

2015

The Digital Vapor Trail: Why Early Digital Assets Merit Special Attention

Chris Muller

Muller Media Conversions, chris@mullermedia.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Muller, Chris (2015) "The Digital Vapor Trail: Why Early Digital Assets Merit Special Attention," *Against the Grain*: Vol. 27: Iss. 4, Article 10.

DOI: <https://doi.org/10.7771/2380-176X.7128>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

The Digital Vapor Trail: Why Early Digital Assets Merit Special Attention

by **Chris Muller** (Muller Media Conversions) <chris@mullermedia.com> www.mullermedia.com

Column Editor's Note: *The following are a series of vignettes from the author's experiences in data-rescue projects.* — **CM**

Let's picture a digital technology jet plane racing along, leaving a vapor trail of incompatible, deteriorating media, undocumented files or file-systems, myriad backup tape and optical disk formats — it's not a good thing.

Some organizations possessing "at-risk data" have created special tools and performed great "data-rescue" projects. Even better, they were scientists and researchers who then proceeded to analyze, publish, re-purpose, etc. But often, the tools for rescuing older digital assets are not readily available to the custodians. Our theme here is that, just as digital data can get lost in the vapor trail, some important *data-rescue capabilities* are also fading away. Later, the author will share some thoughts about creating a broad initiative to preserve and manage data-rescue capabilities.

A Great Pre-Digital Data Recovery Example

12,000 pages of New Amsterdam records sat largely un-noticed in an Albany vault for over 300 years and were then re-discovered in the late 1960s. They were lucky enough to find a great scholar in 17th-century Dutch language and writing style — which we've learned is totally unreadable to modern Dutch speakers. Ongoing translation efforts lead to **Russell Shorto's** wonderful book, *The Island at the Center of the World*. Luckily, those documents had the "Luxury of Languishing." However, with pre-digital information:

- That which remains was recorded on lasting material.
- We can often see what's needed to be read, translated, or copied.
- Optical scanning, OCR, and digital photography are mature and ever-improving tools

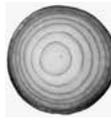
But regarding older digital material:

- If only older magnetic media would last as long as those New Amsterdam records.

- If only we could read arcane media/files without special equipment and software.
- If only some of the tools and skills to deal with those weren't also fading away.

Here's a view of the obstacles to recovery and conversion of legacy data — a bit like peeling back the layers of an onion:

Media Compatibility
Age and Storage Conditions
Recording Method



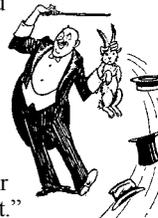
-----*
Operating System/Filing System
Backup, Exchange, or Archiving Software
Application File Structure
Application File Encoding

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

While there are challenges with old media and arcane files, there are also some advantages. Back in the day, computer media was expensive and hard to use — with little in the way of tweets, mp3s, politics, and porn. So the "value density" is often greater than that on much of today's media. Also, a good ratio exists between older media capacity and inexpensive new media.¹

Data-Rescue and Conversion Can Be Fun!

Learning to work with and continue to improve existing equipment and software tools and puzzling out early, sometimes unique data formats are challenges that can be very rewarding. Even better, you get to work with scientists, researchers, historians, and archivists who really understand the underlying value of the data. Here are a few examples that I hope you'll find interesting from our "Tales from the Digital Crypt."



Cryptie #1: Some pre-Whitewater Fun

Think "Clinton and Moscow." We're talking about a recent Secretary of State visiting Russia, right? Nope. In 1992,

Mr. Clinton was a contender for a presidential nomination. There were rumors that as a college student in 1969, he had gone there to protest the Vietnam War. Political rivals got curious and issued a FOIA request. An old State Department tape came to the U.S. Attorney's office in New York. Guess what? There was no metadata (that is, no file format info) — a sadly common occurrence. But the D.A.'s office found some techies who enjoyed hacking away, so that the plaintiffs could receive useful, readable information. This stands out in our memory because of a few other things that happened around the same time...



ASSOCIATED PRESS JANUARY 3, 1991 WASHINGTON—A slice of America's history has become as unreadable as Egyptian hieroglyphics before the discovery of the Rosetta stone. Vast untold volumes of historic, scientific and business data are in danger of dissolving into a meaningless jumble of letters, numbers and computer symbols. Much information from the last 30 years is stranded on computer tape from primitive or discarded systems, unintelligible or soon to be so...

