

HOW THE

# DECENTRALIZED WEB

COULD BRING A

## NEW GOLDEN AGE

FOR LIBRARIES

Matt Zumwalt

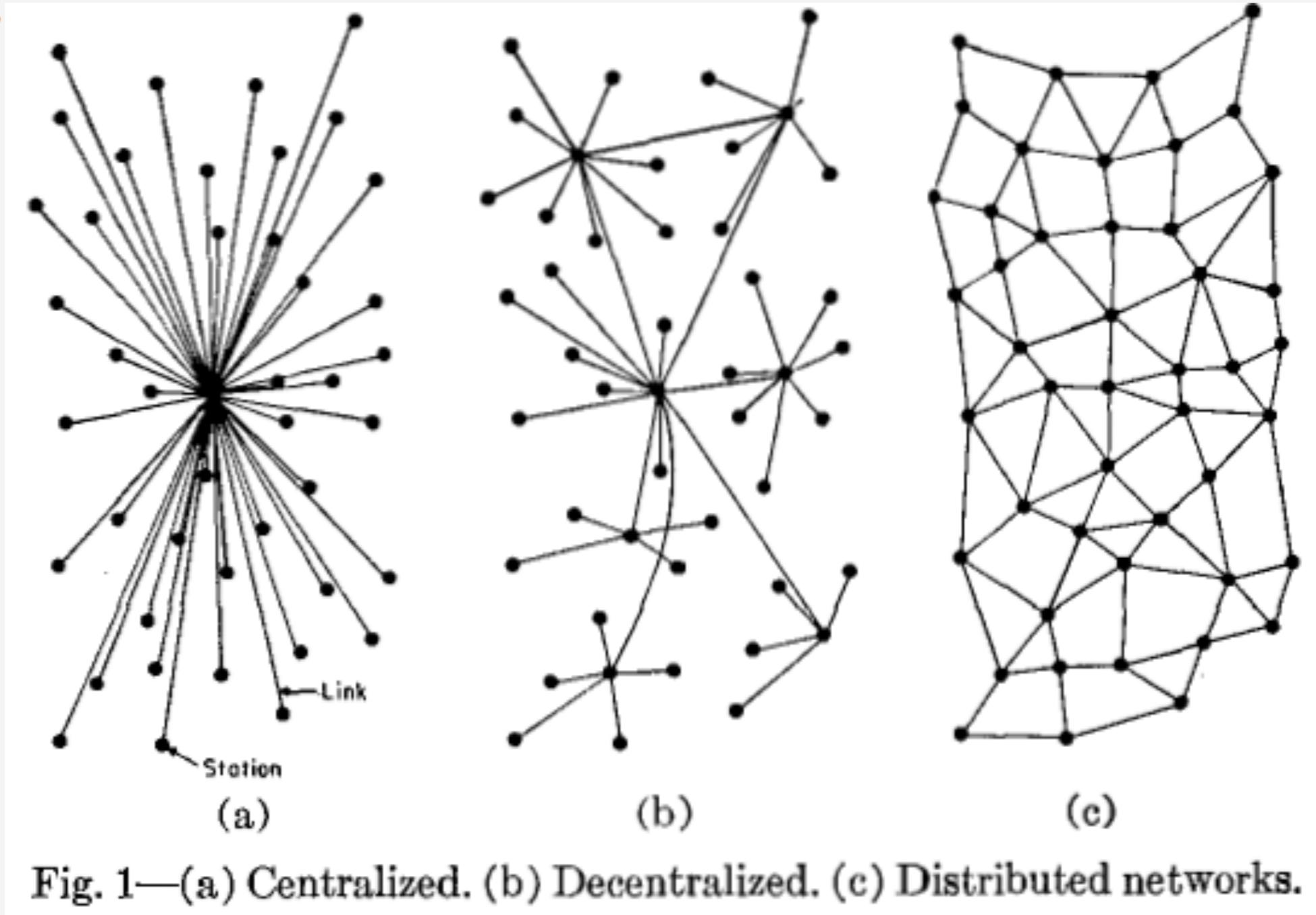
Protocol Labs

Code4Lib 2017

Los Angeles



# DECENTRALIZED WEB



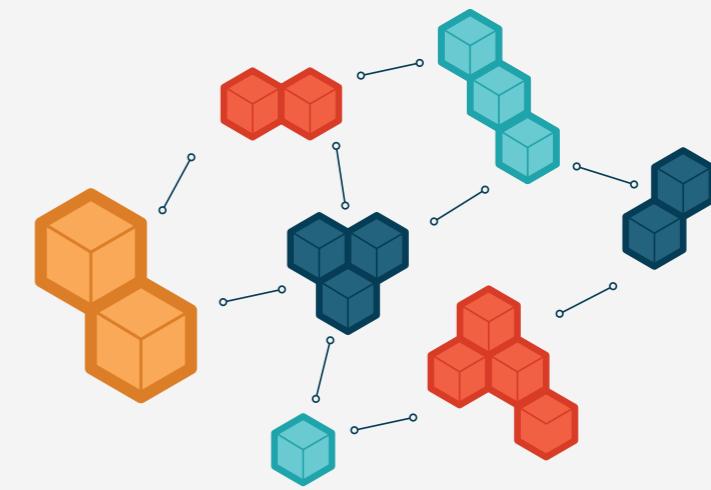
Source: On Distributed Communications Networks, Paul Baran, 1962



