GEORGE SANGER ELECTRONIC MAIL COLLECTION
FINAL REPORT

INTRODUCTION

This report details the work of four University of Texas at Austin School of Information students from Dr. Patricia Galloway's Spring 2010 course, INF 392K Problems in the Permanent Retention of Electronic Records, on the preservation of the George Sanger Electronic Mail Collection. This collection is a part of the George Sanger Papers in the Videogame Archives at the Dolph Briscoe Center for American History. The emails were received from the Briscoe Center in the form of two DVDs of files Sanger archived from the Eudora email client on November 30, 2008 and include files dating back to 1999. The work of the project included:

1. Imaging the DVDs.
2. Accessing the files through Eudora.
3. Inventorying and scanning the files for viruses.
4. Establishing protocols concerning the legal issues related to privacy and intellectual property in the materials.
5. Preparing the original bitstream, MBOX conversion, associated metadata, and documentation for ingest into the University of Texas Digital Repository.

The Dolph Briscoe Center for American History

The Dolph Briscoe Center for American History is an organized research unit and public service component of The University of Texas at Austin. As a leading history research center, it collects, preserves, and makes available documentary and material culture evidence encompassing key themes in Texas and U.S. history.

The UT Videogame Archive is a collection component of the Briscoe Center that seeks to preserve and protect the records of videogame developers, publishers, and artists for use by a wide array of researchers. The Briscoe Center strives to collect materials that not only facilitate research in videogame history, but also to provide materials of interest to those studying communications, computer science, economics, and other academic disciplines drawn to the processes driving the videogame industry. The archive currently includes fourteen collections, including papers, games, and files of the archive's first three donors, Richard Garriott, Warren Spector, and George "Fat Man" Sanger.

Creator Sketch

Since scoring his first video game with *Thin Ice* for the Intellevision in 1983, George "The Fat Man" Sanger (b. 1957) and his "Team Fat" (comprised of fellow game composers Dave Govett, Joe McDermott, and Kevin Weston Phelan) have contributed music to over 250 games, including popular and influential titles such as *Wing Commander* (1990) and its sequel *Wing Commander II: Vengeance of the Kilrathi* (1991), *The 7th Guest* (1992) and its sequel *The 11th Hour* (1995), *Master of Orion* (1993), *Loom* (1990), *Maniac Mansion* (1990) and much of the *Scene-It* game series. The Fat Man and his team also achieved many milestones in game audio, including writing the first general MIDI soundtrack for a game, tracking the first direct-to-MIDI live recording of musicians, and creating the first soundtrack that was considered a selling point for the game.

Actively involved with the game development community, Sanger joined the International Game Developer's Association (IGDA) as a Board Member in 1994 and has also served on the board of advisors for Game Developer Magazine, the Austin Game Developers Conference, and Full Sail's Game Development degree program. Sanger has also played a huge role in re-casting video game audio as a legitimate form of art, successfully campaigning with Chance Thomas for the creation of a Grammy award category for Video Game audio as well as penning *The Fat Man on Audio: Tasty Morsels of Sonic Goodness* (2006), a semi-autobiographical work he describes as "a book about game audio wrapped in a biography wrapped in a philosophy of life."

In 1995, Sanger created Project Bar-B-Q, an interactive music summit held in and around Austin, Texas that has grown into one of the most prestigious and influential conferences in the game audio industry. Inspired by this success, Sanger launched the similar yet more ambitious Project Horseshoe in 2006, an intense "think tank" aimed at solving some of game design's most difficult and pressing problems.

Sanger was born and raised in Coronado, California and has spent the majority of his professional career in Austin, Texas. Recent projects by Sanger and Team Fat include audio for *Spongebob Squarepants: Revenge of the Flying Dutchman* (2002), *The Incredible Hulk* (2008) and slot machine sound design for Multimedia Games.

Email

Electronic mail (email) provides for the exchange of messages over the Internet to other users in possession of an email address. Email refers to both the system of message transportation and to

the messages themselves. Advantages to this form of communication include the fact that it is accessible to both the sender and recipient anywhere with an internet connection, the message is sent much more quickly and is less expensive than with the postal mail, and a message may easily be sent to multiple recipients. Email messages typically contain the following information in addition to the message itself:

- **From**: e-mail address of the sender
- **To**: e-mail address of the recipient
- **Subject**: focus of the electronic message, typically supplied by the sender
- **Cc**: 'Carbon Copy,' informs recipients of the other email addresses to whom the message was sent. All recipients can view these addresses.
- **Bcc**: 'Blind Carbon Copy.' Similar to a carbon copy in that the sender may specify the message be sent to multiple email addresses. However, recipients cannot view the other email addresses.
- **Attached**: Files such as text, sound, image, software, and other complex documents are usually sent as attachments in addition to the message itself.

Archiving email poses particular problems. The content, structure, and context of messages, attachments, and related messages must be maintained in an archive. For this reason, the Collaborative Electronic Records Project (CERP) recommends archiving an entire email account, rather than individual messages. CERP states, "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."

However, saving everything brings with it a myriad of problems. These include unknown and proprietary file formats, deteriorating storage media of email archives, obsolete client software, duplicate files, the sensitivity of personal mail, junk or spam messages, and lack of file order (among others). (Adgent, 2008).

Attachments pose still other preservation challenges. Because of the variety of attachment file formats, the rate at which attachments become obsolete may differ from that of the email messages themselves. Archives must decide whether to undertake assessment of individual attachment formats, a likely time-consuming process, and migrate these materials to more stable formats to preserve access to the material.

Eudora

In 1988, during his tenure at the University of Illinois Urbana-Champaign's Computing Services Office (CSO), software developer Steve Dorner designed the first, freeware version of the Eudora email client. Dorner named his application after author Eudora Welty, referencing her short story "Why I live at the P.O." (1941), because he imagined that email would be like bringing the Post Office to the user. Indeed, the original slogan for the Eudora mail client was "Bringing the P.O. to Where You Live" (Dorner 1994).

Dorner was the first to combine UNIX-based (command line), Internet email with more user-friendly LAN-based systems, and a graphical interface that used little memory; these things together significantly broadened Eudora's appeal and helped pave the way for an entirely new

cohort of email users. Dorner developed the program for Macintosh internet protocol suite (MacTCP).

In 1991, Qualcomm acquired Eudora for internal use, but the company quickly realized Eudora's potential and developed Eudora for Microsoft's Windows operating system. Qualcomm released the first serial version of Eudora freely over the Internet quickly thereafter; this first serial release came in one of two options: Eudora "Light," a freeware option with limited features, or Eudora "Pro," a fully featured commercial application. The company added a "Sponsored" version in 2003, which featured the same functionality as Eudora Pro, but was free and included advertisements. Steve Dorner joined Qualcomm's QUEST business unit in 1993 and continued to work on Eudora's development.

