

https://www.sfexaminer.com/news/technology/internet-archive-future-brewster-kahle-interview/article_083d42dc-7bf8-40a4-a3c8-ebb4052d1dd5.html

SPOTLIGHT

How the Internet Archive is moving beyond its milestone

By Troy Wolverton | Examiner staff writer
Nov 2, 2025



Internet Archive Chairman Brewster Kahle said the nonprofit institution “[wants] to be there” for people who want to create things.

Sebastiaan ter Burg, CC BY 4.0 via Wikimedia Commons

About 30 years ago, Brewster Kahle spearheaded what many might even now consider an outlandish project: archiving as much of the World Wide Web as possible.

But thanks to the efforts of Kahle and his team at San Francisco's Internet Archive, researchers, reporters and curious citizens can see what the early web looked like, review campaign and corporate promises that have since been removed or significantly revised, and compare changes made to government websites by incoming administrations.

The endeavor has been so successful since its 1996 launch that late last month, the archive celebrated the archiving of its 1 trillionth webpage.

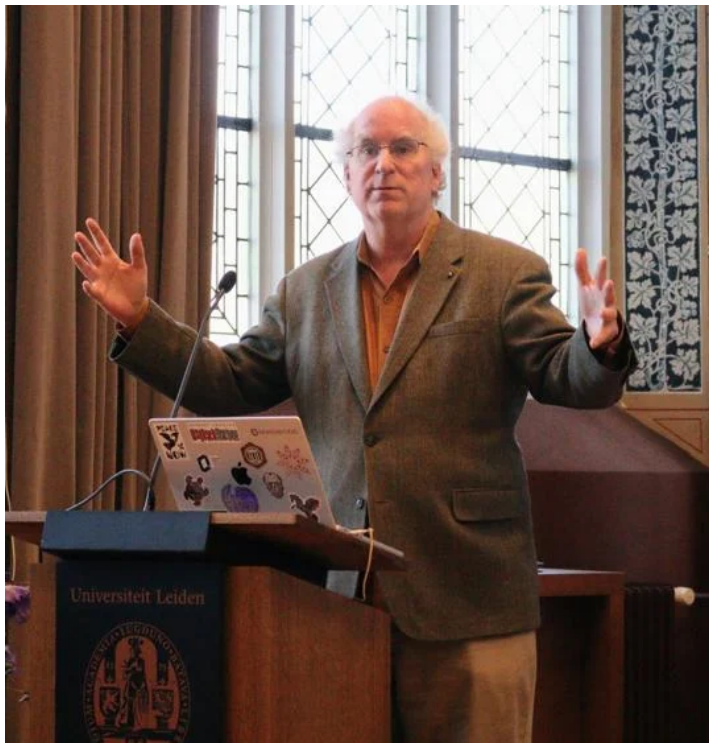
The nonprofit institution that Kahle founded and heads as chairman does more than just preserve the web. It also has extensive digitized collections of books, films, sound recordings, government documents, software and more, derived in many cases from the physical copies of such works sent to it by libraries and other institutions, as well by individuals.

To preserve at least one copy of those physical works it has scanned, the Internet Archive has systematically packaged and catalogued the items, storing them in a collection of warehouses, including in Richmond.

From the perspective of copyright law, the archive's work of wholesale copying of webpages, books and more has always been audacious, legal experts say. In recent years, it's also been contentious. Within the past five years, the institution was sued by both a coalition of major publishing houses and a group of recording-industry titans; it lost the former lawsuit and settled the second, racking up big legal bills along the way.

For his part, Kahle said he sees those battles as examples of a broken system that's made it increasingly difficult for the archive and similar institutions to preserve and disseminate knowledge.

After giving a tour of his institution's Physical Archive in Richmond last month, Kahle spoke with The Examiner about the Internet Archive's mission, its battles with copyright holders, and how artificial intelligence and other new technologies are changing its work. The following conversation has been edited for length and clarity.



Internet Archive founder and Board Chair Brewster Kahle (seen in March): “I have ... faith in people, that they actually want to create, and they will create and do things that are worthy of being in a library. And we want to be there for them.”

