Probing Question: Can we save today's documents for tomorrow?

February 12 2009, By Adam Eshleman

Credit Rob Williams

Even though your grandparents' old photo albums are yellowed and grainy, they're still there for you and your family to enjoy. But will your grandchildren be able to say the same of the digital photo albums you're compiling today? Rapid advances in computer technology have left past hardware and software in the dust. If we're creating documents no one will be able to access, has our Information Age created a digital dark age?

"Perhaps," said Jackie Esposito, Penn State's University archivist. "Electronic forms of information can and probably will be lost because we have no way of accessing older digital files and no way of preserving them," she noted. "If you hand me a floppy disk today, I can't find a

computer to read it. But it's only been 40 or 50 years that computers have existed. From a historical standpoint, that's a short dark age."

For the past two-and-a-half years, Esposito has been working with Penn State's Digital Preservation Steering Committee to guard University records against digital oblivion. "We're looking primarily at materials that were 'born digital,' meaning they were created on the computer and have lived on the computer and may never have been printed out," she explained. "How are we going to make sure these records are still here 50 years from now? How are we going to do that if the average lifespan of hardware and software is 18 months?"

Penn State has already lost some of its non-permanent records as a result of this dilemma, she pointed out. In the early 1980s, the University began using computer systems to manage business and student records. The information was stored on magnetic tapes, which are no longer accessible due to decay and the lack of hardware to read them. Nor is Penn State alone: NASA has reported losing up to 20 percent of the information recorded by the 1976 Viking mission, data also stored on magnetic tape.

As unlikely as it may sound, preserving e-mail and text messages is one of the University's greatest priorities right now, said Esposito. Of particular concern are electronic communications generated by high-profile figures like University President Graham Spanier. "These files will one day become important historical documents," she noted. "If you want to look at the papers of Penn State's first president, Evan Pugh, you come in here and I'll give you a nice little box filled with aging papers. For Graham Spanier's papers I'm going to send you to a server, because they were born digital. It's a different research paradigm."

To ensure that its digital documents are preserved for posterity, Esposito says, "We're looking at systems that are interoperable, meaning they'll

work on any computer system. We're looking at formats like PDF/A." Like the now-standard PDF (for Portable Document Format) that it's based on, a PDF/A file can capture both images and text. But unlike a standard PDF, it doesn't require preinstalled software to be read. Instead, the software is embedded in the file itself. This means a PDF/A will be accessible across a variety of operating systems.

Unfortunately, with technology evolving at a relentless pace, preserving digital files isn't a once and done deal, reminded Esposito. "Five years from now, I'll have to decide what I'm doing with all those PDF/A's to make them readable. And that decision-making process will continue 20 years down the road, by whoever's in charge at that point in time," Esposito said.

As for farther in the future, she's confident that the Information Age will live up to its name. "I think 100 years from now, 30 to 40 percent of this information will be available. That's an acceptable number. I mean, if I look at what records are available from the Middle Ages, it's probably 5 to 10 percent of whatever was created."

In the interim, if you really want to preserve those digital photo albums and e-mail love-notes, Esposito recommends burning them to DVD-R recordable format and making a commitment to reformatting your files once every five years or so. And if you're using rewritable DVDs as your storage medium, it's a good idea to burn new copies every few years, she said. DVDs last only 10 years before the information stored on them begins to corrode.

"There are definitely steps we can take to preserve documents, Esposito said, "but we haven't come up with all the answers yet. I'd call that more of a murky age than a dark age."

Source: By Adam Eshleman, Research/Penn State