In 2006, after almost two decades of success, Qualcomm announced that it would no longer support Eudora Pro. Subsequent versions of the application were open source, and based on, but not in competition with, Mozilla Thunderbird. This new project, led once again by Eudora founder Steve Dorner, was code-named *Penelope*. A beta version was released on July 19, 2007 as Eudora 8.0.1b1.

Eudora Timeline

- 1988- Steve Dorner designs Eudora while working at University of Illinois Urbana-Champaign
- 1990- Eudora released over MacTCP
- 1991- Qualcomm purchases rights to Eudora for internal use

- Serial version of Eudora (1.0) released freely over the Internet (primarily used by universities)
- 1992- Dorner begins working with Qualcomm's QUEST business unit, releases Eudora 1.3 offering free and "enhanced" versions
- 1993- Eudora 1.4 released for free on the Internet and a paid 2.0 for Mac shipped for fee
- 1997- Qualcomm releases Eudora Pro CommCenter to compete in the corporate market
- 2006- (Oct 11) Qualcomm announces that future versions of Eudora will be open source and based on Mozilla Thunderbird
    - Project codename: *Penelope* (to be led by Steve Dorner)
- 2007- (May 1) Paid mode of Eudora is no longer available, but Light and Sponsored modes remain
- (Jul 19) Thunderbird based Eudora 8.0.1b1 (open source beta) is released

Features

- supports POP3, IMAP, and SMTP
- high levels of customizability, much of which is done through Uniform Resource Identifiers (URIs) known as "x-eudora-settings"
- Since the release of Eudora 6.0 in 2003, Eudora has offered an automatic junk and spam filtering feature named SpamWatch
- users can manage multiple email accounts by setting up various "personas"

Files

Eudora creates a number of files and directories upon installation. Additional functions, such as mailboxes, signatures, stationery, and address books also require additional files and directories. Eudora files and directories are as follows:

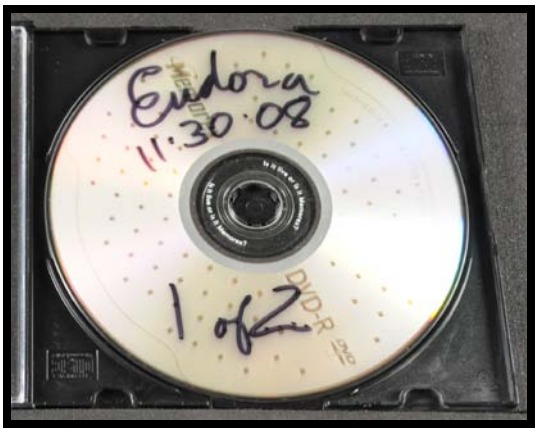
- EudPriv/Ads/AdCache Directory - All ad files downloaded to Eudora are kept in this directory.
- Attach Directory - Incoming attachments are saved in the Attach directory until the user specifies another directory using the Attachment directory button in the Attachment Options.
- DirectoryServices Directory - Eudora uses the DirectoryServices directory to store the dll files for the Directory Services protocols that you use in the Directory Services window.
- Embedded Directory - Eudora uses the Embedded directory to store JPEG image files that you insert into the body of outgoing messages using the Insert Picture... command under the Edit menu. Eudora deletes these files from this directory when the messages containing the images are emptied from the Trash mailbox.
- Filters Directory - Filters are saved in the Filters directory.
- Imap Directory - Eudora uses the Imap directory to store your IMAP mailboxes and messages.
- Nickname Directory (Address Books) - Address Book entries are saved in the Nickname directory, in the default Eudora Nicknames file. Additional Address Book files created by the user are kept under their own name in the Nickname directory.
- Plugins Directory - The EMSAPI plug-ins are kept in the Plugins directory.
- Sigs Directory - The Standard and additional signature files are kept in the Sigs directory. These files are stored with the .txt extension.

- Stationery Directory - The user's stationery files are kept in the Stationery directory. Stationery files are stored with the .sta extension.
- descmap.pce - Mappings between mailbox names and file names are stored in the descmap.pce file.
- Eudora.cnt, Eudora.hlp - The Eudora.cnt and Eudora.hlp files contain, respectively, the table of contents information and the help text for Eudora's online help topics, accessed when the user selects Topics from the Help menu. These two files must be kept in the same directory.
- Eudora.exe - Eudora.exe is the Eudora application executable file.
- Eudora.ini - Your Options information is saved in the Eudora.ini file, along with other information.
- eudora.log, eudorlog.old - Eudora can keep records of all mail transfers. These records are kept in the eudora.log and eudorlog.old files. The eudorlog.old file is overwritten and a new eudora.log file is created when the eudora.log file reaches its approximately 100K maximum size.
- filters.pce - Names and extensions for Eudora filters are saved in the filters.pce file.
- finger.ini, LDAPInit.ini, ph.ini - The finger.ini, LDAPInit.ini, and ph.ini files are used to store settings information for the Finger, LDAP, and Ph protocols used in the Directory Services window.
- in.mbx, out.mbx, trash.mbx - These files hold mail and exist for every mailbox created by the user. These files are in UNIX mail format. Mail folders created by the user are stored as directories with the .fol extension. Mail folders contain mailboxes and other mail folders.

- in.toc, out.toc, trash.toc - These files are the tables of contents of mailboxes and make it much faster for Eudora to access mail. These files exist for every mailbox created by the user.
- lmos.dat - This file contains information about the messages on your mail server. (lmos = leave mail on server.)
- nndbase.toc - This file is the table of contents for nicknames. Extra nickname files are stored in the Nickname directory (see above).
- nndbase.txt - Nicknames are saved in the nndbase.txt file. Note that this file contains the nicknames only, while the files in the Nickname directory (see above) contain the full data for each Address Book entry—which includes the nickname and more.
- Readme.txt - This file contains the Eudora Readme, a text file that contains important, release-current information, and instructions that might not be included in the Eudora User Manual, the Eudora Quick Start Guide, or the Eudora Online Help.
- *.tlx, *.clx - Dictionary information is stored in the .tlx and .clx files. (Eudora 7.1 for Windows User Guide, 2006)

THE COLLECTION

The George Sanger Electronic Mail Collection consists of two DVDs of files Sanger archived from his Eudora email client on November 30, 2008 and includes files dating back to 1999. Sanger used this account for both personal and professional correspondence. The content of the DVDs remained largely unknown until disk imaging and inventory.



Sanger's Use

In email correspondence with the project team, Sanger shared that he did not use additional Eudora features, such as the calendar. Nor were add-ons or x-eudora-settings, discussed under the "features" heading found above, employed by Sanger when using Eudora. Of the three versions of Eudora, "light," "sponsored," and "pro" as explained in the Eudora history section, Sanger's use of Eudora was the "sponsored" version upon its release in 2003. While he did not specifically say so, it may be presumed that he used a free version of Eudora prior to the release of the "sponsored" version as no mention was made of using a paid version of the email client. (G. Sanger, personal communication, March 13, 2010.)

