



# DNA: Is that where we'll be storing data next?

February 3, 2021

There's a compelling belief that the entire current sum of human knowledge could be encoded into 1 kilogram of DNA. In a quest to find out if that's attainable, practically speaking, the Intelligence Advanced Research Projects Activity (IARPA), a research agency within the Office of the Director of National Intelligence, has launched a four-year competition.

There are two research teams. Each is multi-disciplinary, made up of industry, university and research institute members with expertise in biological systems, chemistry, data storage systems, and statistics.

Los Alamos National Laboratory is playing the role of referee, providing the testing and evaluation of the many facets of this challenge and helping researchers refine their work. The winning technologies will demonstrate an end-to-end storage and retrieval workflow, with the potential to provide the United States with an overwhelming intelligence advantage. The success of the program will change the way stakeholders can archive the incredible amounts of information our modern society is generating.

## The predicament

The world was forever changed when the World Wide Web opened to the public 30+ years ago, putting information, communications and products instantly at our fingertips. Today, secure data storage is king, and our digital data load is maxing out current storage systems.

A growing number of public and private sector stakeholders have the need to generate and store Exabyte (EB) (one billion gigabytes) scale data sets. However, the cost of EB storage is very high and carries an extremely large footprint with heavy-duty power and cooling requirements. Facebook's new cold storage facility in Fort Worth, Texas is a 2.6m square-foot facility spanning 150-acres. It is scheduled to be completed in 2022 with a total cost of \$1.5B. However space, power, and cooling requirements are not the only challenge. In addition, synchronization across multiple EB archives is currently impossible.

## The MIST program

To address immense data storage challenges, IARPA has embarked on this futuristic quest: the ability to store data within DNA.

IARPA invests in high-risk, high-payoff research programs that address some of the most difficult challenges in the Intelligence Community. IARPA's Molecular Information Storage Technologies (MIST) program, which began in 2018, aims to develop a storage technology that eventually can scale into the Exabyte regime and beyond. It must meet reduced footprint, power, and cost requirements, without degradation of data. Specifically, it must demonstrate the writing of 1 TB (1TB = 1,000 gigabytes) and reading of 10 TB in 24 hours for \$1,000 operational cost.

The two MIST research teams were awarded contracts in September 2019, and the demonstration of the first phase of the MIST program is set for mid-2021. In this phase, the teams will be showing how their new technologies can provide a path to the ambitious goal of a cost-effective Exabyte storage system.

Depiction of the notional Exabyte MIST system that one day could replace massive data centers.

## **The role of national and military labs**

Los Alamos, a world leader in high performance computing and data storage, is heading the MIST Test and Evaluation (T&E) team that will oversee researchers who are developing a large cold storage/archive using information encoded into DNA.

"The goal of the MIST program is to dramatically advance data storage technology in terms of density and cost. LANL's role is to test the systems that the development teams produce, and evaluate them for aspects such as capacity, density, and cost," Said Tracy Erkkila, Program Manager and lead scientist for the MIST Test and Evaluation Team. "This will be a real game-changer for massive data storage when the metrics of the MIST program are achieved."

Together with Sandia National Laboratories and the U.S. Army Research Laboratory South (ARL-S), the T&E team will review and evaluate deliverables, participate in monthly progress and technical program reviews, and will have an on-site presence during milestone demonstrations, with access to each team's system. The T&E team will then collaborate to develop milestone demonstration test plans and evaluate the researchers' results.

The three laboratories bring a broad array of unique expertise to the program. Los Alamos brings deep experience in genomics and bioinformatics with experts in biochemistry and synthesis. Sandia's expertise for this venture is in microsystems design and fabrication along with extensive microfabrication capabilities, including die level processing. And ARL's expertise lies in nucleic acid synthesis, gene synthesis and assembly, directed evolution, and protein engineering.

## **The science behind molecular storage**

The goal of MIST is to develop "a deployable data storage capability using sequence controlled polymers and then build the necessary devices and information systems to interface with this medium." Researchers, scientists, and subject matter experts will write, store, retrieve, and read data stored in DNA.

Polymers such as DNA can have a stable lifetime of hundreds of years and an information density that is more than 10<sup>5</sup> times higher than that of conventional storage hard drives.

Image courtesy of Nature from the article How DNA could store all the world's data, published on August 31, 2016.

Some researchers believe molecular storage may be an ideal cold archive approach. The benefits of DNA-based storage include high-density storage at low energy, ease of copy, and storage longevity. To do this, existing technology in synthetic biology and DNA sequencing will need to scale up dramatically.

In the figure below, digital information is translated into the four-letter DNA alphabet, and then “written” as custom synthesized DNA. The encoded DNA can be stored very efficiently over a long time. When it comes time to “read” the information, the DNA is sequenced and the digital information is translated back into a computer friendly form.

While the steps needed to perform this “writing” and “reading” do exist, the challenge of the MIST project will be to develop new approaches to provide a many thousand-fold increase over current technologies.

Example of using DNA as physical media in a molecular information storage system. Figure courtesy Luis Ceze, University of Washington.

Learn more about [IARPA's MIST program](#).

**Los Alamos National Laboratory**

[www.lanl.gov](http://www.lanl.gov)

**(505) 667-7000**

**Los Alamos, NM**

Managed by Triad National Security, LLC for the U.S Department of Energy's NNSA