Vera de Kok, CC BY 4.0 via Wikimedia Commons

When you founded the Internet Archive nearly 30 years ago, did you have any sense it would turn out to be the way it is now? What was your vision then? I thought we were only going to have to do the digital side. And I really thought that we’d be able to get the libraries to start having digital collections. Some have, but often they’re really quite small, or they’re special collections. [It’s] not like they’re full collections. So that was a little bit of a surprise. And then there’s the physical side — I didn’t know that we would really want to go off and do that.

The thing that was probably the biggest surprise to me was the publishers became really monopolized — very few, very controlling, and they stopped selling things digitally, which really has [been] a problem

for libraries. They don’t have digital collections anymore. So they’re really being attacked in a very core-to-who-they-are way by the business practices of these very few, monopoly-sized publishers.

When did it become apparent to you that you were going to have to deal with physical archives? In the early 2000s, when we started physically digitizing materials, we found that some of the libraries didn’t want the materials back. They wanted the digital, but they wanted their space back. And because we had the [works] that were digitized, they were in some sense special. We didn’t want to have them throw them away, and we didn’t want to throw them away ourselves.

So I’d say in the mid-2000s, we started trying to figure out how to deal with the scale of the physical archives.

What do you see now as the mission of the Internet Archive? It only has one mission — it’s just a library — which is universal access to all knowledge. That’s all libraries have ever done. And they achieve that through their own collections and through interlibrary loan.

Tell me a little bit about the digital side of things. You digitize things. You have digitized physical objects. You digitize webpages, websites. Websites, television, software.

Your storage must be pretty tremendous at this point. Yes.

And you have redundancy in that, in your digital archives? Yes.

So the big achievement we're celebrating [is] getting 1 trillion web pages by probably over a billion people's voices. So, where corporations are sort of screwing down and being more monopolistic these days, governments are starting to have trouble — people are awesome. People want to share. They want people to know what they know. They'll spend infinite amounts of time working on Wikipedia to go and make sure that it's not constrained into a copyright system that doesn't work very well, that it's open and available.

I have great faith and trust in large-scale human nature that is so trusting and so sharing.



With 1 trillion pages preserved, Brewster Kahle says the Internet Archive is just getting started.

Sebastiaan ter Burg CC BY 4.0 via Wikimedia Commons

You mentioned copyright. You have gotten in trouble with copyright battles. Yes, yes, yes. So there's this question of whether libraries are going to survive. One hundred years ago in the United States, the legislatures and judiciary were very pro-libraries.

Now we have licensing issues. We have the corporations, we've got book bans, we've got defundings, we've got criminalization of librarianship. It's a challenging time in the United States and actually in many countries around the world, as we're going through some political swings.

How much does the archive work with the AI companies that are trying to get data for their models? The AI companies go and grab a lot of data from the Internet Archive, and we try to package that up so that it's easiest on us. So, like the old public-domain books or government webpages from the end-of-term crawls — open collections, we try to make really easy.

So, for instance, a lot of the United States government documents are public domain. We don't put restrictions on that. We ask them to not hit us so hard to make us melt down, but we try to make those available.

All the rest of it — which is a lot of materials — we don't make available to the AI companies, because there's not regulatory clarity.

Has that been taxing to your systems to have the AI companies hitting them? Yes, but it's just another type of use of the Internet Archive — and it's time for us to just kind of live up to it. If it's open, we want people to use it. There are these programs now that know how to go out on the internet and search for it. That's really cool! So, we'd love to be able to rise to that occasion, and we're beefing up our systems all the time.

Are there things you think you'll be archiving in 30 years that you're not archiving now?
Will the site work differently? 3D environments, games, the human experience. And the digital-built experience — how do we go and learn from that? Boy, we really don't know.

Like, what does it mean for a lot of the 3D games that people are doing? Does it make any sense? Or is it just all ephemera that is maybe an experience you want to try to do, but there isn't anything being built? That's what worries me the most, is we're wasting people's time.

In sort of the glory days of print, people would write things down and contribute them to the library. And the early web — and we have 1 trillion web pages — is kind of that ethos.

If we end up with just all ephemera, we're just entertained all day long and watched, that's my version of hell. That's where there is no real role for a library going forward. But I have more faith in people, that they actually want to create, and they will create and do things that are worthy of being in a library. And we want to be there for them.

If you have a tip about tech, startups or the venture industry, contact Troy Wolverton at twolverton@sfxaminer.com or via text or Signal at 415.515.5594.

Troy Wolverton