During the time the emails to which this project pertains were created and received, Sanger used a Windows XP operating system. He did use Eudora on a Mac up to version 7.1. However, to his knowledge, we do not have emails from that time.

At one time, Sanger states that he migrated from Eudora to Firefox (likely Sanger means Thunderbird, Mozilla's email client) 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

Sanger made specific mention of multiple listservs to which he belonged and are found in the email collection: BOSSLEVEL, GANG, VGM, "BBQ," and "Horseshoe." (G. Sanger, personal communication, March 13, 2010.) Each of these will be discussed in turn. Privacy, intellectual property, copyright concerns will be briefly mentioned as they relate to each listserv. A more detailed discussion of these concerns may be found under "Processing the Collection."

- **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 conferences 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. (G. Sanger, personal communication, March 22, 2010.)

PROCESSING

Based on Sanger's description of the collection and initial research into practices for email preservation, our preservation plan was to allow research access to the original files in their original context by ingesting exact copies of the disks into the UT Digital Repository which

could be accessed through Eudora at the Briscoe Center until obsolete. We would also encode the MBX message files in XML, allowing long-term access to the information contained therein. We were aware that conversion of the MBX files to an intermediary format might be necessary before XML encoding would be possible. Our plan for attachments and the other Eudora files was undetermined as our work began.

Our original processing plan for accomplishing these goals was to: 1) Image the DVDs, 2) Access the files through Eudora, 3) Inventory the files and scan for viruses, 4) Establish protocols concerning the legal issues related to privacy and intellectual property in the materials, 5) Encode the emails in an email-specific XML schema (converting the files to an intermediary MBOX format if necessary) and 6) Prepare the original bitstream, XML-encoded emails, associated metadata, and documentation for ingest into the University of Texas Digital Repository. However, those goals evolved over the course of the semester, particularly with regard to the XML-encoding of the emails.

Getting Started

Our first steps in working with the DVDs involved viewing the files on the DVDs. This allowed us to get a feel for the files contained in the collection as we were previously unfamiliar with Eudora. From there, the group researched email file formats, read extensive Eudora documentation about the email client, and sought out previous versions of Eudora. On the advice of Dr. Gallway we made the decision to use the most recent version of Eudora (7.0 for the Windows operating system) to access the files. We also created test email accounts, using

both our own files and a sample of Sanger's emails in order to determine the system's file structure as well as the user experience within Eudora's native environment.

Intellectual Property, Copyright, and Privacy

Given the personal and professional nature of the Sanger Electronic Mail Collection, the group wanted a general understanding of the legal ramifications of creating access to the collection before moving to establish whether the restriction of access to the emails or even the removal of particular files would be necessary.

Intellectual property is defined as the products of human intellect that the law protects from unauthorized use by others. Generally, there are four rights involved: copyright, patent, trademarks, and trade secrets. The rationale behind intellectual property rights is that the creator is induced to create new works by regulating the public's use of such works so that the creator is compensated for his or her efforts.

Copyright is the statutory protection of a creator's work once it has been fixed in a tangible medium. It gives the creator the right to distribute, reproduce, and regulate use of the material for a limited period of time. Works can be sound recordings, software design, graphic arts, and moving pictures in addition to written works.

Exceptions to copyright include fair use and library or archive reproduction in limited amounts. Fair use of a copyrighted work for purposes such as criticism, comment, news reporting, teaching, scholarship, or research is not copyright infringement. Libraries or archives may

reproduce one copy of the work for distribution so long as the work is not copied for a commercial advantage, the work is available to the public or researchers unaffiliated with the institution, and a copyright notice is provided on the reproduction.

Since the creator of a work holds copyright and email is a two-way communication, clearly both Sanger and his correspondents have copyright in the works in the collection. While Sanger holds copyright of his sent messages, his correspondents hold copyright to messages received by Sanger. Further, attachments including photos and sound recordings are also protected under copyright.

Additionally, the video game industry is a for-profit field. Sanger's email includes a number of files that are likely to contain design and music for specific games. For items Sanger created for companies, the company may own the rights to the work. Works made for hire are those done within the scope of employment or when the parties agree in writing that a specifically commissioned work included as part of a larger work is a work for hire. We are uncertain whether Sanger signed express agreements retaining copyright for works created by him. As it is unclear to us whether the creations contained within Sanger's email collection are works for hire, with copyright owned by the company, or Sanger's own works, to which he owns the copyright, the full scope of rights for the items cannot be ascertained. Copyright for some materials included in Sanger's email collection may belong to the company that created the video game, depending upon agreements entered into by Sanger and the company prior to creation of the work.

Privacy rights may protect some information contained within the collection as well. The underlying presumption of privacy law is to protect an individual from having certain private facts disclosed about them. For email, this may include personal data, such as financial information, social security numbers, or any other information that would be highly offensive to the reasonable person and is not of legitimate concern to the public. Additionally, publication of photographs contained as attachments in the collection may implicate privacy rights.

Although Sanger's donor agreement for the entirety of the George Sanger Papers specifies that he retains copyright excepting copies of his materials for research purposes, our official policy going forward with the project was that the Briscoe Center should not allow copying or printing out of any of the email files. Access to the files should be restricted—occurring only at the Briscoe Center on secure computer by appointment supervised by the archivist.

Imaging

With our legal protocols established, we began the actual processing of the collection. The first step was to create access copies of the DVDs so as not to corrupt the original files. We created image files (ISOs) of the DVDs using InfraRecorder (http://infrarecorder.org/). The bitstream that was copied could then be examined without risk of irreparable damage to the original. Through inspection of the files we found the Eudora.exe file, indicating that the files were from a Windows operating system. File properties specified that Sanger employed Eudora version 7. After consulting with Dr. Galloway we decided that providing access to the material in as close as is possible to its original environment is more important than is determining specifically the version of Eudora in use at the moment of creation for each of the collections' 15,844 files.

Therefore, we installed, tested and ultimately used latest version of Eudora (7.1) on a PC running a Windows XP in a virtual environment. The files were extracted from the ISO images and moved to the Eudora system folder within the Windows XP environment. This process allowed us to view the files within Eudora, at which time we began conducting inventory on the files.

Inventory

We generated inventories detailing the file type and size of each individual file as well as a summary of the file formats present using a freeware tool called Simple Directory Analyzer (http://www.simpledirectoryanalyzer.tk/). We also generated a format inventory of the attachments folder alone. The inventories were created by consolidating the files into a single directory on which the software was run. The reports were created as HTML tables, which we converted to tab-delimited plain-text and then Excel spreadsheets.

