

[<< Back to Article](#)

Word Up: Keeping Languages Alive

Kendra Mayfield 11.04.02

When Napoleon's troops discovered a granite slab in 1799 containing Greek and Egyptian hieroglyphic translations of ancient text dating back to 100 B.C., they unearthed more than 1,000 years of history.

Now, a group of scientists and engineers are crafting a modern Rosetta stone that will preserve more than 1,400 of the world's 7,000 languages on a 3-inch [nickel disk](#).

Fifty to ninety percent of the world's languages are predicted to disappear in the next century, according to the [The Rosetta Project](#), a collaborative, open-source endeavor by language specialists and native speakers around the world who are creating a "near permanent" archive of the world's languages.

Developers of the modern Rosetta disk hope they will help future generations recover lost languages that are now on the brink of extinction.

"A broad linguistic database, covering (a) significant percentage of world languages, is a very important research tool for studying language evolution and the specifics of human migration," said project director Jim Mason.

The first version of the disk will be completed this fall.

Given the massive amount of work involved in cataloging the world's languages, adopting an open-source approach made sense, Mason said.

"The range of languages we are working across is much greater than any individual or small group of individuals would be able to deal with," he said.

The results of this "Linux of Linguistics" effort will be available as a [free online archive](#) and as a single-volume reference book, in addition to the disk itself.

The online version contains what may be the world's largest comparable database of words. Users can find words from the 1,445 languages tracked by the group.

Efforts like the Rosetta Project help record and recover moribund languages that are undocumented and typically spoken only by a few elderly people.

"It's extremely important for long-term archiving," said Doug Whalen, founder of the [Endangered Language Fund](#). "(Archiving) is a truly gigantic problem for language preservation."

Instead of using digital storage methods that quickly decay and can become obsolete, Los Alamos Laboratories and Norsam Technologies developed a technology that micro-etches text on a high-density storage disk with a life expectancy of 2,000 years.

"For the long term, there hasn't been anything digital that has had a ghost of a chance of being taken as seriously archival," Whalen said. "Huge chunks of input data have already disappeared."

"(The analog disk) could make a big difference, hopefully sometime in the very far future."

Information will be presented in plain text -- not 0s and 1s. Future generations will need only a 1,000-power microscope to read the microprint.

The disk contains vocabulary lists, phonology, writing systems, sample texts, grammar and numbering systems for each language.

It also displays translations of the first three chapters of the book of Genesis in eight major languages, from Arabic to Swahili.

Eventually the Rosetta Project plans to mass-produce the disks and distribute them around the world.

For now, individuals can [acquire](#) a disk with a \$25,000 donation to [The Long Now Foundation](#).

Project directors hope to bring the cost down to about \$100 per disk. In addition, free disks will likely be distributed to museums and libraries.

Besides preserving the languages themselves, researchers also want to protect their collections of research and other

background information on the endangered languages.

Much of that knowledge is stored using technology that potentially has a short shelf life.

"The computing and recording technologies that are now standard tools in doing field linguistics are changing so quickly that information captured electronically today could cease to be accessible in another decade or two if special care is not taken to ensure that it is archived in stable formats by stable institutions," said Gary Simons, coordinator of the Open Language Archives Community.

To alleviate this problem, the [OLAC](#) is creating a virtual library of global language resources that will use a common, searchable metadata format.

As a participant in the OLAC, the Rosetta Project is adopting these same standards so that native speakers, academics and the general public can easily access its language archive.

But while the Rosetta disk will provide a means for long-term linguistic preservation, some say the project's focus doesn't go far enough.

"The Rosetta Project's disk will certainly last for a long time, but it offers only a static view of textual information," Simons said. "We also need sound recordings and video recordings to have a complete record of a language. And we really want information that can be dynamic and repurposed."

Audio files from native speakers will eventually be included, Mason said.

So will other languages. Next summer, the project will launch a new archive that contains all 4,000 of the world's documented languages.