

[STORAGE](#) > [BACKUP](#)

DNA Storage: Why the Stuff of Life Is Poised to Be the Stuff of Storage

DNA storage may prove to be optimal for archiving files that must be retained but will rarely (if ever) need to be accessed again.

[Karen D. Schwartz](#) | Dec 13, 2018



Sometime in the [next decade](#), you might very well be reconstituting a file you never thought you'd need again from [DNA](#). Yes, that DNA—deoxyribonucleic acid—the stuff of life.

While it seems like the stuff of science fiction, researchers have made a lot of progress over the past several years proving that synthetic [DNA could make the perfect archival storage medium](#) for files that must be retained but may rarely if ever need to be accessed again.

One of the reasons researchers hold out so much hope for DNA as a storage medium is because of its long-term stability. Solid state drives last about five years before they start to degrade, while magnetic disk might last 10 years, magnetic tape 25 years, and optical disk 25 to 35 years. In contrast, DNA can last thousands of years as long as it's kept cool and dry.

“Researchers have been able to show that DNA as a storage medium is very power- and space-efficient. We’re talking terabytes of information in literally grams of cell matter. It’s amazing what can be stored in a small amount of DNA,” said Ray Lucchesi, president of Silverton Consulting.

The research Lucchesi is talking about started several years ago. One of the earliest projects was a Harvard group, which successfully transferred the contents of 53,400 word book and several images into DNA. Several other groups since have successfully encoded and stored millions of bits of data in DNA. [Microsoft Research also is heavily involved in DNA research](#). It has been working steadily with the University of Washington, and the partnership has resulted in several breakthroughs: not only did it manage to encode 200 megabytes of digital data to synthetic DNA, but it recently found a way to add the concept of random access to files stored in DNA. Microsoft also has partnered with Twist Bioscience, which has a silicon-based DNA synthesis platform, to work on long-term data storage solutions for DNA.

Assuming that DNA could be easily writable and readable, with reasonable amounts of access time, the potential for storage is huge.

“Today we are reading megabytes or gigabytes per second off of an SSD. You could probably read 100 bytes or 200 bytes off of DNA in the same amount of time,” Lucchesi said. “That’s orders of magnitude more.”

So how does it work?

“It’s essentially the same idea as today’s storage methods,” explains Richard Hammond, technology director and head of synthetic biology at U.K.-based Cambridge Consultants. “With magnetic tape, you arrange the magnets to represent ones and zeros. With DNA, the core idea is the same; it’s a matter of arranging the information into the medium.”

Challenges Persist

Two of the biggest challenges in making DNA storage a commercial reality are cost and speed. Some estimates put the cost of encoding data at more than \$12,000 per megabyte and \$220 for retrieval. However, DNA synthesis and sequencing costs are already decreasing and, given time, will reduce even further to a palatable level.

“The cost of sequencing is decreasing, so now it’s cheap to read DNA,” said Christophe Dessimoz, a professor at the University of Lausanne who is an expert in this area. Synthesis costs remain relatively expensive, he said, although he expects it to decrease over time.

One company determined to beat the odds is Catalog Technologies, a start-up working with Cambridge Consultants to build a machine capable of encoding DNA data at a speed of 1TB per 24 hours. Hammond says the difference with the Catalog model is the way data is encoded into the DNA.

“The traditional approach involves writing the data directly into the DNA and creating the DNA one base at a time. That’s slow and

expensive,” he explained. “Compare what Catalog has done to a movable type printing press: They have a bunch of standardized letters—in other words, short small pieces of DNA—and they combine the bits of DNA together in the right order.”

By connecting pre-existing pieces of DNA instead of creating DNA from scratch, Catalog will be able to reduce the number of assembly steps, increasing the speed, reducing energy consumption and ultimately, reducing cost.

Catalog envisions its machine as part of a data archiving service offering. An organization would transfer its data to Catalog, which would input the digital data stream into its machine, process it, and use that information to assemble pieces of DNA. The result is a tube with powder—dry DNA with the organization’s encoded information. When an organization requests access to some of the data, Catalog will take it out of storage, re-suspend it to liquid form and run that liquid through its DNA sequencer. It would then convert it back to digital form and run through the inverse of the original algorithm. The result? The original data, in the original format.

The process is complicated, and Catalog doesn’t expect a commercially available offering for a few years. But Hammond hopes that eventually it will help DNA become a cost-effective, fast, reliable way of storing cold, archival data.

In essence, Dessimoz agrees. Most likely he says, organizations will have some combination of long-term DNA storage and short-term solid state storage. It’s just a matter of time.