Due to this article, along with radio commentaries by **Charles Osgood** and others, the issue of "data at risk" got a lot of attention. One of the people mentioned was the iconic **Dr. Ken Thibodeau** of the **National Archives and Records Administration**. At my wife's urging I contacted **Dr. Thibodeau**. Although we had many years of experience in media conversion, we had little exposure to government records. Unfortunately, we did not hear back from **NARA** for over a year.

In the meantime, there was revived Congressional support for digital preservation. It also turned out our new president (**Mr. Clinton**) was a strong supporter of the **NARA** and records preservation² and approved new funding support.

Helping NARA to Build that Rosetta Stone

If you enjoy history, you can imagine how exciting it would have been to be interviewed in **FDR's** "fireside chat" room. In fact, it was nowhere near a fireplace! **President Roos-**

continued on page 20



Chris Muller has had the pleasure of working with researchers and archivists for many years, performing the "rescue" and conversion of digital legacy data. Long, happy dealings with **NARA**, **IPUMS**, state archives, and several universities have made his work very enjoyable. His professional strengths include creating and enhancing recovery/conversion software, puzzling out arcane file formats, and dealing with now-unusual media such as 9-track tape reels, as well as writing proposals and managing project teams. He is also proud to be a volunteer member of the **CODATA Data-at-Risk Task Group**, the **InterPARES Trust PaaST Team**, and the **RDA Data Rescue** interest group. 🐰

The Digital Vapor Trail ... from page 18



evelt had used a small studio (now a conference room) in the depths of the NARA building on Pennsylvania Avenue. Sub-

sequently, NARA issued a contract for the development and support of something called “APS,” the Archival Preservation System, beginning a happy 14-year relationship between a small company in New York and great digital preservationists in DC.

Cryptie #2: Some Watergate Fun

In the early 1970s, a mainframe computer was used to store **President Nixon’s** appointment calendar and notes. Twenty-five years later, the only copy of the information was contained on an old tape reel. Historical researchers wanted access to the information, but they had two problems: (1) the tape was in danger of decay, and (2) the data format had never been identified. They asked their techies to convert the data and create a new program for researchers. Luckily, the analyst had done work with Vietnam military



records and noticed similarities in the data structure (from a program known as NIPS). Of course, we all hoped there would be Watergate secrets revealed, which did not prove true. But it was interesting to baseball fans like us to see that the first visitor to the

Nixon Whitehouse was **Hank Aaron**.

Cryptie #3: More Whitewater Fun

In the 1980s an attorney for the **Rose Law Firm** in Little Rock, Arkansas did legal work for a real estate project named “Castle Grande.” By early 1994, all related paper and computer records had mysteriously vanished (or were “vacuumed”³). The firm requested data recovery. Obviously, the perpetrator was not a big fan of digital preservation — but he did not realize that those disks are not so easily erased.



IPUMS and Crypties from Around the World

IPUMS-International⁴ collects, analyzes, privatizes, and publishes population data from around the world. Tapes and disks arrive from many countries, such as Bangladesh, Egypt, Kenya, Mali, Mexico, Nepal, Pakistan, Peru, Romania, Santo Domingo, Sudan, and Turkey.

One of our most interesting projects involved six weeks in Dhaka, at the **Bangladesh Bureau of Statistics**. Very capable people, they had been confronted with daily power outages and other problems, making it impossible to store legacy tapes in an optimal way. Many of their tapes suffered from decomposition. A system was installed to read and convert the files, plus tape cleaners and training for **BBS Staff**.



One lesson learned through our work with **IPUMS**: *there is a world-wide need for data rescue*. But many nations do not have the financial resources to put such projects at high priority. That’s why groups like **IPUMS** with funding from **NSF** and others often reach out to help. In other cases, collections of “at-risk data” have yet to be officially noticed.

