Breaking News

Fragile digital data in danger of fading past history's reach

Source: *The Atlanta Journal-Constitution* (6-7-06)

It is estimated that in the next three years, humanity will generate more data than it has in the past 1,000 years --- nearly all of it digital.

Many of the records that once allowed historians to study a society's history --- from personal correspondence to government documents --- may be slipping, irretrievably, into the digital ether.

"It's a major historical problem and presents, potentially, a major political problem," says Air Force historian Eduard Mark. "Someday it will erupt. There will probably be a major loss to history."

The nation's greatest institutions of record-keeping, the Library of Congress and the National Archives, have recently begun to create new electronic archives to catalog and preserve for posterity the billions of digital records they now receive.

But it's an overwhelming goal, one that, despite today's breadth of information technology, isn't quite possible to meet --- yet.

"We need to preserve digital information in such a way that it will be intelligible in 100 years," says Abby Smith, a consultant with the Library of Congress' digital preservation program. But she and others say that when it comes to preserving digital records, a solution has yet to be found.

"In retrospect, this will look like a period of the digital dark ages," she says. "There's information which we're producing now which we will not be able to save for the future."

'Very ephemeral'

The problem is that, compared to the sturdy format of paper and books, digital information is extremely fragile, disappearing as software becomes obsolete, hardware breaks down and viruses wipe out volumes.

"Digital media can be very ephemeral. They can decay," says Anne Okerson, of the Council on Library and Information Resources. "For example, will a Word or Word Perfect document still be readable in 10 years, several versions later? Mine aren't ... how about a CD? Doubtful."

Even if the media on which information is saved endure for years, what

happens when the technology to extract and read it becomes obsolete?

Jon Prial, IBM's vice president of content management, asks, "If something is saved digitally now, the question becomes, can I save a CD somewhere for 1,000 years? If I can, will there be something to play it on?"

Obsolete formats

At the National Archives, staffers are already experiencing such problems. The government began storing key military records, such as flight details, on computers as early as the Vietnam War era, says Kenneth Thibodeau, director of its Electronic Records Archives program. Today, those records of every flight in Vietnam "are sitting in obsolete tapes in an obsolete format," says Thibodeau.

He cites another example: The National Archives has always preserved drawings of Navy ships.

"But there aren't drawings anymore. All the records are digital. For information on the structure of one ship, you're talking hundreds of millions of computer files," he says.

Since the average Navy ship is kept afloat for 50 years, there's a strong chance that engineers or anyone else needing structural information about the ships may be out of luck in a mere 20 years if those electronic files become corrupted or inaccessible, Thibodeau says.

"The Navy has a digital preservation problem. After all, what does the Navy know about what computer-assisted design programs will be used 20 years from now? They just know it will be different. The aerospace industry has the same problem. It's also true for designs of power grids and bridges."

James J. McSweeney, regional administrator for the National Archives Southeast Region in Jonesboro, worries that future historians will miss out on the experience of sifting through original physical records and that a palpable element of discovering history may be lost.

"In Atlanta we have 115,000 cubic feet of historically valuable documents, dating back to 1716. We've got everything from the Tuskegee syphilis records to Rosa Parks' original arrest records," McSweeney says. "Nothing can reproduce the experience, for researchers, of coming here and working directly, hands-on, with those records. That's the type of thing you're not going to get from a Google-type search."

Gaps possible

But the most chilling implications might be for the historical record.

"The difference between someone 25 years from now trying to sort out how we became involved in Iraq, vs. someone today studying the Cuban missile

crisis, say, is that for the future historian, the records could be much less comprehensive, and there could be much fewer of them," says Mark, the Air Force historian.

The irony is that far more records are now being created --- another problem for those racing to save and archive them. Thibodeau compares the correspondence the National Archives received from the Nixon administration --- 40 million pages --- to the estimated 100 million e-mails alone it expects to receive from the Bush administration.

"We've got to build a system that never becomes obsolete, even though we assume that each separate piece of hardware within the system will eventually become obsolete," says Thibodeau. "[And] it's got to grow to accommodate new stuff that I don't even know about yet."

Despite the massive challenges, Thibodeau expects a first version of such a system to be online for the public by 2009. Even then, though, the Electronic Records Archive and similar systems will be works in progress for a long time, he says.

"There's a lot more information that's being captured than ever before," he says. "If we do figure out how to preserve this stuff, people will be in good shape to research the second half of the 20th century. But by and large, no one's yet been able to figure out how to preserve this stuff for the next 25 years."

ShareThis

Posted on Wednesday, June 7, 2006 at 10:14 AM | Comments (0) | Return

Home Newsletter Submissions Advertising Donations Archives Internships About Us FAQs Contact Us All Articles