# ENDANGERED DATASETS



# DATA RESCUE



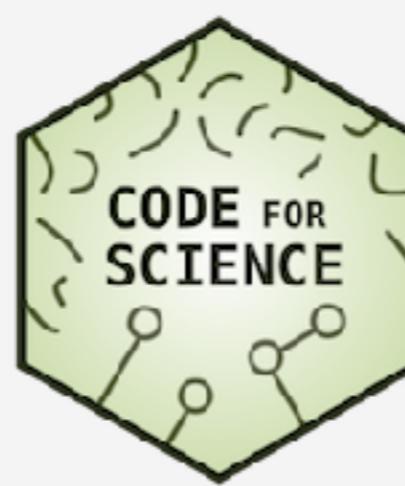
# STORING DATA TOGETHER

THE MOVEMENT TO  
DECENTRALIZE DATA  
AND HOW LIBRARIES CAN LEAD IT



ORGANIZATIONS + TECH

# DATA RESCUE



ORGANIZATIONS + TECH

# DATA RESCUE

[datarefuge.org](http://datarefuge.org)



[Internet Archive](http://archive.org)



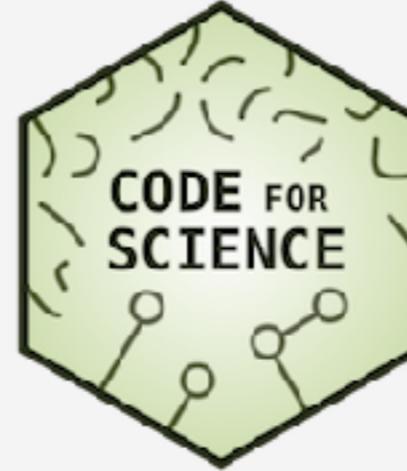
[EDGI](http://edgi.org)



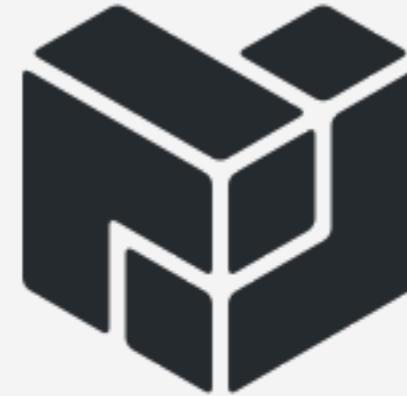
[Climate Mirror](http://climate-mirror.org)



[Code for Science](http://codeforscience.org)



[Protocol Labs](http://ProtocolLabs.com)



## ORGANIZATIONS + TECH

# DATA RESCUE

### [datarefuge.org](#)

- CKAN instance
- Amazon S3
- LOCKSS ?

### [Internet Archive](#)

- **Web Crawlers**, EoT Harvest
- Nomination Tool
- **Wayback Machine**

### [EDGI](#)

- Hackathons
- **Agency Guides**
- Tools & Scripts
- **Dataset Monitoring**

### [Climate Mirror](#)

- [climate.daknob.net](#)

### [Azimuth Backup](#)

- [http://math.ucr.edu/home/baez/azimuth\\_backup\\_project/](http://math.ucr.edu/home/baez/azimuth_backup_project/)

### [Code for Science](#)

- **Dat**
- max ogden's crawls
- **Project Svalbard**

### [Protocol Labs](#)

- IPFS
- **ipfs-pack**
- **ipfs-cluster**



# DECENTRALIZING DATA

This is a **grassroots effort to decentralize government data**. It would be **foolish to re-centralize** on some new organization/location.



