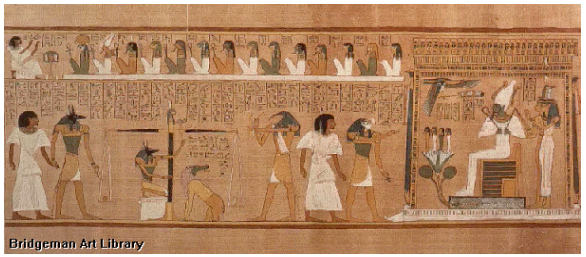
These are slides from a presentation given by Chris Muller at the Data-at-Risk session of the Library of Congress sponsored ***Digital Preservation 2014 Conference*** in Washington DC.

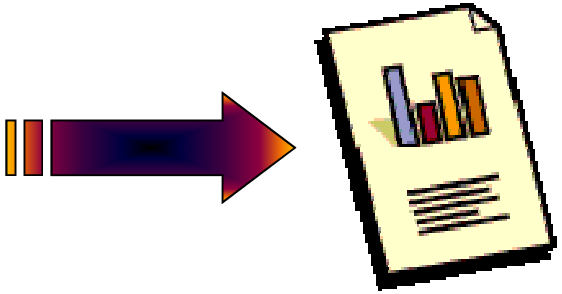
Why early digital assets merit special attention.

We just heard a great talk focused on the need to rescue and preserve pre-digital scientific research.

Now let's discuss why our older digital treasures are often at greater risk of loss, along with some fun examples of data rescue and thoughts about preserving capabilities as well as data.

Many of our pre-digital records are visual in nature but often difficult to transcribe and translate.

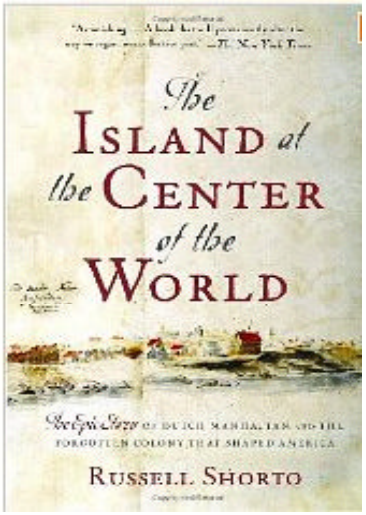




However, we've have three things going for us:

1. Often it was recorded on lasting material.
2. We could see what needed to be read/copied.
3. Optical scanning, OCR and digital photography are mature and ever-improving tools, with many folks skilled in their use.

One great example of pre-digital preservation and translation.



12,000 pages of New Amsterdam records sat in an Albany vault for over 300 years. Re-discovered in 1973, leading to Russell Shorto's wonderful book *"Island at the Center of the World"*.

Luckily, those pages had the "Luxury of Languishing"*.

However…

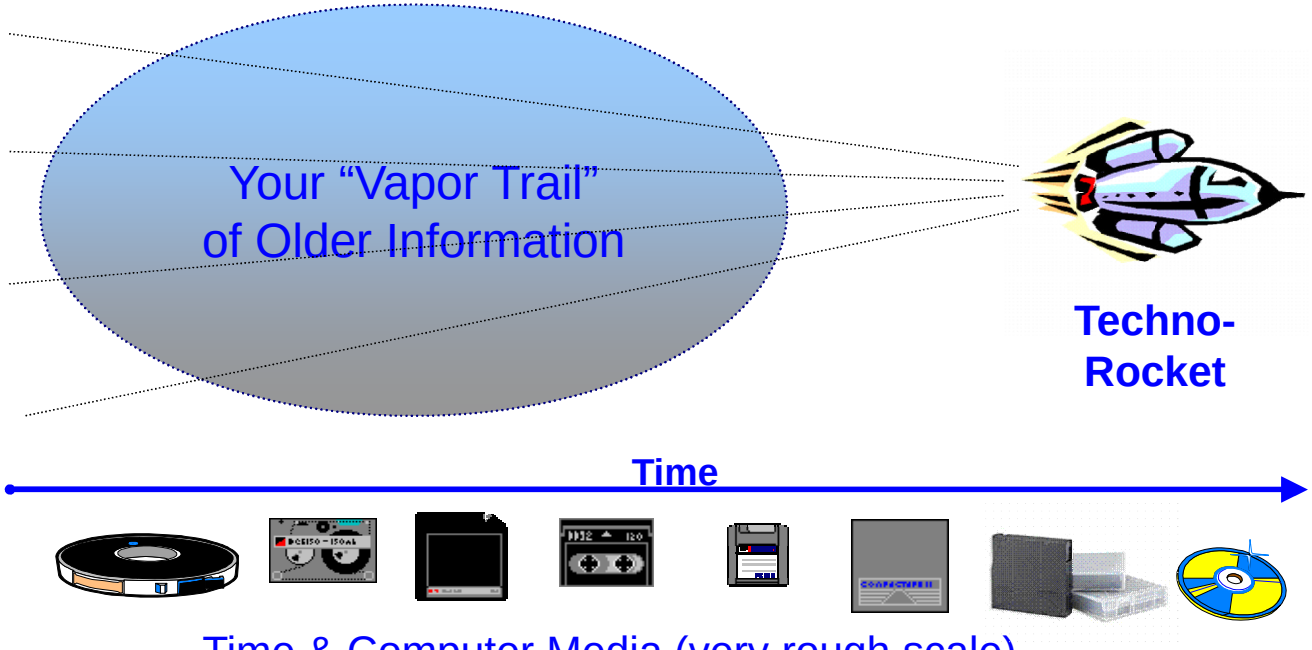
*LoL, as we say. ☺

With digital material, we are losing all three of those former advantages.

