MANAGING TOMORROW'S RECORDS TODAY: AN EXPERIMENT IN ARCHIVAL PRESERVATION OF ELECTRONIC MAIL

CAROLE ELIZABETH NOWICKE

ABSTRACT: Encouraged by the speed of transmission and ease of use, many organizations are relying heavily on electronic mail for internal and external communications. What are the implications of electronic mail for the archivist? An experiment was undertaken by the Navy Laboratories History Program to examine use of electronic mail within the Navy Laboratory community and attempt to preserve and archivally manage this evanescent form of communication.

The Report of the Committee on the Records of Government stated in 1985 that the United States was in danger of losing its memory. That conclusion has been frequently quoted in conjunction with archival concerns over the fate of information stored in computerized records. The fear simply put is that the records' custodians are "erasing tomorrow's records today."[1] While the first archival concern should be the quality of the information contained within electronic formats, most studies of electronic records have focused on the physical structure of data on magnetic tape or disks, conservation practices, or the difficulty of retaining hardware, rather than on the value of the information recorded.

This article focuses on the quality of information within an electronic communications system. It will describe an in-house study of the use and archival management of an electronic mail (E-Mail) system used in the office of the Director of Navy Laboratories (DNL), report on results of the study, and draw a few tentative conclusions.

What is electronic mail? Electronic mail is a system whereby people who share a computer system, or have access to an electronic data link, are able to send each other messages. The fundamental advantage of E-Mail is that it is fast. Messages are received within minutes, whereas surface mail can take over a week to go across the country. Electronic mail systems are also easy to use. Most are menu-driven, leading the user through a step-by-step set of instruc-

tions. Figure 1 is a representation of a menu from an IBM PROFS communication system.[2] Many systems have override capacities that allow more experienced users to bypass the menus and move directly to commands (such as "talk" or "snd" typed at a prompt sign), saving the user time and providing him or her a greater sense of power and control. Local computers are linked together through national computer networks similar to telephone long distance lines.

In the Navy research and development community, scientists and engineers can communicate with their peers in the same facility, Department of Defense agencies, government-owned laboratories, or university laboratories working on government contracts, through the Navy Laboratory Computer Network (NALCON), the Defense Advanced Research Projects Agency Network (ARPANET), or other services.[3]

Why study electronic mail? The Navy Laboratories History Program had three specific reasons for undertaking the project. First, the history program was interested in the impact of E-Mail on the historical records of naval research and development. What was being communicated over the electronic mail network that was not documented any other way? Use of computer resources by history program staff itself, and paper copies of electronic mail messages sent to the Navy Laboratory Archives as "historic" items, made it obvious that some substantive information was being transmitted this way. Second, electronic mail messages met the criteria for Title 44 U.S. Code definition of official government records, "regardless of physical form or characteristics."[4] Therefore, an obligation existed to establish some controls over information in the E-Mail system. Finally, DNL headquarters was concerned about the lack of management