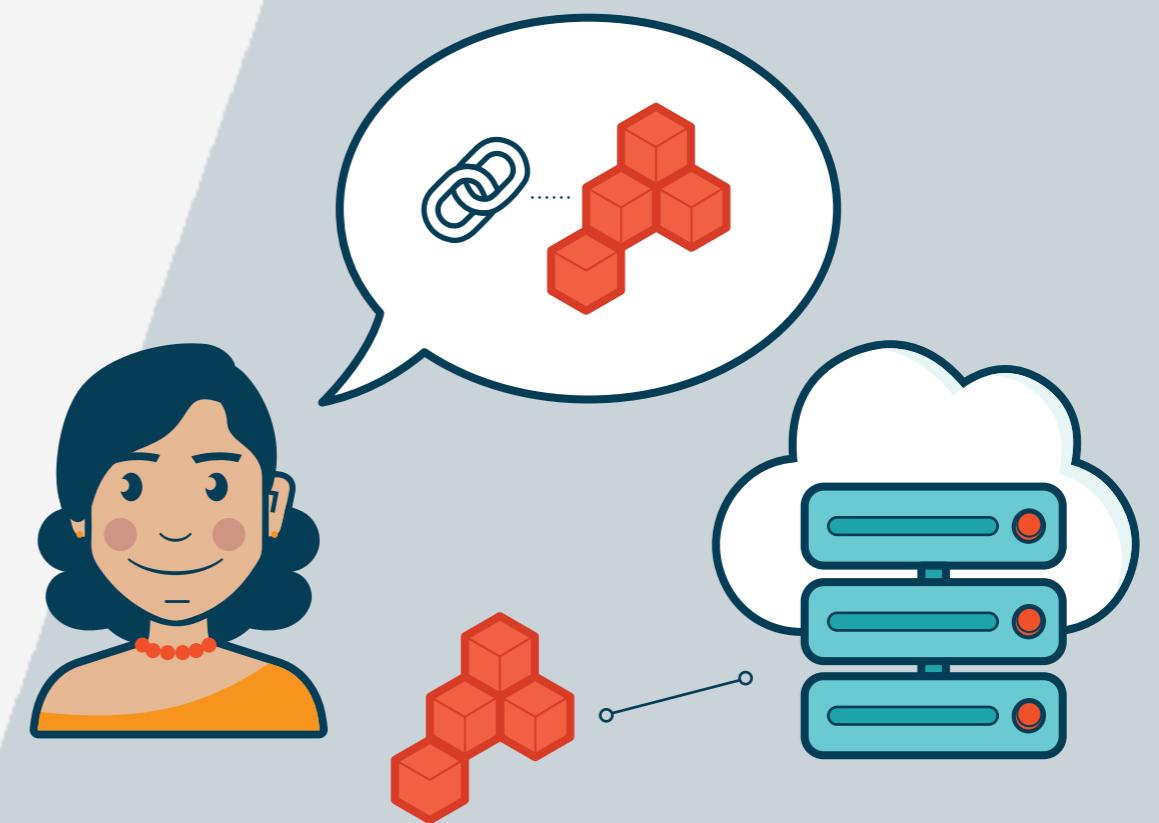
# PRECARIOUS WEB



# THE LOCATION-ADDRESSED FARCE

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

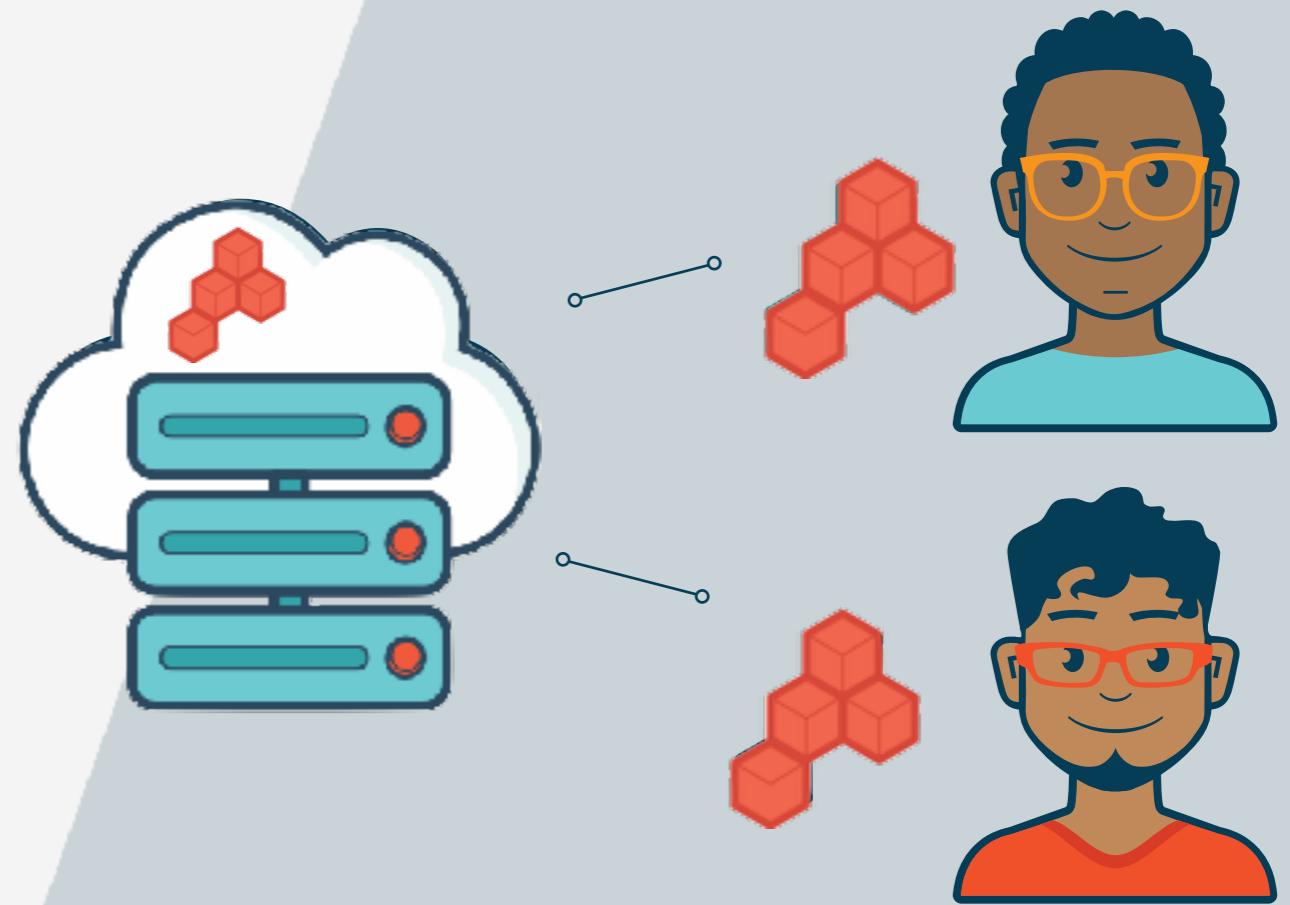
- Someone puts a file on a server and passes around an http link that points to the server



