

Saving the “Fat Man”: The George Sanger Eudora Email Archive



**LAUREN ALGEE, CANDIDATE -
M.S. INFORMATION SCIENCE
C.A.S. PRESERVATION ADMINISTRATION**

**THE UNIVERSITY OF TEXAS AT AUSTIN
SCHOOL OF INFORMATION**

APRIL 30, 2010

Project Background



The Fat Man

Sounds Incredible



THE UT
VIDEOGAME ARCHIVE



THE DOLPH BRISCOE CENTER FOR AMERICAN HISTORY
THE UNIVERSITY OF TEXAS AT AUSTIN

Project Team

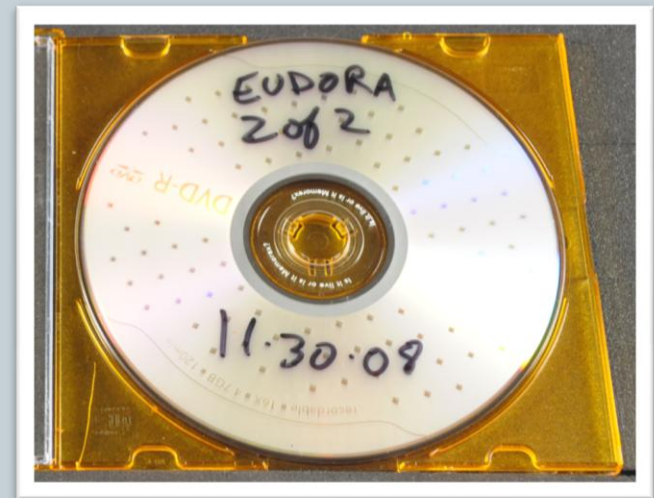
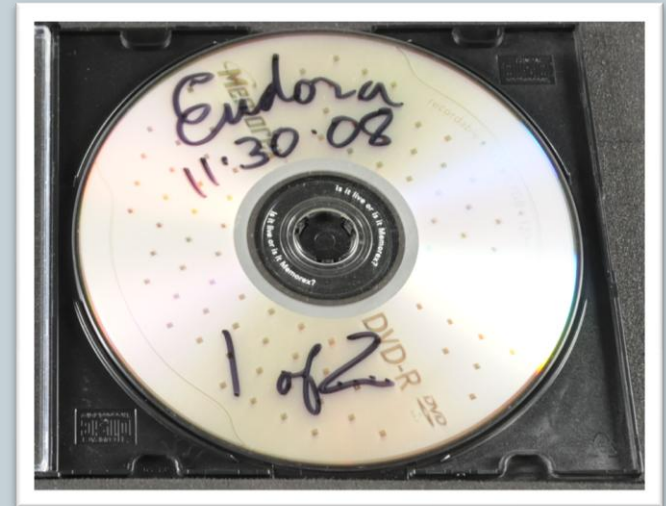


- Lauren Algee, Stephanie Bordy, Elizabeth Hilkins, Chris Latham
- DBCAH Digital Archivist, Zach Vowell
- Dr. Patricia Galloway
- University of Texas Digital Repository

What we were handed...



- 2 DVDs of files archived from Eudora email client on November 30, 2008 and dating back to 1999
- Client used for personal and professional correspondence.



About Email



- Electronic mail (email) provides for the exchange of messages over the Internet to other users in possession of an email address.
 - ✦ From
 - ✦ To
 - ✦ Subject
 - ✦ Cc
 - ✦ Bcc
 - ✦ Attached

Preserving Email as an Archive



“1) the sheer volume precludes using scarce archival resources to preserve each message and document its contextual relationships; and
2) the value of preserving email messages ‘in situ’ resolved issues of original order and overall metadata and documentation.”

- Collaborative Electronic Records Project
(CERP)

The Problem(s) with Email



- Unknown and proprietary file formats
- Deteriorating storage media of email archives
- Obsolete client software
- Duplicate files
- Sensitivity of personal mail
- Junk or spam messages
- Lack of file order

(Adgent, 2008)

- Attachments
 - Rate of obsolescence may differ from that of the messages
 - Migration for a myriad of formats?

About Eudora



- Developed in 1988 by Steve Dorner at University of Illinois Urbana-Champaign.
- Named after Eudora Welty story “Why I live at the P.O.”
- In 1991, Qualcomm acquired Eudora for internal use, developed Eudora for Windows and began offering “Light” and “Pro” versions
- Now in version 8.0.1 beta
- High levels of customizability
- Sanger archive includes all Eudora client operating files, in addition to actual messages
 - .toc and .mbx

Sanger's Use



- Did not use additional Eudora features, such as the calendar or add-ons or special settings.
- Used a free version of Eudora until switching to the "sponsored" version of Eudora upon its release in 2003.
- Used a Windows XP operating system during the time span of the emails.
- At one time, Sanger migrated from Eudora to Firefox back to Eudora then to MS Outlook. He copied folders from one directory to another and renamed files; the process was both "sloppy" and "brutal" according to Sanger. The time frame of these migrations is unclear.