Old Professors’ Stashes (OPS)

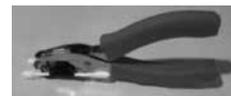
“I’m ready to retire, and those old data tapes are my legacy!” The professor or his colleagues take a new look at his data and



realize that the old health data, combined with population and climate information, may well produce remarkable new insights. Rediscovered stashes like this are cropping up all the time! Of course, they’re not always from an old professor, but it’s a fun way to characterize valuable scientific or historical data that, at least for a while, have been overlooked. Here’s a quick look at three data recovery cases:

Physical Science Case — This data was from a forestry genetics project that began some 35 years ago. The intent was to study various timber species and see how they would do if planted at different altitudes, longitudes, and moisture levels. Later, a key player realized that the details gathered by the original study could be re-purposed to gain insight into “tree migration,” a process that normally takes centuries. This could enable researchers to know which species will do best under certain aspects of climate change, and even to assist the migration process. Floppies were from older **Apple drives**⁵ in a variety of formats, but the real challenge was that the diskettes were “flippy floppies,” with each side written as a separate volume. Currently available floppy drives require index holes. And when a diskette is flipped over, the hole is no longer in the right spot. To do the flip side we had to punch extra holes in the diskette casing without scratching the floppy. The “*highly sophis-*

ticated” tool we designed is pictured to the right.



Social Science Case — Like the first-side chat room experience, picking up well-packed census tapes at a Consulate certainly adds some flair to a project. The micro-data had been backed up to tape in what on the surface seemed to be ANSI standard format. But somehow, a type of compression had been applied to the individual records. Similar to the **Nixon** tape experience, we were lucky enough⁶ to have seen that compression before.



Cultural History Case — **Dr. Frank Siebert** dedicated much of his life to safeguarding the Penobscot Native American language. After many interviews with native speakers he created a full dictionary using special software, stored on diskettes that were unfortunately not at all compatible with standard file systems. Character encoding was also morphed for unique display hardware. We had the honor to work with the **American Philosophical Society** and **Maine Folk Life Center** on the project to ensure the preservation of this legacy treasure.

Thoughts on Preserving Data Rescue Capabilities

There are some great data rescue organizations addressing specific needs. One example: the **International Environmental Data Rescue Organization (IEDRO)**, focused on recovering older environmental observation data. They have leadership with the know-how to evaluate the scientific value and also perform the physical recovery and digitization.

Other groups have focused on certain things such as cataloguing the myriad PC file types and in some cases creating open-source conversion software. Another example, the **CODATA Data-at-Risk Task Group**, has several projects, one being creation of an inventory of important scientific data at risk of loss. And of course, there are computer museums to preserve old systems — some working, some not.

But, to our knowledge, there is not yet an organization to support the *Preservation of Data Rescue Capabilities* for older digital assets (that vapor trail). Such a group would be of support to scientists, historians, librarians, and archivists. It would collect details of capabilities and projects across government, academia, and commercial entities and stay in touch, ensuring that the tools don’t just fade away when projects end or certain staffers retire. In some cases, it would collect and manage equipment and software that would otherwise be lost. Do you agree that such an

continued on page 22

Audio-Visual Collection ... from page 21

The Video Lab supports many of the industry's obsolete formats as well as many of today's High-Definition video tape/disk-based formats, including 3/4" U-matic, 1" Type C open reel, 1/2" EIA open reel, Betamax, BetaCam, Hi8mm, VHS, S-VHS, D2, DVCPRO, Digital BetaCam, BetaCam SX, HDCAM-SR, D5, DVCPRO-HD, XD-CAM, and others. Similarly to audio media, Video Preservation Specialists spend much time prepping, cleaning, and repairing video media for safe and transparent playback for digital capture.

Overall, the Audio-Video Preservation Laboratory at the **National Archives** is a well-equipped and expertly staffed facility, trained in the preservation and reformatting of obsolete and difficult formats while keeping in step with the latest technologies, formats, processes, and technical and archival standards.

Motion Picture Preservation Laboratory

Whenever possible, **NARA** strives to retain three copies of motion picture records — a Preservation copy that is contained in deep storage, an Intermediate copy that is readily accessible to fulfill vendor requests, and Reference copies to be served in the research room.

To accomplish this mission the collection is preserved and made accessible by staff of the Motion Picture Preservation Laboratory. The Film Preservation Lab is located at **Archives II** and is one of the last fully operational photochemical labs in the country. A small staff of five is responsible for several major activities — preservation, technical processing, vendor inspection, creating access copies, and digital restorations. Staff expertise ranges from traditional film editing and production, photochemical lab operations, digital editing and post production, library science, archival science, and film preservation specific training. Among all staff members we have over 100 years of experience.

