E-mail Metadata In A Post-Armstrong World

Jason R. Baron
U.S. Department of Justice
Civil Division, Federal Programs Branch
901 E Street N.W., Rm. 1040
Washington, D.C. 20530
phone: (202) 514-4336
fax: (202) 616-8470
jason.baron@usdoj.gov

ABSTRACT: In light of a series of landmark judicial decisions over the past decade involving the federal government's electronic recordkeeping policies and practices, the Archivist of the United States and the Executive Office of the President (EOP) have confronted the issue of how best to maintain and preserve on a long-term basis office automation records, including e-mail and word processing documents, in an electronic format. In substantial part, the case law in this area fundamentally involves a set of assumptions concerning the nature of what constitutes a "complete" record under the U.S. federal records laws, including consideration of the importance of contextual elements of electronic records (such as transmission and receipt data) which may be considered "metadata." At the same time as the courts and the Executive branch have struggled with this issue, the archival profession has been moving ahead in providing more systematic thinking about the management of electronic records, including functional requirements which include contextual metadata elements, with some of this work already incorporated into best practices in government. Additionally, Internet-driven standards and protocols promise to play a future role in defining the nature and scope of what forms of e-mail and other records will be considered appropriate for preservation as part of the permanent electronic records of the U.S. government.

1.0 INTRODUCTION

Archivist John Carlin has recognized that "[e]lectronic records pose the biggest challenge ever to recordkeeping in the Federal government and elsewhere," and federal agencies "want to manage their electronic records . . . more effectively." As just one example, the Department of Health and Human Services (HHS), with 55,000 e-mail users, estimates that it generates over 1 million e-mail messages a day, taking up 18 gigabytes of temporary storage space. On a yearly basis, based on a conservative multiple of 100 (accounting for over 300 separate reporting agencies and departments of the Federal government), this estimate yields e-mail traffic approaching 36.5 billion messages per year (100 x 365 million). Clearly, even if not all messages warrant "record" status under the federal records laws (as discussed below), the Federal government nevertheless must confront the fact that this volume of e-mail (with attachments) presents formidable records management issues regardless of an agency choice of paper and/or electronic recordkeeping systems.

Based on a series of landmark judicial decisions over the past decade involving the federal government's recordkeeping policies and practices, the Archivist of the United States and the Executive Office of the President (EOP) have confronted the issue of how best to maintain and preserve on a long-term basis office automation records, including e-mail and word processing documents, in an electronic format. In substantial part, the case law in this area fundamentally involves a set of assumptions concerning the nature of what constitutes a "complete" record under the U.S. federal records laws, including
consideration of the importance of contextual elements of electronic records (such as transmission and receipt data) which may be considered "metadata." Most federal agencies continue to operate recordkeeping programs which place primary reliance on paper-based recordkeeping systems for the long-term preservation of office automation records such as generated on e-mail and word processing systems. However, the coming emergence over the next several years of document management programs in the marketplace which have been certified as meeting federal recordkeeping requirements, including legally required metadata elements, will allow agencies that have a legitimate business need to do so to elect to retain electronic versions of e-mail and other office automation records, without having to devote substantial resources to customization of existing in-house, proprietary e-mail systems.

2.0 LITIGATION BACKGROUND

On January 19, 1989, the last day of the Reagan Administration, a federal court in Washington, D.C. granted a temporary restraining order to preserve a collection of PROFS backup tapes from the National Security Council (NSC) and EOP's Office of Administration, in a lawsuit brought by several individuals and nonprofit associations captioned Armstrong v. Bush and filed as a Federal Records Act (and Freedom of Information Act) challenge. The PROFS tapes contained, among other things, electronic mail messages of Oliver North concerning the Iran-Contra affair, transmitted over the NSC's internal e-mail system. By a later estimate, the original grouping of 392 backup tapes contained over 7,000,000 pages of materials constituting e-mail notes, documents, and calendars captured on the PROFS system dating from the mid-1980s through 1989. Following an appeal on certain threshold jurisdictional issues, and immediately subsequent to the November 1992 U.S. election, the lawsuit was subsequently expanded to cover ten or more additional e-mail systems in use by staff of various components of the EOP, so as to include within the scope of the lawsuit Bush Administration e-mail messages captured on backup dates created after November 20, 1992. On January 6, 1993, as amended on January 11, Judge Charles R. Richey issued an injunction against the EOP (including the NSC), holding that based on the characteristics of the proprietary e-mail systems in place within the EOP and NSC at the time, the defendants' written records policies directing that hard copies of e-mail messages be printed out as the sole means of preserving e-mail records was arbitrary, capricious and contrary to the U.S. federal records laws. This holding was affirmed on later appeal to the U.S. Court of Appeals for the District of Columbia. On July 14, 1994, the EOP adopted new recordkeeping procedures ensuring the preservation of federal record e-mail messages, with their attachments, in electronic form.