# THE LOCATION-ADDRESSED FARCE

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

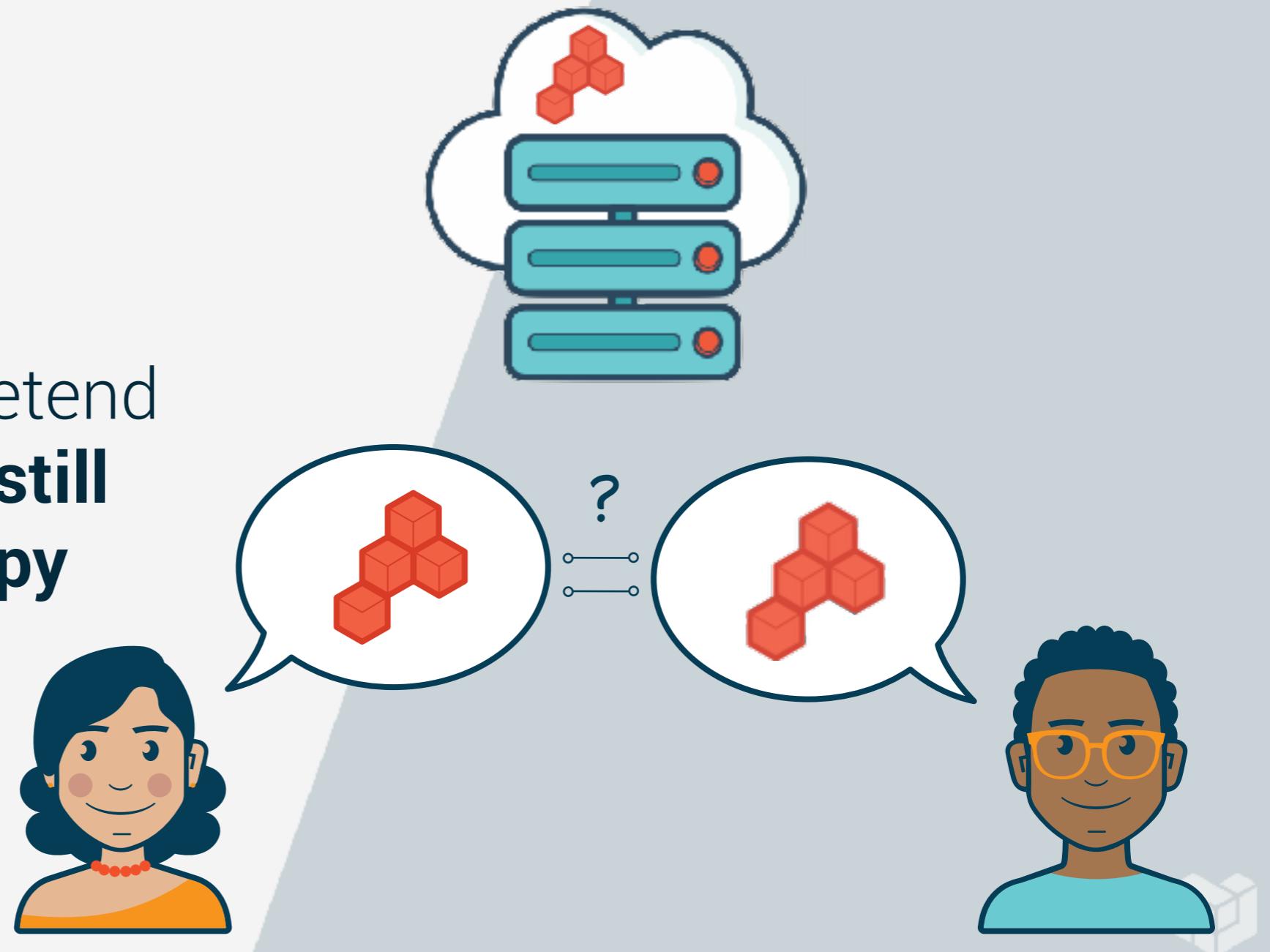
- Many people download the file. Now there are many copies of the file – all of the downloaded copies