The archive consists of 15,844 files in 91 directories. The following table shows the file formats of which there were at least 10 in the Sanger Electronic Mail Collection. The full inventory of file formats can be found in Appendix B. Due to its size, the full inventory of individual files was ingested separately into the Sanger Electronic Mail Collection in the UT Digital Repository as an RTF file.

	ALL		ATTACHMENTS		NOT ATTACHMENTS	
File Type	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
.xls	347	2.14	347	4.2	0	0.00

.png	327	2.02	94	1.14	233	0.03
.mp3	320	1.98	320	3.87	0	0.00
.txt	254	1.57	208	2.52	46	0.01
.dat	242	1.49	183	2.21	59	0.01
.wav	185	1.14	185	2.24	0	0.00
.msf	119	0.73	0	0	119	0.01
.ics	88	0.54	88	1.06	0	0.00
.dll	87	0.54	23	0.28	64	0.01
.bmp	80	0.49	75	0.91	5	0.00
.vcf	75	0.46	75	0.91	0	0.00
.mov	62	0.38	62	0.75	0	0.00
.jpeg	45	0.28	10	0.12	35	0.00
.mid	40	0.25	40	0.48	0	0.00
.pce	37	0.23	0	0	37	0.00
.tiff	35	0.22	32	0.39	3	0.00
.bik	31	0.19	31	0.37	0	0.00
.rar	31	0.19	31	0.37	0	0.00
.ini	28	0.17	19	0.23	9	0.00
.cf	25	0.15	0	0	25	0.00
.lst	23	0.14	0	0	23	0.00
.htm	22	0.14	20	0.24	2	0.00
.html	22	0.14	22	0.27	0	0.00
.ppt	22	0.14	22	0.27	0	0.00
.rtf	22	0.14	22	0.27	0	0.00
.cer	20	0.12	0	0	20	0.00
.wmv	19	0.12	19	0.23	0	0.00
.fdf	17	0.1	17	0.21	0	0.00
.tif	17	0.1	17	0.21	0	0.00
.exe	16	0.1	9	0.11	7	0.00
.smk	13	0.08	13	0.16	0	0.00
.kmz	12	0.07	12	0.15	0	0.00
.act	10	0.06	0	0	10	0.00
.inf	10	0.06	0	0	10	0.00
.log	10	0.06	3	0.04	7	0.00
.opt	10	0.06	0	0	10	0.00
.p7s	10	0.06	10	0.12	0	0.00
.xml	10	0.06	6	0.07	4	0.00

Virus Scan

To determine the dangers of using the files both on the project team's personal computers and the Briscoe Center's computer that would allow access for researchers, we scanned all of the

imaged files with ClamWin Antivirus Portable version 0.95.3. One-hundred-eighteen files were found to be infected. All but six of the files identified as infected are located in the Attachments folder. One file was in the Trash folder and another was in the Junk folder. The full ClamWin report can be found in Appendix A.

The group elected to keep these viruses in our preservation copy of the Sanger Electronic Mail Archive for ingest into the UT Digital Repository for two reasons. First, should someone wish to study email viruses, specifically those from the 1990s and early 2000s, in the future, leaving these viruses would allow such an endeavor. Furthermore, no virus scan is infallible. Thus, subsequent research may determine that some of the items identified as viruses do not pose a threat but are valuable components of the Sanger archive.

The scan and resulting report do allow us to make the Briscoe Center and Digital Archivist, Zach Vowell, aware of the potential threats that will reside on their access computer and so that they will properly safeguard their network against these identified files.

Scope and Contents

The sheer volume of emails, concern over the large number of viruses found in the virus scan, as well as issues on one computer following the opening of some files, caused us to decide against conducting a survey of the contents of the email beyond what is provided by Sanger's file names. From the file and folder names in the file inventory list, one can establish that Sanger conducted both professional and personal business through this email.

Conversion to MBOX

Before XML encoding, we wanted convert the email message files from the proprietary Eudora MBX format to the standard email format MBOX. The MBOX file is more sustainable as it is not client-specific and was additionally required by most email conversion tools, including XML parsers. A number of translation tools were considered, including Aid4Mail and Emailchemy. Ultimately, we used Emailchemy version 9.9.2, offered by Weird Kid Software for the conversion process at Dr. Galloway's recommendation and based on her previous experience. The conversion was a success with the exception of four files (out of 794) that the program was unable to turn into MBOX.

The Search for XML

Although XML markup is not essential to the project, we included it as a goal for long-term preservation of the emails. Like all file formats, Eudora's MBX will not last forever, and possibly even sooner that the death of the Eudora company may come the end of retroactive support for Eudora files created by previous versions of the software. XML-encoding of the actual email messages (the MBX files) creates a human-readable, and sustainable open format for viewing the material. Various open source tools for harvesting data and reading emails in XML already exist that will allow the messages to be searched by header information and be converted to PDF or HTML.

We began our project with very limited knowledge of both the CERP and EMCAP research groups and their work to create tools to convert email to XML for long-term preservation,

including conversion software, viewers, and XML schemas. We assumed that one of these well-tested tools from reputable research institutions and archives would also work for the Sanger email.

The Collaborative Electronic Records Project (CERP) was joint a research project of Smithsonian Institution Archives and the Rockefeller Archive Center, concluded in 2008, to develop, test, and share the technology to preserve digital documents, specifically email, with other non-profit organizations. The Preservation of Electronic Mail Collaboration Initiative (EMCAP) is an ongoing effort by the North Carolina State Archives, the Kentucky Department of Library and Archives, and the Pennsylvania State Archives to develop tools for the conversion of email to XML and then to provide access and training for the tools developed. The two research projects have created a shared XML Schema for email account preservation.
After imaging our discs and beginning to test the files using MBOX conversion tools, we again revisited the issue of XML and came to the realization that, despite the wonderful work and documentation of EMCAP and CERP, their tools and schema would not work for the Sanger email.

EMCAP was developed for the purpose of records management in an institutional setting, and thus the conversion and access tools intended for ongoing use by the email creator through their email client. Additionally, the documentation on and training for these tools was Microsoft-Outlook based. CERP was developed with goals more closely matched to those of the Sanger email project and was tested on Eudora email at the Rockefeller Archive Center. However, after downloading the administrative instructions for the CERP Parser, our team was uncomfortable

with the level of technical knowledge required for its operation. Furthermore, the CERP project was concluded in 2008 and thus offers no active support for its software.

Discouraged by the incompatibility EMCAP and CERP tools with our project, we began looking into other conversion tools and found Xml Electronic Normalising for Archives (Xena). Xena is free and open source software developed by the National Archives of Australia to aid in the long term preservation of various formats of digital records by converting them to XML. Though Xena is not only for the conversion of email it has been used by the National Archives of Australia on email archives. There is little documentation on its XML schema, and we could find no publication about its usage; however, it was user-friendly and also our best chance at encoding the emails with archivally-developed conversion tool and XML schema.

