# Workshop on Open Geospatial Science and the Decentralized Geospatial Web: Summary & Call to Action

**Background & Motivation**

Approximately 30 specialists gathered in person on the 3rd and 4th of April 2024 at the University of Maryland, College Park to participate in a workshop on open geospatial science and the decentralized geospatial web. The event was organized by the EASIER Data Initiative and hosted by the Department of Geographical Sciences, with support from the International Center for Innovation in Geospatial Analysis and Earth Observation (the GeoCenter) and the Filecoin Foundation for the Decentralized Web (FFDW). Across the two days, there was a series of presentations, short lightning talks, discussion sessions, and networking breaks. Participants represented three broad areas, including open science, geospatial technology, and the decentralized web, as well as those whose specialization overlapped in one or more of these areas. Figure 1 presents a visualization of these main workshop themes, which were central to motivating and organizing the event.

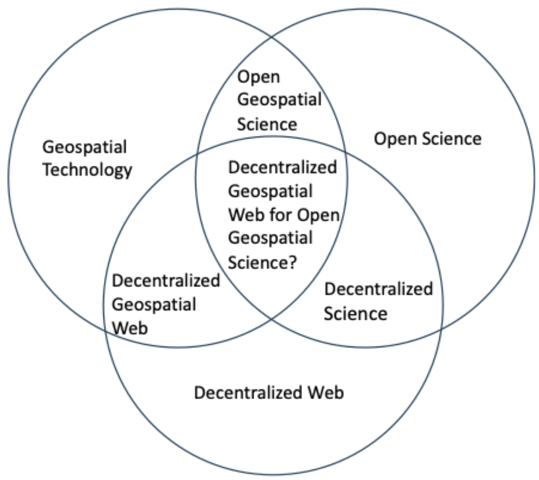


Figure 1: Workshop themes.

Each of the three themes relates to a set of technologies, protocols, practices, and tools. For example, geospatial web technology has been developed over several decades and includes an array of conventions that allow geospatial data and services to be stored, shared, and communicated through the web. Open science has emerged as a pillar of modern research for increasing the accessibility of science and the associated artifacts, often to increase reproducibility and replicability. Meanwhile, the decentralized web consists of a relatively new collection of technologies and protocols for providing or consuming information without centralized infrastructure or control to promote digital freedom and prevent single points of failure.

The motivation and organization of the event are further brought into focus by examining the overlapping sections of the circles in Figure 1, which highlights topics that emerge from the interconnection of these three themes. Open geospatial science is enabled by the cross-fertilization of geospatial technology (and of course geospatial science) and open science. Web technologies such as cloud-optimized file formats for large datasets (i.e., satellite images), mapping services, and spatiotemporal metadata allow geospatial scientific artifacts to be preserved and shared. At the same time, the geospatial sciences are actively adopting and adapting tools and practices from the open science paradigm, such as version-controlled repositories, interactive notebooks, and open-access publishing. These trends underlying open geospatial science are perhaps the most established portion (top) of Figure 1, while bringing the relatively recent decentralized web paradigm (bottom) into the fold yields two more emerging areas of interest. One is decentralized science (bottom right), and the other is the decentralized geospatial web (bottom left). Decentralized science is a recent movement to incorporate decentralized technology into the open science stack and reimagine aspects of the scientific enterprise, such as funding mechanisms and data stewardship. In comparison, the decentralized geospatial web is perhaps even more nascent and is a term that has only been introduced in the last year or so despite the existence of some projects that have been conducting work in the area for some time. At present, the decentralized geospatial web may be best characterized as a loose amalgamation of efforts to make location information more accessible, trustworthy, and persistent through decentralized technology.

In particular, the notion of a distinct decentralized geospatial web is something the EASIER Data Initiative has been advocating for, both through its own projects and by highlighting other projects, in order to help form an ecosystem at this nexus. It is this energy that the workshop intended to illuminate in tandem with the decentralized science and open geospatial science movements and their constituent components. All three of these topics are at the center of Figure 1 and culminate in the question of how a decentralized geospatial web can support open geospatial science. However, before that question can be answered, a foundation was needed in each of the three main themes and their cross-fertilizations. The workshop started to lay this foundation and catalyze discussions and connections.