# THE LOCATION-ADDRESSED FARCE

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

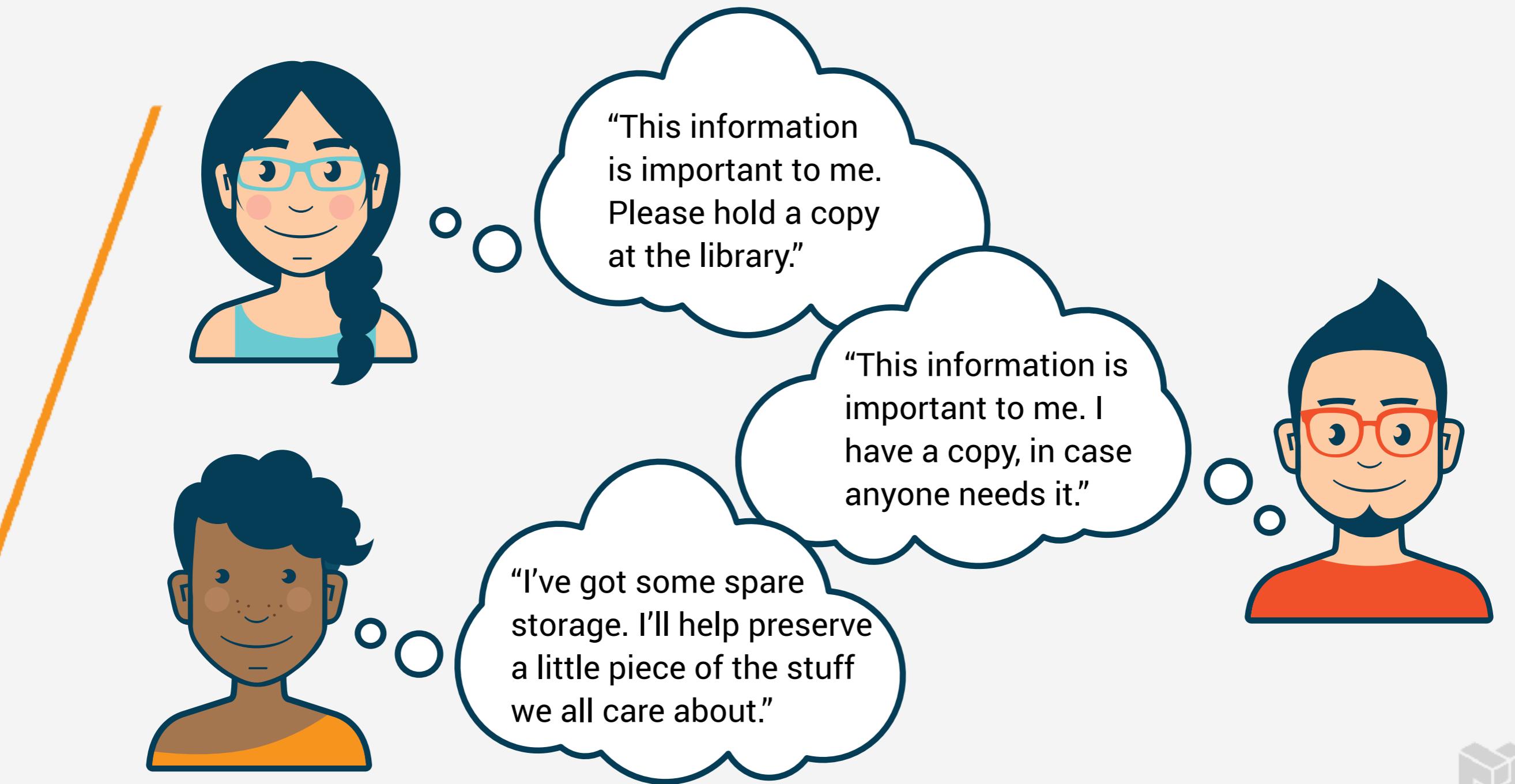
- ... But we pretend that there's **still only one copy**