Mbox2xml 2.02.6 (http://tools.elit.nl/mbox2xml.php) was also considered. Elit, a software development company in the Netherlands, created the tool as a backup archiver for Mozilla Thunderbird. Mbox2xml was regarded as a last resort as its orgins are not archival and we were unable to ascertain the level of XML markup it would output.

We tested both Xena and Mbox2xml on the files. When a small number of files were converted, both programs created XML formatted files. Xena was our first choice of the two programs as it offered more robust mark-up than Mbox2xml, but upon attempting conversion of the entire collection, Xena was unable to complete the process. The National Archives of Australia suggested that the problem was due to memory limit issue and offered a solution. While the Xena team's suggestion resulted in more success with our second attempt, Xena again was

unable to process a specific file and failed to finish conversion of the collection. Xena converted everything up to the file that was not normalized, at which point the process ceased.

With this turn of events, we tried once again to use Mbox2xml. However, this program had similar results. This program also stopped progress mid-way through conversion. Processing each mailbox separately may have resulted in some success, but given the limited information supplied by this particular tool as well as the size of the collection and time needed to convert each file independently, the group rejected using this approach as only one week remained before the project deadline.

TAR of Files

Before ingest we decided to create a tarball for each of our sets of Sanger emails, which included the original bitstreams, and the converted MBOX files. Collecting each set of files in a single TAR file will ensure that the files are kept together and preserve their file information and directory structures. We created the TAR file by placing each set into a single folder and files using the open-source program 7-Zip (http://www.7-zip.org). The files were left uncompressed to avoid loss of data.

We did encounter problems reading and copying from the DVDs themselves on certain computers. Once we were able to read the disks we did not have any difficulties imaging. However, one stubborn file gave permission errors when we attempted to extract it from the disk images for TAR-making. Through trial and error of different copying techniques, we finally succeeded and the file (Rolls.fol/Rolls Social Listserve) is now part of the original bitstream tarball deposited in the UT Digital Repository.

A note indicating that Emailchemy was unable to convert four of the MBX files to MBOX is included in the MBOX TAR. The note reads as follows:

> **Sunday, April 18, 2010**
> **The original Eudora e-mail was converted to standard .mbox format using Emailchemy (http://www.weirdkid.com/products/emailchemy/).**
> **The following four mailbox files were unsuccessfully converted:**
> **aa-MGAM q thru zNU.fol/Rocket Reels.mbx**
> **Junk.mbx**
> **Out.mbx**
> **Trash.mbx**
> **Chris Latham**

Metadata

UT Digital Repository Elements
dc.title
dc.description
dc.date.created
dc.creator
dc.description.department
dc.language
dc.publisher
dc.rights
dc.format.original
dc.subject
DBCAH Metadata Schema 1.1
dc.type.original
dc.format.container
dc.format.codec

The metadata for this collection includes all of the elements automatically generated by the University of Texas Digital Repository and additional elements from the Dolph Briscoe Center for American History Metadata Schema 1.1. Both are based on Qualified Dublin Core. The table below identifies the metadata elements used as well as their source. Metadata was created for each of the tarballs, as well as the project documentation.

Ingest

Given the size of the files to be ingested and UT Digital Repository's limits, we were unable to ingest directly into the UT DR. Instead, the files were placed on the School of Information

server to be ingested by Zach Vowell. The following items, along with corresponding metadata, were placed on the server for subsequent ingest: a TAR file of the disk image titled George Sanger Electronic Mail Collection Eudora files and a TAR file of the material converted to MBOX titled George Sanger Electronic Mail Collection MBOX files. We were able to ingest copies of our report directly into the repository. This was done on May 7, 2010. The items ingested were PDF and RTF copies of this report, titled Processing the George Sanger Electronic Mail Collection: Final Report, a PDF of our final presentation (presented at the Society of Southwest Archivists and to the class) and a RTF file containing the file inventory list, titled George Sanger File Inventory List. Additionally, the PDF and RTF documentation files as well as a PDF of our final presentation were ingested into the Pacer repository on May 7, 2010.

OUTCOMES AND RECOMMENDATIONS

Technical Aspects

The group included the TAR file of the original bitstream to be downloaded from the University of Texas Digital Repository for each use. This will ensure the authenticity of the material for each user. The Dolph Briscoe Center for American History will need to download the Eudora client to act as a viewer for the complete Eudora files. The most up-to-date version of the Eudora email client will permit the user to see the email files, messages, and attachments within the Eudora environment. This may be found at http://www.eudora.com/.

XML

As noted in the Processing section above, the XML conversions attempted by this team were unsuccessful. Given the size of the collection, it is likely that the tools were simply ill-equipped

to deal with the amount of material. We have contacted the National Archives of Australia, developers of Xena, to report our issues with the tool as they have been very communicative with us regarding conversion of the collection. Future research and development of XML converters may reveal a tool that is able to process collections of this volume. In this event, conversion of the MBOX files to XML would be beneficial for the continued maintenance of the collection.

Copyright, Intellectual Property, and Privacy

While this group has not ascertained the full scope of material contained in the Sanger Electronic Mail Collection, it is clear even from perusing file the file inventory that messages and attachments within the archive are likely to invoke issues of copyright, privacy, and intellectual property. In all probability, Sanger's correspondents are unaware that the files have been donated to the Briscoe Center. These correspondents retain all copyright related to the messages and attachments they sent to Sanger. Privacy concerns are also likely to arise in the email messages and attachments. A number of emails appear to contain resumes, clearly including sensitive personal information of these employees or potential employees. Additional concerns implicate intellectual property issues due to the numerous files referencing to Sanger's professional work and projects in which he was involved. Ideas and files in the collection may have been created and belong to Sanger, may have been shared by Sanger's colleagues and belong to them, or may belong to the companies employing Sanger and/or his colleagues.

The Sanger Email group, in collaboration with Briscoe Center's Digital Archivist Zach Vowell, has decided that the collection will be restricted in the University of Texas Digital Repository

and will only be accessible from within the Dolph Briscoe Center for American History by appointment. Additionally, users will not be permitted to make copies from the collection. (Z. Vowell, personal communication, March 22, 2010.) This suggestion would involve negotiating approaches with respect to the material to which Sanger holds copyright, while relying on fair use to supply material of Sanger's correspondents for private study. In this case, researchers would be responsible for clearing permission with the copyright holder before publication. The date of copyright expiration will vary depending on the individual file and would need to be determined for each individual message and attachment in the collection. Until the entire collection can be verified as being in the public domain (which will be quite a long time from now), access should be limited to a controlled environment, in this case the Center for American History.

