

The Washington Post

Saving Our Digital Heritage

By Jim Barksdale and Francine Berman
Wednesday, May 16, 2007

It is commonly agreed that the destruction of the ancient Library of Alexandria in Egypt was one of the most devastating losses of knowledge in all of civilization. Today, however, the digital information that drives our world and powers our economy is in many ways more susceptible to loss than the papyrus and parchment at Alexandria.

An estimated 44 percent of Web sites that existed in 1998 vanished without a trace within just one year. The average life span of a Web site is only 44 to 75 days. The gadgets that inform our lives -- cellphones, computers, iPods, DVDs, memory cards -- are filled with digital content. Yet the lifetime of these media is discouragingly short. Data on 5 1/4 -inch floppies may already be lost forever; this format, so pervasive only a decade ago, can't be read by the latest generation of computers. Changing file and hardware formats, or computer viruses and hard-drive crashes, can render years of creativity inaccessible.

By contrast, the Library of Congress has in its care millions of printed works, some on stone or animal skin that have survived for centuries. The challenges underlying digital preservation led Congress in 2000 to appropriate \$100 million for the Library of Congress to lead the National Digital Information Infrastructure and Preservation Program, a growing partnership of 67 organizations charged with preserving and making accessible "born digital" information for current and future generations.

Some of the crucial programs funded by NDIIPP include the archiving of important Web sites such as those covering federal elections and Hurricane Katrina; public health, geospatial and map data; public television and foreign news broadcasts; and other vital born-digital content.

Unfortunately, the program is threatened. In February, Congress passed and the president signed legislation rescinding \$47 million of the program's approved funding. This jeopardizes an additional \$37 million in matching, non-federal funds that partners would contribute as in-kind donations.

Some of the projects that were to be funded include preservation of important government records at the state level, such as legislative data and court records. Another new project at risk, "Preserving Creative America," is an initiative with commercial producers of creative content, such as digital film, music, photography, other forms of pictorial art and even video games.

We have seen what happens when valuable public data are inadequately preserved, lost or not available when needed. For example, the original, raw data from the 1960 Census were stored on a state-of-the-art UNIVAC computer. When the Census Bureau turned the data over to the National Archives in the mid-1970s, UNIVAC computers were long obsolete. Much of the information was eventually recovered, but at a huge cost. Raw data from early satellite probes, including the Viking mission to Mars, pre-1979 Landsat images of Earth and high-resolution images of the moon, have been lost for similar reasons.

Advertisement » Your Ad Here



Current estimates are that in 2006, 161 billion trillion bytes -- 161 exabytes -- of digital data were generated in the world -- equivalent to 12 stacks of books reaching from the Earth to the sun. In just 15 minutes, the world produces an amount of data equal to all the information held at the Library of Congress. While it is unrealistic to think that we will be able to preserve all the data produced solely in digital form, NDIIPP convenes top experts to help decide which at-risk content is most critical and how to go about saving it.

Responsible preservation of our most valued digital data requires answers to key questions: Which data should we keep and how should we keep it? How can we ensure that we can access it in five years, 100 years or 1,000 years? And, who will pay for it?

The importance of developing sensible plans to preserve our digital heritage cannot be minimized. We can't save it all, nor do we want to. It's also critical that we agree on how to save this data. In the next 100 years, we will go through dozens of generations of computers and storage media, and our digital data will need to be transferred from one generation to the next, and by someone we trust to do it.

The National Digital Information Infrastructure and Preservation Program provides a good start, and Congress has an opportunity to restore \$21.5 million requested by the Library of Congress to continue the program and sustain the partnerships needed to fulfill the critical task of preserving our nation's important born-digital information.

It would be a national and a global shame if our most valuable born-digital knowledge, like the ancient holdings at Alexandria, were lost forever.

Jim Barksdale is the former chief executive of Netscape Communications Corp. and is an executive member of the National Digital Information Infrastructure and Preservation Program Advisory Council. Francine Berman is director of the San Diego Supercomputer Center at the University of California at San Diego. She holds the High Performance Computing Endowed Chair at UCSD's Jacobs School of Engineering.

[View all comments](#) that have been posted about this article.

Ads by Google

Digital Archive Services

Professional Digital Archiving, Cost efficient, High Quality, Learn More
Aptaracorp.com/DigitalArchive

MA Historic Preservation

Limited residency program Nationally known faculty
www.goucher.edu/mahp

LG Computer Drives

External & Internal Drives W/ Enormous Storage Capacity!
www.LGE.com

© 2007 The Washington Post Company