# STORING DATA TOGETHER



# PEERS COORDINATING



CORE CONCEPT

# CONTENT ADDRESSING

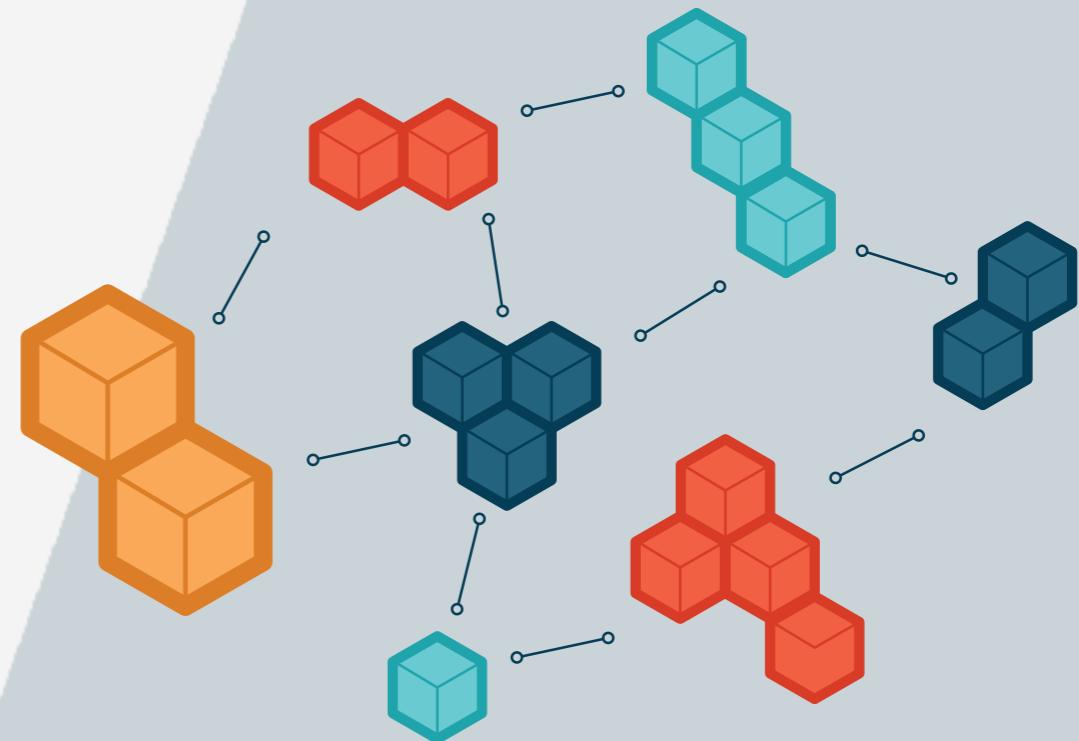
**Benefit:** If anyone on the network has a copy of the content you will be able to find and retrieve it



CORE CONCEPT

# CONTENT ADDRESSING

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



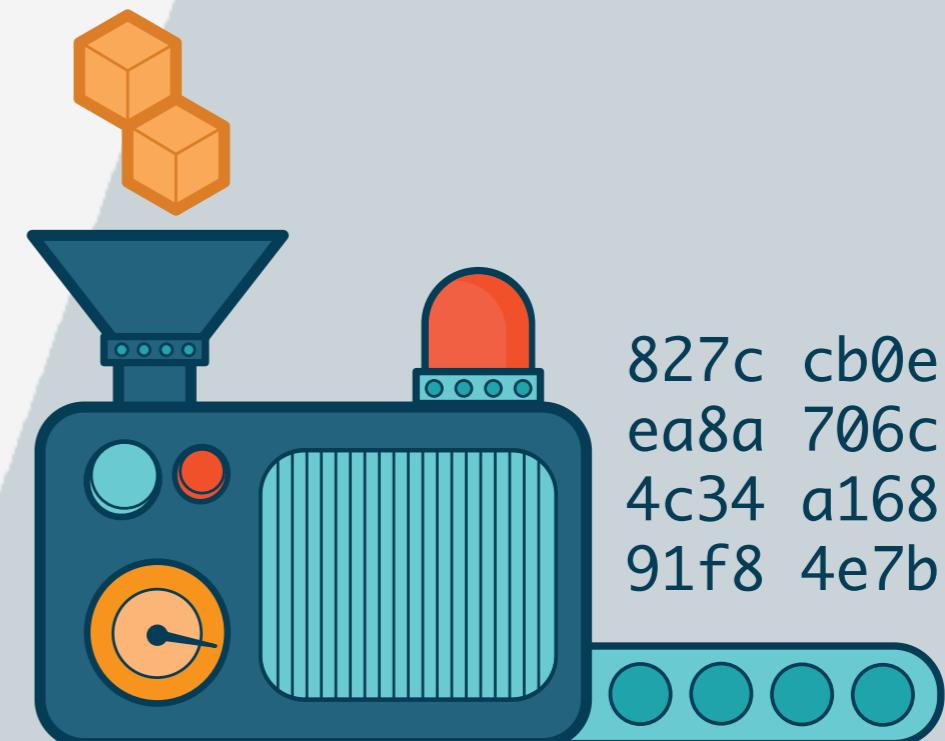
\*also important to consider who you got the link/address from.



