

# The Forever Disc

Can our digital data represent us a million years from now?

**O**UR ANCESTORS SPEAK TO US across millennia through paper, canvas, and stone; the natural world encompasses a narrative billions of years old. Yet data on magnetic and optical media fade within decades. If we have information worth passing on, how can we better store it for future generations? Could it outlast even our species itself; if so, how can we make its meaning evident to its recipients?

These are questions faced by digital preservationists who are starting to design data storage mechanisms that can survive centuries, millennia, and beyond. In doing so, they reveal the surprising possibility that we may be able to embed messages in our bodies and our descendants' bodies to last for millions of years, under the right circumstances.

## The Scale of Time

Digital preservationists are typically charged with maintaining materials for years or decades, depending on need. A business' client records, for example, might only be useful as long as the relationship is active; the law may additionally require the business to save those records for a few more years. Those with longer-term needs often find themselves rolling their own solutions based on expected storage length, available resources, and the type of data to be stored.

For help, those in the United Kingdom and Ireland can turn to the Digital Preservation Coalition. According to executive director William Kilbride, its 40-odd large organizational members seek to preserve items for as long as they have "business use"—which, for some, is measured in centuries. "Our membership includes national archives and libraries," he said. "They exist to maintain public access to culture and heritage."

Across the Atlantic, Leslie Johnston, chief of the Repository Development Center at the U.S. Library of Congress,



The Long Now Foundation's Rosetta Disk.

says she faces similar challenges. "We very much think of preservation long term, partly because there is a concept of federal agencies as being 'for the life of the Republic.' But we also have to be pragmatic about technology."

Part of that pragmatism has meant saving digital data even when access to it is not guaranteed, as is often the case for digital files. "We do not currently maintain either a physical or an emulation environment for the use of those files," Johnston said, while pointing to the Library's analogous collection of vintage audio and video players. "To interpret a file you might need the source code, and the compiled code, and the operating system, and the hardware, and libraries, and all the other software dependencies ... that is potentially a huge collection. We have conversations about that, but it is not our cur-

rent position."

One way to keep digital information "alive" is to continually migrate it to current formats and media, a process the Digital Preservation Coalition's Kilbride called "more like a relay race than a marathon." Said Kilbride, "There are some extreme cases where data needs to be managed for a very long time. For example, the challenges facing managers of radioactive waste from medical, power, and military facilities reach into many thousands of years. It is not realistic to expect a record to last that long, but as long as the current generation of records managers can pass the baton to the next generation, their job is done."

For some, the drive for long-lasting information is as much philosophical as practical. The Long Now Foundation—founded by Stewart Brand, who created the influential online service The WELL,

along with Thinking Machines' Danny Hillis—concerns itself with projects to “help make long-term thinking more common.” One of them is Long Server, an “overarching program for Long Now’s digital continuity software projects,” including an in-the-works file-format conversion system. According to executive director Alexander Rose, “For us, the major element is an awareness one. No industry has really tried to make people aware that their data is ephemeral ... I would defy you to open up a current file 200 years from now in digital format.”

### The Medium Is the Message

The Long Now Foundation’s Rosetta Project explicitly avoids the digital compatibility issue by micro-inscribing more than 13,000 pages of information, in more than 1,500 languages, on a palm-size disk (a zoomable online version may be seen at <http://rosetta-project.org/disk/interactive>). The data is like a printed page: analog, albeit legible only through a microscope. In this way it uses essentially the same technology as the original 2,200-year-old Rosetta Stone, as did the “Pioneer Plaques” launched skyward on spacecraft in 1972 and 1973.

The Rosetta Disk is made of solid nickel; the Pioneer Plaques were made of gold-anodized aluminum. As part of his thesis, University of Twente doctoral student Jeroen de Vries explored other materials in an attempt to create an even more durable data-carrying substrate. He ultimately devised an optical disc of tungsten coated with a clear layer of silicon nitride, whose heat resistance he demonstrates on video by cooking an egg on it. He estimates the QR codes he inscribed on it would still be readable after a million years at a temperature of 200 degrees Celsius; however, he acknowledged “the disc is actually very fragile, because it is made of crystalline silicon, so it breaks very easily along the crystal lines.”

