DARPA Molecular Informatics Program: Molecular Data Storage and Computing

Anne Fischer, Ph.D.
Program Manager
Defense Sciences Office

Library of Congress Designing Storage Architectures

September 10, 2019
Motivation: We’re creating vast amounts of data, much of which is not stored or analyzed.
Rapid progress in DNA-based data storage, now content addressable

**2013:** First practical demonstration – 739 kB
European Bioinformatics Institute

Text, Image and Audio

Write
0101110011...

...0101110011...

Read
0101110011...

Dr. Martin Luther King’s “I Have a Dream” speech

**2016:** 200 MB of data stored at ~10^{17} bytes/mm^3
University of Washington/Microsoft

2016: First random access demonstration from a set of images

**2019:** Encoded features of 1.6M images*
DARPA Molecular Informatics
University of Washington/Microsoft

*Distribution A: Approved for public release; distribution is unlimited.
Molecular data storage is not all about DNA

Selection of molecular substrate depends on factors such as density, stability, scalability, accessibility, security and exploitability -- application dependent.
Diverse approaches to storage of digital data in molecules

DNA

- **Washington**: Synthetic DNA sequences

- **Illinois**: Enzymatic nicking of natural DNA

Molecular mixtures

- **Brown**: Large synthetic library of small molecules

- **Harvard**: Small library of peptide mixtures
Moving to practical scales

DNA

Molecular mixtures

End-to-end automation

Exploiting existing infrastructure