Explorations in Very Long-term Archiving

Laura Welcher Ph.D.

Designing Storage Architectures for Digital Collections
March 27-28, 2023
Library of Congress, Washington DC

(1)

Analog & Digital

(2)

Rosetta Disk & Wearable
Version: Microscopic
Electroforming in Nickel

(3)

(4)

(5)

5D Optical Data Storage
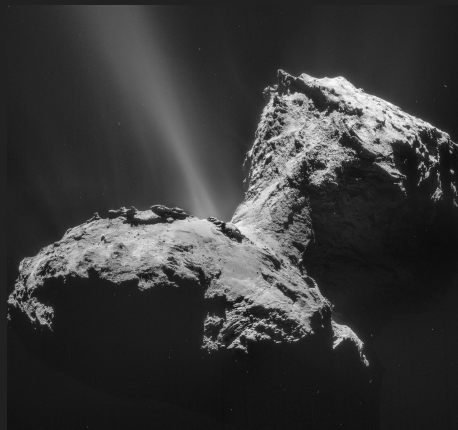Manipulation of Quartz
Crystal Nanostructures

(6)

Memory of Mankind in Salt Mines

(9)

Barbarastollen Microfilm Archive

On- & Off-world Sites

(7)

Rosetta Disk on Comet 67P

(8)

5D Foundation Trilogy on Tesla Roadster
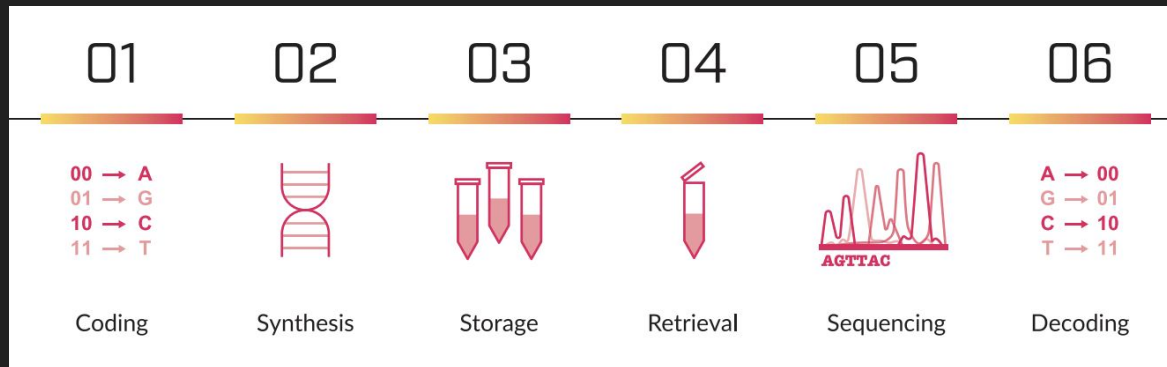
(10) Norsam HD-Rosetta Technology

(11) Microsoft's Project Silica

Real World Applications

(12) Data Storage in DNA
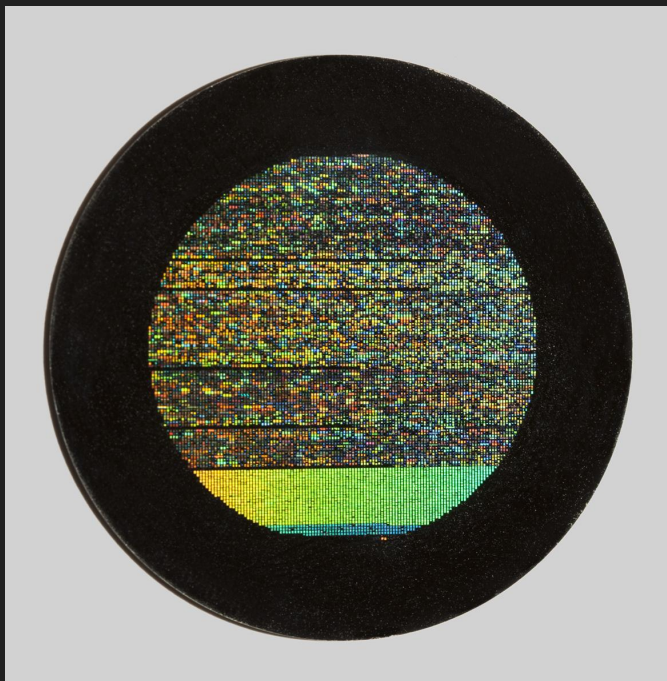
Thought Experiment

(13)

A Futuristic Archive

Your decoding challenge:
- ACGT > Binary 1s and 0s
- Binary data encodes text
- Text is in Unicode
- Writing is the glottographic encoding of some language
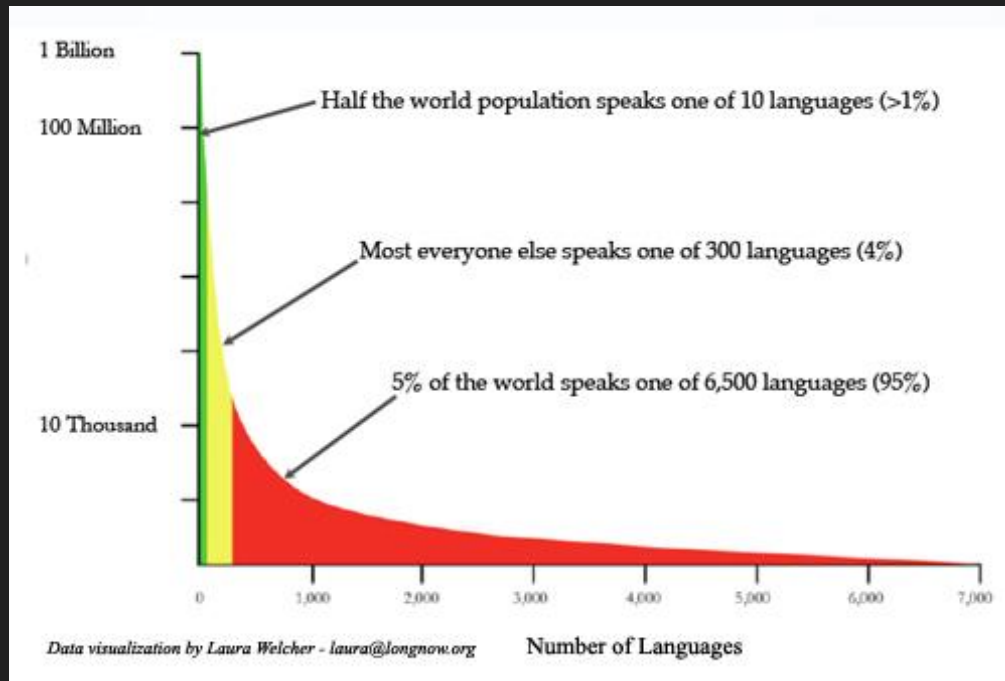- The fundamental encoding is human language itself

Archiving Endangered Language Data

Rosetta Disk V. 1.0

The Long Tail of Human Languages

Lessons from Long-Term Archives

- Material
- Medium - analog or digital?
- Method
  - What are all of the layers of encoding?
  - Have you minimized barriers to decoding?
- Meaning
  - The primary goal, but hardest part!
  - Metadata and contextualization
  - Human language itself is a fundamental encoding

Thank you!

The ideas presented in this talk are based on the following publication:

Buszard-Welcher, Laura. **"Linguistic Data in the Long View."** In Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister, eds. The Open Handbook of Linguistic Data Management. MIT Press, 2022.

Image Credits