

=

$\overline{X}$
LONG NOW

Ideas



A digital document dissolves into a Dark Age. Illustration by Casey Cripe

DIGITAL DARK AGE

Shining a Light on the Digital Dark Age

Without maintenance, most digital information will be lost in just a few decades. How might we secure our data so that it survives for

generations?

by **ADRIENNE BERNHARD**

Jul 19, 02023

The Dead Sea scrolls, made of parchment and papyrus, are still readable nearly two millennia after their creation — yet the expected shelf life of a DVD is about 100 years. Several of Andy Warhol's doodles, created and stored on a Commodore Amiga computer in the 01980s, were forever stranded there in an obsolete format. During a data-migration in 02019, millions of songs, videos and photos were lost when MySpace — once the Internet's leading social network — fell prey to an irreversible data loss.

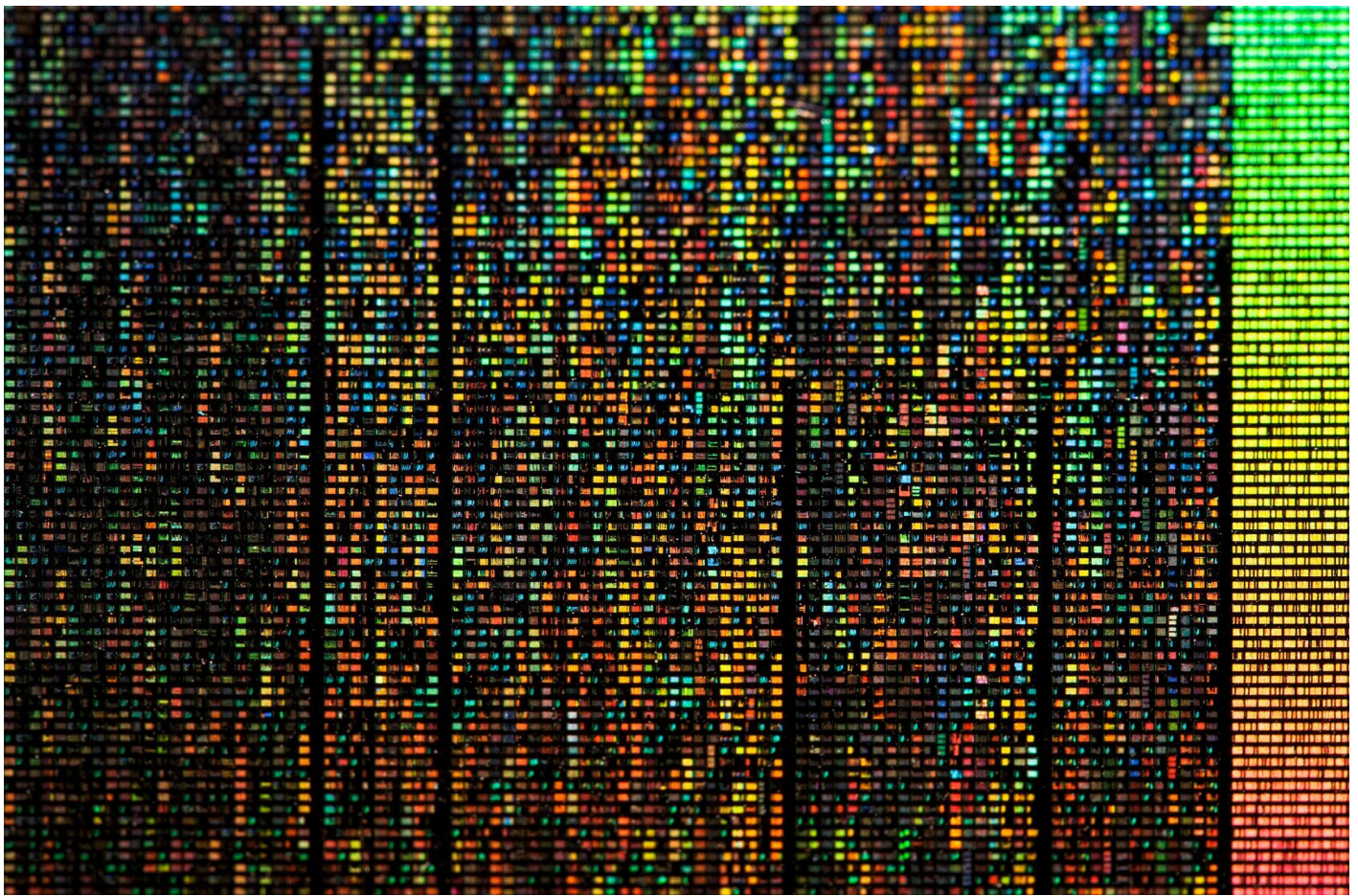
It is easy to grow complacent about data — to grant the digital a permanence that we deny to the physical. Making countless exact copies of a digital document is trivial, leading most of us to believe that our electronic files have an indefinite shelf life and unlimited retrieval opportunities. In fact, preserving the world's online content is an increasing concern, particularly as file formats (and the hardware and software used to run them) become scarce, inaccessible, or antiquated, technologies evolve, and data decays. Without constant maintenance and management, most digital information will be lost in just a few decades. Our modern records are far from permanent.