Attachments

The time constraints of this project did not allow for research into the migration of attachments or action toward that end. As a result, all attachments are available only in their original formats in the TAR of the original bitstreams deposited in the UT Digital Repository. However, the email collection contains attachments in a variety of file formats, both proprietary and open source, as evidenced by the Full Inventory Type Comparison found in Appendix B. Further, as revealed by the file inventory some of the attachments have do not have well-formed file names or easily identifiable formats. Given the array of file formats, there is a great risk that the attachments will become obsolete, if they are not already. The migration of these attachments to open source, supported alternative formats to ensure their future functionality should be considered for a future INF 392K class project.

Appendix A: ClamWin Antivirus Report

Scan Started Tue Apr 06 20:53:12 2010

\aa-MGAM aa-Other CategoriesNU.fol\aa-non game specific.fol\Audio Asset List.mbx: Trojan.Agent-28012 FOUND
\attach\abrechnung.zip: Trojan.Agent-62899 FOUND
\attach\abrechnung1.zip: Trojan.Agent-62899 FOUND
\attach\abrechnung2.zip: Trojan.Agent-62899 FOUND
\attach\abrechnung3.zip: Trojan.Agent-62899 FOUND
\attach\abrechnung4.zip: Trojan.Agent-62899 FOUND
\attach\abrechnung5.zip: Trojan.Agent-62899 FOUND
\attach\abrechnung6.zip: Trojan.Agent-62899 FOUND
\attach\Anhang.zip: Trojan.Agent-57253 FOUND
\attach\Exclusive.rar: Trojan.Fakealert-532 FOUND
\attach\Exclusive1.rar: Trojan.Fakealert-532 FOUND
\attach\Exclusive2.rar: Trojan.Fakealert-532 FOUND
\attach\Exclusive3.rar: Trojan.Fakealert-532 FOUND
\attach\Exclusive4.rar: Trojan.Fakealert-532 FOUND
\attach\FatPipe.exe: Trojan.Agent-28012 FOUND
\attach\Fees_2008-20094.zip: Worm.Autorun-1775 FOUND
\attach\Fees_2008-20095.zip: Worm.Autorun-1775 FOUND
\attach\Fees_2008-20096.zip: Worm.Autorun-1775 FOUND
\attach\Instruction.zip: Suspect.DoubleExtension-zippwd-9 FOUND
\attach\INVOICE_8761277.zip: Trojan.Zbot-2535 FOUND
\attach\INVOICE_87612771.zip: Trojan.Zbot-2535 FOUND
\attach\INVOICE_87612772.zip: Trojan.Zbot-2535 FOUND
\attach\Invoice_UPS.zip: Trojan.Zbot-2539 FOUND
\attach\Late.Night.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night1.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night10.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night2.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night3.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night4.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night5.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night6.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night7.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night8.rar: Trojan.Fakealert-532 FOUND
\attach\Late.Night9.rar: Trojan.Fakealert-532 FOUND
\attach\Mahnung.zip: Trojan.Dropper.Rechnung-1 FOUND
\attach\Mahnung1.zip: Trojan.Dropper.Rechnung-1 FOUND
\attach\Mahnung2.zip: Trojan.Dropper.Rechnung-1 FOUND
\attach\power.zip: Trojan.Fakealert-zippwd FOUND
\attach\power1.zip: Trojan.Fakealert-zippwd FOUND
\attach\power2.zip: Trojan.Fakealert-zippwd FOUND
\attach\power3.zip: Trojan.Fakealert-zippwd FOUND

\attach\pussy.zip: Trojan.Downloader-56516 FOUND
\attach\pussy1.zip: Trojan.Downloader-56516 FOUND
\attach\Rechnung1.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung2.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung3.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung4.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung5.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung6.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung7.zip: Trojan.Dropper.Rechnung FOUND
\attach\Rechnung8.zip: Trojan.Agent-57253 FOUND
\attach\Rechnung9.zip: Trojan.Agent-57253 FOUND
\attach\Statment_details.zip: Suspect.DoubleExtension-zippwd-9 FOUND
\attach\Statment_details1.zip: Suspect.DoubleExtension-zippwd-9 FOUND
\attach\ticket_983992.zip: Trojan.Agent-40967 FOUND
\attach\ticket_9839921.zip: Trojan.Agent-40967 FOUND
\attach\tits.rar: Trojan.Agent-54534 FOUND
\attach\tits1.rar: Trojan.Agent-54534 FOUND
\attach\tube.zip: Trojan.Fakealert-532 FOUND
\attach\tube1.zip: Trojan.Fakealert-532 FOUND
\attach\tube10.zip: Trojan.Fakealert-532 FOUND
\attach\tube11.zip: Trojan.Fakealert-532 FOUND
\attach\tube12.zip: Trojan.Fakealert-532 FOUND
\attach\tube13.zip: Trojan.Fakealert-532 FOUND
\attach\tube14.zip: Trojan.Fakealert-532 FOUND
\attach\tube15.zip: Trojan.Fakealert-532 FOUND
\attach\tube2.zip: Trojan.Fakealert-532 FOUND
\attach\tube3.zip: Trojan.Fakealert-532 FOUND
\attach\tube4.zip: Trojan.Fakealert-532 FOUND
\attach\tube5.zip: Trojan.Fakealert-532 FOUND
\attach\tube6.zip: Trojan.Fakealert-532 FOUND
\attach\tube7.zip: Trojan.Fakealert-532 FOUND
\attach\tube8.zip: Trojan.Fakealert-532 FOUND
\attach\tube9.zip: Trojan.Fakealert-532 FOUND
\attach\UPS.zip: Trojan.Spy.Zbot-12 FOUND
\attach\UPS1.zip: Trojan.Spy.Zbot-12 FOUND
\attach\UPSIn87122.doc.zip: Trojan.Zbot-2489 FOUND
\attach\UPSInfo.zip: Trojan.Zbot-3347 FOUND
\attach\UPSINVOICE.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE1.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice77179.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice771791.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice771792.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice8761.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice87611.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice8771.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_019002.zip: Suspect.Bredozip-zippwd-7 FOUND

\attach\UPSInvoice_0190021.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_7791028.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_79971.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_8000073.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_80000731.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_87612.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_876121.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_876178.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_8761781.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_8765122.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_87651221.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSINVOICE_88087.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_89076152.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_890761521.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_90001.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_9008612.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_997612.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPSInvoice_9976121.zip: Suspect.Bredozip-zippwd-7 FOUND
\attach\UPS_INVOICE.exl.zip: Suspect.Bredozip-zippwd-5 FOUND
\attach\UPS_INVOICE1.exl.zip: Suspect.Bredozip-zippwd-5 FOUND
\attach\UPS_INVOICE_9871.zip: Suspect.Bredozip-zippwd-5 FOUND
\attach\UPS_INVOICE_98711.zip: Suspect.Bredozip-zippwd-5 FOUND
\attach\Video.rar: Trojan.Fakealert-532 FOUND
\attach\Video1.rar: Trojan.Fakealert-532 FOUND
\attach\Video2.rar: Trojan.Fakealert-532 FOUND
\attach\Video3.rar: Trojan.Fakealert-532 FOUND
\Conferences and TripsNU.fol\Edinburgh-fringe-etc.mbx: Trojan.Bagle.BN FOUND
\e-bayNU.mbx: HTML.Phishing.Bank-1 FOUND
\Junk.mbx: Email.Trojan.GZC FOUND
\Rolls.fol\Rolls Social Listserve.mbx: WScr.Unsafe.D FOUND
\Trash.mbx: Email.Trojan.GZC FOUND
---------- SCAN SUMMARY ----------
Known viruses: 753956
Engine version: 0.95.3
Scanned directories: 91
Scanned files: 15844
Infected files: 118