The approaches of de Vries, the Rosetta Project, and the Pioneer Plaques’ creators all are to freeze information on a static medium. This approach is comparatively limited: the oldest deliberate, data-bearing artifact is perhaps a cave painting 40,000 years of age, a mere toddler in the grand scheme of things. If we are to present ourselves to descendants beyond that, we will need a better medium.

**DNA storage has another advantage over traditional media, it could be made self-replicating within a living being, and be passed on to offspring through natural reproduction.**

### The Living Legend

One of nature’s more-robust solutions is in DNA, the chemical strands found in all living beings. DNA exists in several forms, most notably comprising long strings of nucleotides that appear in sequences unique to each individual. The DNA sequence of a creature—its genome—remains nearly constant throughout its life, and is evident in bones and other remains after death. Although the half-life of DNA can be as short as 500 years, a complete genome has been decoded from archeological samples as old as 700,000 years.

Some DNA defines a creature’s essential traits, such as eye color and propensity for certain diseases. However, there is also a considerable amount of nonfunctional “junk” DNA in which additional information could be encoded. Researchers have written in those spaces since 1988, when they successfully encoded a 35-bit graphic into the DNA of *E. coli* bacteria. In 2012, a team led by Harvard Medical School genetics professor George M. Church surpassed previous storage records by putting his 5.27-megabit book “Regenesis,” comprising 53,426 words, 11 images, and one JavaScript program, on a string of nucleic acids. Church believes his team’s accomplishment is only the beginning, saying this storage medium could theoretically store over 450 million terabytes per gram of DNA.

DNA storage has another advantage over traditional media: it could be made self-replicating within a living being,

## ACM Member News

**BIAS, SUBJECTIVITY MAKE BOTH HUMANS, AI, INTELLIGENT: STEIN**



“I think of myself as an applied philosopher and a computer science education

innovator,” says Lynn Andrea Stein, professor of computer and cognitive science and a founding faculty member at Franklin W. Olin College (Olin) in Needham, MA.

Stein graduated cum laude in computer science from Harvard and Radcliffe colleges; she obtained her master’s degree and doctorate in computer science from Brown University. Artificial intelligence and robotics have been areas of specialization throughout her teaching career.

In 2000, Stein left her position as associate professor at Massachusetts Institute of Technology (MIT) to help found Olin, an innovative undergraduate engineering college that espouses hands-on engineering and design projects rather than a traditional core-curriculum teaching model. “It was a once-in-a-lifetime opportunity to start a college from scratch and think about what education could and should look like,” Stein says. “I researched how to understand and create intelligent systems and create systems in which people can learn well.”

Stein believes bias and subjectivity make both humans and AI systems intelligent. “To discriminate, in its original sense of differentiation, is a fundamental necessity for human beings and AI,” she says. “In that sense of discrimination, we don’t want to rid ourselves of the ability to differentiate.”

She applies these concepts to her work in AI and robotics. The robotic systems her research group builds bridge the gap between low-level behaviors associated with robotics and levels of cognition that better approximate human thinking.

—Laura DiDio

and be passed on to offspring through natural reproduction. According to Church, “We have already put our synthetic DNA into a living organism. Every time the message-carrying cell replicates, it does not know where the synthetic part begins and the natural part ends. You could essentially design self-ish DNA with your message, in which case almost all the progeny get it.”

Even so, Church believes “dead” DNA from archaeological sites might be a better data-storage bet than such “living” DNA. “Some DNA information has survived for three billion years,” he said, “but that DNA had information of great importance to the survival of the species. Your message, if it is an arbitrary message, will not be of great importance to the cell, and so will not last as long.”

### On Chemicals: Active, Sleeping, and Awakened

DNA is of the chemistry of life, and as such participates in the cycles of birth, reproduction, and death. Other chemicals are comparatively stable in nature, leading a group of researchers at Palacký University in the Czech Republic to consider two other ways we might encode information in molecules.

