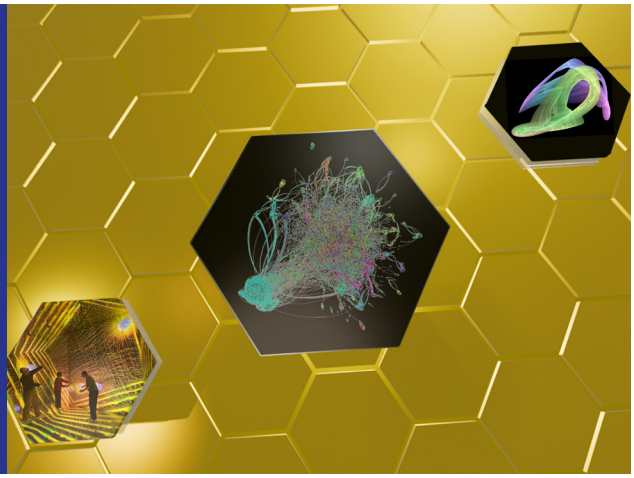
Big Change for Information Exchange

ELEANOR HUTTERER



The archive that rewired research and launched a revolution

In 1991, Los Alamos physicist Paul Ginsparg had a problem with how research findings were shared, or not, within the scientific community. New findings were slow to publish, and access was exclusive, hindering creativity and collaboration. And he wasn't alone; it was increasingly apparent that physicists needed a faster and more equitable way to share results.

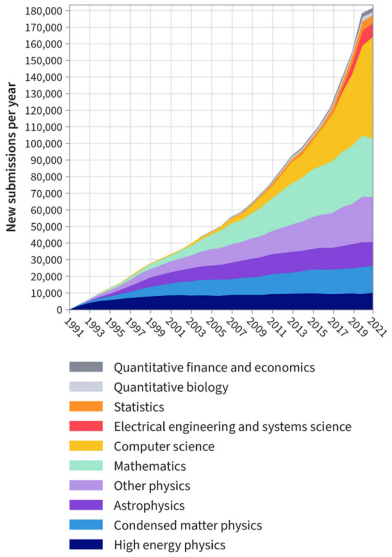
"There was something unfair about a system where there was a privileged loop where people in the know were passing around this information and benefiting from it, whereas people out of this privileged loop, at second tier institutions or abroad, were not able to benefit from these communication channels," Ginsparg recalls.

The internet existed but it was not yet the infrastructure of civilization that it is today; it was a tool for specialists, a network of networks connecting researchers in academia, industry, and government. Its users were people who knew how to use command-line interfaces, mainly scientists, engineers, and computer hobbyists. Research results were shared via email mailing list, then downloaded.

"Disc space on accounts was very limited," says Ginsparg, "so it became natural to think about having some kind of central repository." He was an early adapter of the internet and saw in it an opportunity to create a research-sharing platform that would revolutionize scientific discourse.

At Los Alamos, he built and launched the world's first scientific preprint server, a place on the internet where people could share their work without waiting for the slow and formal process of journal publication. Researchers still needed to publish in journals, but the preprint server let them share their completed manuscripts with colleagues, keeping new ideas moving and fostering collaboration, while waiting for the journal to perform the due diligence known as peer review. Ginsparg called it the "LANL preprint archive," and he expected one submission every three days or so. In the first six months, it got closer to three submissions per day, and what began as a workaround for slow publishing became the foundation of a new paradigm: knowledge as a shared, global resource.

Today it's called "arXiv" (pronounced "archive") and has received



Rise and growth of the arXiv preprint server, 1991–2021, as measured by submission numbers. What started as an online preprint server for physics research quickly became a central hub of science communications across disciplines and now hosts over 2.8 million articles. ArXiv, and the open access movement it sparked, revolutionized how scientific information is shared worldwide and remains an essential tool for discovery. Credit: arXiv.org.

2.8 million submissions over the past 34 years, which averages to 230 a day. Though no longer housed at Los Alamos (the server and Ginsparg moved to Cornell University in 2001), arXiv now hosts a range of scientific topics and remains the central hub for preprint sharing, particularly in physics and mathematics.

The immediate popularity of arXiv proved that scientists were willing to share openly with one another and that self-archiving could work on a global scale. By the late 1990s, there were other online archives popping up, and the open access movement was growing. But each new repository used its own database structure and metadata format, so users couldn't easily search across them or aggregate their contents. There needed to be some standardization, so in 2001, leaders from arXiv along with several likeminded colleagues launched the Open Archives Initiative (OAI), with the objective of defining a simple, universal protocol that any archive could use to manage metadata in a standardized way, thus making all the new digital repositories interoperable.

The OAI's Protocol for Metadata Harvesting (OAI-PMH) became the standard for the open access movement and still runs under the hood of most modern repositories and open science archives. Even as newer technologies slowly replace the OAI-PMH, its legacy persists in the baked-in assumption that metadata should be openly exchanged.

"What began as a workaround for slow publishing became the foundation of a new paradigm: knowledge as a shared, global resource."

"This is a great example of the Lab's role in discovery," says Bill Priedhorsky, retired Los Alamos physicist and Laboratory Fellow. "The Laboratory starts a lot of things, but we don't necessarily finish them—that's not our job. Los Alamos scientists explore the frontiers of their fields, get big new ideas started, then hand things off once they're tractable."

ArXiv democratized access to research results and accelerated scholarly communication, sparking the open access movement. What sets arXiv apart is that it's preprint—straight from the researcher to the scientific community, with formal publication to follow. Submissions are vetted by moderators to ensure they meet a minimum standard but are not evaluated for correctness or scientific rigor. Users accept the mutability of preprint in exchange for speed and transparency.

Many major discoveries from across the sciences first appeared on arXiv. Examples include "Attention Is All You Need," the paper that launched AI into the public consciousness; and "*Observation of Gravitational Waves from a Binary Black Hole Merger*," the paper that announced the first ever observation of ripples in spacetime that Einstein predicted a century earlier. And in time-sensitive scientific circumstances, like the start of the COVID-19 pandemic, when new information was arriving each day, the immediacy of preprint far outstrips the polish of peer review.

Today's internet is rife with archives and repositories, some doing similar jobs to arXiv, but none are doing it better. It's the original. Strong and simple and standing the tests of time, it's the cast iron skillet of scientific communication platforms—there are newer flashier versions, but the original remains the favorite. It's the first place many researchers go when seeking new information.

"It just grew and grew and took over the world," says Priedhorsky. "It's the one archive that rules them all."