Obstacles to data preservation are generally divided into three broad categories: *hardware longevity* (e.g., a hard drive that degrades and eventually fails); *format accessibility* (a 5 ¼ inch floppy disk formatted with a filesystem that can't be read by a new laptop); and *comprehensibility* (a document with an long-abandoned file type that can't be interpreted by any modern machine). The problem is compounded by encryption (data *designed* to be inaccessible) and abundance (deciding what among the vast human archive of stored data is actually worth preserving).

The looming threat of the so-called “Digital Dark Age”, accelerated by the extraordinary growth of an invisible commodity — data — suggests we have, as Roy Rosenzweig put it in a 2003 article for *The American Historical Review*, “somehow fallen from a golden age of preservation in which everything of importance was saved.” In fact, countless records of the past have vanished. The first Dark Ages, shorthand for the period beginning with the fall of the Roman Empire and stretching into the Middle Ages (00500-01000 CE), weren't actually characterized by intellectual and cultural emptiness but rather by a dearth of historical documentation produced during that era. Even institutions built for the express purpose of information preservation have succumbed to the ravages of time, natural disaster or human conquest. (The Library of Alexandria is perhaps the most famous example.)

Digital archives are no different. The durability of the web is far from guaranteed. Link rot, in which outdated links lead readers to dead content (or a cheeky dinosaur icon), sets in like a pestilence. Corporate data sets are often abandoned when a company folds, left to sit in proprietary formats that no one without the right combination of hardware, software, and encryption keys can access. Scientific data is a particularly thorny problem: unless it's saved to a public repository accessible to other researchers, technical information essentially becomes unusable or lost. Beyond switching to analog alternatives, which have their own drawbacks, how might we secure our digital information so that it survives for generations? How can individuals, private corporations and public entities coordinate efforts to ensure that their data is saved in more resilient formats?

Organizations like The Long Now Foundation are among those working to combat the Digital Dark Age (Long Now in fact coined the term at an early digital continuity conference in 1998), drawing on open-source software, coordinated action across platforms, transparency in design, innovative technologies, and a long view of preservation. From thought experiments and industry analysis to more concrete projects, these organizations are imagining preservation on a massive time scale.



The Rosetta Disk is just one possible way that we can preserve linguistic data for future generations.

Consider the Rosetta Project, Long Now's first exploration into very long-term archiving. The idea was to build a publicly accessible digital library of human languages that would be readable to an audience 10,000 years hence. "What does it mean to have a 10,000-year library?" asks Andrew Warner, Project Manager of Rosetta. "Having parallel translations was enough to unlock the actual Rosetta Stone, so we decided to construct a 'decoder ring,' figuring that if someone understood one of the 1,500 languages on our disk, they could eventually decipher the entire library." Long Now worked with linguists and engineers to

microscopically etch this text onto a solid nickel disk, creating an artifact that could survive millennia. More symbolic than pragmatic, the Rosetta Project underscores the problem of digital obsolescence and explores ways we might address it through creative archival storage methods. “We created the disk not to be apocalyptic, but to encourage people to think about our more immediate future,” says Warner.

READ Kevin Kelly's [02008 Long Now essay](#) on the ideas behind the Rosetta Disk.

READ Kurt Bollacker's [02010 American Scientist article](#) on the technical challenges of data longevity and how the Rosetta Disk tries to address them.

READ Laura Welcher's [02022 paper](#) on Rosetta, archives, and linguistic data.

