Decentralized Web
webinar series by

METROPOLITAN NEW YORK LIBRARY COUNCIL

INTERNET ARCHIVE

Web

LIBRARY FUTURES

Resource guide

session 02

Decentralized Storage

Thursday, February 24, 2022

Libraries, archives, and other institutions such as museums have long served as stewards of humanity's cultural heritage. For many centuries, they have been dealing with the many challenges associated with preserving and protecting the world's cultural, artistic, and intellectual assets for future generations.

Today we find ourselves in a once-in-a-millennia shift from the primacy of physical artifacts to digital. As we move into an era in which more and more of our cultural heritage is digital, how do we preserve it?

Some of the challenges include:

Keeping materials safe

How do we preserve digital content in a way that protects it from loss, destruction, or decay?

Keeping materials accessible

How do we guarantee a persistent access point, ensuring that the materials can be found and discovered?

Keeping materials trustworthy

Can we provide genuine, verifiable content that users can trust, regardless of where it is located or accessed?

Scale

What if you want to keep everything, and you don't want to weed materials to make room?

Long-term sustainability

How do you ensure long term preservation, access, and verifiability?

Cost

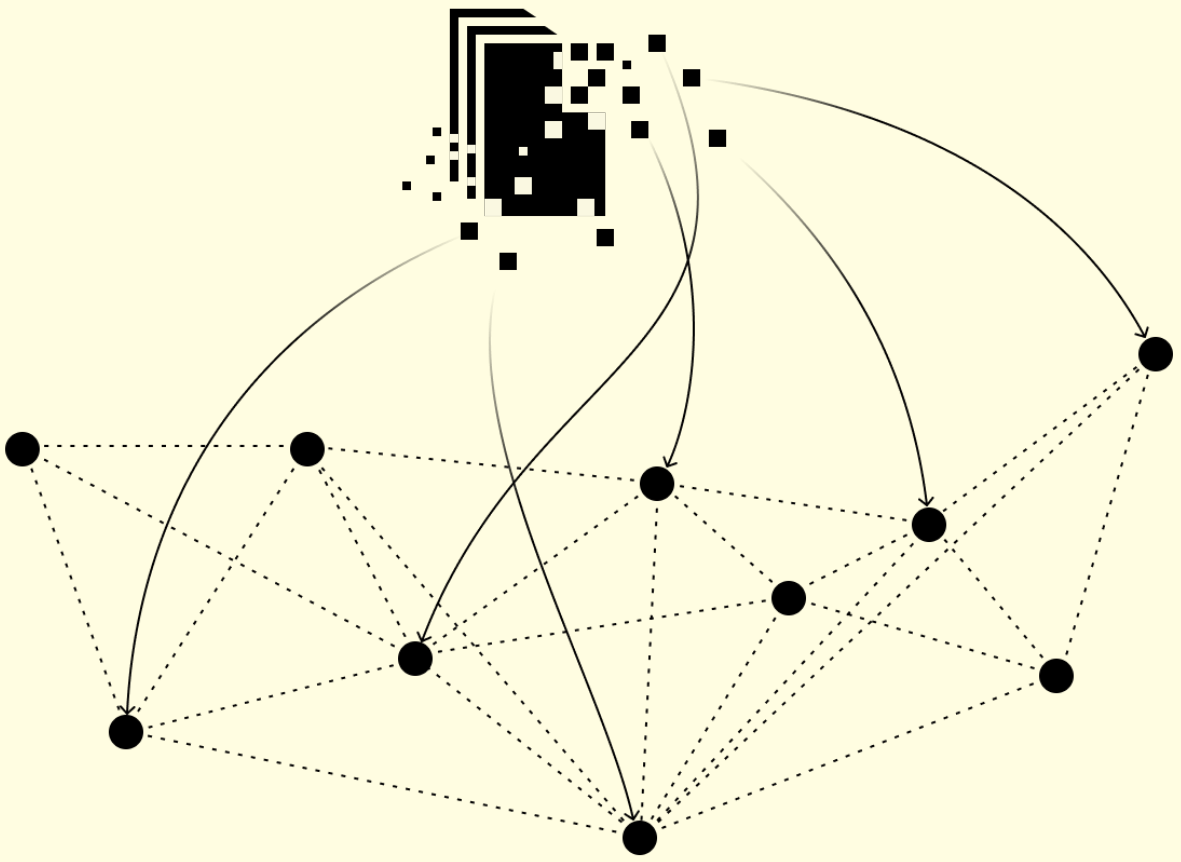
How can you provide these services in the long-term, at a cost these institutions can afford?

The potential of decentralized storage

What if we could stash multiple, certified copies of your collections all over the globe in a way that makes content findable and retrievable by everyone who needs it?

Decentralized storage[1] can be visualized as a network of P2P (peer-to-peer) servers that store materials across a global network of storage nodes. That means that a truly decentralized storage system is one where both *storage location* and *storage management* are decentralized. A decentralized storage management system is not under control of a single actor, or a small, exclusive number of them, but rather shared across all the users in the network.

[1] → To review what we mean by 'decentralized', please take a look at the resource guide for our first Webinar session, where we give a brief introduction to the concept.



These servers can work collaboratively to store multiple copies of the data assets across many different locations.

What are the advantages of decentralized storage?

