

# 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

# Analog & Digital



(2)

Rosetta Disk & Wearable  
Version: Microscopic  
Electroforming in Nickel



(3)



(4)



(5)

5D Optical Data Storage  
Manipulation of Quartz  
Crystal Nanostructures



(6)

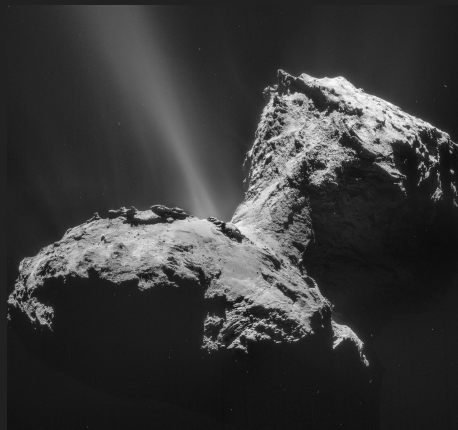
# On- & Off-world Sites

Memory of Mankind in Salt Mines



(9)

Barbarastollen Microfilm Archive



(7)

Rosetta Disk on Comet 67P



(8)

5D Foundation Trilogy on Tesla Roadster



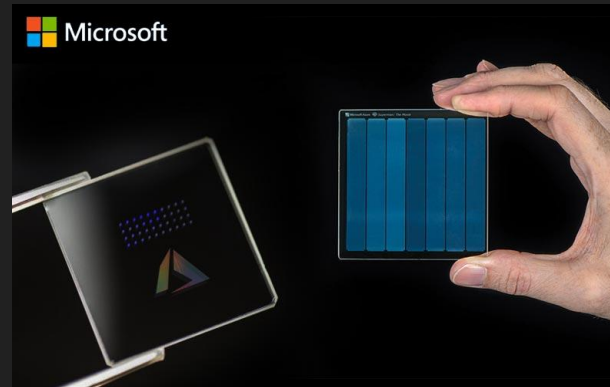


Norsam HD-Rosetta Technology

(10)

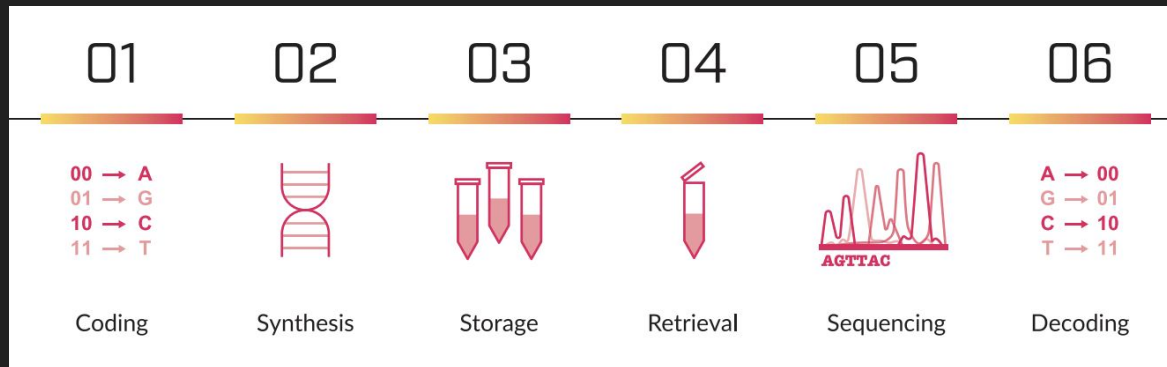
# Real World Applications

(11)



Microsoft's Project Silica

(12)



Data Storage in DNA

# Thought Experiment

(13)