More specifically, the goals of the workshop were fourfold. First, to provide an overview in the above areas to the participants and those interested in later following the content. This was done through a series of presentations and lightning talks from participants, some of whom were experts in one area, others who were working at the overlap of two areas, and a handful with some experience in all three. Many of the presentations were recorded and are available online[1]. The second goal was to build on this knowledge exchange through discussions across areas to fortify existing connections, build new bridges, and identify common challenges. A third goal was to begin bringing the various areas together to assess the state of the emerging decentralized geospatial web and how it might support open geospatial science. Finally, the fourth goal was to initialize a community for these topics to continue melding and to spark new collaborations within these spaces going forward. To help guide the workshop, some additional background context and overarching questions were developed ahead of time and shared with participants, which are provided below.

*Open science has emerged as a top priority for ensuring the robustness of the scientific enterprise, especially as research becomes increasingly dependent on computational workflows and large heterogeneous data sources. This is especially true for the geospatial sciences that are dependent upon massive amounts of location-based data, such as remotely sensed images, demographic profiles, environmental field samples, climate simulations, land parcel and building footprint records, the status of telecommunication and transportation infrastructure, and geotagged social media posts. Data sources are continuously being combined and accumulated to conduct complex analyses in support of decision-making where location matters. Some examples include responding to natural disasters and extreme weather, developing climate change mitigation and adaptation strategies, promoting sustainable, just, and optimized cities, and ensuring national security.*

*The full spectrum of "openness" is diverse and typically includes the transparent use of data and methods, access to resources and training, and inclusivity within the scientific process. Furthermore, the openness of science*

[1] https://www.youtube.com/playlist?list=PLPTtwBCoBxZuaj0xPcvkTIBzBeHPJyPBj

*is directly related to the important issues of reproducibility and replicability (R&R). A lack of R&R is associated with: (i) the inability to access the resources necessary for computational data-intensive workflows; (ii) the inability to interact with research artifacts for veracity and educational purposes; (iii) the inability to obtain the same or similar results for the same or similar data and methods. One result is that the pace of scientific progress is stifled because resources are not widely available, and it is challenging to learn from and build on one another's work. A recent report by the US National Academy of Sciences underscores the gravity of the issue and recent attention has highlighted the additional complexity of R&R in the geographical sciences. Great efforts have been put towards increasing R&R in the geospatial sciences and science more generally, yet the exponential growth in both the volume of data and the demand for computing resources raises new challenges for ensuring the geospatial sciences remain open.*

*In particular, decentralized practices and technology, such as web protocols, blockchain, and smart contracts, can increase the openness of the geospatial sciences and facilitate reproducibility and replicability. For instance, content-addressing helps avoid 'link rot' by allowing files to be referenced by a unique hash based on the content of a file itself, rather than location, so that the file can be sustainably referenced on the web regardless of where it is stored. Content-addressing can also ensure file integrity because a change in a file also changes the hash used to reference it. Distributed systems for sharing files based on content-addressing facilitate a more connected and optimal user-to-user network of content delivery rather than a hub-and-spoke system of file requests. Meanwhile, blockchain and smart contracts make it possible to incorporate transparent, trustless, and consensus-based layers into the data ecosystem. However, many decentralized technologies are still nascent and their adoption, especially in the geospatial domain, is still limited. Thus, an important and timely question is "How might an emerging decentralized geospatial web support the geospatial sciences to become more open, transparent, reproducible, and robust?".*

*Towards answering this question, this workshop will bring together domain experts from the geospatial sciences, open science best-practices, and the decentralized web ecosystem to better understand the current state-of-the-art and how these three areas can come together to imagine possible future pathways for the next generation of open geospatial science.*

*Some focus questions for the event include:*
*1. What is the current state-of-the-art in open science practices?*
*2. What are the unique challenges and opportunities for making geospatial science and technology more open?*
*3. How can the decentralized web contribute to open science and society more broadly and specifically the geospatial sciences and aspects of society?*
*4. What are the primary hurdles that need to be overcome to develop a decentralized geospatial web in support of open geospatial science and technology?*

## Summary of Content

To kick off the first day of the event, the first two presentations started with some core decentralized web concepts and potential applications for geospatial data. Carson Farmer (Textile) introduced decentralization and decentralized systems to provide an overview of some opportunities associated with the decentralized web and how they can provide a path to a more trustworthy and permissionless geospatial web. Next, Volker Mische (independent researcher, previously Protocol Labs) introduced content-addressing in comparison to location-addressing, which is a core technology underlying the decentralized web. The presentation highlighted how content-addressing facilitates data persistence and verification and what that could mean for geospatial data, such as satellite images of Earth's surface that are composed of trillions of pixels and constantly changing over time.

