Frozen Data

written by Jeremy Huggett | 20/11/2020

Original image by Noel Bauza from Pixabay

There was a flurry of interest in the technical press during the summer with the news that GitHub had placed much of the open source code it held into an almost improbably long-term Arctic archive (e.g. Kimball 2020; Metcalf 2020; Vaughan 2020). GitHub's timing seemed propitious: in the midst of a global pandemic, with wild fires burning out of control on the west coast of the USA and elsewhere, and with upgrades to the nearby Global Seed Vault recently finished after being flooded as a consequence of global warming.

The Arctic World Archive was set up by Piql in 2017 and situated in a decommissioned mineshaft deep within the permafrost near Longyearbyen on the Svalbard archipelago. The data are stored on reels of piqlFilm (see Piql 2019, Piql nd), a high-resolution photosensitive film claimed to be secure for 750 years (and over 1000 years in cold low-oxygen conditions) and hence require no cycle of refresh and migrate, unlike all other forms of digital archive. The film holds both analog (text, images etc.) and digital information, with digital data stored as high resolution QR codes. Explanations of how to decode and retrieve the information are included as text at the beginning of each reel that can simply be read by holding it up to a light source with a magnifying glass, and Piql claim that only a camera/scanner and a computer of some kind will be required to restore the information in the future which means that the archive outlives any technology used to store the data in the first place.

What particularly grabbed my attention was a quote from GitHub's Vice President for Strategic Programs, Thomas Dohmke, who pushed my archaeological buttons by saying

> … it's worth preserving open-source software for a thousand years, in the same way mankind

> has preserved the Roman Forum, the Taj Mahal, the Bodleian Library. All those artefacts of human history tell us something about who we are and how we have developed. (Stevenson 2021, 34).

Perhaps perversely, it struck me that this statement might carry connotations that Dohmke hadn't appreciated. For instance, the analogy seems open to challenge. We can hardly claim these monuments have survived their journey from the past unscathed and unaltered by people and nature, and their damage and decay over time is much more akin to the risks of data loss, corruption, or unwarranted manipulation associated with the traditional refresh and migration cycle than the new long-term archiving offered by piqlFilm (Piql nd., 2).

There was also an Interesting and purely coincidental juxtaposition with a discussion led by Monika Stobiecka (2020) within the latest issue of *Archaeological Dialogues* on the digitisation and physical replication of the Palymyra Arch, which she and others argue reproduces western colonial attitudes with little apparent consideration for the Syrian people in whose territory the original monument sat. The replica represented a series of political statements, "hijacked by the imperial countries, 'civilized' and possessed thanks to their powerful technological tools, to finally become an artefact of ideological discourse" (Stobiecka 2020, 121; see also Meskell 2020, Rico 2020). In turn this reminded me of a blog post by Tim Hitchcock (2016 – ironically now missing but available via the Internet Archive!) in which he argued that the transfer of public analogue archives into the digital realm has had the effect of making Western data and values hyper-available and authoritative, and flattened out the range and diversity of human experience. This selection bias consequently supports a continuation of cultural homogenisation and cultural hegemony. It's difficult to get a clear impression of the range of deposits currently in the Arctic World Archive, but the examples provided, with some exceptions, have a strong western European flavour to date. In the absence of any information about policies or costs, for example, it risks leaving an impression of a technically, financially, and politically privileged perspective belonging to only a sector of humanity about what is valued or considered worthy of preservation.

Then there's the appearance of the well-worn trope of 'backing up heritage' in digital form. The Arctic World Archive website talks of holding an impressive collection of digital artefacts: "home to manuscripts … masterpieces from different eras (including Rembrandt and Munch), scientific breakthroughs and contemporary cultural treasures"; for example, the Vatican Library has deposited 500 "digitally preserved" manuscripts. The overall collection is inevitably somewhat eclectic given the archive has only been open for deposits since 2017 but there is an evident confusion in what Rico (2020, 125) describes as a pervasive rhetoric: the idea that archiving *is* saving. As she points out, we may preserve information about a monument, but that isn't the same as supporting its existence in the real world, within its associated context of people, landscape and history, and the equivalent is equally true in the way that digital manuscripts, artworks and the like are removed from their contexts of creation and use.

