Very Long-Term Backup

by Kevin Kelly

Paper, it turns out, is a very reliable backup medium for information. While it can burn or dissolve in water, good acid-free versions of paper are otherwise stable over the long term, cheap to warehouse, and oblivious to technological change because its pages are "eye-scanable." No special devices needed. Well-made, well-cared for paper can last 1,000 years easily, and probably reach 2,000 without much extra trouble.

We can not say the same for digital storage. Pages stored on plastic DVDs are neither stable over the very long term, nor readable over the long term. Unless digital information is ceaselessly migrated from one fading medium to another new one, it will quickly cease to be accessible. Two decades ago the floppy disk was ubiquitous. Most personal digital information then was stored on this format. Today, any information stored only on a floppy disk is essentially gone. Imagine the incompatibility of today's DVD in 1,000 years.

As durable as paper is, its inherent limitations in storing digital data are clear. Pity the person who would need to find something if the only backup of the web was a paper printout that filled several airline hangers. What we need are media that have the durability of paper and the accessibility of a floppy disk (or better!).

This problem of long-term digital storage seemed a crucial hurdle for any civilization trying to act generationaly. How could a society think in terms of centuries unless there was a reliable way to transmit and store its knowledge over centuries? This puzzle was the focus of a conference hosted by Long Now in 1998, dedicated to technical solutions for Managing Digital Continuity. At this meeting Brewster Kahle of the Internet Archive suggested a new technology developed by Los Alamos labs, and commercialized by the Norsam company, as a solution for long term digital storage. Norsam promised to micro-etch 350,000 pages of information onto a 3-inch nickel disk with an estimated lifespan of 2,000 -10,000 years.

Might it be possible to etch an entire library onto a set of disks? It might be worth trying. All we needed was a finite data set that a society might want to have backed up.

During a Long Now field trip to a southwest archeological site, the idea of a modern Rosetta Stone came up — a backup of human languages that future generations might cherish. At a winter retreat in 1999, Long Now board member Doug Carlston suggested that for the parallel common text of this modern Rosetta Stone we should

use the book of Genesis, since it was most likely already translated into all languages already. We hatched a plan to produce a 3-inch non-corroding disk which contained at least 1,000 translations of Genesis and other linguistic information about each language.

Following the archiving principle of LOCKS (Lots of Copies Keep 'em Safe) we would replicate the disk promiscuously and distribute them around the world with built in magnifiers. This project in long term thinking would do two things: it would showcase this new long-term storage technology, and it would give the world a minimal backup of human languages. We thought it might take a year to do.

Long story short, it took eight years. Last night at a ceremony at the Long Now museum in Fort Mason, one of five prototype disks Rosetta disk was presented to the Oliver Wilke Foundation, a Frankfurt-based linguistic center, who help support the project. The disk is 3 inches in diameter, and mounted beneath a glass hemisphere.

One side of the disk contains a graphic teaser. The design shows headlines in the eight major languages of the world today spiraling inward in ever-decreasing size till it becomes so small you have trouble reading it, yet the text goes on getting smaller. The sentences announce: "Languages of the World: This is an archive of over 1,500 human languages assembled in the year 02008 C.E. Magnify 1,000 times to find over 13,000 pages of language documentation."

This graphic side of the disk is pure titanium. A black oxide coating has been added to the surface. The text is etched into that, revealing the whiter titanium. This bold sign board is needed because the pages of genesis which are etched on the mirror-like opposite side of the disk are nearly invisible.

This business side of the disk is pure nickel. Picking it up you would not be aware there were 13,500 pages of linguistic gold hiding on it. The nickel is deposited on an etched silicon disk. In effect the Rosetta disk is a nickel cast of a micro-etch silicon mold. When the disk is held at the right angle the grid array of the pages form a slight diffraction rainbow. You need a 750-power optical microscope to read the pages.

The Rosetta disk is not digital. The pages are analog "human-readable" scans of scripts, text, and diagrams. Among the 13,500 scanned pages are 1,500 different language versions of Genesis 1-3, a universal list of the words common for each language, pronunciation guides and so on. Some of the key indexing meta-data for each language section (such as the standard linguistic code number for that language) are displayed in a machine-readable font (OCRb) so that a smart microscope could guide you through this analog trove.

Our hope is that at least one of the eight headline languages can be recovered in 1,000 years. But even without reading, a person might guess there are small things to see in this disk.

All this took eight years because back in 2000 the Norsam technology could not handle the size of our library, and there was in fact, contrary to our assumptions, no library of already completed Genesis translations. There was no central depository of language information, either. So in order to gather 1,000 translations of Genesis and related linguistic information for those 1,000 language, Long Now created the Rosetta Project.

Heading the project was artist/linguist Jim Mason, who ran Rosetta at first like it was an art project. Which it kinda was. Working under the radar of the academic linguistic community, Mason began collating and scanning all known versions of Genesis, and later regional and ethnic creation stories in native languages. He collected maverick linguists and bridged the feudal factions in the academic linguistic community. Under Mason the project quickly morphed from art project into a major linguist initiative. Mason steadily won the support of the world-wide professionals as the Rosetta website grew into the "All Language Archive." Eight years down the road, after major NSF grants and other funding the Rosetta Project now has a unified (such as it exists) set of information in 2,300 languages. At several points in its evolution, the Rosetta's tiny non-profit offices were crammed with dozens of grad students scanning pages of wonderfully obscure languages as fast as paper could move. Over 100 people contributed work in the office and thousands more on the website. The intention all along has been to cram this all language archive onto a few disks. Or a tiny cube. Or maybe, art project at the core, etched onto a long wall.

This is a Long Now project, which means it is okay if it takes a while. It took 8 years to gather the scanned Genesis texts. During that time Norsam perfected their production. Now we have a disk.

But it was not the very first disk. That one is in space. In 2004 the Rosetta Space Probe was launched by the European Space Agency. This small craft was created to land on a comet in 2014. Before it blasted off, the ESA contacted us because we share names. They asked if we'd like to mount a version of the disk on their probe. Of course we would! We had manufactured a pure nickel disc with a subset of 6,000 pages of language translations, which was mounted on the payload section of the probe.

So assuming the mission continues well, in 2014 the Rosetta Probe will land on Comet 67P/Churyumov-Gerasimenko, where it will measure the comet's molecular

composition. Then it will remain at rest as the comet orbits the sun for hundreds of millions of years. So somewhere in the solar system, where it is safe but hard to reach, a backup sample of human languages is stored, in case we need one.

Or you can have one on earth, if you want, acting as an additional node in the distributed archive. There are still two disks available from this prototype run. Currently, for all its high techness, each disk is hand crafted, and so they have a corresponding high hand-crafted cost: $25,000. Contact the office if you are interested in caretaking an archive of all languages. Long Now hopes to produce additional copies in the future, so that these small globes will be scattered across the world in nondescript locations; that way at least one will survive their 2,000-year lifespan.

There's a small hidden cavity inside the globe where owners can inscribe their name, with room and encouragement to have the next owners inscribe theirs. This is a multi-generational device. As Oliver Wilke said when he picked up his glass sphere last night, "This is one of the most fascinating objects on earth. If we found one of these things 2,000 years ago, with all the languages of the time, it would be among our most priceless artifacts. I feel a high responsibility for preserving it for future generations."