In response to the holding in Armstrong, on August 28, 1995, the National Archives and Records Administration (NARA) issued final regulations governing the preservation of e-mail in electronic mail systems. As revised, NARA regulations included distinct definitions of what constitutes an "electronic information system" versus an "electronic recordkeeping system." The former term was intended to include information on a "live" office automation system, including existing proprietary e-mail and word processing systems; the latter term was intended to exclude such systems, unless they complied with specified functional requirements. Contemporaneously, NARA also issued a revised version of General Records Schedule 20, governing Electronic Records, which contained government-wide disposition authority for the deletion of e-mail and word processing records from live electronic information systems, provided that such records had been transferred to an agency recordkeeping system (including either paper, microform, or electronic recordkeeping systems). Of special note, GRS 20, Item 14, covering email records, expressly included a provision to the effect that
transmission and receipt information accompanying e-mail records must be preserved along with the e-mail records themselves when transferring the records to existing agency recordkeeping systems.\(^{(14)}\)

In December 1996, a group of plaintiffs (including some of the Armstrong plaintiffs) sued the Archivist and the EOF, challenging the legality of the 1995 version of GRS 20 in a case captioned Public Citizen v. Carlin. In October 1997, U.S. District Court Judge Paul Friedman ruled that the Archivist had exceeded his authority under the Federal Records Act (specifically, the Records Disposition Act of 1943, as amended in 1945), in promulgating a "general records schedule" of government-wide applicability which allowed for wholesale destruction of the electronic versions of the government's electronic mail and word processing records, without requiring that agencies individually submit records schedules justifying the disposition of these records.\(^{(15)}\) As part of its holding, the court expressed the view that the electronic content of electronic mail, word processing documents, and spreadsheets may contain information not preserved in the print-out record or may "retain features unique to their medium."\(^{(16)}\) An order accompanying the opinion declared that GRS 20 was null and void. The government has appealed from this and a subsequent order of the district court. At the time of this writing, the Carlin case has been briefed and argued on appeal, and is currently awaiting decision from the U.S. Court of Appeals for the District of Columbia Circuit.

3.0 THE ARMSTRONG CORE METADATA SET

At the core of the Armstrong holdings is the view expressed by both the district and appellate courts that with respect to proprietary e-mail systems such as utilized by the EOF, the paper and electronic versions of e-mail messages may differ in crucial respects, and therefore separately qualify for treatment as "federal records" (as opposed to the view that after printout and preservation in a paper system, the electronic versions of email messages may be considered to be presumptively nonrecord). As a corollary, where paper and electronic versions of email records differ, the electronic versions of such records must be considered separately for disposition under the federal records laws.

A "record" under the Federal Records Act (FRA) may include "machine readable [e.g., electronic] materials . . . regardless of physical form or characteristics," which satisfy a two-pronged test: the item must be (i) made or received by an agency of the United States government under Federal law or in connection with the transaction of public business," and (ii) "preserved or appropriate for preservation by that agency . . . as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of data in them."\(^{(17)}\) Under longstanding NARA regulations, records are "preserved" or "appropriate for preservation" when they are filed, stored or otherwise "systematically maintained" by an agency in agency recordkeeping systems.\(^{(18)}\)