This is perhaps related to what Dohmke sees as the importance of the GitHub Arctic archive: not so much for the open source software themselves, but the way in which in 1000 years' time it will provide insights into how the software was written, how people collaborated, how languages etc. developed, and the range of practices and relationships (Stevenson 2021, 35-6). A similar case can be made for the artworks, stories, images, manuscripts etc. stored in the Arctic World Archive. But a

history of software development based on the GitHub archive alone would be an inaccurate one; likewise, the image of the past presented by the larger archive in the future will always be incomplete, regardless of the quality and longevity of its preservation. A problem not unfamiliar to archaeologists, of course.

It's interesting to think of the potential future archaeological response, and in this light to read the GitHub archive user guide on the piqlFilm, which was prepared by a panel of linguists, archivists, historians, and librarians to give future explorers a chance to decode their discovery. What is immediately apparent is how difficult it is to communicate the present to the future, how complex it is to try and ensure concepts current today are capable of being understood by indeterminate discoverers in the distant future and deal with the range of uncertainties and barriers to understanding. So, for example, inclusion of the Universal Declaration of Human Rights in 500 languages is seen as a form of Rosetta Stone to ensure that if English is no longer the lingua franca, the information and instructions are still capable of being decoded.

In many respects, the challenges faced by Piql and the Arctic World Archive are no different from those of more traditional archives with their 'warmer' data: issues surrounding selection, scope, and so on. But their much longer term storage media, almost 'file and forget' approach, in some respects exacerbates these questions with their ring of confidence disguising the very real issues underlying attempts to capture, and preserve snapshots of human culture and practice. And there is a notable absence of accessible information which are key components of any traditional archive – the lack of a collection policy, the absence of accreditation, no details of management or governance arrangements, and no charging policy or rates, for instance. Joining the Digital Preservation Coalition might be a good start.

> It is 3020. The newly opened door to the mysterious vault swings open. Our intrepid explorers shine their torches along the abandoned mine, walking a few hundred metres to the large fireproof container where, a thousand years ago, [an] … archivist … carefully placed a reel of film on a shelf … Carefully the explorers remove their gloves and open the lid of the canister … Now what? (Stevenson 2021, 36).

References

Hitchcock, T. (2016) 'Privatising the Digital Past', *Historyonics*, http://historyonics.blogspot.co.uk/2016/06/privatising-digital-past.html – available at https://web.archive.org/web/20170311194350/http://historyonics.blogspot.com/2016_06_01_archive.html

Kimball, W. (2020) GitHub Has Stored Its Code in an Arctic Vault It Hopes Will Last 1,000 Years. https://gizmodo.com/github-has-stored-its-code-in-an-arctic-vault-it-hopes-1844420340

Meskell, L. (2020). 'Hijacking ISIS. Digital imperialism and salvage politics', *Archaeological Dialogues*, 27(2), 126–128. doi: 10.1017/S1380203820000252

Metcalf, J. (2020) GitHub Archive Program: the journey of the world's open source code to the Arctic. https://github.blog/2020-07-16-github-archive-program-the-journey-of-the-worlds-open-source-code-

to-the-arctic/

Piql (2019) Brochure (ver 2.0) https://www.piql.com/resource/piql-services/

Piql (nd) What We Do Behind the Scenes
https://www.piql.com/resource/piql-technology-behind-scenes/

Rico, T. (2020). 'The second coming of Palmyra. A technological prison', *Archaeological Dialogues*, 27(2), 125–126. doi: 10.1017/S1380203820000240

Stevenson, D. (2021) 'Deep-freeze data that will last 1,000 years', *PC Pro* 315 (Jan 2021), 32-36.

Stobiecka, M. (2020). 'Archaeological heritage in the age of digital colonialism', *Archaeological Dialogues*, 27(2), 113–125. doi: 10.1017/S1380203820000239

Vaughan, A. (2020) 'World's most essential open-source code to be stored in Arctic vault', *New Scientist* 3276 (4 April 2020)
https://www.newscientist.com/article/2238586-worlds-most-essential-open-source-code-to-be-stored-in-arctic-vault/