The first involves chirality, or “handedness,” wherein certain molecules can be expressed in left-hand (S) or right-hand (R) forms known as “enantiomers;” these could take the place of ones and zeros in digital documents. According to lead researcher Jan Petr, “We have examples of chiral molecules in our bodies, and they are stable for millions of years. The synthesis of chiral molecules can be induced by light, so it would be much like writing a

DVD.” However, he warned, such molecules could also be affected by light in the environment; he calls for further research into molecules whose chirality is easy to induce, but which are stable in the natural environment.

Petr’s group also explored the possibilities of storing information in mixtures of chemicals that are essentially dormant until a certain event occurs, such as exposure to oxygen; then, its messages would become apparent through the resulting reaction, possibly by forming 3-D structures. According to Petr, “It is a lot like cooking; you open the fridge, grab something, and cook it. You have taken information from the fridge, where it was in stasis.”

### Nature Always Wins

In an episode of the animated comedy series “Futurama,” a time traveller finds himself separated from a friend by nearly a billion years. The one in 3050 A.D. strategically damages the ceiling of a cave; the pattern of her laser blasts causes dripping water to spell out a message in slowly growing stalagmites, which her friend in 1,000,000,000 A.D. reads. This fanciful scenario underscores a real-world truth: successful long-term storage solutions must take nature into account, and work with it.

The field continues to explore materials both human-crafted and biological, while also examining how nature has passed messages through the ages thus far. For every piece of data a geologist, geneticist, or archaeologist deciphers, there is a potential revelation of how that message was written, how it survived, and whether we might use a similar method.

The media and storage methods we choose could make an enormous difference in the legacy we leave behind. As Human Document Project co-creator Andreas Manz said, “We believe that our ancestors were cave-men, particularly in Europe, because we found fossil bones and paintings in caves. Now, humans at the time would probably rather live in a nice, cozy valley with wine grapes and the lakeside and whatnot than in a cave, but if they die in those locations that are nice and sunny, then those bones will not survive; you do not find fossilized bones along the road or in the forest, only under very special conditions. So the information we have is biased by how the information is stored.” □

### Further Reading

- de Vries, J.* (2013). *Energy barriers in patterned media*. Doctoral thesis, University of Twente.
- Elwenspoek, M. C.* (2011). *Long-time data storage: relevant time scales*. *Challenges*, 2(1), 19-36.
- Petr, J., Ranc, V., Maier, V., Ginterová, P., Znaležiona, J., Knob, R., & Ševčík, J.* (2011). *How to Preserve Documents: A Short Meditation on Three Themes*. *Challenges*, 2(1), 37-42.
- Church, G.M., Gao, Y., and Kosuri, S.* (2012). *Next-generation digital information storage in DNA*. *Science* 337.6102: 1628-1628.
- Rothenberg, J.* (1998). *Ensuring the longevity of digital information*. *International Journal of Legal Information*, 26, 1.

Tom Geller is an Oberlin, Ohio-based science, technology, and business writer.

© 2014 ACM 0001-0782/14/05 \$15.00

## Milestones

# ACM Bestows A.M. Turing, Infosys Awards

ACM has announced the names of this year’s recipients of two of its most prestigious awards.

The A.M. Turing Award, ACM’s most prestigious technical award, is given for major contributions of lasting importance to computing.

This year’s ACM A.M. Turing Award is being presented to Leslie Lamport, a principal researcher at Microsoft Research

Silicon Valley, “for fundamental contributions to the theory and practice of distributed and concurrent systems, notably the invention of concepts such as causality and logical clocks, safety and liveness, replicated state machines, and sequential consistency.”

The ACM-Infosys Foundation Award in the Computing Sciences recognizes personal

contributions by young scientists and system developers to a contemporary innovation that, through its depth, fundamental impact and broad implications, exemplifies the greatest achievements in the discipline.

This year’s ACM-Infosys Foundation Award is being presented to David M. Blei of the computer science department of

Princeton University, “for fundamental contributions to artificial intelligence, in particular the development of topic models and their applications.

*Communications* will provide in-depth coverage of both award recipients in upcoming issues, beginning with interviews with Lamport in the June issue.

— Lawrence Fisher