As originally framed by the district court, the electronic versions of e-mail contained "qualitatively different" information of "tremendous historical value in demonstrating . . . what officials knew, and when they knew it."\(^{(19)}\) Thus, recordkeeping guidance which merely required the printing out of hard copies of email messages constituting federal records where the paper printouts lack certain transmission and receipt data accompanying the email message itself was insufficient for purposes of compliance with the federal records laws. Importantly, the district court held that not all e-mail messages constitute "records;" agencies are only required to pay proper attention to the management of the subset of e-mail that do constitute "records" within the above FRA definition.\(^{(20)}\) For such records, the Armstrong core set of transmission and receipt "metadata" information which the court held was legally required to be
captured consists of:

1. <an intelligent representation of name of sender>
2. <an intelligent representation of name of all recipients, including on distribution list of "cc's">
3. <the date of transmission>
4. <the date/time of an acknowledgement of receipt, only where requested by sender>.\(^{(21)}\)

Thus, for example, e-mail messages on the NSC PROFS system sometimes identified senders and recipients in only cryptic fashion, *e.g.*, from "OLN," rather than "Oliver North," to a personal group under the name "List A," rather than to a list of identifiable recipients. Hard copy printouts would fail to contain the additional information presumed to be captured somewhere by the proprietary software (either in user directory lookup tables, personal group lists, or linked to the main body of the e-mail message in accompanying pages such as "Info" pages).\(^{(22)}\)

In affirming the district court, the D.C. Circuit stated that paper printouts "do not affect the record status of the electronic materials unless the paper versions include all *significant* material contained in the electronic records,"\(^{(23)}\) deeming the electronic versions to be "at most, kissing cousins" of their paper counterparts.\(^{(24)}\) The appellate panel variously referred to e-mail hard copy printouts as "dismembered," "amputated," or "lopp\[ed\] off," without missing transmission and receipt information, which in its view was "integral," "fundamental, and "meaningful," to the preservation of a complete electronic record under the FRA.\(^{(26)}\)

In a footnote, the Court of Appeals stated its assumption that the missing sender/recipient contextual information would necessarily be in some fashion incorporated, bound up with, or linked to the electronic version of the e-mail record at the time of transmission.\(^{(27)}\) However, the appellate panel went on to state that even if user directories or personal group distribution lists which specify complete sender and recipient information are perceived to be independent records, they are to be preserved in situations where the email record itself fails to provide complete contextual information.\(^{(28)}\)

Finally, the *Armstrong* appellate court did not decide whether "parallel documents" maintained in different records systems in different media could be considered as "copies" subject to "unobstructed destruction."\(^{(29)}\) Thus, the *Armstrong* holding does not determine whether an agency's preservation of the *Armstrong* transmission and receipt core metadata in paper form justifies treatment of electronic versions of the same records as uniformly disposable once recordkeeping copies have in fact been preserved. (The legal issue left open in *Carlin* may be resolved in the pending *Carlin* case.)

The 1995 NARA e-mail regulations faithfully track the *Armstrong* core transmission and receipt data elements.\(^{(30)}\) These regulations, which set out the "unique aspects" of electronic mail, require that e-mail records residing on a "live" system be in turn placed in some form of agency recordkeeping system (paper or electronic),\(^{(31)}\) "as the best means to preserve the content, structure, and context of the electronic records."\(^{(32)}\)

**4.0 EOP’S IMPLEMENTATION OF THE ARMSTRONG DECISION**

As the result of an accommodation reached with plaintiffs in the *Armstrong* litigation, as of July 1994 all
EOP components covered under the court’s orders voluntarily chose to put into place one or more systems of electronic recordkeeping, whereby electronic mail messages (including word processing attachments) have been captured at the time of both transmission and receipt for purposes and maintained in electronic format. The most comprehensive of these systems, known as "ARMS" (the "automated records management system"), is maintained by EOP’s Office of Administration, which manages e-mail for several of EOP’s federal components. USTR and NSC have continued to maintain their own separate and internal e-mail systems.