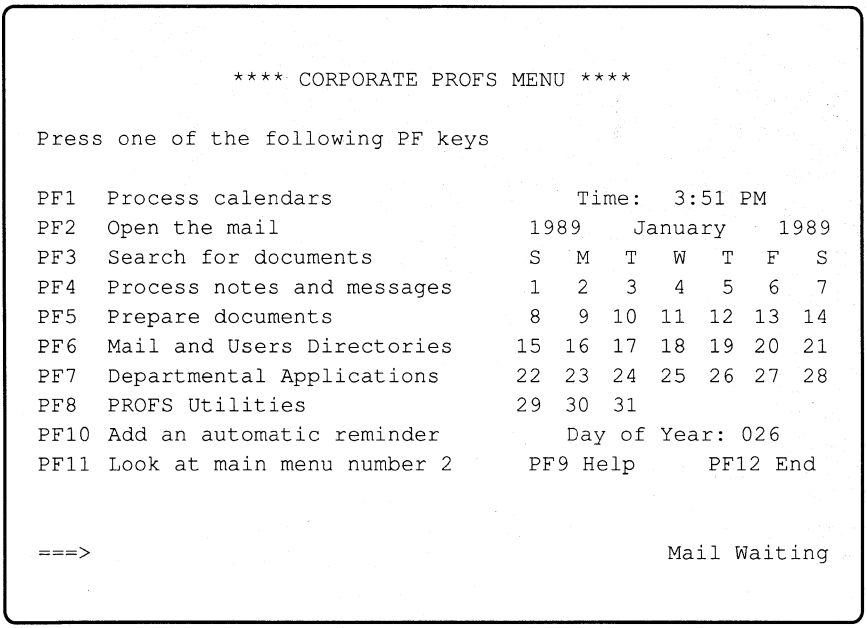

Figure 1

control or accountability for E-Mail. This interest by top level management provided a third reason to devote staff time and resources to the experiment.

The data collected were the incoming and outgoing electronic mail messages of a group of volunteers from the offices of the Director of Navy Laboratories from February 1985 through December 1985. The number of participants varied from nine to fourteen through the course of the experiment.

The methodology employed in the experiment was fairly simple. The history program staff inserted commands in the "log in" files of DNL personnel instructing the host computer to copy the person's incoming and outgoing messages and send them to the history program mail box at the end of the business day.[5] The methodology thus exploited the system's automated data processing capabilities to copy and transfer messages without rekeyboarding information. The methodology was invisible to participants and did not impede the information flow.

When a user logs in and looks into an electronic mailbox, he or she sees a list of messages received, as in Figure 2.

A single Navy laboratory electronic mail message would appear on a user's screen as shown in Figure 3.

The "headers" of the message ("date," "from," "subject," "to") perform the same function as the structured format of official Navy correspondence and provide the same points of entry for access and file management. These subject headings are not thesaurus controlled, but neither are they standardized in official correspondence. In theory, Navy correspondence is filed according to the Navy Standard Subject Indentification Code (SSIC), but most offices disregard

```
1   25 Apr 89   19:22    mann        USS EMORY LAND
2   26 APR 89   08:25    hunnicut    CNM Monthly Highlights
3   26 Apr 89   09:15    mailer      undeliverable mail
4   26 Apr 89   10:03    webern      VIP visit
5   26 Apr 89   11:25    grogg       Fraud, Waste & Abuse
6   26 Apr 89   13:42    dreyfus     DNL oral history
7   26 Apr 89   16:30    burke       let's have lunch
8   26 Apr 89   17:15    mcgovern    CONGRESSIONAL INQ.
9   26 Apr 89   17:32    hiestand    Zupko memo
```

Figure 2

instructions and arrange records by chronological, alphabetic, or case file method.

When the history program staff received the messages, they were examined, and those that were determined to have permanent value were retained, while the remainder were deleted.[6] Total mail received each day would average 100-150 messages; of this number ten percent or less would be retained. The entire processing procedure would generally take under one hour per day, dependent on number of messages, length of messages, and whether any format editing needed to be performed.[7]

The headers of the messages selected for retention were then read into a database that allowed Boolean searching and retrieval. A predefined command took the existing headers, standardized the format, and added three more elements, "actor," "action," and "action date."[8] If a message required some type of response, these three additional data elements were altered to indicate this.

The database was made of unit entries with the following data fields:

actor:	[Who is assigned action (when appropriate)]
action:	[Brief statement of what is required]
act_date:	[Deadline for action (when appropriate)]
m_date:	[Date messsage was sent (when appropriate)]
fr:	[Author of the message]
subj:	[Subject statement assigned message by author]
for:	[Who received the message, including both the original "to" line and the original carbon copy ("cc") line. For forwarding messages, includes the original and new distribution list]

```
Date:    26 Apr 89 15:43 EDT
From:    J. Carr@NOTS.ARPA
Subj:    National Register Nomination
To:      M. Grogg <grogg@DTMB.ARPA>


----------------------------------------
Michael--

Please send a copy of the package you prepared
for the Ft. Nonsense nomination ASAP. I have to do
much the same for two of the structures here.

                  Thanks!


                  Jim
```

Figure 3

Only the headers and not the entire message were entered into the database due to storage limitations. Some hand-editing of the file was occasionally necessary due to differences in formats, computers, software, and operating systems. The entire electronic mail record was retained, and filed by year and month.

