



# How an Open Science Commons Project Built on Web3 Infrastructure Empowers Community Discovery

Brought to you by  
 Filecoin  
Foundation



## Key Highlights

- Opscientia (OpSci) is an **open source DeSci project** leveraging the power of Web3 to grow, sustain, and continuously improve a decentralized science movement to drive progress in scientific research, making it easier for scientists around the world to share data. The project is focused on data sharing, research fellowship, community building, and automation for knowledge curation.
- The OpSci community's Commons project is a data commons using [Filecoin Plus](#) to provide **free archival, indexing, search, and discovery for neuroimaging datasets**. Commons is a permanent home for large open access data repositories such as Open Neuro, ABIDE, and Neurovault; with more being indexed.
- In exchange for storing their scientific data in OpSci's free archives, users must prove academic credentials on-chain using [Verse](#), an OpSci dApp. Published data must pass an automated validation that checks whether data conforms to a reproducibility standard. The goal of the OpSci project is to **incentivize reproducible data publishing with impact certificates** to fund researchers that perform open science practices.

- OpSci contributors have indexed 500TB of high quality open access data into the project's archives via the Filecoin network. The community has **grown to almost 700 contributors**, where 40% self-identify as researchers.
- OpSci is designed to facilitate a peer-to-peer science data commons for web-native scientists. The Filecoin network, IPFS, and other decentralized web technologies are key components of this vision.
- OpSci was one of the **earliest DeSci DAOs** with a strong academic contributing core. OpSci has been utilized to fund a number of community events, starting with the Open Web Fellowship, the first DeSci conference at ethDenver, and later helping prop up the DeSci.eth community fund.

***"Thousands of petabytes of valuable data and observations on human health, economic activity, social dynamics, and the universe and our impact on it are siloed in outdated storage systems."***

## Designing Incentives for Open Source Practice

OpSci is a **Decentralized Autonomous Organization, or DAO**, founded by neuroscientist and researcher Shady El Damaty, Ph.D. A data engineer by necessity, El Damaty has experienced firsthand the cumbersome process of locating, accessing, downloading, and working with neuroimaging data in collaborative digital spaces.

Frustrated by the state of software tools for web-native scientists, he set off with a mission to **nucleate a community of scientists and web developers** to co-develop solutions that would unlock data silos, encourage collaboration, and streamline science funding. OpSci addresses key pain points experienced by scientists through three projects:

**1) Verse:** a decentralized science identity directory that ports off-chain credentials with DeSci dApps,

**2) Fellowship:** a permissionless research fellowship program that matches trainees with mentors, and

**3) Commons:** a data commons built with incentives for sharing high-value and reproducible scientific datasets using the Filecoin network and IPFS.

OpSci Society is an autonomous scientific community that **mints NFTs to track scientific impact following the Hyper Certificate specification**. These on-chain discoverable and non-transferable tokens will live on impact marketplaces, and funders can browse the activities of verified academic identities, fellowship projects, and published datasets. In this way, individuals, proposals, completed projects, and published datasets can be funded both proactively and retroactively based on their history of impact and expectations for future contributions.

OpSci is architected to create a digital society of web-native scientists that **steward automated knowledge foundries and funding is directly linked to verifiable scientific impact**. OpSci's vision is to bootstrap tools where any community can bootstrap a scientific society of experts that collectively govern over standards, choose their criteria for rigor and excellence, and own their own data.

### The Challenge: “Science is Broken”

The productivity of web-native scientists is limited by availability of data, reproducibility of results, poor incentives for collaboration, and non-existent coordination tooling. Data shows that **less than 10% of data collected by publicly funded research in the U.S. is published** alongside their papers and, on average, 70% of researchers have been unable to replicate their colleagues' work across disciplines, including physics, biology, and psychology. El Damaty writes:

# Today's centralization of scientific research and development data is hindering society's ability to achieve life-changing and life-saving breakthroughs.