In choosing to implement the Armstrong holding by management of electronic versions of e-mail in their electronic form, and given the enormous volumes of e-mail traffic generated by even the tiny portion of agency staff within the EOP, various EOP components have confronted the need for designating the record status of e-mail messages in an effort to reduce their long-term management burdens. EOP components have done so by customization of existing proprietary software to provide for the embedding of record status metadata (including by means of a separate field or label), designating individual e-mail messages as "records" or "nonrecords." Even prior to the main Armstrong case holding, as of June 1992 the NSC had put into place a system of categorizing e-mail by record status. Before transmission of any e-mail message, NSC staff members are required to designate whether particular messages constitute "records" or "nonrecords." The electronic version of e-mail messages designated as records have been streamed to a central recordkeeping account for retention under applicable law. In the ARMS system, users are also prompted to designate the record/nonrecord status of e-mail; however, should they neglect to do so, the default designation is "record." External e-mail received over the Internet and sent to ARMS users also is tagged with "record" status.

Additionally, EOP components have made substantial efforts to customize existing proprietary software (including All-in-1, Lotus Notes, cc:Mail, and WordPerfect) so as to bind certain Armstrong core transmittal and receipt information to individual e-mail records. (For example, using Word Perfect, the separate "Info" screen containing the full name of the sender and full distribution lists of recipients has been married up to the original e-mail message for maintenance in the USTR recordkeeping system.) Acknowledgements of receipts are captured and maintained in electronic form as separate e-mail records in all of EOP’s systems.

5.0 BEYOND ARMSTRONG: UTILIZING STANDARDIZED METADATA FOR PRESERVING THE GOVERNMENT’S E-MAIL RECORDS

The Armstrong precedent recognizes as a matter of law that some core data set of contextual information associated with e-mail records is appropriate for preservation to ensure that complete records are maintained by government. This legal holding is in line with more recent efforts by the archival profession to provide conceptual foundations for the design and implementation of electronic recordkeeping systems covering a broader spectrum of electronic records (textual and nontextual in nature). For example, one of the objectives of the first phase of the ongoing research project on electronic records sponsored by the University of British Columbia (UBC) has been "to establish what a record is in principle and how it can be recognised in an electronic environment." The UBC project proposes that complete records created in an electronic environment must include certain elements of intellectual form, which serve to expand upon the Armstrong core metadata set. The very differing conceptual architecture in what is known as the "Pittsburgh Project" includes a set of functional requirements, production rules, and metadata specifications for electronic recordkeeping systems.
In collaboration with the UBC project, a task force at the U.S. Department of Defense identified requirements for records management applications, resulting in the subsequent issuance in November 1997 of DoD Standard 5015.2-STD, a Design Criteria Standard For Electronic Records Management Software Applications.\(^{(39)}\) The DoD standard provides for a "minimum set of baseline functional requirements" consistent with governing statutes and regulations that purport to be "applicable to all records management applications regardless of organizational and site-specific implementations."\(^{(40)}\)

As part of a more robust set of functional requirements, section C2.2.3 of DoD Standard 5015.2 incorporates the *Armstrong* core metadata set, in providing for general rules for the filing of e-mail, including: first, that e-mail records shall be treated as any other records subject to the requirements of Standard 5015.2 (C2.2.3.1); second, that record management applications shall capture and automatically store transmission and receipt data as identified in an accompanying table, and shall not allow editing of this metadata (C.2.2.3.2); and third, that e-mail attachments shall be stored or linked with e-mail (C.2.2.3.3). With respect to the *Armstrong* core element of acknowledgements of receipts, the DoD Standard also expressly states that record management applications shall provide supporting links between supporting and related records and related information such as "notes, marginalia, attachments, and electronic mail return receipts" (C.2.2.15). The DoD standard assumes record status for documents subsumed within the scope of the records management application, and allows for labelling of record categories with appropriate time, and/or event dispositions categories (C.2.2.5.2).

In conjunction with issuance of the DoD Standard, DoD has also put into place a process for testing and certifying proprietary software products for compliance with the standard, as set out in a register.\(^{(41)}\) A register of certified products is listed on DoD's website along with summary information. Also, as part of a continuing collaborative partnership between DoD and NARA, in November 1998, Archivist John Carlin wrote to DoD stating that NARA does endorse the DoD standard as establishing baseline requirements for managing electronic records, subject to certain qualifications which remain to be addressed in either the next iteration of the standard or in supplemental guidance.\(^{(42)}\)

It is presently unknown how long it may take before DoD-certified software is tested for utilization across a spectrum of non-defense federal departments and agencies. However, assuming a business need exists to implement electronic recordkeeping, the DoD standard and certification process provides a step along the road towards future implementation of document management products in the workplace which meet federal records law requirements.

