



Leveraging Filecoin Network for Greater
Visibility, Access, and Storage of
Genomic Data



Brought to you by
**Filecoin
Foundation**



Key Highlights

- GenRAIT provides scientists with the platform they need to analyze genomic data, generate scientific breakthroughs, and experiment at scale, all in one unified ecosystem. In the GenRAIT ecosystem, scientists retain full autonomy and flexibility with regards to their data and analytics processes.
- GenRAIT will initially host over 100 terabytes of genome research data on the Filecoin network and enable scientists and researchers to upload their own genomic files they then can retrieve, analyze, and put into action.
- GenRAIT currently uses the open-source interface Estuary to upload genomic data to the Filecoin network, leveraging the power of blockchain-based technology as a cost-effective storage solution. For GenRAIT, the Filecoin network is a Web3 solution that will scale with its data.

Data storage needs in the biotech space

In the 20 years since the first mapping of the human genome, there has been an explosion of genomic data, with an estimated 40 billion gigabytes of genomic data generated every year. Data centers have struggled to accommodate these high volumes of data and continue to run up against high storage capacity levels and skyrocketing prices. Several scientific giants have explored how to handle this enormous stockpile of information, but the issue remains unsolved – there is no unified ecosystem that enables scientific enterprises to securely store, access, share, analyze, and use the unique data in this vertical.

Founded in December 2021 by Taylor Capito, Razib Khan and Santanu Das, GenRAIT is a platform that puts the architecture, tools, and data scientists need at their fingertips in order to analyze genome information, make scientific breakthroughs, and experiment at scale. This platform serves as an ecosystem for scientists and researchers with the end-to-end experience, covering data management through advanced analytics, and the GenRAIT platform provides autonomy unlike other analytics platforms available today. By leveraging decentralized storage and other web3 technologies, the GenRAIT architecture promotes and enables users to securely and scalably share data across the community.

The GenRAIT team envisions a world in which genomic data can be used by everyone as an integral part of healthcare throughout all of the major life moments: fertility, birth, illness, prevention, cures, and death. The company's leadership team believes that genomics can unlock the future of personalized medicine and wellbeing, and that GenRAIT provides a solution that opens the door for wider access to genomic data. Unlike existing solutions, GenRAIT isn't leaning into a black-boxed approach where researchers lose authority over their data once it's uploaded. On the GenRAIT platform, all users maintain autonomy throughout the process so that the data remains accessible and actionable.

The amount of data in the scientific community is exponentially growing, and each human genome is tens – if not hundreds – of gigabytes per individual file, forcing scientists to compress the raw files or only save a subset of the genome – which is a substantial loss. Often, in order to use the same data, scientists have to resequence (if they're lucky enough to still have a sample), or the data has to be thrown out.

By some estimates, 80% of a researcher's time is often lost to "data wrangling."

By some estimates, 80% of a researcher's time is often lost to "data wrangling," leaving only 20% to conduct research and analysis, so a large focus of GenRAIT's platform is providing a one-stop shop to store that data. The GenRAIT founders believe that by providing the tools scientists need to store and analyze the data, scientists can focus on the research and innovation.

Enabling access to proactive information rather than reactive

Taylor Capito, CEO and co-founder of GenRAIT, was inspired to make a change in the genomics industry after a series of health scares at a young age. She grappled with inflammatory health and autoimmune issues for years. Over the course of four orthopedic surgeries by the age of 20 and multiple hospital visits, Capito took it upon herself to do some research about why this was happening. After many long hours meeting with doctors and doing her own research and testing, she formed hypotheses on how her genes interacted with certain triggers. She undertook corrective lifestyle changes to improve her health. It worked, and some eight years later, she is healthy, her test results are normal and her outlook on life has completely changed.

Capito feels that she had access that others do not due to her scientific background, research experience, and family of physicians. And, while she was able to go through her data herself, she knows there are others with similar health issues who haven't been so lucky. After Capito proactively took corrective action for her condition, she realized the vast potential for individuals to learn more about their health challenges if the power of genomics data could be unlocked.