Three products were created from the data collected for the experiment. The first was a paper print-out of the electronic mail, totaling 2596 pages. The paper copy was generated to discover the physical volume of the correspondence, and also because the system was not failsafe; in point of fact, it "lost" the mail for March 1985.[9] The second product was fourteen cards of Computer Output Microfiche (COM), comprising three indices created from a sorted file of the database and the messages for each month printed on a separate card. The indices were sorted by "from" and date, by "subject" and date, and by date alone. The COM was generated to see if it was a practical method for access to the record in the archives. The last item produced was the complete file of the mail, plus the indices, on one roll of nine-track 1600 bpi ASCII tape.

During the first few months of the experiment, two classes of E-Mail were defined; "Routine Mail" which was anything sent, and "Executive Mail" which required some type of action, generally a response by electronic mail. "Executive Mail" status was assigned by the person sending the message, in this case, the DNL or his designee. "Executive Mail" involved statements of official policy, official assignments, major commitments of personnel or funds, significant changes, etc. It was correspondence that required action or close attention by top management, and carried a special marking to designate its status (**EM**). When there was a question of validity, the authenticity of the message could be confirmed by telephone (no electronic signature or other verification system being extant).[10]

The electronic mail messages retained, including both Executive Mail and Routine Mail, were analyzed for subject content. Only 12.6% of these messages had been designated Executive Mail by their creators, but the definition had not been applied for the entire year of the study. Executive Mail messages were also examined to ascertain "turn-around" time.

The subject categories shown in Table 1 were generalized from the subject headings selected by the senders of the messages. In this example only one subject was counted per letter. "Inspections and investigation" letters concerned scheduled inspections, investigations of waste, fraud, and abuse, and investigations by military and civilian law enforcement agencies. "Personnel matters" E-Mail dealt with general or specific personnel management issues, awards to employees, or monthly employment statistic reports. "Research and program management" messages were statements of research policy, concerns about specific programs, or discussions of so-called "technology base issues." "Corporate planning" was a narrowly defined area concerning the Navy Laboratories' corporate planning process and "mission and function" statements for each of the laboratories. "Meetings and conference" messages, as the category suggests, were agendas, attendance lists, speeches, and other information relating to meetings within the Navy laboratory community. "Financial management" E-Mail concerned military construction monies (MILCON), general funding, procurement, commercial activities (i.e., contracting out), and so on. The category of "routine office support" consisted of messages with too many subjects to be

TABLE 1
MESSAGE SUBJECT CATEGORIES

Category	Number	Percentage of total (rounded)
Inspections and Investigation:	153	11 %
Personnel Matters:	150	11 %
Research and Program Management:	380	29 %
Corporate Planning:	102	8 %
Meetings and Conferences:	120	9 %
Financial Management:	107	8 %
Routine Office Support:	324	24 %
Total	1336	100 %

accounted for (such as those designated "miscellaneous" or "several items" by their creator), or topics occurring in only a few messages.

Time constraints placed on the laboratories to answer inquiries were in many cases impossible to meet when sending the reply by surface mail. The time limits on replies in the sample shown in Table 2 were from a smaller group of messages than the total retained, as the "Executive Mail" status was not devised until well into the experiment, and because the status was only applied to outgoing headquarters messages.

Subjects discussed in the messages retained in the experiment were to some extent indicative of the positions of the individuals participating in the experiment. Because most of the participants were in the DNL division concerned with general laboratory management rather than facilities management, long-range planning, or financial resources management, there are a lower number of communiques regarding finance and infrastructure than would be expected.

The history program found the retained electronic mail to be extremely valuable in documenting the evolution of Navy laboratory management policy. During the course of the experiment, management oversight of the laboratory changed three times—from the Naval Material Command (NAVMAT), to the Chief of Naval Research (CNR), and finally under the cognizance of the Space and Naval Warfare Systems Command (SPAWAR). Because of the rapidity of the reorganizations and the nature of communications between the parties involved, the electronic mail record will be one of the primary sources of documentation for the history program.