### 6.0 RECORDKEEPING METADATA IN A WEB-BASED WORLD

In the decade since 1989, when the *Armstrong* case was filed, the federal workplace has seen mass-deployment of e-mail at the desktop. Coupled with this has been the transformation of the office environment due to the advent of the Internet and particularly the World Wide Web. Federal procurement of state-of-the-art proprietary office automation software allows a growing number of Federal agency staff to have the desktop capabilities of e-mailing documents created in HTML (hyper-text markup language) format, for placement in internal agency intranets, or communicating externally with the world utilizing such features as e-mail and word processing documents with embedded hypertext links. On the near-term horizon is the expected arrival of applications which will
provide for true integration of e-mail, voice mail and fax transmissions. Federal staff may choose to be
communicating with their offices via the Internet by means of portable devices, including personal
digital assistants and other means of wireless transmission. Portions of government may seek out routine
means of employing desktop video conferencing and video mail for business needs. The explosion of
new forms of communication affecting the Federal sector pose profound challenges in keeping up with
federal records laws requirements.

With the coming dominance of the Internet and the Web, a future program for standardizing federal
recordkeeping metadata elements in e-mail and other office automation products may need to confront
compatibility with open, international standards and protocols which involve forms of document
creation. These would include the RDF (Resource Description Framework) standard in development by
the World Wide Web (W3) Consortium, for the purpose of providing a general metadata framework for
the Web using XML (Extensible Markup Language), as well as the PICS (Platform for Internet Content
Selection) rating system.\(^{(43)}\)

Against this backdrop of exponential change and growth in the quantity and quality of data, as a useful
first step in confronting recordkeeping obligations in differing Federal information environments
(including but not limited to Web-based software platforms), consideration should be given to adopting
the strategy of employing a simple metadata label, field, or tag, similar to that presently used by the EOP
for designation of e-mail records, so as to identify subsets of data (in metadata registries or elsewhere)
which are appropriate for preservation as federal "records." Utilizing front-end metadata designations
arguably will greatly assist the federal user in meeting future obligations in the area of managing
electronic data as records under the federal records laws.

7.0 CONCLUSION

Judicial holdings in the Armstrong case have recognized that electronic records may contain additional
data elements than their paper counterparts which are appropriate for long-term preservation as federal
records of the U.S. government. The core data elements identified in Armstrong form a subset of the
metadata elements which depending on the business needs of federal agencies may be necessary to
provide adequate context for electronic records. The EOP's implementation of Armstrong provides one
important model of how agencies may adopt front-end metadata management for their e-mail records.
DoD's standards for document management functionality are a further step towards a standardized
metadata set for the e-mail records of the federal government.

8.0 ENDNOTES

grs20/state921.html>.

2. Declaration of Dr. Neil J. Stillman, Deputy Chief Information Officer for HHS, dated March 17, 1998
(filed on behalf of the government in Public Citizen v. Carlin, No. 96-2840 (D.D.C.)) (available from
author).

3. The following summary of the Armstrong lawsuit only purports to deal with a subset of the issues
which together consumed seven years of active litigation ending in 1997, including but not limited to
access to the e-mail records on backup tapes under the Freedom of Information Act, the legal status of
the National Security Council, the continued preservation of backup tapes under stipulations and orders
still in effect, and the initial jurisdictional basis of the lawsuit. The author notes that legal analysis of the merits of the central court holdings in *Armstrong* and subsequent case law is beyond the scope of this article, and nothing in this article is intended to be construed as an endorsement of prior judicial characterizations of the government’s position in *Armstrong* or in subsequent litigation.


6. See *Armstrong v. Executive Office of the President*, 807 F. Supp. 816 (D.D.C. 1992). Various additional components of the EOP subsumed within the scope of the Federal Records Act, included, for example, the Office of Management and Budget, the Office of the U.S. Trade Representative, the Office of National Drug Control Policy, and the Office of Science and Technology Policy.