"Thousands of petabytes of valuable data and observations on human health, economic activity, social dynamics, and the universe and our impact on it are siloed in outdated storage systems. This data is inaccessible to search engines, stored in arcane schemas known only to a few, and likely never to be utilized. Just think about it – **over 80% of raw scientific data collected from the 1990s is estimated to be lost forever** due to deprecated technology and inadequate archival infrastructure. Today, the likelihood of finding a dataset falls by 17% year-after-year, beginning three years after a paper is published."

Knowledge communities are often isolated within institutional data spaces. Because of this, scientific findings do not replicate in real-world applications and, most importantly, the economic value from knowledge creation does not flow back to scientific communities. All of this has led to a broken, hyper-competitive culture in academia that precludes collaboration. All the while, **the world has never been more in need of immediate scientific solutions.**

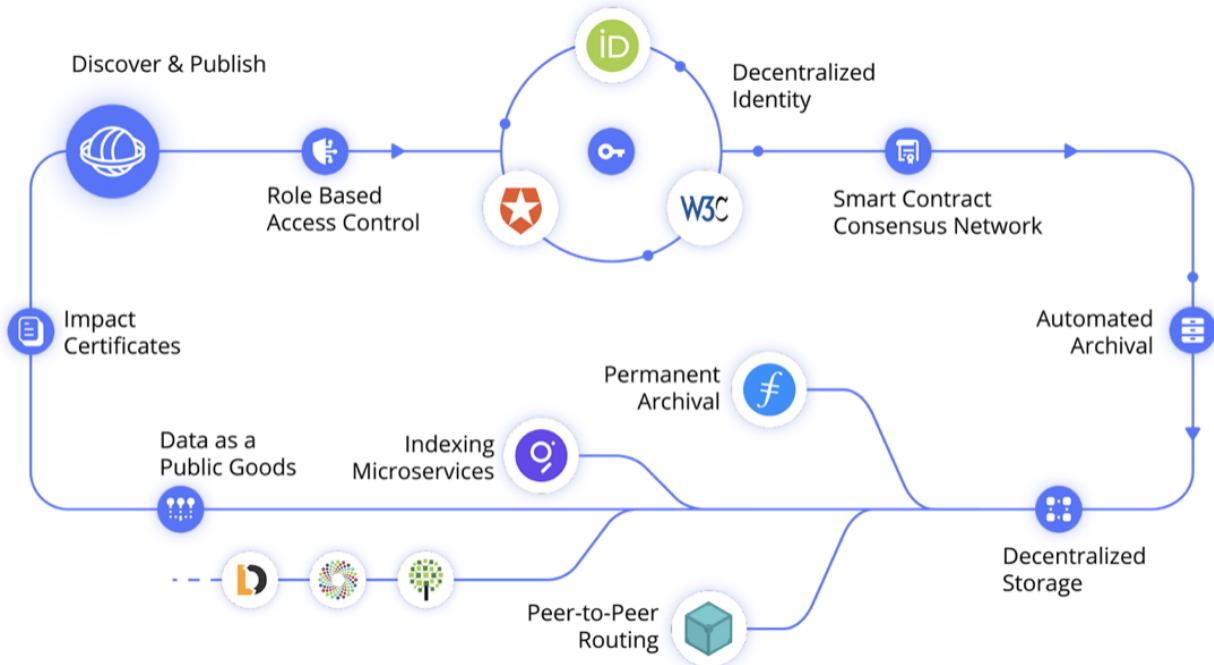
And the disparities are growing.

The current system creates exclusive communities of researchers at institutions with proprietary instruments, siloed data, and specialized infrastructure models that are monopolizing modern scientific insights. Citizen scientists outside of the status quo — the grassroots of the scientific community — are denied access to instruments of information-gathering and coordination. Today's centralization of scientific research and development data is **hindering society's ability to achieve life-changing and life-saving breakthroughs.**

***"The practice of deliberately restricting access to scientific data limits our society's rate of innovation precisely when we have never had so many problems that require scientific innovation to solve" – Shady El Damaty, Ph.D., Founder of Opscientia***

While centralization was critical for achieving national strategic goals like NASA space launches or expedited vaccine development, the current system does not scale to meet the challenges of today. **OpSci can close the gap as an environment for decentralized, transparent, open source, immutable, community-owned, and governed applications.** These tools can be used to empower users with private virtual identities for self-sovereignty over their data.