1. If only our magnetic media would last as long as those New Amsterdam records.
2. If only we could read those arcane files and media without special equipment and software.
3. If only the equipment, software and skills to make sense of that data weren't slowly fading away.

Technology's Vapor Trail

(not a good sort of Cloud Computing)



Time & Computer Media (very rough scale)

There are challenges, as mentioned, but there are some advantages:

Back in the day, computer media was expensive and difficult to use--little in the way of tweets, mp3s, politics etc. The value density is greater than that on much of today's media.

A good ratio exists between the capacity of old media and inexpensive, durable new media. An example: the content of 9,000 mainframe cartridges fit on one inexpensive hard drive. Future backup efforts are tiny compared to reading all those tapes.

Obstacles to Reading and Conversion.

It's a bit like peeling back the layers of an onion:

- Media Compatibility
- Age and Storage Conditions
- Recording Method

- -*

- Operating System/Filing System
- Backup, Exchange, or Archiving Software
- Application File Structure
- Application File Encoding

*Virtual tape and disk copies can protect the bits and bytes, allowing time to tackle the other layers as time and funding permit.

Why must we be aggressive regarding older digital materials?.

A. We, as a community, need to preserve not just data, but the tools and skills to "rescue" those data collections as they continue to turn up.
B. It's fun! Learning to work with (and continue to improve) existing equipment and software tools; puzzling out early, sometimes unique data formats; these challenges can be very enjoyable.
C. Even better, working with scientists, researchers, historians and archivists who really understand the value of these digital assets.

Some fun examples follow…

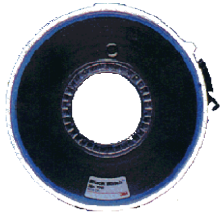
Some pre-Whitewater fun.

Clinton + Moscow = ?

Today, we'd tend to think of something like this:

But in 1969, a young fellow had visited Moscow, and now (1992) he was a contender for a presidential nomination. Political rivals curious. FOIA request. Old State Department tape.

Problem: un-documented file format (a sadly common occurrence). But the D.A.'s office found some folks who enjoyed hacking away.



This was fun, but our real exposure to the world of digital preservation and the great people involved in it was yet to come. (See next slide.)

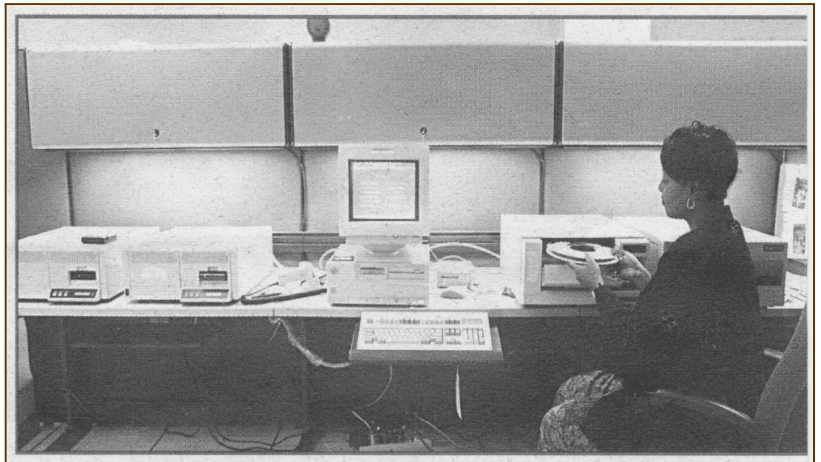
Later that year, NARA issued an RFP.

(Interviewed in the FDR "fireside chat" room. ☺)

NARA uses tape conversion system to preserve records

The National Archives and Records Administration has stepped up its preservation of electronic records with the installation of a $510,000 records maintenance system. Responsible for preserving the government's electronic mail and other data stored in digital formats, NARA's Center for Electronic Records in 1992 contracted **Muller Media Conversions** of New York, to develop and install…

…GOVERNMENT COMPUTER NEWS, AUGUST, **1994**



Sharon Smoot, a staff member from the Center for Electronic Records, is copying accessioned records onto archival-quality magnetic tape as part of the Center's ongoing preservation.

- Award-winning (actually, Ms. Eaton →)
- Contract extended continually (16 yrs)
- Ongoing enhancements
- MMC retains certain software rights.

