

Preservation Action Plan: Email Records

National Archives and Records Administration (NARA)

DRAFT 20190801
Template: 201907

Electronic Record or Digital Surrogate Types and Associated Formats

A document created or received via an electronic mail system, including brief notes, formal or substantive narrative documents, and any attachments, such as word processing and other electronic documents, which may be transmitted with the message.

Essential Characteristics of this Record Type

Appearance characteristics are not generally essential to email, with the possible exception of text (and which is generally “inherited” from the primary record type of text, the general category of textual records). Layout might also be considered in certain circumstances, as specified below.

Behavior/User interaction with email does not generate a characteristic that can be preserved or replicated after the fact.

Name	Definition	Function Description
Font	The appearance of the message(s).	Font information inclusive of bolding, italics, etc., would apply. Color, orientation, etc. would not apply. This applies only to the messages; attachments are handled as separate objects.
Layout	Distinct layout in which more than just the discernable data elements themselves provide meaning	Overall visual layout or structure of highly recognizable email “formats” such as State Department cables, EOP ARMS email, or Defense Messaging System messages.

Message body mark-up		Markup may be necessary to preserve, depending on the content and layout of the message.
----------------------	--	------------------------------------------------------------------------------------------

e

Name	Definition	Function Description
Structure	Proprietary internal structure of a commercial email format such as MS-Outlook or Lotus Notes. Includes links and connections to other types within the same system (calendar, meetings, etc.)	During transformations, the specific constructs of the proprietary format should be preserved. Markup may be necessary to preserve, depending on the type and format of the message.
Attachments (if present)	An identifier that indicates one or more attachments (including inline attachments) are associated with the email.	
Original Order	The organization and sequence of records established by the creator of the records. Possible strong association to email, e.g., maintenance of email in folders that indicate separate mailboxes, time periods sent/received, etc.	Preservation of original order for email would most likely come into play in circumstances where email is transferred in distinct folders or directories that reflect a clear operational structure; for example, where email is segregated and transferred by mailbox or username; or where segregated and transferred by blocks of time, such as by week or month. In such circumstances, that structure would be named as an essential characteristic and should be preserved.

Behavior

Name	Definition	Function Description
------	------------	----------------------

User Interaction	N/A	User interaction with email does not generate a characteristic that can be preserved or replicated after the fact.
------------------	-----	--------------------------------------------------------------------------------------------------------------------

Context

Name	Definition	Function Description
Subject	A short string that may identify the topic of the message. The subject line may be blank, indicate the content of the email to which the Sender is replying, or contain other information.	The Subject may provide qualitative information that indicates the message purpose. Additionally, it may provide a simple method to sort several emails into a thread when used in conjunction with the received date.
Keywords (if present)	Words and phrases that may summarize the content of the message. Keywords may be created by the person or software application that creates, receives the message, or archives the message.	
Originator		Contains several interdependent parts, outlined below.
Local-part	The user account of the Agent assigned by a mail provider. The local-part is identified by alphanumeric characters prior to the @ symbol of an email address.	Establishes the provenance (and as a result support or contradict its authenticity) of message by identifying the user account that was used to transmit the message.
Domain-part	The host or domain name used by a DNS to indicate the mail provider that handles the email message.	Establishes the provenance (and as a result support or contradict its authenticity) of message by identifying the domain from which the message originated.
Domain-literal	The IP address of the source or destination domain.	Establishes the provenance (and as a result support or contradict its authenticity) of message by identifying

		the machine address from which the message originated.
Display name (if present)	A plain text indication of the Agent's name.	Establishes the provenance (and as a result support or contradict its authenticity) of message by identifying the name of the Agent specified for the mail account.
Relationship	The relationship that the Agent has with the email message, e.g. creator, sender, recipient (primary, CC, BCC).	Establishes the provenance (and as a result support or contradict its authenticity) of message by identifying how each agent relates to the email.
Message-ID	A unique identifier created by the domain from which the email originated that is embedded within the email header. The message-id is found in received emails and is not present in local emails.	The message-id should be used when attempting to understand the relationship between two or more emails that constitute a thread. It is beneficial when the subject line has changed. However, it has only limited use when handling a single email.
msg-id	A globally unique identifier used as a subfield in the Identification, Resent, References, and Trace fields	
Sent-date	The date and time that an email was completed by a Creator and/or transmitted by the Sender, or received by a Recipient.	The sent-date is obtained from the system settings of the sender's machine. It may indicate the datetime in which an idea was expressed. However, there is the potential that the datetime has been accidentally or deliberately altered, which may result in the value being untrustworthy.
Received date	The date and time that an email was received by the recipient's host.	Indicates the datetime that an email was received. However, it does not confirm that the email was downloaded or read by a recipient.
Resent	Resent fields are be added to any message that is reintroduced by a user into the email transport system.	Each of the resent fields corresponds to a particular field used elsewhere. For instance, the "Resent-Date:" field corresponds to the "Date:" field and the "Resent-To:" field corresponds to the "To:" field.
References	Used when creating a reply to a message. The References fields are used to identify a "thread" of	Present in child messages on email threads; the fields will contain the contents of the parent's "References:"

