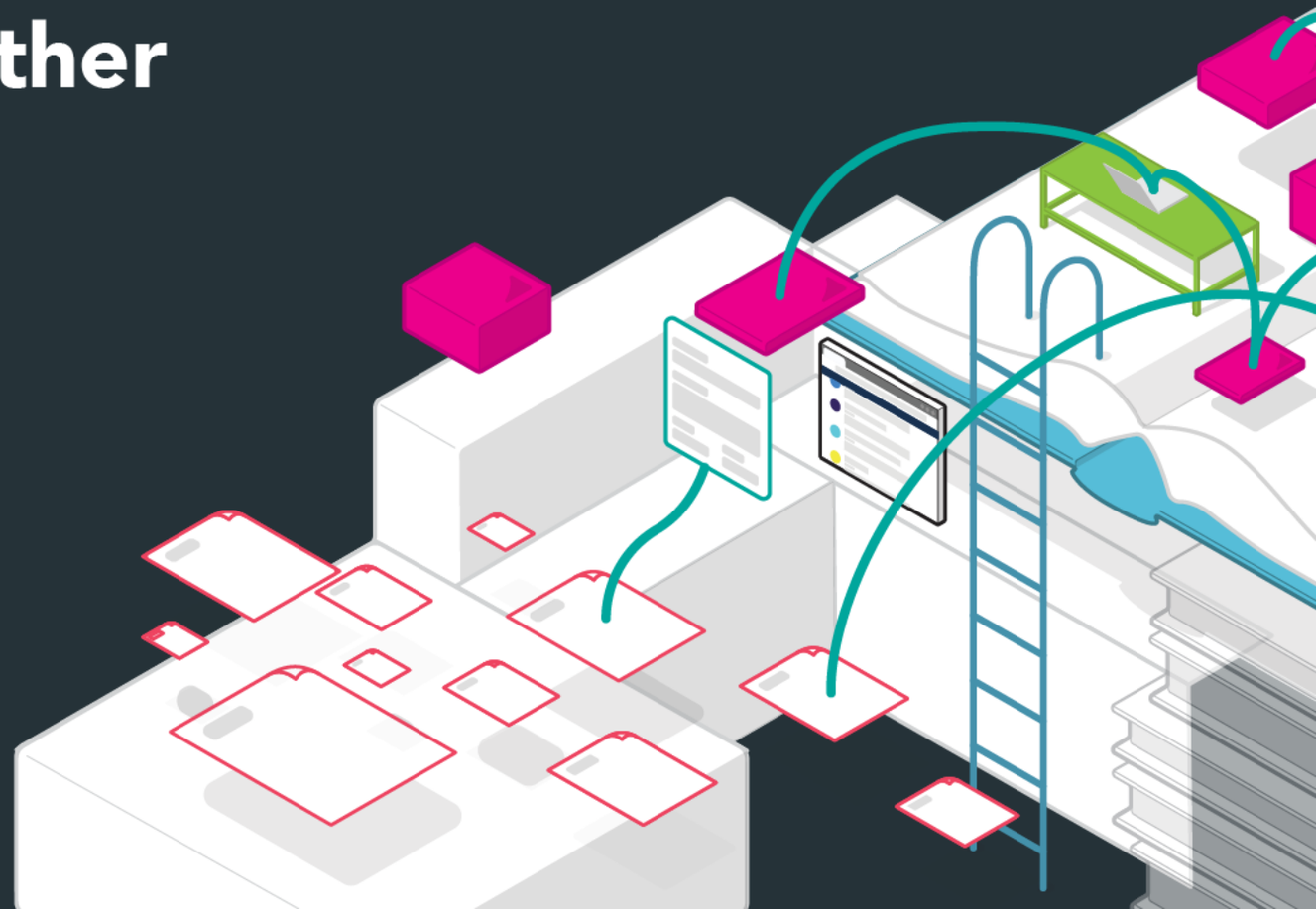
4/1019

WHAT DO THESE HAVE IN COMMON?