TABLE 2
RESPONSE REQUIRED TIMES
FOR 169 "EXECUTIVE MAIL" MESSAGES

Reply Time	Number	Percentage of total (rounded)
1-6 Days	103	61 %
7-21 Days	41	24 %
22-59 Days	25	15 %
Over 60 Days	0	0 %
	169	100 %

Electronic mail is the nearest written equivalent to the correspondence of the pre-World War II era, when decision makers committed their thoughts, feelings and judgments to discursive prose in official letters.[11] The letters conveyed information for an immediate purpose; they were not written with an eye to history, nor did they serve the purpose of most official letters today—after-the-fact confirmation of decisions already reached.[12]

In this study, some of the most active users of the electronic mail system were the laboratory technical directors and commanding officers. As trained scientists and engineers, these officials had no qualms about personal use of computers. Their electronic mail records retained some of the spontaneity and directness of their nineteenth century predecessors' handwritten notes, representing an improvement over official Navy correspondence which is written in "a characteristic Navy style which conveys every personal action as an abstract state of mind."[13]

Electronic mail in the Navy laboratory community also replaces, in part, information "lost" in telephone calls. Because the laboratories are located in four time zones, electronic mail is a popular medium for communication—avoiding "telephone tag" and waiting because all the AUTOVON circuits to Washington are busy.[14] Electronic mail gives the illusion of privacy with none of the constraints of official correspondence. While any business conducted over an official government electronic mail network is in fact official, the feeling about E-Mail is that it provides the "weight of official pronouncement with the deniability of a telephone call."[15]

During this study the archivist could witness the rampant proliferation of records as letters with multiple addressees would be copied and passed on. With the aid of the computer, multiple copies without additional significant comments could be identified and deleted quickly. The archivists' processing time could have been pared down further with more sophisticated computer programming that would, for example, automatically remove duplicate messages and reformat messages to conform to the E-Mail configuration of the host system.

Having determined that important information was flowing through the electronic mail system, the history program attempted to develop a way to capture it for the archives. Of the three methods tried—paper, computer output microfiche, and a data tape—in the best of all possible worlds and with a larger universe of information (for example, the electronic mail of all of the Navy laboratories) the nine-track tape would be the preferred form for the retention of electronic mail. The problems that this can create, however, were made apparent by a decision to change the host computer system's DBMS (Database Management System), requiring all databases to be moved to a new DBMS. This particular change represented an inconvenience rather than a loss of records, but it was a reminder that both software and hardware change. A change like this entails defining the data fields in a new database management system, usually stripping the existing data files of control characters into another file, inserting control characters used by the new system, and then reading the file up into the new database.[16] The nine-track tape, stripped of control characters to make it software independent can, however, be run on a variety of machines using many different DBMS.

For a small number of records, as in this experiment, the most cost effective storage format would be the COM version. With only fourteen cards, a manual inspection for an item is faster than loading a tape and implementing a Boolean search. For the records of an entire laboratory, obviously the magnetic tape or disk version would be more useful.

The experiment was regarded as successful by management and the history program, proving that electronic mail was valuable, could be captured, managed, archivally processed, and stored easily and economically. The program was simple and straightforward, and initial limitations were accepted. The system was developed in-house taking advantage of technical expertise of staff members. In-house development also allowed low cost and low risk experimentation, and quick implementation of the system. Since the electronic mail system was already in operation, the only costs for the experiment were labor and storage. The personnel costs consisted of the labor of the historian in creating the command files and defining the database, the daily processing time for the archivist, and the time of a computer programmer to produce the COM and ASCII tape. Storage costs were for both active memory and off-line storage of the host computer.

One facet of electronic mail this study did not address was how use of the system affected communications between departments. The lack of a formal structure for routing allows for crossing of departmental boundaries, and the lack of hierarchical distinctions allows any user to contact any other user directly. Some unlikely "pen pals" developed on the host system as proficient users would become interested in the programs, or in the use personnel in other departments were making of the computer. Some of this is directly attributable to the "friendliness" of the Unix operating system, where certain commands tell who is using the system, what they are doing, how long they have been doing it, and where they belong in the organization.[17]

This study helped to document a time of drastic changes for Navy laboratory management. Many of the messages would have been printed out and saved by the senders, some would have been deleted, but never would the collective information have been gathered and organized.

The author suspects that the Navy laboratory community is not unique in its use of electronic mail, and hopes that others will examine electronic network communication within their organizations.