01 Resiliency

As the LOCKSS (Lots of Copies Keeps Stuff Safe) Network has shown in the Web 2.0 context, distributed, redundant storage can be a key to security and resilience. That is, when materials (e.g. books, records, data…) are stored in multiple copies across different locations and jurisdictions, they are less likely to be completely lost. In practice, this means that decentralized storage can help keep data safe from state or corporate censorship, natural disasters, or just the normal cycle of institutions ceasing to exist.

02 Persistence

In a peer-to-peer system, each item is assigned a unique, immutable identifier or hash that can help track all the copies of an item, no matter where those copies are located. But for storage to be truly persistent, there needs to be a mechanism that guarantees that sufficient copies of the data will be maintained in the network, ideally forever.

While a decentralized storage network does enable persistence, it does not by itself *guarantee* it. Guaranteed persistence would require that the key nodes in the storage network somehow *commit*

to keeping those assets over the long term. Lacking some kind of coordination, e.g. via contracts, a large enough system may perhaps provide enough redundancy to expect some (or even most) of the data to be stored persistently, but not all of it.

Data persistence in decentralized storage *may* be possible through a combination of incentives, shared protocols, and storage marketplaces where available servers can be matched with users seeking long-term storage for their data. Filecoin is a good example of a decentralized storage system that implements persistence through protocols, contracts and incentives. Other examples of decentralized storage systems include Airweave, Storj and Skynet. Different systems make different trade-offs between decentralization, performance, convenience, and cost, and we encourage you to compare them if you're considering using decentralized storage. You can find out more about them in the Recommended Resources section below.

03 Self-certification

Unlike physical artifacts, digital materials are much easier to alter. Indeed, in this era of "deep fakes," how do we make sure our materials are authentic?

In other words, how can users be sure that a 1999 copy of Kissthisguy.com that they downloaded via IPFS from a random server in Cochabamba, Bolivia, is an unadulterated copy of the real thing?

The key difference in decentralized storage is that each item is assigned a unique, immutable hash. These persistent identifiers serve two important functions. They allow data assets to be found regardless of where they are stored (as we saw above). And they certify their provenance as “an authentic copy” of the original file. If the file’s content changes, so does the identifier.

In other words, this means that even in an era of deep fakes and fake news, self-certification should allow you to trust the contents you are retrieving, regardless of where or how it reaches you.

04 Interoperability

When libraries’ digital collections are saved on a decentralized storage network, they have the potential to grow into a collaborative, authenticated, collaboratively hosted world, one that seeks to bridge the distances and boundaries between unique, sometimes isolated or siloed library collections, making them available to users across the globe. We envision a decentralized, global web of knowledge that would integrate different materials, from multiple sources and in multiple formats, coexisting and referencing each other (e.g. via links or annotations).

Recommended resources

Decentralized Storage: Preserving Humanity's Greatest Assets. Arkadiy Kukarkin and Jonathan Dotan at the DWeb Meetup September 2021.

Introducing the Starling Lab, a short video explaining how the Shoah Foundation and Starling Lab are using decentralized technologies to store human rights videos.

What is Arweave? Wiki article that describes Arweave.

Introduction to Skynet, video featuring Daniel Helm & Manasi Vora, March 2021, starting at 1:14:30

Preservation Principles. Principles for robust data preservation by the LOCKSS (Lots of Copies Keep Stuff Safe) Project.

Try it out!

Get 150GB of Free Encrypted Decentralized Cloud Object Storage with Storj.

Skynet Basics. GitBook. Skynet Guide, 2021
Reuters visual guide to Blockchain technology. June 2018.

Dive deeper

‘Preserving Digital History: How to Close the Web’s ‘Memory Hole’ a16z Podcast with Brewster Kahle, Sam Williams, Alex Pruden, and Zoran Basich

“Exploring the Decentralized Web. Episode 3: Data and User Control.” A video by the Filecoin Foundation for the Decentralized Web

“Storing High-Performance Computing Data Using Storj Decentralised Cloud Storage.” A recent report by Antonin Portelli from the University of Edinburgh.

“Security and Privacy Benefits of Decentralized Cloud Object Storage.” A white paper on Storj by TechTarget Inc.

“Storage Endowment.” Short wiki article about Arweave.

“Requirements for Digital Preservation Systems.” Research article by Rosenthal et. al. (2005).

Community resources

GetDWeb.net - web site of the DWeb Community, a global network of meetup groups working to build a better web, following these core principles

Redigest - Monthly newsletter by Redecentralize.org

Stories from the Decentralized Web - Medium Channel with event recaps, articles & reposts of fundamentals of the Decentralized Web

DWeb Community Calendar

Past and upcoming webinar sessions

Jan 27	**The Decentralized Web: An Introduction** Watch the recording →
Feb 24 4 pm EST	**Using Decentralized Storage to Keep Your Materials Safe** Register →
Mar 31 4 pm EST	**Keeping Your Personal Data Personal: How Decentralized Identity Drives Data Privacy** Register →
Apr 28 4 pm EST	**Goodbye Facebook, Hello Decentralized Social Media? Can Peer-to-Peer Lead to Less Toxic Online Platforms?** Register →
May 26 4 pm EST	**Decentralized Apps, the Metaverse and the “Next Big Thing”** Register →
June 30 4 pm EST	**Ethics of the Decentralized Web & Uses for the Law, Journalism and Humanitarian Work** Register →