Data scanned: 12739.94 MB
Data read: 6239.89 MB (ratio 2.04:1)
Time: 8531.818 sec (142 m 11 s)

Completed

Appendix B: Full inventory of file types with comparison between "Attachments" folder and all other directories.

	ALL FILES		ATTACHMENTS		NOT ATTACHMENTS	
File Type	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
.xls	347	2.14	347	4.2	0	0.00
.png	327	2.02	94	1.14	233	0.03
.mp3	320	1.98	320	3.87	0	0.00
.txt	254	1.57	208	2.52	46	0.01
.dat	242	1.49	183	2.21	59	0.01
.wav	185	1.14	185	2.24	0	0.00
.msf	119	0.73	0	0	119	0.01
.ics	88	0.54	88	1.06	0	0.00
.dll	87	0.54	23	0.28	64	0.01
.bmp	80	0.49	75	0.91	5	0.00
.vcf	75	0.46	75	0.91	0	0.00
.mov	62	0.38	62	0.75	0	0.00
.jpeg	45	0.28	10	0.12	35	0.00
.mid	40	0.25	40	0.48	0	0.00
.pce	37	0.23	0	0	37	0.00
.tiff	35	0.22	32	0.39	3	0.00
.bik	31	0.19	31	0.37	0	0.00
.rar	31	0.19	31	0.37	0	0.00
.ini	28	0.17	19	0.23	9	0.00
.cf	25	0.15	0	0	25	0.00
.lst	23	0.14	0	0	23	0.00
.htm	22	0.14	20	0.24	2	0.00
.html	22	0.14	22	0.27	0	0.00
.ppt	22	0.14	22	0.27	0	0.00
.rtf	22	0.14	22	0.27	0	0.00
.cer	20	0.12	0	0	20	0.00
.wmv	19	0.12	19	0.23	0	0.00
.fdf	17	0.1	17	0.21	0	0.00
.tif	17	0.1	17	0.21	0	0.00

.exe	16	0.1	9	0.11	7	0.00
.smk	13	0.08	13	0.16	0	0.00
.kmz	12	0.07	12	0.15	0	0.00
.act	10	0.06	0	0	10	0.00
.inf	10	0.06	0	0	10	0.00
.log	10	0.06	3	0.04	7	0.00
.opt	10	0.06	0	0	10	0.00
.p7s	10	0.06	10	0.12	0	0.00
.xml	10	0.06	6	0.07	4	0.00
.ico	9	0.06	9	0.11	0	0.00
.vwm	8	0.05	8	0.1	0	0.00
.wma	8	0.05	8	0.1	0	0.00
0	7	0.04	1	0.01	6	0.00
0	7	0.04	1	0.01	6	0.00
.url	7	0.04	7	0.08	0	0.00
.ogg	6	0.04	6	0.07	0	0.00
.3g2	5	0.03	5	0.06	0	0.00
.out	5	0.03	5	0.06	0	0.00
.tlx	5	0.03	0	0	5	0.00
.ca	4	0.02	4	0.05	0	0.00
.cpp	4	0.02	4	0.05	0	0.00
.eps	4	0.02	4	0.05	0	0.00
.exx	4	0.02	4	0.05	0	0.00
.mod	4	0.02	4	0.05	0	0.00
.mpg	4	0.02	4	0.05	0	0.00
.rcf	4	0.02	4	0.05	0	0.00
.rcv	4	0.02	4	0.05	0	0.00
.s3m	4	0.02	4	0.05	0	0.00
.swf	4	0.02	4	0.05	0	0.00
.tmp	4	0.02	0	0	4	0.00
.torrent	4	0.02	4	0.05	0	0.00
.wpd	4	0.02	4	0.05	0	0.00
.wps	4	0.02	4	0.05	0	0.00
0	3	0.02	0	0	3	0.00
.-2	3	0.02	0	0	3	0.00
.aif	3	0.02	3	0.04	0	0.00
.avi	3	0.02	3	0.04	0	0.00
.cnt	3	0.02	0	0	3	0.00
.csv	3	0.02	3	0.04	0	0.00
.docx	3	0.02	3	0.04	0	0.00
.eif	3	0.02	0	0	3	0.00
.emz	3	0.02	0	0	3	0.00
.hlp	3	0.02	0	0	3	0.00
.m4a	3	0.02	3	0.04	0	0.00

.mp4	3	0.02	3	0.04	0	0.00
.mso	3	0.02	0	0	3	0.00
.nnt	3	0.02	0	0	3	0.00
.php	3	0.02	0	0	3	0.00
.psd	3	0.02	3	0.04	0	0.00
.reg	3	0.02	3	0.04	0	0.00
.snd	3	0.02	3	0.04	0	0.00
.temp	3	0.02	3	0.04	0	0.00
.wmz	3	0.02	0	0	3	0.00
.clx	2	0.01	0	0	2	0.00
.epi	2	0.01	0	0	2	0.00
.inx	2	0.01	2	0.02	0	0.00
.ldif	2	0.01	0	0	2	0.00
.mht	2	0.01	2	0.02	0	0.00
.old	2	0.01	0	0	2	0.00
.p7b	2	0.01	0	0	2	0.00
.pps	2	0.01	2	0.02	0	0.00
.qcl	2	0.01	0	0	2	0.00
.sfk	2	0.01	2	0.02	0	0.00
.tbx	2	0.01	0	0	2	0.00
.wss	2	0.01	2	0.02	0	0.00
.xap	2	0.01	2	0.02	0	0.00
.xee	2	0.01	2	0.02	0	0.00
0	1	0.01	0	0	1	0.00
0	1	0.01	0	0	1	0.00
0.01	1	0.01	0	0	1	0.00
0.2	1	0.01	0	0	1	0.00
. batman	1	0.01	0	0	1	0.00
. interesting offe	1	0.01	1	0.01	0	0.00
.¨	1	0.01	0	0	1	0.00
.¨__ÜÇƒˆ	1	0.01	0	0	1	0.00
.__dll__	1	0.01	1	0.01	0	0.00
._woman	1	0.01	1	0.01	0	0.00
.- start bonus, play 1 hour	1	0.01	1	0.01	0	0.00
.{pic_type}	1	0.01	1	0.01	0	0.00
.a20607	1	0.01	0	0	1	0.00
.a22215	1	0.01	0	0	1	0.00
.a34607	1	0.01	0	0	1	0.00
.a43621	1	0.01	0	0	1	0.00
.a45129	1	0.01	0	0	1	0.00
.a75873	1	0.01	0	0	1	0.00
.a77627	1	0.01	0	0	1	0.00
.a86859	1	0.01	0	0	1	0.00