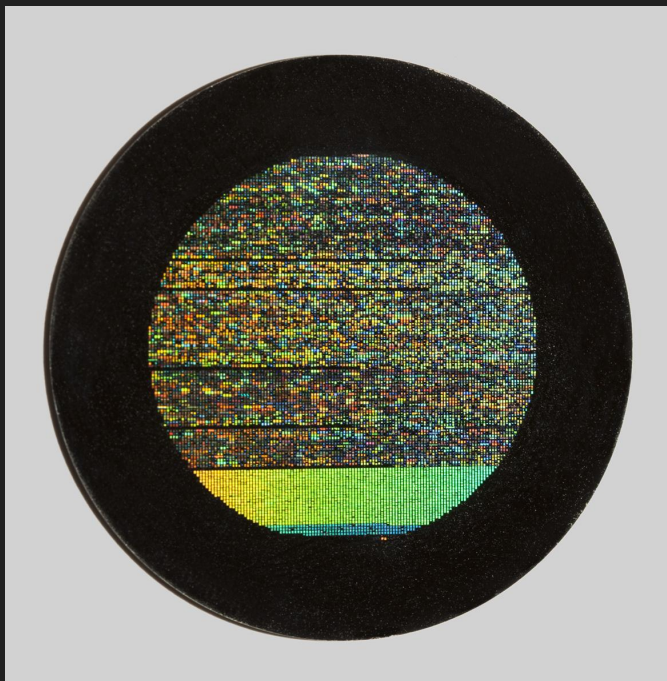
A Futuristic Archive

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

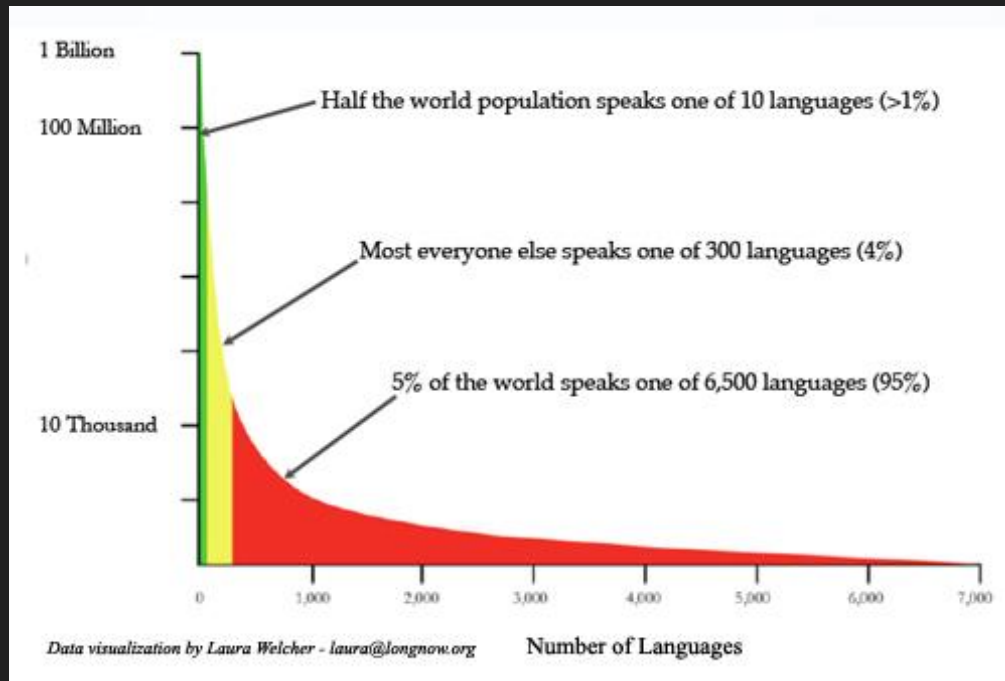
# Archiving Endangered Language Data

(14)



Rosetta Disk V. 1.0

(15)



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.

**DOI:** <https://doi.org/10.7551/mitpress/12200.003.0014>



# Image Credits

1. Rosetta Stone photo © Hans Hillewaert / [CC BY-SA 4.0](#)
2. The Rosetta Disk © The Long Now Foundation
3. The Rosetta Wearable Disk © The Long Now Foundation
4. Peter Kazansky's Optical Storage Disk (Photo © The Long Now Foundation)
5. Peter Kazansky and Laura Welcher (Photo © The Long Now Foundation)
6. Der Token des Memory of Mankind Projektes vor Hallstatt by Martin Kunze License: [CC BY-SA 4.0](#)
7. Comet 67P/Churyumov–Gerasimenko by ESA/Rosetta/NAVCAM, [CC BY-SA IGO 3.0](#)
8. Issac Asimov in Space (Photo © The Arch Mission)
9. Der Barbarastollen by Jörgens.Mi/Wikipedia, Licence: [CC-BY-SA 3.0](#)
10. HD-Rosetta (Photo © Norsam Technologies)
11. Microsoft's Project Silica (Photo © CIO Bulletin)
12. Digital Data to DNA Pipeline in [An Introduction to DNA Storage](#), DNA Data Storage Alliance, June 2021
13. Futuristic Archive (image generated by DALL·E 2023-03-19 13.00.51)
14. Rosetta Disk v 1.0 © The Long Now Foundation
15. The Long Tail of Human Languages © Laura Welcher