8. See *Armstrong v. Executive Office of the President*, 1 F.3d 1274 (D.C. Cir. 1993). On January 15, 1993, the government’s emergency motion for a partial stay of the district court’s Order was granted by the U.S. Court of Appeals, allowing EOP components to delete e-mail messages from their live operating systems, so long as all record information was preserved in identical form on backup tapes. Order of January 15, 1993, Civ. No. 93-5002 (D.C. Cir.).


12. See 36 C.F.R. 1234.22 (text documents generally); 1234.24(b)(1) (e-mail).


14. *Id.* at 44,649.


16. *Id.* at 14.

17. See 44 U.S.C. 3301.


20. *Id.*, at 340-41.

21. *Id.*, at 341.
22. *See* Armstrong, 1 F.3d at 1280 & 1285 n.8.

23. *Id.,* at 1283 (emphasis added).

24. *Id.*

25. *Id.,* at 1285, 1286.

26. *Id.,* 1277, 1284-87.

27. *Id.,* 1285 n.8.

28. *Id.*

29. *Id.,* at 1284.

30. *See* 36 C.F.R. 1234.24 (standards for managing electronic mail records). As interpreted by NARA, the Armstrong requirement to capture acknowledgments of receipts ("read receipts") does not translate into a more general requirement to capture information on when every e-mail record is "opened" or read by one or more recipients. Rather, the requirement is confined only to situations where the sender has a business need for obtaining confirmation of receipt, equivalent to the mailing of a letter by certified mail, return receipt requested. *See* 36 C.F.R. 1234.24(a)(3).


34. As the result of an important, separate aspect of the Armstrong litigation, the D.C. Circuit later ruled that the NSC is a component of the White House which solely advises and assists the President and thus creates records within the scope of the Presidential Record Act (PRA), rather than the FRA. Armstrong *v. EOP,* 90 F.3d 553 (D.C. Cir. 1996), *cert. denied,* 117 S. Ct. 1842 (1997).

35. Also, the ARMS system maintains multiple copies of e-mail records in various "buckets" designated by EOP component. In some cases, due to the fact that the e-mail system includes components covered by the PRA and the FRA, the identical e-mail message may be a presidential record in the sender’s repository and a federal record in the recipient’s repository, or vice versa.


37. Template 6 of the UBC Project states that "All complete electronic records, whether textual or non-textual, must include the following elements of intellectual form:

1. <chronological dates of transmission and receipt>
2. <topical date (i.e., mention of place document made and/or transmitted)>
3. <originating address>
See <http://www.slais.ubc.ca/users/duranti/tem6.htm>. (The Template provides complete a full
description of the data elements described above.) The UBC project goes on to include additional
elements as part of its discussion of ensuring the reliability and authenticity of electronic records. See
Templates 7 & 8, <http://www.slais.ubc.ca/users/duranti/tem7.htm> and
<http://www.slais.ubc.ca/users/duranti/tem8.htm>, respectively.

38. See <http://www.lis.pitt.edu/~nhpr>. The Pittsburgh Project introduces the concept of multiple levels
of metadata, all of which comprise and are considered bound to an electronic record object. These
potential forms of metadata are denominated by the terms "handle layer," "terms and conditions layer,"
"structural layer," "contextual layer," "content layer," and "use history layer." See id.


40. Id., C.1.2 (Limitations). Of note, the DoD Standard expressly provides for a definition of metadata
as "[d]ata describing stored data; that is, data describing the structure, data elements, interrrelationships,
and other characteristics of electronic records." DoD Standard 5015.2-STD, AP1.39. The definition also
references "Record profile data," where Record profile is in part defined as "[i]nformation (metadata)
about a record that is used by the RMA [Record Management Application] to file and retrieve the
record." Id., AP1.50.


42. Letter of Archivist John W. Carlin to Arthur L. Money, Senior Civilian Official, Assistant Secretary
of Defense (Command, Control, Communications and Intelligence), dated Nov. 18, 1998.

43. See generally <http://www.w3.org/TR/NOTE-rdf-simple-intro>.

9.0 AUTHOR'S NOTE AND ACKNOWLEDGEMENTS

Mr. Baron has appeared as Justice Department counsel of record in Armstrong v. Executive Office of the
President and Public Citizen v. Carlin. The author wishes to acknowledge the helpful comments of Bill
Leary, Miriam Nisbet, and Gary Stern in preparation of this paper.