CORE CONCEPT

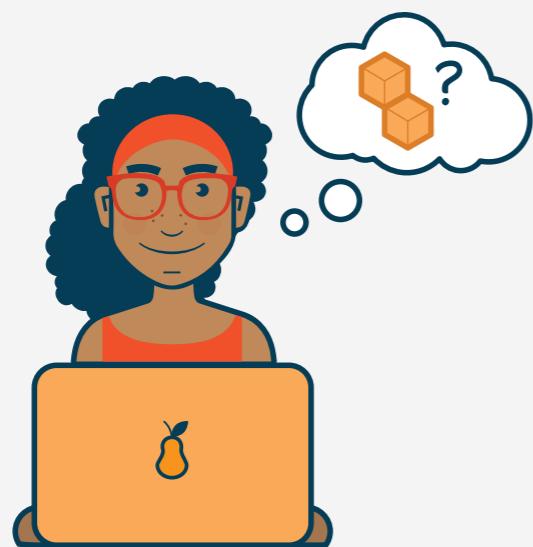
# CONTENT ADDRESSING

**How:** identify content by its cryptographic hash

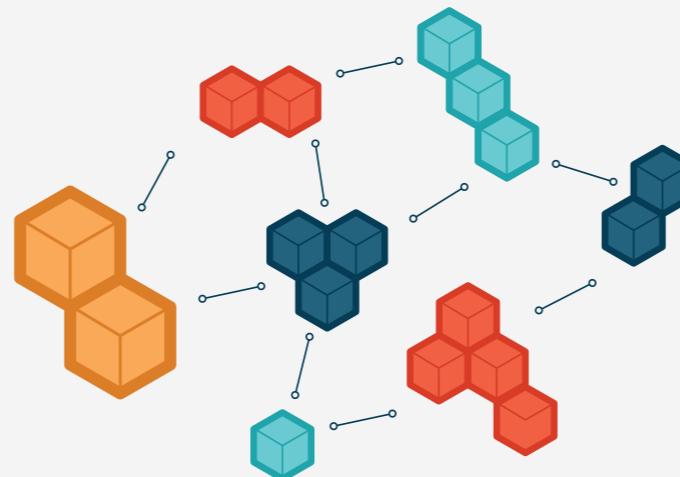


## CORE CONCEPT

# 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 a file 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.



# ACCESSION WITHOUT UPLOAD



# ACCESSION FEEDS OF DATA



**"Please pin all the versions of my dataset. The version history is in the feed at <dweb://ipns/jack/dataset1>.**

I'll keep updating it when there are new versions of my data."

"No problem. We've pinned all the versions in that feed and will grab any new versions when they show up."



# COLLABORATIVE CATALOGING



# DISTRIBUTED NETWORK OF PEOPLE AND DEVICES

“Today, instead of going to the library building to find and read books and journal articles, students and faculty access, organize, and read scholarly content on their own devices.”

- A Vision for MIT Libraries (2017)



# DECENTRALIZED WEB

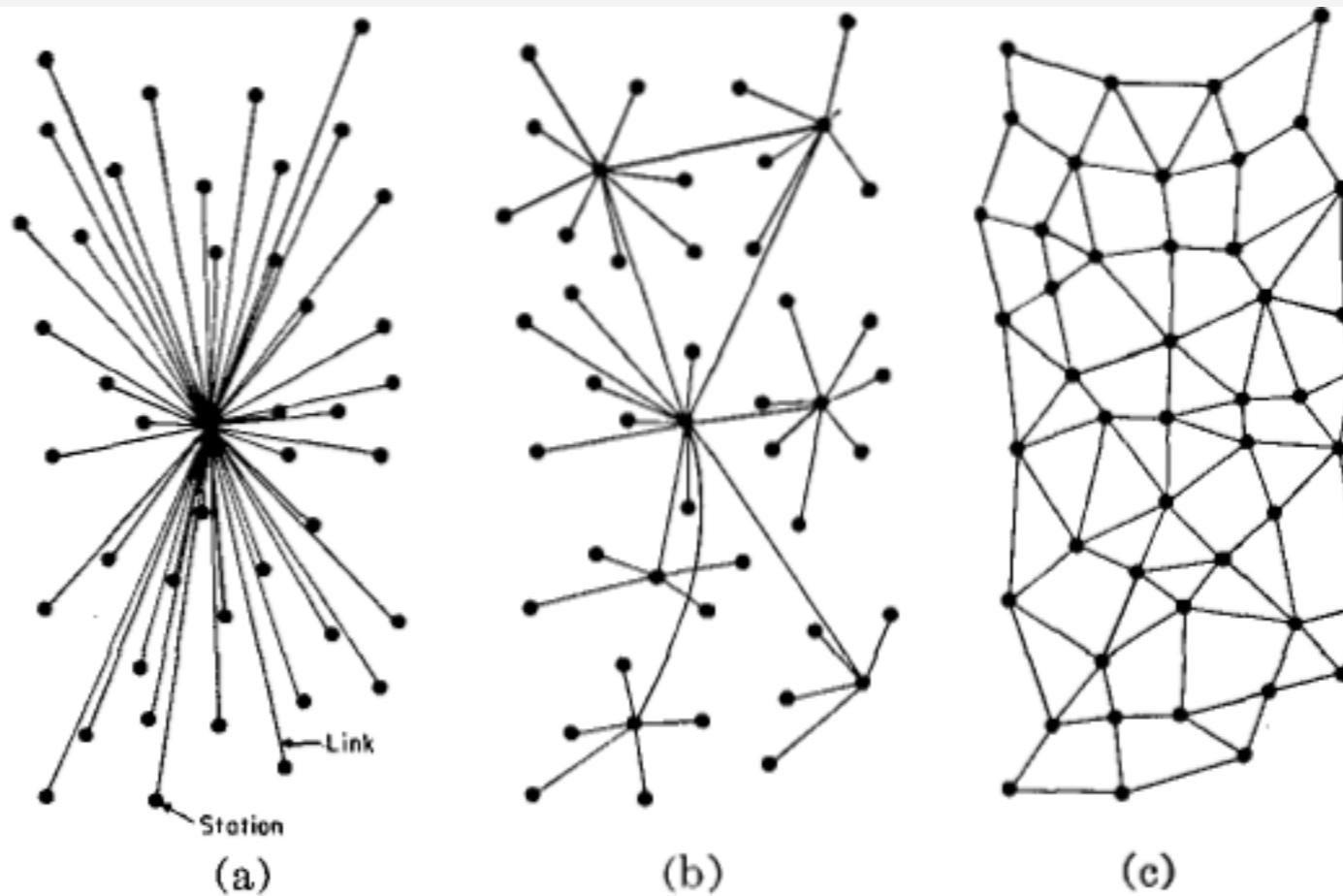


Fig. 1—(a) Centralized. (b) Decentralized. (c) Distributed networks.