Listservs



- Videogames = private creativity
- “Very Sensitive”
 - ✦ **BOSSLEVEL** is an exclusive list for game industry executives. The list was explicitly a "do not share" listserv.
 - ✦ **GANG** is the Game Audio Network Guild, a nonprofit organization supported by paid memberships. According to Sanger, GANG is particularly litigious and has previously threatened him with legal action.
 - ✦ **VGM** is a listserv for video game musicians.
 - ✦ **BBQ** and **Horseshoe** are both annual meetings hosted by Sanger. Both operate under an "oath of secrecy or blabbing" requirement. If something is published, all persons and companies need to approve it first; alternatively, the lists should be given credit if something results from them.

Challenges



- **Preservation**
 - Migration to additional formats
 - ✦ Originals
 - ✦ MBOX
 - ✦ XML
 - Authenticity
- **Legal**
 - Intellectual Property
 - Privacy
- **Access**
 - Research – where and how?
 - Description

Initial Goals



- Allow access to originals in original context (Eudora) as long as possible AND encode messages in XML for long-term preservation
1. Image the DVDs
 2. Access the files through Eudora
 3. Establish protocols concerning the legal issues related to privacy and intellectual property in the materials
 4. Inventory the files and scan for viruses
 5. Encode the email files in an email-specific XML schema (converting the files to intermediary format if necessary)
 6. Prepare the original bitstreams, intermediaries, XML, metadata, and documentation for ingest.

Intellectual Property



- Since the creator of a work holds copyright and email is a two-way communication, both Sanger and his correspondents hold copyright on messages and attachments in the collection (including Listservs).
- No copying or printing out of any files.
- Restricted access - only at the Briscoe Center on secure computer by appointment supervised by the archivist

Imaging and Access



- Create access copies of original files to avoid corruption
- Providing access to all of the material in the most recent version of Eudora possible RATHER THAN determining version of Eudora in use at the moment of creation for each 15,844 files.
- Created image files (ISOs) with InfraRecorder (<http://infrarecorder.org/>)
- Eudora.exe indicated files were from a Windows system and File properties told us Eudora version 7 was used.
- To access the files, we ran Eudora (7.1) Windows XP.

Inventory & Viruses



| File Type | ALL | | ATTACHMENTS | | NOT ATTACHMENTS | |
|-----------|-------------|-------|-------------|-------|-----------------|------|
| | Files Count | % | Files Count | % | Files Count | % |
| .jpg | 4696 | 28.98 | 1870 | 22.62 | 2826 | 0.36 |
| .gif | 2690 | 16.6 | 517 | 6.25 | 2173 | 0.27 |
| .zip | 1333 | 8.23 | 1322 | 15.99 | 11 | 0.00 |
| .doc | 1183 | 7.3 | 1182 | 14.3 | 1 | 0.00 |
| .pdf | 840 | 5.18 | 838 | 10.14 | 2 | 0.00 |
| .toc | 836 | 5.16 | 0 | 0 | 836 | 0.11 |
| no suffix | 798 | 4.93 | 334 | 4.04 | 464 | 0.06 |
| .mbx | 794 | 4.9 | 0 | 0 | 794 | 0.10 |

- 15,844 files in 91 directories
- 118 files infected – KEEP

XML and Email Preservation



- **.mbx will not last forever, and Eudora may end of retroactive support for earlier files even sooner**
 - create a human-readable, and sustainable open format for viewing the material
 - tools for harvesting data and reading already exist
 - searchability
 - conversion to PDF or HTML
- **Intermediary file needed – convert to MBOX**
 - Emailchemy version 9.9.2

XML Trial and Error(s)



- Collaborative Electronic Records Project (CERP)
 - Smithsonian Institution Archives (SIA) and the Rockefeller Archive Center (RAC)
- Preservation of Electronic Mail Collaboration Initiative (EMCAP)
 - North Carolina State Archives, the Kentucky Department of Library and Archives, and the Pennsylvania State Archives
- XML Electronic Normalising for Archives (XENA)
 - National Archives of Australia
- Mbox2XML

Outcome



- To be ingested into UT Digital Repository:
 - Original bitstream TAR
 - MBOX TAR
 - Documentation
- With Qualified DC metadata
 - dc.title
 - dc.description
 - dc.date.created
 - dc.creator
 - dc.description.department
 - dc.language
 - dc.publisher
 - dc.rights
 - dc.subject

Recommendations to the Briscoe Center



- **7.1 Eudora on secure computer for access**
 - At the Center
 - By appointment
 - No copies
 - Original TAR to be downloaded for every use
- **XML**
 - Check for future advancements
- **Assess attachments**