For films requiring preservation the film is assessed by hand, and specialists repair damage, measure the film for shrinkage, determine the level of vinegar syndrome for acetate materials, and assess the overall preservation risk for each reel. The film is then preserved photochemically by creating a new copy on

Jason Love is the Supervisory Audio-Video Preservation Specialist in the Audio-Video Preservation Laboratory at the **National Archives** in College Park, MD. His experience in the Audio and Video field spans over 20 years, which includes work in commercial recording facilities, live concert sound, audio-video system design, repair and installation, sales of professional recording and sound reinforcement systems, facility design and implementation, audio-video preservation and digitization, and participation on technical standard committees, and he is a member of the **Audio Engineering Society**. 🌿

polyester-based film stock using one of our five film printers, developing the film, and performing quality control on the new copy.

The Film Preservation Lab staff assists the archival unit by performing technical processing inspection. Once the archivists have determined which titles to accession from Federal Agencies, the film is sent to the Lab for the Preservation Specialists to determine which copies of the film should remain at **NARA** if there are multiple copies or versions of each title. The most original copy and/or the copy in the best physical shape will be retained as the Preservation copy; the second best copy will be retained as an Intermediate Copy, and if there is a serviceable film print that will be retained as the Reference Copy. In addition to making these determinations, staff are also able to address preservation concerns as soon as the film is brought into the collection.

Not all titles have multiple copies, but **NARA** makes every effort to provide the public with the records they need access to. In order to make content available, **NARA** provides Preservation copies to vendors for other institutions or production companies to gain access to the records. Staff in the Lab inspect all of the reels for vendor requests before they are sent to an approved facility and after the originals are sent back. This ensures that reels in need of preservation are not sent to a vendor and that no damage has occurred while the film has been out of **NARA's** custody.

To increase access to the collection the Film Preservation Lab staff creates access copies that are served in the Research Room and/or uploaded to the *National Archives Catalog*, You Tube, and Amara. All 35mm and 16mm films are captured in High Definition (HD) at 1920 x 1080 resolution. For long-term storage we retain uncompressed HD AVI files in our tape library, and for ease of access we create HD MP4 (H.264) files that are stored on local network attached storage.

As **NARA** lacks a dedicated infrastructure to deliver files to the research room, Lab staff create DVDs that researchers can view and make copies of in the research room. Certain titles that are of high public interest or are part of our WWI & WWII scanning project are also uploaded to **NARA's You Tube** channel and into **Amara**, where individuals can tag and transcribe content to make titles more accessible or available in multiple languages.

The last major activity done in the Film Preservation Lab is restoration projects. While photochemical means are used to preserve the holdings, our digital restorations are meant to provide enhanced access to the public to provide them with a glimpse of how a film would have looked at the time it was originally shown. 35mm film is scanned in at 4K, and most 16mm film is scanned in at 2K. We capture DPX files, which is a file format designed to capture the intrinsic qualities inherent in film. Once the film is captured we utilize our restoration software to correct for shifts in color and exposure and to remove dirt, dust, scratches, flicker, and other defects introduced through time and use. We capture the audio using specially developed equipment to handle deteriorated magnetic tracks or optical tracks and use a different software program to reduce hiss, noise, clicks, or pops. The separate image file and audio file are then synched and brought into yet another software program to create digital cinema copies that can be shown in theaters. To date **NARA** has restored *Let There Be Light*, *The Negro Soldier*, *The March*, *The True Glory*, and *Nine from Little Rock*.

To learn more about the Motion Picture and Audio and Video Collections at **NARA** and the types of work done here we encourage interested parties to visit our blog, preservation guidelines, or view *Out of the Dark: Bringing Films to Light at the National Archives*. 🌿

The Digital Vapor Trail ... from page 20

effort should be created? If so, should it be a new, distinct non-profit organization, or perhaps a broad initiative among entities already concerned with related matters?

If you'd like to learn more about this or send comments, we'd like to hear from you! Contact us at <chris@mullermedia.com>. 🌿

Endnotes

1. One example: the content of 9,000 government mainframe cartridges fit on one inexpensive USB hard drive, so future backup/migration efforts are tiny compared to reading all those old tapes.
2. Just don't ask about **Sandy Berger's** visits to **NARA**.
3. A handwritten note from a Whitehouse staff meeting several weeks before the records disappeared contained the phrase "vacuum Rose records" (apparently a total coincidence).
4. Part of the **Minnesota Population Center**.
5. Those particular drives didn't rely on track index holes.
6. Can't resist this quote. 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."