



With a “hello,” Microsoft and UW demonstrate first fully automated DNA data storage



•

Written by

- [Jennifer Langston](#)

Published

[March 21, 2019](#)

Researchers from Microsoft and the University of Washington have demonstrated the first fully automated system to store and retrieve data in manufactured DNA — a key step in moving the technology out of the research lab and into commercial datacenters.

In a simple proof-of-concept test, the team successfully encoded the word “hello” in snippets of fabricated DNA and converted it back to digital data using a fully automated end-to-end system, which is described in a [new paper](#) published March 21 in Nature Scientific Reports.

DNA can store digital information in a space that is orders of magnitude smaller than datacenters use today. It's one promising solution for storing the exploding amount of data the world generates each day, from business records and cute animal videos to medical scans and images from outer space.

Microsoft is exploring ways to close a looming gap between the [amount of data we are producing](#) that needs to be preserved and our capacity to store it. That includes developing algorithms and molecular computing technologies to [encode and retrieve data in fabricated DNA](#), which could fit all the information currently stored in a warehouse-sized datacenter into a space roughly the size of a few board game dice.

“Our ultimate goal is to put a system into production that, to the end user, looks very much like any other cloud storage service — bits are sent to a datacenter and stored there and then they just appear when the customer wants them,” said Microsoft principal researcher [Karin Strauss](#). “To do that, we needed to prove that this is practical from an automation perspective.”

Information is stored in synthetic DNA molecules created in a lab, not DNA from humans or other living things, and can be encrypted before it is sent to the system. While sophisticated machines such as synthesizers and sequencers already perform key parts of the process, many of the intermediate steps until now have required manual labor in the research lab. But that wouldn't be viable in a commercial setting, said Chris Takahashi, senior research scientist at the UW's [Paul G. Allen School of Computer Science & Engineering](#).

“You can't have a bunch of people running around a datacenter with pipettes — it's too prone to human error, it's too costly and the footprint would be too large,” Takahashi said.

For the technique to make sense as a commercial storage solution, costs need to decrease for both synthesizing DNA — essentially custom building strands with meaningful sequences — and the sequencing process that extracts the stored information. Trends are [moving rapidly in that direction](#), researchers say.

Automation is another key piece of that puzzle, as it would enable storage at a commercial scale and make it more affordable, Microsoft researchers say.

Under the right conditions, DNA can last much longer than current archival storage technologies that degrade in a matter of decades. Some DNA has managed to persist in less than ideal storage conditions for tens of thousands of years in mammoth tusks and bones of early humans, and it should have relevancy as long as people are alive.

The automated DNA data storage system uses software developed by the Microsoft and UW team that converts the ones and zeros of digital data into the As, Ts, Cs and Gs that make up the building blocks of DNA. Then it uses inexpensive, largely off-the-shelf lab equipment to flow the necessary liquids and chemicals into a synthesizer that builds manufactured snippets of DNA and to push them into a storage vessel.

When the system needs to retrieve the information, it adds other chemicals to properly prepare the DNA and uses microfluidic pumps to push the liquids into other parts of the system that “read” the DNA sequences and convert it back to information that a computer can

understand. The goal of the project was not to prove how fast or inexpensively the system could work, researchers say, but simply to demonstrate that automation is possible.

One immediate benefit of having an automated DNA storage system is that it frees researchers up to probe deeper questions, instead of spending time searching for bottles of reagents or repetitively squeezing drops of liquids into test tubes.

“Having an automated system to do the repetitive work allows those of us working in the lab to take a higher view and begin to assemble new strategies — to essentially innovate much faster,” said Microsoft researcher Bichlien Nguyen.

The team from the [Molecular Information Systems Lab](#) has already demonstrated that it can store cat photographs, great literary works, [pop videos](#) and archival recordings in DNA, and retrieve those files without errors in a research setting. To date they’ve been able to store 1 gigabyte of data in DNA, besting their [previous world record of 200 MB](#).

Video Player

00:03

00:07

To store data in DNA, algorithms convert the 1s and 0s in digital data to ACTG sequences in DNA. Microsoft and University of Washington researchers stored and retrieved the word “hello” using the first fully automated system for DNA storage.

The researchers have also developed techniques to [perform meaningful computation](#) — like searching for and retrieving only images that contain an apple or a green bicycle — using the molecules themselves and without having to convert the files back into a digital format.

“We are definitely seeing a new kind of computer system being born here where you are using molecules to store data and electronics for control and processing. Putting them together holds some really interesting possibilities for the future,” said UW Allen School professor [Luis Ceze](#).

Unlike silicon-based computing systems, DNA-based storage and computing systems have to use liquids to move molecules around. But fluids are inherently different than electrons and require entirely new engineering solutions.

The UW team, in collaboration with Microsoft, is also developing a programmable system that automates lab experiments by harnessing the properties of electricity and water to move droplets around on a grid of electrodes. The full stack of software and hardware, nicknamed [“Puddle” and “PurpleDrop”](#), can mix, separate, heat or cool different liquids and run lab protocols.

The goal is to automate lab experiments that are currently being done by hand or by expensive liquid handling robots — but for a fraction of the cost.

Next steps for the MISL team include integrating the simple end-to-end automated system with technologies such as PurpleDrop and those that enable searching with DNA molecules. The researchers specifically designed the automated system to be modular, allowing it to evolve as new technologies emerge for synthesizing, sequencing or working with DNA.

“What’s great about this system is that if we wanted to replace one of the parts with something new or better or faster, we can just plug that in,” Nguyen said. “It gives us a lot of flexibility for the future.”

Top image: Microsoft and University of Washington researchers have successfully encoded and retrieved the word “hello” using this new system that fully automates DNA storage. It’s a key step in moving the technology out of the lab and into commercial datacenters.

Related to DNA storage:

- Listen: [Microsoft Research podcast with Dr. Karin Strauss](#)
- Read: [Microsoft and UW researchers set record for DNA storage](#)
- Read: [UW team stores digital images in DNA — and retrieves them perfectly](#)
- Read: [Twist Bioscience announces Microsoft purchase of its synthetic DNA for digital data storage research](#)
- Visit: [Microsoft Research: DNA Storage](#)