Indeed, the idea of a 10,000 year timescale captures an anxious public imagination that’s very much grounded in the present. How will technological innovation such as generative artificial intelligence change the way we interpret (and govern) the world? What threats — climate change, pandemics, nuclear war, economic instability, asteroid impacts — will loom large in the next decamillennium? Which information should we preserve for posterity, and how? What are the costs associated with that preservation, be they economic, environmental or moral? “Long

Now will be a keeper of knowledge,” says Warner. “But we don’t want to end up erecting a digital edifice in a barren wasteland.”

As personal computing and the ability to create digital documents became ubiquitous toward the end of the 20th century, preservation institutions — museums, libraries, non-profits, archives — and individuals — were faced with a new challenge: not simply how to preserve material, but *what* to collect and preserve in perpetuity. The astonishingly rapid buildup of data has led to a problem of abundance: digital information is accumulating at a staggering rate. Scientific, medical and government sectors in particular have amassed billions of emails, messages, reports, cables and images, leaving future historians to sift through and categorize an enormously capacious collection — one projected to exceed hundreds of zettabytes by 02025. The endless proliferation of junk-content, both human and A.I.-generated, makes sifting through all this data even trickier.

The term of art for this sorting is known as *appraisal*: deciding whether or not something should be preserved and what resources are required to do so. “One ongoing challenge is that there’s so much to preserve, and so much that can help diversify the archival record and reflect the society we live in,” says Jefferson Bailey, Director of Archiving & Data Services at the

Internet Archive, a digital library based in San Francisco that provides free public access to countless collections of digitized materials.

Servers housed at the Internet Archive's San Francisco headquarters.

The organization began in 1996 by archiving the Internet itself, a medium that was just beginning to grow in use. Like newspapers, the content published on the web was ephemeral — but unlike newspapers, no one was saving it. Now, the service has more than two million users per day, the vast majority of whom use the site's Wayback Machine, a search function for over 700 billion pages of internet history. Educators interested in archived websites, journalists tracking citations, litigators

seeking out intellectual property evidencing or trademark infringement, even armchair scholars curious about the Internet's past, all can take a ride on this digital time travel apparatus, whose mission is “to provide Universal Access to All Knowledge.” That includes websites, music, films, moving images and books.

WATCH Brewster Kahle's 02011 Long Now Talk, "Universal Access to All Knowledge," on how the Internet Archive, deemed "impossible" when he founded it in 01996, became an essential part of the web.

WATCH Jason Scott's 02015 Long Now Talk, "The Web in an Eye Blink," which details Scott's project at the Internet Archive to save all the computer games and make them playable again inside modern web browsers.

Some of the services provided by the Internet Archive enable institutions to document their own web properties or older materials, but the frequency of this backup varies. “We archive many widely-read and frequently-changing websites multiple times a day, whereas other sites may only be archived once or twice a year,” says Bailey. Web scrapers allow for easy archiving, but then there's the problem of storage. A single copy of the Internet Archive library collection occupies more than 99 petabytes of server space (the organization stores at least two copies of everything). Safely storing any long-lived data requires significant emissions, energy and cost.

The Internet Archive also takes multi-century preservation energy costs into account: “We own and operate and run all our own data centers,” says Bailey. “That’s partly because we’re an archive and don’t want to be dependent on corporate infrastructure, and partly because we can then run less expensive climate control operations.”

Microsoft Research’s Project Silica storage device.

Microsoft’s Project Silica aims to tackle the problem by providing an alternative to traditional magnetic media, which degrades over time. The company created a low-cost, durable WORM media (an acronym for Write Once, Read Many, which means that data is written to a storage

medium a single time and cannot be erased or modified). According to Robert Shelton of *TechTarget*, WORM storage is considered immutable, and “plays a pivotal role in meeting data security and compliance requirements and protecting against ransomware and other threats.” Project Silica’s media is also resistant to electromagnetic frequencies: stored in quartz glass, the lifetime of data can be extended to tens of thousands of years. This has important implications for sustainability, because users can leave data in situ, eliminating “the costly cycle of periodically copying data to a new media generation.”

The Arctic Code Vault in Svalbard.