Shifting gears towards highlighting some organizations and initiatives, Rachel Opitz from the Open Geospatial Consortium (OGC) shared how the organization creates innovation through standards and community-building. This includes maintaining findable, accessible, interoperable, and reusable (FAIR) data, creating portable and transparent services, and assuring the integrity and truth of data and services. In particular, the Open Science Persistent Demonstrator (OSPD) is an OGC sponsored collaborative project working to connect geospatial platforms and enable unified workflows across platforms. The Filecoin Foundation for the Decentralized Web was then introduced by Brynn O'Donnell, which is working to preserve the digital commons and demonstrate the benefits of decentralized storage, with a focus on global needs. The presentation evoked questions about data access and ownership, stressing the need for resilient infrastructure that better serves society. Next, Juan Fernando Martinez (Columbia University) provided an overview of the NASA Transform to Open Science (TOPS) initiative to accelerate scientific discoveries by broadening participation and facilitating the adoption of open science best practices. TOPS is cultivating a broad open science curriculum with the objective of incorporating Earth science data and geospatial applications.

Continuing the focus on the scientific enterprise, Eric Olsen discussed how the Center for Open Science (COS) is working to increase the integrity and reproducibility of research through policy and infrastructure, while Erik Van Winkle (DeSci Labs) shared ongoing work to develop a decentralized persistent identifier (dPID). Both presentations recounted challenges and pitfalls with the current scientific workflow, along with tools and practices that can help make scientific artifacts more robust and interactive. Through new incentives, technology, and culture, it becomes possible to envision a more open and sustainable pipeline for conducting science, sharing results, and evaluating evidence over time.

The remainder of the first day consisted of examples of efforts to create and use open geospatial data, services, and tools, demonstrating their importance in science and society. First, Douglas Rao (NC State University) discussed the need for AI-ready data and the role of open geospatial data and providers in the pipeline to clean, curate, and annotate datasets. Community-driven efforts through the Earth Science Information Partners (ESIP) are enabling collaboration to improve the quality, accessibility, and documentation of AI-ready datasets. In addition, Qiusheng Wu contributed some demonstrations of open source tools for interacting (e.g., accessing, processing, analyzing, exporting) geospatial data on the web with a focus on satellite imagery. Felipe Montealegre-Mora (UC, Berkeley) also contributed a synopsis of how the open science toolkit can be used to investigate environmental justice, which provides tutorials for learning and teaching as part of the NASA TOPS. Subsequently, Christopher Tucker (GeoRobotix) led an overview of OpenSensorHub, providing an example of an OGC API in the wild. Leveraging the Connected Systems API, OpenSensorHub facilitates the connection and integration of distributed sensors, their deployment, and the associated data streams. Ryan King (FOAM.space) then introduced the FOAM crypto-native protocol for proof-of-location, which uses a decentralized network of radio beacons for conducting geolocation. This provides secure location proofs and an alternative to GPS that overcomes some of its limitations. Complementing this, Kiersten Jowett (University of Melbourne) also contributed an outline of proof-of-location concepts and technology and a summary of ongoing research to understand the challenges and opportunities of decentralized proof-of-location.

During the second day, activities started with additional talks pertaining to open geospatial science issues and examples. First, Shayna Solis (Navteca) presented on the need for conversational AI for geospatial science applications. One example involved asking conversational questions about the Parker Solar Probe, which is connected to a real-time mission API, and another example included being able to ask questions about information from across multiple sources while visualizing real-time data in the context of emergency response and disaster relief. Second, Ziheng Sun (George Mason University) highlighted the need for workflow reproducibility and introduced Geoweaver as a new tool for increasing scientific productivity within teams. Third, Anna Burzykowska (European Space Agency) shared an overview of issues and applications of distributed ledger technologies for Earth observation. In particular, there is a need to account for the provenance chain, including the full manifest of data, when using AI in order to protect against deepfakes, processing errors, and data corruption. Though blockchains can provide permanent records and a trust layer for sharing information, they may not always be required or appropriate, such as when it is not straightforward and timely to build functionality around them or when central authority is considered a system or organizational feature.