# Opscientia Ecosystem



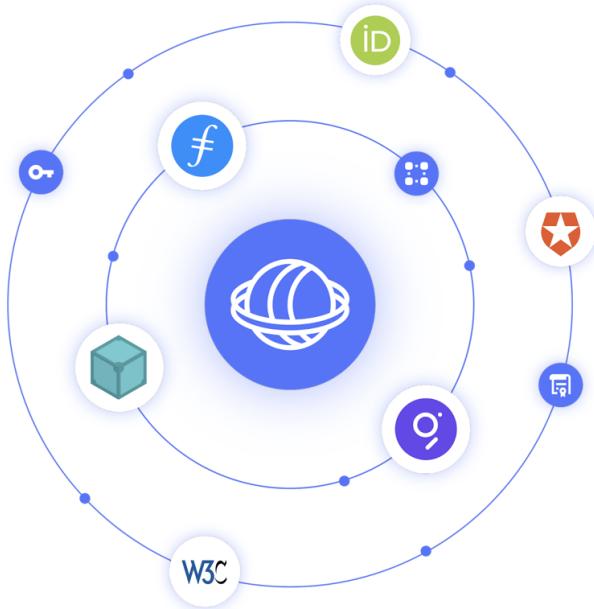
## The Solution: A Web3 Data Commons for Science

OpSci Commons is an interoperable federation and data commons for science. Commons components are intended to serve as modules that can be used by other developers to create science web-apps running on an open data layer. As a central part of the stack, **IPFS and the Filecoin network provide content-addressability, programmable incentives for data storage**, and bandwidth that scales with global adoption.

OpSci's peer-to-peer science data commons powered by Filecoin can provide a resilient digital fabric that aligns a decentralized community around reproducible science practices. The project has already crawled and **indexed more than 500TB of Web2 data onto the Filecoin network** and has built a pipeline to archive more open access datasets.

OpSci contributors discovered during archival that some datasets were “falling off the map,” disappearing from their original source, due to the high transiency of the web. As the Commons archival pipeline chugs along, we can expect that the **Filecoin network becomes the *de facto* archive** for copies of open science data with no active maintainer due to slashed funds or infrastructure failures.

OpSci’s public data commons runs on Estuary, a tool to upload public data onto the Filecoin network and pin it to IPFS, promoting easier access for individuals on the front end. **Users running their own IPFS node can upload public data onto the Filecoin network** and pin it to IPFS themselves. This type of decentralized, peer-to-peer dynamic for data creators and collaborators opens the door for a more practical and effective data-sharing platform for scientific web applications such as compute markets, permissionless journals, and data lockers.