Another option is to store data in Earth’s most remote places. GitHub created an Arctic Code Vault, a data repository preserved in the Arctic World Archive (AWA), a very-long-term archival facility 250 meters deep in the permafrost of an Arctic mountain. Such cold storage is meant to last at least 1,000 years and shield data from the outside world’s effects. Arctic archiving strategies could have potential as much of the world warms. Eliminating the need for costly HVAC filtration or cooling

systems, building for redundancy, putting in place multiple failovers, storing data in remote or unconventional locations — these are certainly laudable measures, but more will need to be done to mitigate the enormous ecological impact of all this storage. Streaming, cryptocurrency, and day-to-day Internet use already make up nearly 4% of global CO2 emissions — a statistic that rivals the carbon footprint of the aviation industry. Sustainability must be a paramount concern of digital preservation technologies.

For individual users worried about data preservation, the answer may lie in a switch to better systems with more robust methods of encoding, or investing in mechanisms that permit a kind of time travel to older formats. Archive-It, Internet Archive’s web archiving service, offers subsidized and grant-funded services to organizations, including over 1,500 libraries, archives, governments and nonprofits, enabling users to create a personalized cultural heritage storage system. There’s also Permanent Legacy Foundation, a cloud service backed by a nonprofit that allows consumers to permanently store and share their digital archives with loved ones, community members and future generations. “The most distinctive component of Permanent is our legacy planning feature,” says Robert Friedman, Executive Director at Permanent. “We have a very robust vision for memorialized content —

what people can do with their archives once they're no longer able to maintain them." For a one-time fee, Permanent will convert a user's files (documents, video, image, audio) to more durable formats; a Microsoft Word document might be converted to a PDF, for example. "The purpose is to make sure there exists at least one copy of every file that isn't tied to a proprietary format," says Friedman.

READ our [02020 profile](#) of the Permanent Legacy Foundation and its goal to preserve individuals' digital legacies for future generations.

Not being tied to a corporate entity or the commercial cloud means many of these organizations can have a direct stake in the sustainability of their operations. "If you need your data available consistently, then you have to keep running the operation at idle," says Friedman. But for many people, their data will never be needed again after death, in which case, those personal archives can be sealed and deleted. "We don't strip you of your rights," he adds, "but we do need to know what to do with that data, including whether to destroy it." Permanent is also conservative in its appraisals, and its subscription model encourages individuals to be selective about what they choose to preserve. (Fees collected by Permanent are used to fund the ongoing cost of storage and operations.)

"This challenge is not something that one organization is going to solve: it has to be a collective effort," says Bailey. Federal agencies, public libraries, research institutions and private corporations, each with different budgets

and mandates, must come together to address the issue of storage and associated concerns. And, as technology accelerates and emulation of older systems becomes more difficult, Digital Dark Age preppers may need to consider not just the survival of data, but the possibility that it may be misunderstood. Getting future readers to parse what we've preserved, and to make sense of that information, is an existential challenge without easy answers.

As Rosenzweig puts it: "Because digital data are in the simple lingua franca of bits, of ones and zeros, they can be embodied in magnetic impulses that require almost no physical space, be transmitted over long distances, and represent very different objects (for instance, words, pictures, or sounds as well as text)." Barring the invention of some extraordinary universal translation system, making sense of these binary codes without knowing how they were intended to be read is impossible.

Also consider online disinformation, spread either by humans or artificial intelligence. Many digital archives already struggle to exclude or annotate content, not to mention living websites like Wikipedia, Twitter or Facebook, which must be continuously monitored for harmful or fraudulent content. "If you exclude fake news, does that mean you run the risk of a non-representative archive?" asks Bailey. "We are open to removal if there's a reason, and we do it."

Public and private organizations and policy-makers will need to decide what content should be restricted, embargoed, or scraped from the web

entirely, and design robust digital repositories that can weather the ravages of time. People with limited resources must also be ensured the ability to preserve their data; long-term storage should be designed for equity as well as for sustainability. Finally, we must be judicious in what we store and preserve, creating a meaningful collection for future historians. Our future — and our past — depend on it.

A CORRECTION WAS MADE ON OCT 23, 02023: An earlier version of this article failed to properly attribute its source material, including sections from Roy Rosenzweig's 02003 article for *The American Historical Review*, "Scarcity or Abundance? Preserving the Past in a Digital Era." We regret the error.

Join our newsletter for the latest in long-term thinking

SUBSCRIBE

IDEAS

"the long now"

Pace Layers framework

Digital dark age

TALKS

Podcast

YouTube

Attend live

PROJECTS

10,000-year clocks

Labs

The Interval

Pace Layers

Nevada Bristlecone Preserve

Rosetta

PEOPLE

Board

Staff

The Council

Community

Careers

DONATE

Join

Shop

Member Center

Contact

