

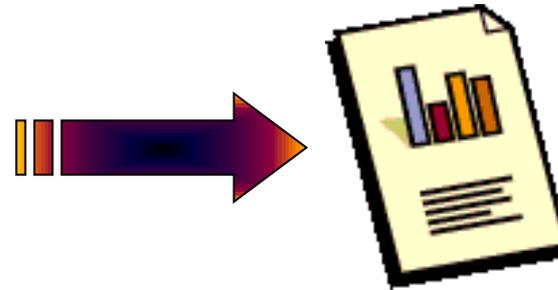
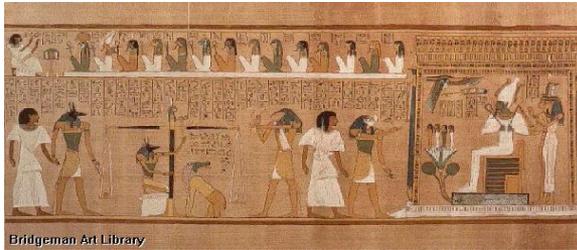
*Data Rescue Now*

# *Why early digital assets merit special attention.*

A look at the reasons why many of our older digital treasures are threatened with extinction, along with some fun examples of data rescue.

*Why DARTG and others need to Raid those Lost Archives ASAP.*

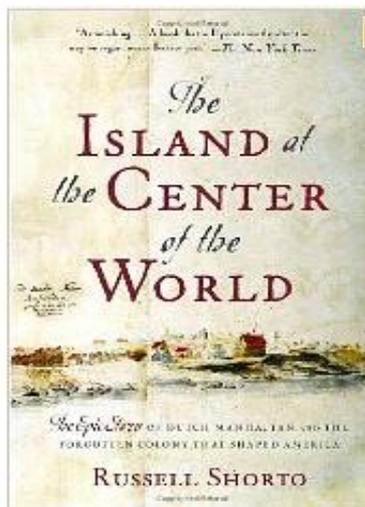
*Many of our pre-digital records are visual in nature but have been difficult to transcribe and translate.*



*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 people having been trained in their use.*

One great example of visual record scanning,  
preservation and translation.

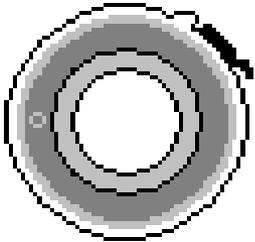


12,000 pages of New Amsterdam records sat in an Albany vault for over 300 years. Re-discovered in 1973, leading to Russell Shorto's wonderful book *"Island at the Center of the World"*.

Luckily, the medium was capable of lengthy survival. Those pages had the "Luxury of Languishing".

However...

*With digital material, we have lost (or are losing) all three of those former advantages.*