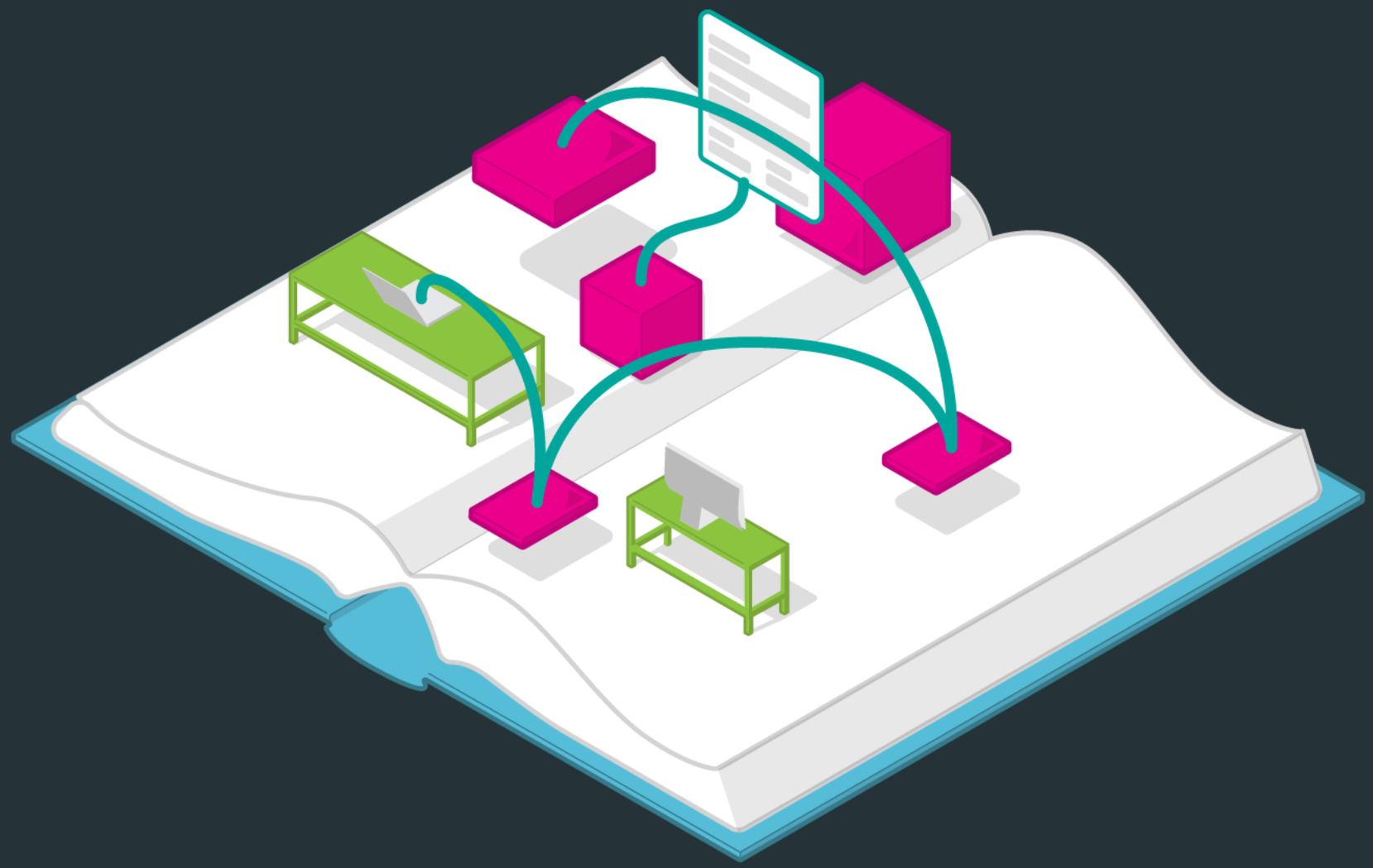
Communities

Harvesting
Montioring
Storing
Analyzing
Rescuing

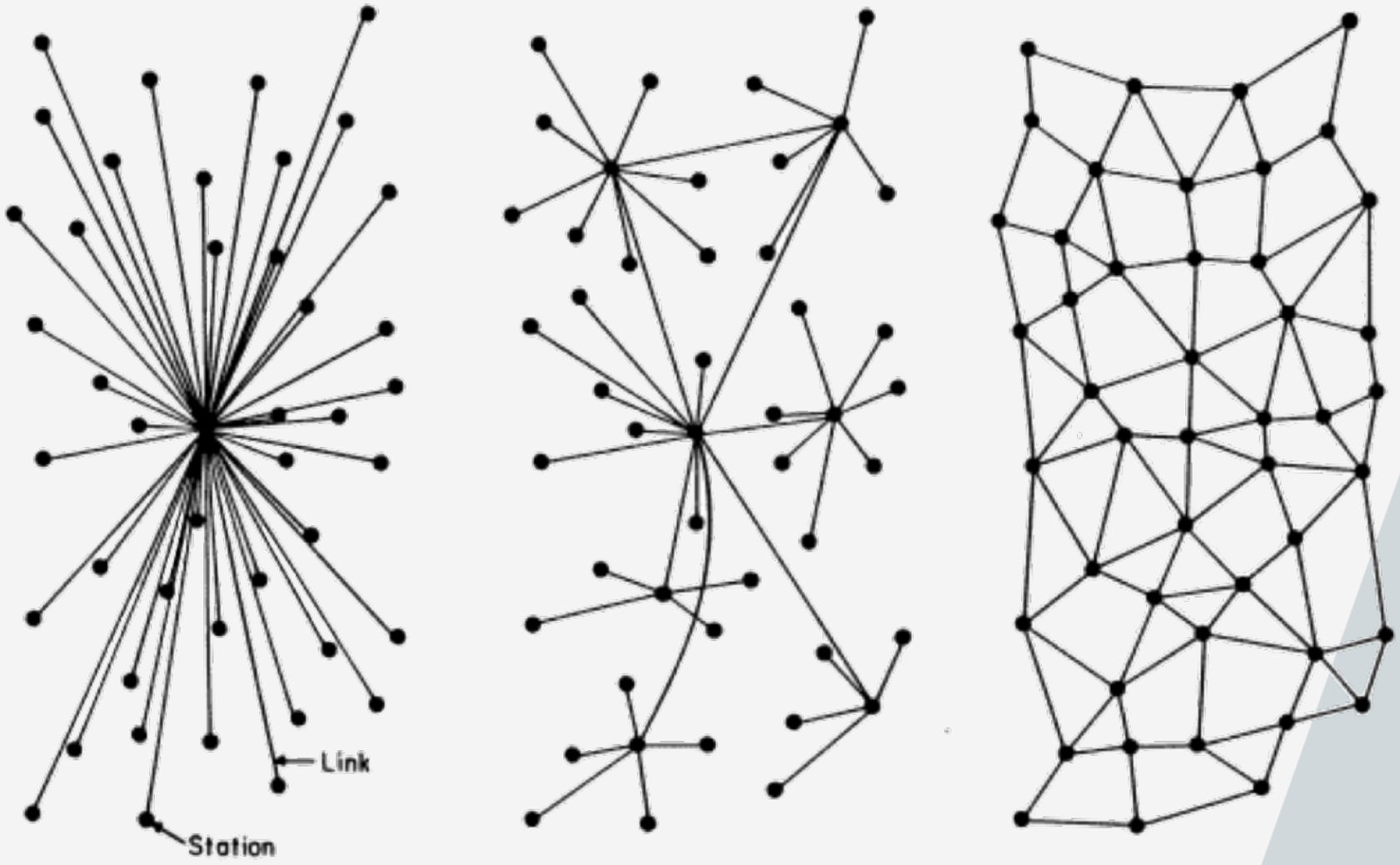
Data, Together



The Public Record



DISTRIBUTED WEB MENTALITY

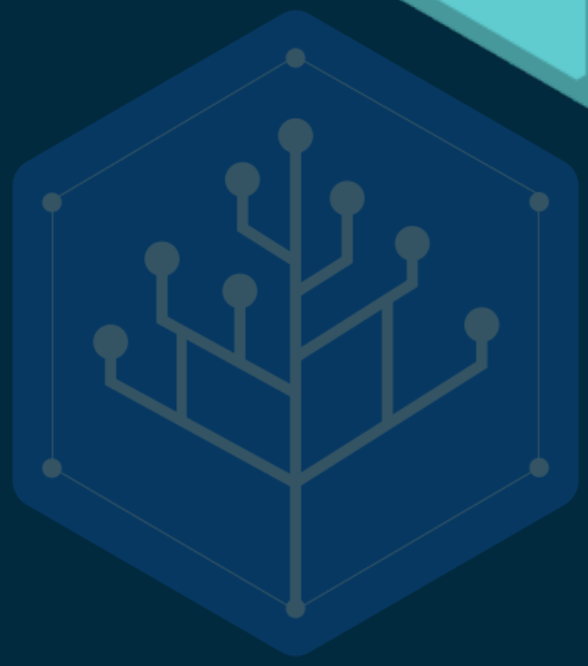


BECOME “MERELY” PEERS — VALUED PEERS,
WITH THE RESOURCES OF INSTITUTIONS.

IPFS:

CONTENT-ADDRESSED PROTOCOL TO REPLACE HTTP

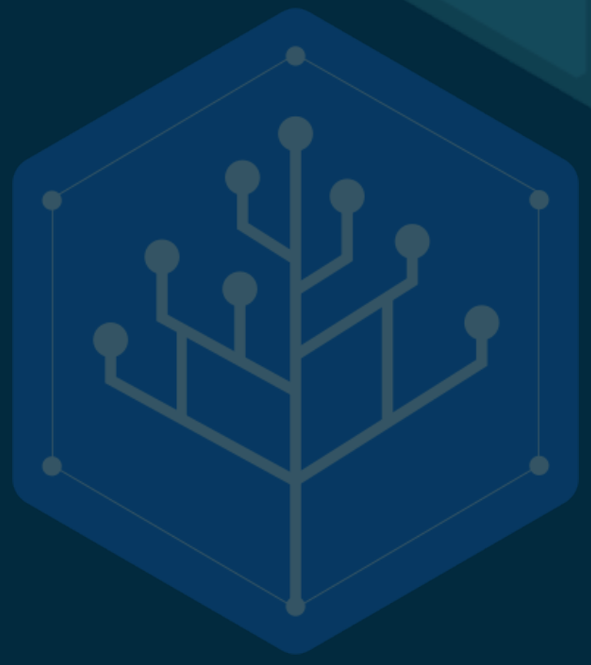
ipfs.io



FILECOIN:

TOKEN-POWERED DECENTRALIZED STORAGE NETWORK

filecoin.io



IPLD:

ABSTRACT DATA MODEL FOR ALL HASH-LINKED DATA STRUCTURES

ipld.io

DATA TOGETHER

https://datatogether.org

https://ipfs.io

Matt Zumwalt
@flyingzumwalt
Protocol Labs