ABOUT THE AUTHOR: Carole Elizabeth Nowicke was Archivist of Navy Laboratories at the David Taylor Naval Ship Research and Development Center in Carderock, Maryland, from 1983 to 1986. She has also worked at the Henry Ford Museum and the U.S. Marine Corps Museum.

NOTES

1. *Report of the Committee on the Records of Government* (Washington, D.C.: The Committee, 1985), 1. Also, Thomas E. Brown, "Erasing Tomorrow's Archives Today" (Paper delivered at the Annual Meeting of the Society of American Archivists, Chicago, Illinois, 28 August 1986).
2. PROFS (Professional Office System) 5664309 © Copyright IBM Corp., 1983, 1987.
3. Anthony V. Cinotta, Larry Avarunin, and Petro Matula, *Naval Electronic Mail Service Users Guide* (Bethesda, Md.: David W. Taylor Naval Ship R&D Center, 1982), passim.
4. 44, United States Code, 3301.
5. The host computers were DEC VAX 11/780 machines using the Unix operating system. David K. Allison, then Historian of Navy Laboratories, devised and programmed the electronic mail management system.
6. Theodore Scheilenberg, *The Appraisal of Modern Public Records,* National Archives Bulletin no. 8 (Washington, D.C.: National Archives and Records Service, 1956), 6-8, 22-7.

   Many of the deleted records were duplicates, with all personnel receiving copies listed on the distribution line, or in an electronic "forwarding" message. This pattern of multiple copies was common in the paper correspondence records of this office, as it is in many organizations. Other records deleted included copies of printed Navy or Department of Defense instructions not originated in the laboratories, military personnel fitness reports, notices from the host computer about system shut-downs, "fortune cookie" or "message of the day" or "word of the day" messages, notices of social activities, blood drives, fire drills, etc., on the grounds of the Navy laboratory where the host computers were located, and "undeliverable mail," messages from the E-Mail system with copies of letters sent with incorrect E-Mail addresses or when the receiver's system was down.
7. This included printing out the retained messages to determine the physical volume.
8. These three elements were used for the management portion of the experiment to create daily "tickler files" to keep track of staff assignments.
9. The database entries were of course saved, and the text was still available in the paper print-out.
10. David K. Allison, Director of Navy Laboratories, Guidance Letter #4, "Communication Using Electronic Mail," 1985. Records of the Director of Navy Laboratories, Navy Laboratory Archives.
11. David K. Allison, unpublished letter, 12 July 1985. Records of the Director of Navy Laboratories, Navy Laboratory Archives.
12. David K. Allison, letter, 12 July 1985. David Garraty, *The Nature of Biography* (New York: Knopf, 1957), 199.
13. Rodney Carlisle, unpublished draft for *Center of Excellence: A History of the David W. Taylor Naval Ship Research and Development Center,* 41. Navy Laboratory Archives.
14. Automated Voice Network, a military "Watts line." Clark A. Elliott, "Communications and Events in History: Toward a Theory for Documenting the Past," *American Archivist* 47 (Fall 1985):361.
15. Colin Cherry, *On Human Communication* (Cambridge, Mass.: Massachusetts Institute of Technology, 1957), 21. Jerry S. Kidd, conversation with author, College Park, Md., 26 July 1986.
16. An entry in a file with control characters would look something like this: ^GDate: 21 Sep 85: 14:30 EDT^GFrom: c.hiestand@Ft.Custer.ARPA^GSubj: Command History^GTo: t.mann@china.lake@ARPA^G.
17. This ability to cross boundaries is not seen as positive in all quarters. In "Office Automation at the White House" (*Computers and People,* January-February 1984, p.19), Craig Fuller reports on this viewpoint:

    > We have had a spirited debate at times over the question of just who should be allowed to use an electronic mail system. Some fear the worst—all sorts of "back channel" communication coming through the electronic mail system. Well, I can see little reason to restrict the use of an important management tool because of fear that several people will start to communicate with me electronically.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse.
I am interested in electronic ordering through the following system(s):

☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem

☐ Other (please specify)_____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name_____

Title_____

Institution/Company_____

Department_____

Address_____

City_____ State_____ Zip_____

Phone (_____)_____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106