

# Buried GitHub archival storage to last 1,000 years

July 20 2020, by Peter Grad

---



Credit: Pixabay/CC0 Public Domain

If civilization ever succumbs to global warming, nuclear annihilation, an unrelenting pandemic or a martian invasion, some future civilization—or alien life form—might well still be able to reconstruct today's computers

and compose advanced machine learning algorithms ... and perhaps even play a round of Doom while they're at it.

That's because GitHub Inc. announced today that it has delivered a clone of all open source code stored on its servers to a vault buried 820 feet beneath an Arctic mountain.

Partnering with long-term storage company Piql, GitHub copied data onto 186 reels of silver halide microfilm. Storing 8.8 million pixels on each reel, the film cache tops out at 21 terabytes of data.

No one has to worry about quality issues due to heat. The repository will reside not too far from the North Pole, situated deep into the permafrost of a decommissioned coal mine on the island of Svalbard in Norway. GitHub expects the data will last 1,000 years.

There is a presumption—possibly an overly optimistic one—that should civilization remain basically intact through the year 3020, they will have computer-type devices that can read the stored archives. Included among the data will be blueprints for coding languages, software development, machine learning, web operations, essential computing principles and more.

Developers of the future hopefully will at least have a scanner on their latest iPhones; most of the repository data will be QR-encoded. By 3020, perhaps we'll have decoders embedded in our fingers. If not, the authors also included human-readable documentation—basically lengthy read.me files—covering the basic principles of computing.

"Reading, decoding and uncompressing this data will require considerable computation itself. In theory it could be done without computers, but it would be very tedious and difficult," GitHub states in documentation included on all archival data reels. "Our expectation is

that you didn't need our definitions of software, computer, and other terms. We imagine you have computers of your own, probably vastly more advanced than ours, and possibly fundamentally differently architected. Once you understand the overview and guide below you will easily be able to access all of the data."

They continue, "However, it's possible that you have inferior computers to ours, or even no computers at all. In case of that eventuality, we have prepared an uncompressed, unencoded, human-readable reel of data which we call the Tech Tree. The Tech Tree contains information about our fundamental technologies, our computers, and our software, in the hopes that, over time, you will be able to use this knowledge to recreate computers that can make use of the [open source software](#) in this archive."

The Arctic repository is one of several projects aimed towards long-term preservation of global open source data launched by GitHub. The company is working towards that goal with the Long Now Foundation, the Internet Archive, the Software Heritage Foundation, Arctic World Archive, Microsoft Research, the Bodleian Library and Stanford Libraries.

Microsoft has undertaken a project to eventually engrave all of the Arctic repositories onto fused quartz glass platters etched with a femtosecond laser. Such lasers emit pulses measured in one quadrillionth of a second. The platters can withstand electromagnetic interference, water and heat. As such, GitHub, a subsidiary of Microsoft, says Project Silica should ensure the repository lasts 10,000 years.

Which means when coronavirus, Donald Trump and The Bachelorette will be too distant to rank even a footnote in history, whatever state of affairs the world finds itself in at that point in time, Doomguy and the demons from Hell may still be alive and well and a part of the culture of

the Class of 12020.

© 2020 Science X Network

Citation: Buried GitHub archival storage to last 1,000 years (2020, July 20) retrieved 26 April 2024 from <https://phys.org/news/2020-07-github-archival-storage-years.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.