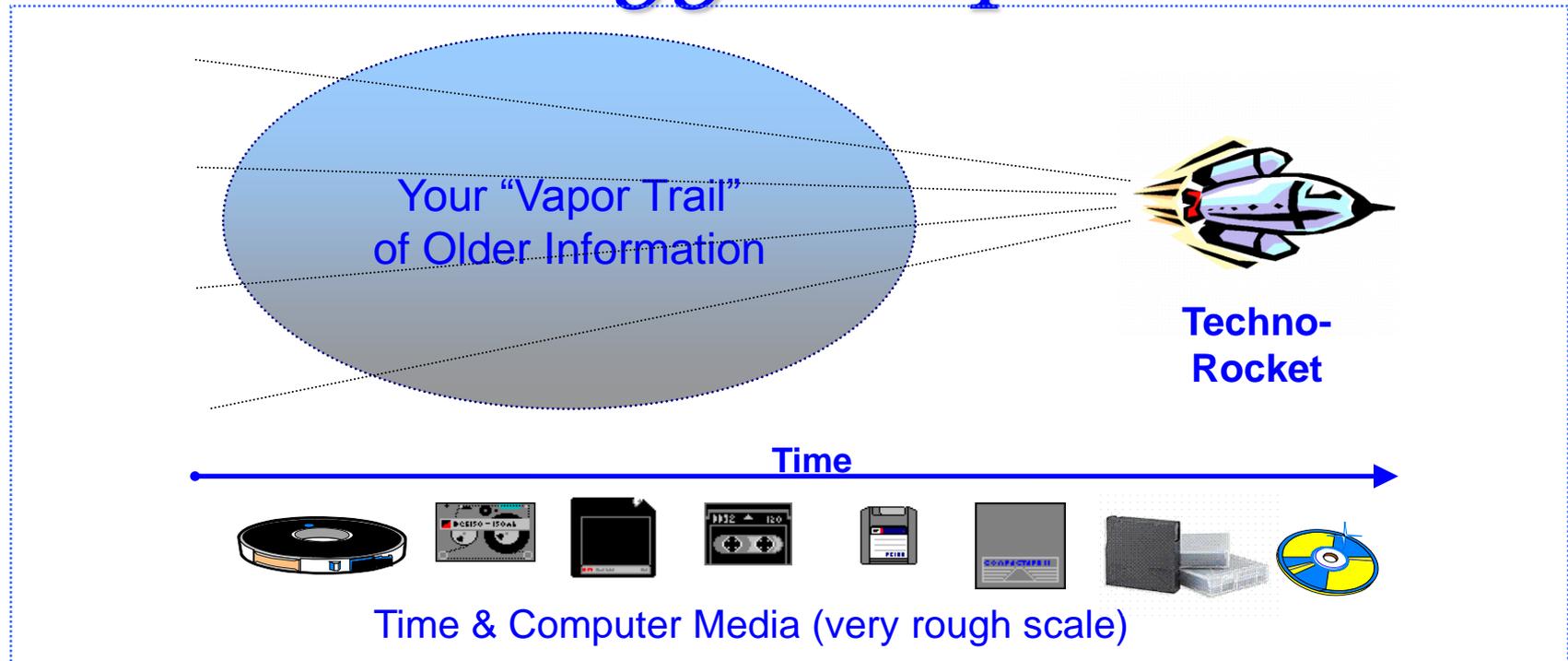


If only our magnetic media would last as long as those ancient hieroglyphs...

If only we could read the information on them without special equipment and software...

If only we weren't slowly losing the tools, skills and equipment...

# Technology's Vapor Trail



- The "vapor trail" includes lots material of of lasting value. (little in the way of tweets, mp3s and politics)
- Shelf life is dramatically lower than most pre-digital material.
- The ability to handle such media/file formats is becoming increasingly rare.

## One Advantage:

A good ratio exists between capacity of old media and inexpensive, durable new media. For example, the content of 9,000 "legacy" mainframe tape cartridges at US DoI fit on a single, inexpensive hard drive. The work to make a backup copy of the drive is tiny compared to the original effort to read those 9,000 tapes.

# *Obstacles to Preservation and Future Access.*

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

# *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. Even better, working with scientists, researchers, historians and archivists who really understand the value of to these digital assets. Some fun examples to 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 his party's presidential nomination.



Political rivals curious. FOIA request.  
State Department mainframe tape.  
Problem: un-documented file format.



(Turned out the new president was a strong supporter of digital preservation.) 😊

# *Some real Whitewater fun.*



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 records of it had mysteriously vanished ("vacuumed") from the firm's systems.

The firm does due-diligence; requests data recovery. The perp did not realize that good old Wang 8 inch floppies are not so easily erased.



Shame on you for thinking yet another presidential candidate could be involved.

# *Some Watergate fun.*



Early 1970's, 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) format not figured out for 25+ years.

Asked to puzzle it out and make a new database and program for researchers. Old DB similar to that used for Vietnam era military records ("NIPS".)



## **Major Discoveries!**

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

# *Beverly Hills High School Investigation*

**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.



# IPUMS

International

Another source of real enjoyment: the Minnesota Population Center and IPUMS.org. They collect, analyze and 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*



# *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 tapes OK, but many suffered from decomposition as shown. Plan: install read-convert system plus tape cleaners, training for BBS Staff. Initial goal: recover the 1981 Census (200 tapes, 95% recovered.)



# *Recovery/ Conversion Equipment*

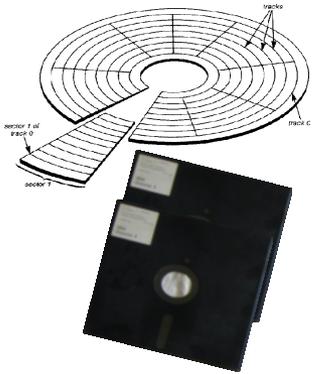


# *The Penobscot Dictionary*

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.



Many diskettes not compatible with standard operating systems; char encoding was morphed to use unique display hardware.



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



# *The Old Professors' Stashes*



Ready to retire or move on to another university—those old tapes are my legacy!

Or maybe they've left, and others realize that the health data, re-purposed and combined with population and climate information may well produce remarkable insights.

Re-discovered stashes like this are happening all the time! All kinds of stories, if you have a few more hours. 😊

# *Other fun jobs (mostly OP)*

- *Physical Science:*
  - 40-year-old crystallography research
  - Canadian Forestry (flippy floppies)
- *Social Science:*
  - SPSS "Version 1" (pre 1985)
  - Reagan Era Surveys (DEC RTST-11)
- *Cultural:*
  - Packard Humanities Institute
  - New York Philharmonic

The tools and skills for rescuing older digital assets apply across the boundaries of science and culture. A valuable 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 LoL)*