Drawing from her experience in the technology industry, Capito wanted to make a difference and change the course for predictive and preventive medicine. She recognizes the power genomic data holds, and intends for GenRAIT to unlock that potential for scientists. To do this, Taylor also plans to make the data more accessible. This is where Web3 comes in.

Enabling secure storage detection

GenRAIT built its own Web3 management layer, but found that traditional cloud storage providers carry a high price for storing hundreds and hundreds of terabytes of data, not to mention the large egress fees. Searching for a Web3-based solution, GenRAIT found the Filecoin network. The founders immediately saw that storing on the Filecoin network was affordable, even for biotech-size files, and a promising solution for storing this life-altering data.

Kicking off its work with decentralized storage, GenRAIT uploaded non-sensitive community data to the Filecoin network, which it will later expand to more high-stakes genomic research data. The GenRAIT platform is designed to save scientists and researchers time and money, while providing end-to-end security and accessibility. The Filecoin network will be GenRAIT's default archive for non-sensitive genomics and scientific research data, providing a solution for the company to store the data at scale.

"We want scientists to keep their data and not have to delete it; and we also want to create a system to empower them to use their data, not disincentivize the use of genome data like so many of the current solutions do," says Capito. "To enable that system, we see solutions like Filecoin as groundbreaking for scaling the storage of large data files like the human genome."

"To enable that system, we see solutions like Filecoin as groundbreaking for scaling the storage of large data files like the human genome."

GenRAIT is currently onboarding over 100 terabytes of data on the Filecoin network and hopes that, in the future, individuals and healthcare providers can upload files that they can then retrieve and put into action. This first 100 terabytes is just the beginning of GenRAIT's work leveraging the Filecoin network, and GenRAIT predicts there will be much more after that.

Capito says: "We have enabled a seamless data ingestion, with communication between customers, Filecoin, and the cloud, in a way that's secure and encrypted. We see storage solutions like Filecoin as enablers of a future where data can grow. We believe that researchers need such a modernized storage solution to incentivize them to store their data enabling them to use it properly for the richest and most robust experimentation. We see a strong future for the Filecoin network and think that the future is bright for Web3 with its scaling potential for data storage."

Meet The Team

GenRAIT has a founding team with a diverse set of complementary skills that are uniquely tailored to this domain covering product, business, science (core customer understanding), and technology. The CTO and CEO of this team have a track record working together and solving interesting problems.



Taylor Capito, CEO

Experienced in identifying gaps and ideating solutions in technology and product development. She has led initiatives in big data and productivity and brought together the right people to build and launch them. As stated in the article above, more than just passion but necessity drives her to solve this problem.



Santanu Das, CTO

Tenured career as an engineering leader in ML/AI and infrastructure with experience at NASA, Verizon, Palo Alto Networks, and Cisco. Santanu has led teams, built products, and architected systems all with a focus on discovery. He is passionate about data and discovery, and he wants to democratize access to the relevant data and technologies scientists need, empowering them to make breakthroughs.



Razib Khan, CXO

Accomplished geneticist - worked at Embark, Family Tree DNA, and several other genetics enterprises - with large industry reach as a writer and public figure in the genomics community. As a geneticist, Razib has lived through many of the challenges GenRAIT hopes to solve. He believes that solving these issues unlocks a brighter future for humanity.

A Partner For The Long Haul

GenRAIT has already seen success with the Filecoin network for its open-source and nonsensitive scientific research data use cases, and the possibility to now affordably onboard large amounts of data. GenRAIT sees a strong future for growth on the Filecoin network, and other Web3 technologies, as the company moves into more sensitive and healthcare use cases.

To learn more about Filecoin, storage providers and developer use cases, please visit: Filecoin.io or fil.org.

“We continue to find more relevant use cases in the genomics field for hosting data on Filecoin. The potential here is huge and it just shows how decentralized technologies are the future for enabling emerging technologies like genomics to grow and flourish,” says Capito.



Leveraging Filecoin Network
for Greater Visibility, Access,
and Storage of Genomic Data

Learn more about the Filecoin Ecosystem



fil.org



[@FilFoundation](https://twitter.com/FilFoundation)



filecoin.io/slack/