IAC/IRM honors federal I/T leaders

Fynette Eaton of the National Archives and Records Administration's Center for Electronic Records receives a ***GSA Technology Excellence Award*** for developing the Archival Preservation System to store records and capture data about their physical and technical attributes…

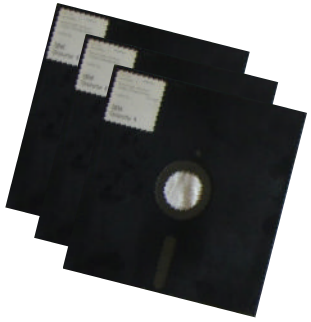
…GOVERNMENT COMPUTER NEWS, JUNE, **1996**

(That new president turned out to be a strong supporter of digital preservation.) ☺

Some real Whitewater stuff.

In the 1980's at a law firm in Little Rock, Arkansas, someone did legal work for a real estate project named "Castle Grande". By early 1994, all related records had mysteriously vanished (been "vacuumed") from the firm's systems.

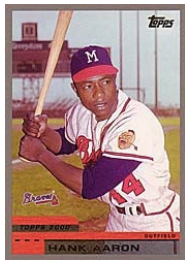
The firm does its due-diligence, and requests data recovery. Obviously, the perp was not a big fan of digital preservation—and he did not realize that good old Wang 8 inch floppies are not so easily erased. (Sorry, no details permitted.)



Some Watergate fun.

In the early 1970's, a Whitehouse mainframe used a proprietary database format to store presidential appointment calendar and notes. Two problems:
(a) the tape was in danger of decay, and
(b) data format not figured out for 25 years.

Asked to puzzle it out and make a new database and program for researchers. Luckily*, the analyst had done some work with Vietnam era military records and noticed similarities in the data structure.



Major Discoveries!
For instance, the first appointment after taking office was with Hank Aaron.

*Re this sort of "luck": A young Lee Trevino was asked if he felt lucky to have gone from a poor Latino teenage caddy to a world-famous golf pro. "Yup," he said "and the more I practice the luckier I get."

Beverly Hills High School

A famous & beautiful litigator (see that movie?) decided that oil wells on school property had been endangering student health since 1970.

In the school basement for decades were dozens of Vydec floppy disks with student records from the late 70's.

Anyone old enough to remember Vydec? A really weird onion. By 2003 ability to read them had almost vanished. (Not physically compatible with standard 8" drives—spins in the opposite direction!)

IPUMS International

A source of real enjoyment: the Minnesota Population Center and IPUMS.org. They collect/analyze/publish population data from around the world. It becomes apparent that the need for data-rescue is even greater in the developing world. Tapes and disks coming in from such places as:

- ***Bangladesh***
- ***Egypt***
- ***Kenya***
- ***Mali***
- ***Mexico***
- ***Nepal***
- ***Pakistan***
- ***Peru***
- ***Qatar***
- ***Romania***
- ***Santo Domingo***
- ***Sudan***

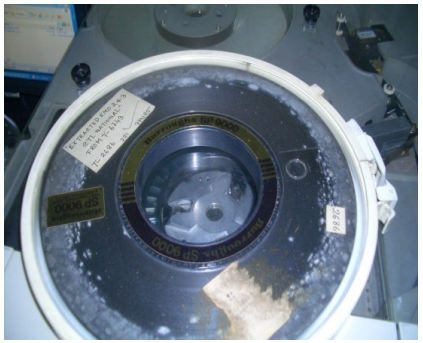
All kinds of stories. One involves the Peru Census of 1981; picking up well-packed tapes at the NY Consulate -→

The next page shows one of the most interesting projects…

Dhaka –

at the Bangladesh Bureau of Statistics.

Very capable people, with other pressing duties; they had been confronted with daily power outages and other problems making it impossible to store legacy tapes in an optimal way. Many suffered from decomposition.

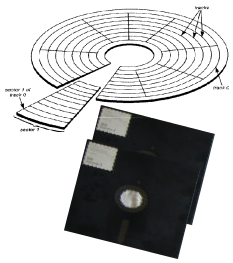


Plan: install read-convert system plus tape cleaners, training for BBS Staff. Initial goal: recover 1981 Census microdata.

There's a world-wide need for data rescue which can lead to many fascinating projects—presuming that focus and funding are available.

Penobscot Dictionary

Dr. Frank Siebert dedicated much of his life to the preservation of the Penobscot language. Many interviews with native-speakers, then created detailed dictionary in a rare format.



He left many diskettes not compatible with standard operating systems; char encoding was morphed for unique display hardware.

American Philosophical Society and Maine Folk Life Center undertook a project to ensure the preservation of this cultural treasure.

Working with great organizations like APS and MFLC to preserve the legacy of someone like the brilliant Dr. Siebert makes data-rescue a real pleasure. An honor to participate.

Old Professors' Stashes ("OPS")

Ready to retire or move on to another university—those old tapes are my legacy! Let's have another look.

Or maybe others realize that (for example) the health data, re-purposed and combined with population and climate information may well produce remarkable insights.

Re-discovered stashes like this are cropping up all the time!

All kinds of stories, if you have a few more hours. ☺

The tools and skills for rescuing older digital assets apply across the boundaries of science and culture. A valuable NPO resource can be created to work with scientists, historians, librarians and archivists.

Data Rescue:

If it's worth doing, it's worth doing now.

(No Luxury of Languishing)