The final stretch of the workshop included a series of lightning talks to outline ongoing work to build decentralized geospatial web capabilities. Zheng Liu (University of Maryland) highlighted work by the EASIER Data Initiative to merge the hierarchical data structures used on the decentralized web with the hierarchical nature of discrete global grids to make geospatial data more accessible, robust, and efficient, especially for retrieving relatively small portions of the global datasets that are growing in size and popularity (i.e., Overture Maps points of interest). Then, Matthew Nanas (University of Maryland) shared work by the EASIER Data Initiative to integrate traditional geospatial web metadata and decentralized web metadata to facilitate the discoverability, retrieval, and sharing of geospatial data. This is being achieved through a combination of APIs and a geovisual dashboard that allows users to explore and interact with geospatial data stored on the decentralized web through familiar patterns. Next, Danny Gattas (Astral) introduced efforts to create decentralized spatial registries and proof-of-location protocols to facilitate geospatial computing natively on the decentralized web. These foundational technologies will allow secure and trustless transactions, with applications in shipping and logistics, transportation zoning regulation, and even local governance. Finally, Jefferson Sankara (FFDW) provided updates on the Interplanetary Consensus (IPC), a framework to enhance blockchain transactions for scalable computing and applications on the Filecoin network. Among other things, this will enable the efficient storage and management of geospatial datasets.

**Some Synthesis & Insights**

Through lively discussion sessions and questions and answers after each presentation, there were a number of emerging insights during the workshop. The first of these is that regardless of whether individuals represented the geospatial web, open science, the geospatial technology, or some combination thereof, there seemed a broad consensus that there is an enduring need: (a) to protect data integrity, provenance, and access; and (b) for inclusive, transparent, and reproducible science. It was perhaps inspiring on the one hand to experience such agreement between the various areas in Figure 1 and understand all the work being done towards these goals but also sobering on the other hand to understand how much still needs to be done on this front to bring it all together. This goal is perhaps more important than ever within the context of the rapidly advancing artificial intelligence revolution.

Another topic across the two days was an attempt to unpack some of the potential advantages and challenges of the decentralized web. One of the central features of the decentralized web is the ability to make interactions permissionless – no centralized approval is typically needed to participate in decentralized systems, and it therefore advocates for the ability of individuals to freely echo information or data that they find valuable. This freedom is often restricted in reality by financial burdens, technological limitations, intellectual property laws, culture, and power dynamics. For example, in science, data is often not shared in order to protect resources or out of fear of losing credit. Meanwhile, many publications are not owned and controlled by scientists, making it difficult for them to disseminate their work, make incremental revisions, or even retract a contribution. Other times it could be too expensive in terms of cost or time to curate and preserve scientific artifacts. The decentralized web does not offer immediate solutions to all of these issues, but it does provide some technological innovations and alternative ideas about how to organize people and interactions that could improve science. Peer-to-peer sharing systems, hashing, and distributed ledger technologies are examples of technologies that allow information to flow more freely. Decentralized autonomous organizations (DAOs) build on these technologies and incorporate consensus mechanisms to facilitate community-based management of resources. It becomes possible to envision the maintenance of datasets and even entire research projects without one individual holding all the power or responsibility. A related thought experiment might wonder what science might look like if all scientists were allowed to freely share or endorse data, publications, or communications without any restrictions, but they could sponsor only a limited quantity of content. We might see the most important and impactful assets emerge in contrast to their noisier and less innovative counterparts while encouraging fewer overall contributions, both of which could be beneficial for science and academia. Of course, this raises important questions about who gets to be a scientist, how impact is defined, and what is considered valuable.

These discussions also sparked debate about how decentralized systems could also be used for spreading misinformation and reinforced the need to complement them with authoritative sources of information. Analogies were drawn to the proliferation of GIS software and browser-based cartography tools that have made it easier than ever for novice map production. This has vastly democratized mapping, though there still exists the need for authoritative maps, such as official evacuation routes or weather forecasts. OpenStreetMap was cited as an example from the geospatial realm where data is collected in a decentralized manner (i.e., crowdsourced) and while the individual data points may contain some error, bias, or misinformation, it has nevertheless produced a data source that is typically trustworthy overall and a valuable source of information, especially when no authoritative sources are available.