## TWO SENSES OF CENTRALIZATION

TECHNICAL: Network Architecture

SOCIOECONOMIC: Who gets to be a “peer”



# DISTRIBUTED WEB MENTALITY

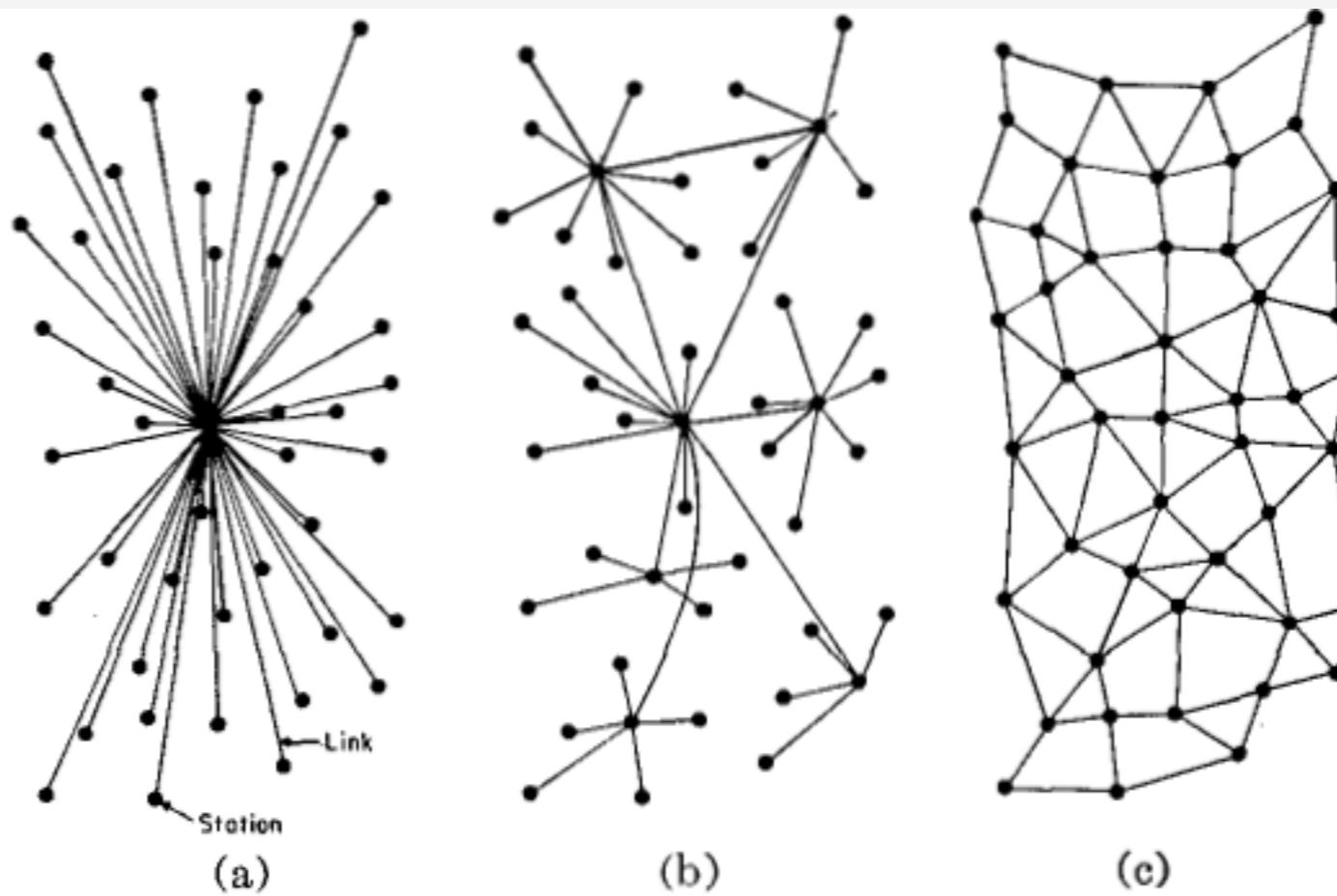


Fig. 1—(a) Centralized. (b) Decentralized. (c) Distributed networks.

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



IT'S OK TO BE "MERELY A PEER"

# A LIBRARY'S WORK SPEAKS FOR ITSELF.



# STORING DATA TOGETHER #DWebLAM

<https://ipfs.io>

<https://datatogether.org>

Matt Zumwalt  
@flyingzumwalt  
Protocol Labs