.acq	1	0.01	0	0	1	0.00
.ai	1	0.01	1	0.01	0	0.00
.asps=5176697856	1	0.01	0	0	1	0.00
.aspsite=s11tprstats	1	0.01	0	0	1	0.00
.b34607	1	0.01	0	0	1	0.00
.bak	1	0.01	0	0	1	0.00
.bdb	1	0.01	1	0.01	0	0.00
.bin	1	0.01	1	0.01	0	0.00
.c34607	1	0.01	0	0	1	0.00
.cda	1	0.01	1	0.01	0	0.00
.co	1	0.01	1	0.01	0	0.00
.com	1	0.01	0	0	1	0.00
.com)	1	0.01	1	0.01	0	0.00
.d34607	1	0.01	0	0	1	0.00
.dll-rename just dll	1	0.01	1	0.01	0	0.00
.dllp	1	0.01	1	0.01	0	0.00
.dmp	1	0.01	0	0	1	0.00
.dxx	1	0.01	1	0.01	0	0.00
.e34607	1	0.01	0	0	1	0.00
.gid	1	0.01	0	0	1	0.00
.gz	1	0.01	1	0.01	0	0.00
.idx	1	0.01	0	0	1	0.00
.isu	1	0.01	0	0	1	0.00
.it	1	0.01	1	0.01	0	0.00
.jan12b	1	0.01	1	0.01	0	0.00
.jpe	1	0.01	1	0.01	0	0.00
.mdi	1	0.01	1	0.01	0	0.00
.mpeg	1	0.01	1	0.01	0	0.00
.msi	1	0.01	1	0.01	0	0.00
.npr	1	0.01	1	0.01	0	0.00
.ocx	1	0.01	0	0	1	0.00
.ott	1	0.01	1	0.01	0	0.00
.pcm	1	0.01	1	0.01	0	0.00
.pcx	1	0.01	1	0.01	0	0.00
.pfs	1	0.01	1	0.01	0	0.00
.pm$	1	0.01	1	0.01	0	0.00
.pub	1	0.01	1	0.01	0	0.00
.ram	1	0.01	1	0.01	0	0.00
.rege	1	0.01	1	0.01	0	0.00
.regs	1	0.01	1	0.01	0	0.00
.renametoexe	1	0.01	1	0.01	0	0.00
.rms	1	0.01	1	0.01	0	0.00
.rns	1	0.01	1	0.01	0	0.00
.rwz	1	0.01	1	0.01	0	0.00

.s	1	0.01	0	0	1	0.00
.sdb	1	0.01	0	0	1	0.00
.sfap0	1	0.01	1	0.01	0	0.00
.sit	1	0.01	1	0.01	0	0.00
.sxw	1	0.01	1	0.01	0	0.00
.t	1	0.01	1	0.01	0	0.00
.template	1	0.01	0	0	1	0.00
.tip	1	0.01	0	0	1	0.00
.vsd	1	0.01	1	0.01	0	0.00
.w__a_™¨Δ¨_	1	0.01	0	0	1	0.00
.w__a_™¨Δ¨__Ü	1	0.01	0	0	1	0.00
.wdb	1	0.01	1	0.01	0	0.00
.xex	1	0.01	1	0.01	0	0.00
.zi_	1	0.01	1	0.01	0	0.00
TOTAL	**16202**		**8267**		**7935**	

REFERENCES

Adgent, Nancy and Lynda Schmitz Fuhrig. (2008). Digital Dilemmas: Archiving e-mail. Sleepy Hollow, NY and Washington DC: The Collaborative Electronic Records Project. Retrieved April 23, 2010 from: http://siarchives.si.edu/cerp/CERP%20SAA%20Research%20Forum%202008_CC.pdf

Adgent, Nancy and Lynda Schmitz Fuhrig. (2009). The Collaborative Electronic Records Project Summary. Sleepy Hollow, NY and Washington DC: The Collaborative Electronic Records Project. Retrieved April 23, 2010 from: http://siarchives.si.edu/cerp/CERP_project_summary_122008_CC.pdf

Boudrez, Filip; Van den Eynde, Sofie. (Aug 2002). DAVID: archiving e-mail. Retrieved April 24, 2010 from: http://www.expertisecentrumdavid.be/davidproject/teksten/Rapporten/Report4.pdf

Collaborative Electronic Records Project. Retrieved April 24, 2010 from: http://siarchives.si.edu/cerp/

Copyrights, 17 U.S.C. §§102-201 (2002).

Dorner, Steve. (Spring 1994). Dorner Speaks: A conversation with the creator of Eudora. *Enterprise Times*. Retrieved March 17, 2010 from: http://www.ibiblio.org/wpercy/welty/html/dorner.html

Engst, Adam. (2001). Eudora 4.3 public beta adds free usage modes. Retrieved March 14, 2010 from: http://www.tidbits.com/tb-issues/TidBITS-509.html#lnk2.

Legal Information Institute. Intellectual Property. Retrieved April 23, 2010 from: http://topics.law.cornell.edu/wex/intellectual_property

Luening, Erich. (Dec 8, 1997) New Eudora more than just email. Retrieved April 4, 2010 from: http://news.cnet.com/New-Eudora-more-than-just-email/2100-1001_3-206092.html

North Carolina State Archives. Preservation of Electronic Mail Collaboration Initiative. Retrieved April 24, 2010 from: http://www.records.ncdcr.gov/EmailPreservation/default.htm

Paradigm Project Workbook on Digital Private Papers. (2006). Intellectual Property Rights. Retrieved April 23, 2010 from: http://www.paradigm.ac.uk/workbook/legal-issues/copyright-archives.html

Thomas, Jo. (January 21, 1997). For inventor of Eudora, great fame, no fortune. Retrieved March 14, 2010 from: http://www.nytimes.com/library/cyber/week/012197eudora.html.

Qualcomm. (2006). Eudora 7.1 for Windows User Guide. Retrieved April 23, 2010 from: http://www.eudora.com/download/eudora/windows/7.1/Eudora_71_User_Manual.pdf

Qualcomm. (2008). Historical Backgrounder. Retrieved March 14, 2009 from: http://www.eudora.com/presskit/backgrounder.html#name.