The example of OpenStreetMap provides an excellent segue to another notion that arose repeatedly, which is that it is useful to think of decentralization as a spectrum rather than a binary attribute of systems. For example, OpenStreetMap provides an example of decentralized data collection and maintenance. However, centralized technology may be used for storage and dissemination, and the project governance is neither strictly centralized nor decentralized (there is an overarching foundation and local chapters). This highlights a need to think more generally about the entire data lifecycle and critically evaluate the degree of decentralization that might be preferable, useful, or feasible in each circumstance. In some instances, a fully decentralized system, including technology, organization, and governance, may be an option while in others, some centralization may play an important role. And it is vital to consider the range of possibilities in between and through different combinations of these dimensions. Adopting a spectrum perspective could also prove useful in science and research where publishing and retraction typically exist in a rigid binary state due to the centralized organizations and processes underlying academic publishing and institutional incentives. Scientific contributions could alternatively be assessed continuously and in a modular fashion through systems of attestations and partial approvals that allow work to incrementally gain or lose credit over time. This type of system could perhaps help create a more flexible and equitable scientific enterprise, operating instead of or in tandem with the traditional system.

There were also some general challenges that were identified that surfaced. One such challenge is how to facilitate the adoption of new tools and practices as academia typically evolves relatively slowly and there are a host of reasons that researchers might be hesitant. This includes the resource investment needed to learn new workflows, the desire to prioritize new scientific inquiries, or apprehension to share artifacts such as code that have not been professionally developed. Another challenge is that technology often moves faster than policy can be developed to help incorporate innovations. Deliberate incentives are required to overcome these challenges, and this again raises the need to design applications and tools for open science that incorporate new technology into familiar interfaces.

**Final Remarks**

This workshop provided the scaffolding to bring various topics together and lay the foundation for understanding the decentralized geospatial web and how it might support open geospatial science. It served as a chance for various groups to connect and educate each other, which is significant because there was previously limited interaction despite the existence of some overlapping goals and values. It is therefore important to capitalize on this momentum and provide additional opportunities for these communities to continue interacting and deepening the initial connections established here.

The decentralized web seems poised to make contributions to the entire scientific workflow, from data collection, sharing, and storage to publishing and peer review. For example, the built-in verifiability afforded by hashing has clear implications for defending against data corruption and this is especially important for sensitive geospatial data. Similarly, the benefits of distributed ledgers like blockchain may play a central role in documenting geospatial data provenance as it is collected and consumed by users and algorithms. More decentralized systems for establishing credibility and reputation could also be helpful in addition to these technologies. Some downstream benefits have already been alluded to, including the democratization of scientific inquiry, more flexible and adaptive dissemination, and increasing the robustness, accessibility, and trustworthiness of data systems. Since geospatial data makes up a large portion of all data, there is a need for a distinct decentralized geospatial web to support these efforts.

At least two directions were identified that could be constructive towards fortifying the decentralized geospatial web ecosystem and expanding the community. First, a focus on developing and adopting a core set of technologies, tools, and protocols to mirror, complement, or extend existing geospatial technology. The geospatial web is already mature and comprehensive, making it prudent to build directly upon it where possible. To the extent that geospatial web users can participate in the decentralized web using familiar interfaces, this greatly increases the potential to bridge the two communities and is a strategy that the EASIER Data Initiative has been pursuing. Of course, it is also necessary to build completely new technology that is native to the decentralized web. In this regard, a very exciting feature of the decentralized geospatial web is the development of more diverse, secure, and flexible methods for proof-of-location. Second, there is a strong need to educate and conduct outreach to those outside of the workshop participants, both within the identified circles and beyond. It is critical to be able to build the decentralized geospatial web with input from diverse participants. One of the most important tasks is to continue disseminating the innovations of the decentralized web. And at the same time, it is perhaps equally important to incorporate the expertise of geospatial technologists and open science advocates. Without pursuing these bridges and expanding the community, it could be challenging to see these innovations adopted and the benefits of the decentralized web realized. We therefore plan to continue pushing the boundaries and invite those interested in learning more or contributing to reach out and join us!