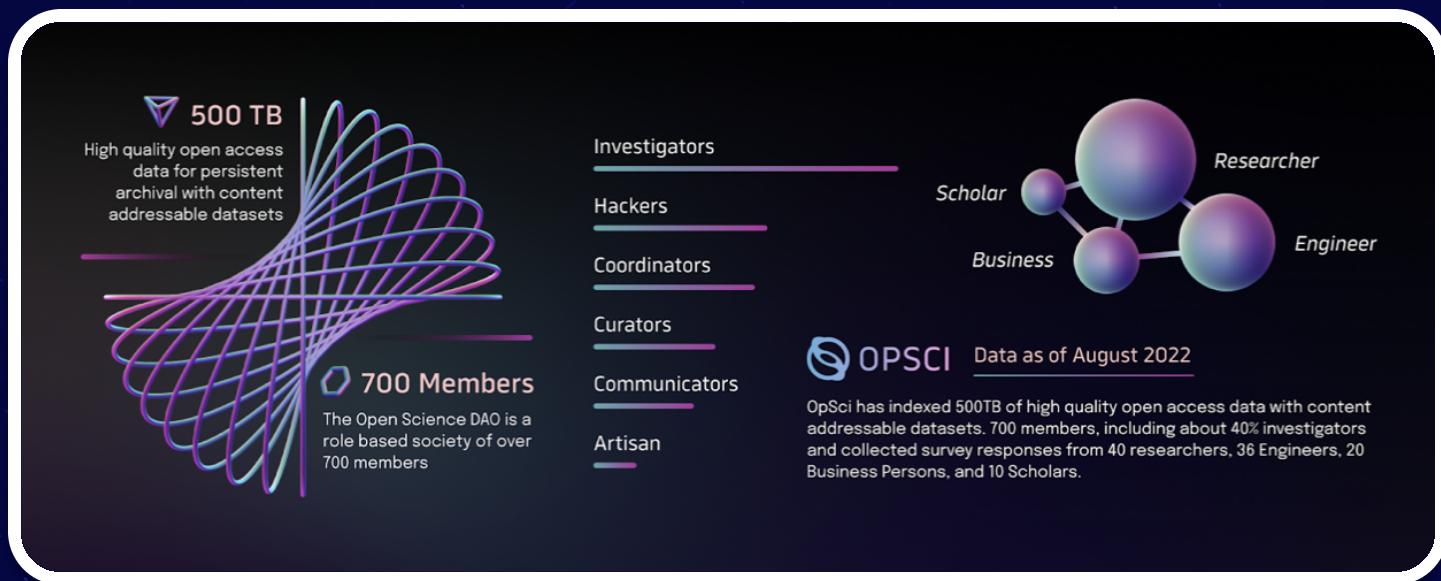
## Looking ahead

OpSci contributors are working to scale the project to include individual research labs with extra storage hardware to run Estuary nodes and join the network as a node. Contributors are also focused on developing ways to collect funds that can be redistributed back to OpSci researchers and programs on the decentralized platform.

**“Research labs generally want to be able to allow others to work with their data sets, so the ability to provide views on top of that data through science web apps for unaffiliated researchers will be an absolute game changer by unlocking those silos for data” – El Damaty**

Estuary helps create a more accessible interface to store data on the Filecoin network, and makes storing data simpler by managing replication tasks and automating much of the workflow.

Additionally, once **OpSci incorporates production-ready data layers**, others can add “application layers” to operate on top of the data and provide new views. DeSci builders can create functions like in-browser 3D rendering, or hooks into interactive digital lab notebooks, where anyone with a verified pseudonymous science identity can import a dataset and run a verifiable compute job on it.



## The Results

As of August 2022, **OpSci has indexed 500TB of high-quality, open access data with content addressable datasets**. The community has grown to almost 700 members, including about 40% that self-identify as researchers, and collected 114 interviews and/or survey responses from researchers. Additionally, 53 academic identities have been verified using OpSci's Verse platform.

## What's Next?

OpSci is Web3's first decentralized data commons and onboarding hundreds of terabytes of open access data. Contributors are working with researchers to build a **free solution for depositing data at the end of a grant**. For example, the National Institute of Health has a data sharing mandate for all grants they fund but researchers don't have an easy way to satisfy this directive yet.

OpSci contributors are also exploring ways to continue advancing the decentralized research network that's formed within the wider DeSci community by scaling up the decentralized research fellowship program, impact certificates, and more sophisticated mentor/mentee matching.

Finally, **OpSci will transition into a Professional Web3 DeSci Society**. This will center around establishing open science practices serving as an entry point to scientists, setting standards in the context of dWeb technology, exploring novel funding mechanisms, conducting verifiable and reproducible research projects on OpSci Commons, and developing research proposals by fellows.



# How an Open Science Commons Project Built on Web3 Infrastructure Empowers Community Discovery

To learn more about how to get involved with OpSci,  
please visit **opsci.io**.

Learn more about the Filecoin Ecosystem