	conversation.	fields (if any) followed by the contents of the parent's "Message-ID:" fields (if any)
In-Reply-To	Used when creating a reply to a message. The "In-Reply-To:" fields are used to identify the message (or messages) to which the new message is a reply.	The "In-Reply-To:" fields will contain the "Message-ID:" fields of the message to which this one is a reply (the "parent message"). If there is more than one parent message, then the "In-Reply-To:" fields will contain the contents of all of the parents' "Message-ID:" fields. If there is no "Message-ID:" field in any of the parent messages, then the new message will have no "In-Reply-To:" field.
Trace-field	Indicates the route that the email took to travel from the sender to the recipient and when it occurred. A repeatable value consisting of an optional "Return-Path" field and one or more "Received" fields.	The trace fields are external to the control of the sender and recipient and, therefore may be thought more trustworthy than the Sent and Received date for validation.

Structure

Name	Definition	Function Description
Structure	Proprietary internal structure of a commercial email format such as MS-Outlook or Lotus Notes.	During transformations, the specific constructs of the proprietary format should be preserved. Markup may be necessary to preserve, depending on the type and format of the message.
Attachments (if present)	An identifier that indicates one or more attachments are associated with the email.	
Original Order	The organization and sequence of records established by the creator of the records. Possible strong association to email, e.g., maintenance of email in folders that indicate separate mailboxes, time periods sent/received, etc.	Preservation of original order for email would most likely come into play in circumstances where email is transferred in distinct folders or directories that reflect a clear operational structure; for example, where email is segregated and

		transferred by mailbox or username; or where segregated and transferred by blocks of time, such as by week or month. In such circumstances, that structure would be named as an essential characteristic and should be preserved.
--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Current NARA Transfer Guidance for this Record Type

- Preferred Transfer Formats
 - EML (individual messages)
 - PST (folders)
 - MBOX (individual messages or folders)
- Acceptable Transfer Formats
 - XML (individual messages)
 - MSG (individual messages)

<https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#email>

Current NARA Public Access/Reference Format(s) for this Record Type

This Plan references existing public access file formats for electronic records at NARA, determined with a survey of the available public access formats in the National Archives Catalog. These references do not represent recommended public access formats under NARA policies. They are intended for informational purposes only.

Reference Format: The current preferred reference format for supplying Federal email records to researchers is PDF or the native plain text format.

Public Access Format: A portion of Federal and Presidential email records from the holdings are available through the National Archives Catalog as PDFs and CSV or unstructured ASCII text.

Electronic Mail Format

NARA Format ID: NF00170

Extension(s)

- eml

Format Documentation

IMF

Internet Message Format (IMF) is the top-level standardized ASCII-based syntax required by SMTP for all email message bit streams used by a message transfer agent, sometimes referred to as a mail transfer agent or MTA, when moving messages between computers. IMF is standardized by RFC 5322. EML is an extension of the IMF standard.

Specification: <https://www.ietf.org/rfc/rfc5322.txt>

EML (Individual email messages)

EML is an extension for an email message saved to a file in the Internet Message Format protocol for electronic mail messages. It is the standard format used by Microsoft Outlook Express as well as some other email programs. Since EML files are created to comply with IMF specification RFC 5322, EML files can be used with most email clients, servers and applications. EML files typically store each message as a single file (unlike MBOX which concatenates all the messages from a folder into one file), and attachments may either be included as MIME content in the message or written off as a separate file, referenced from a marker in the EML file.

Specification: <https://www.ietf.org/rfc/rfc2822.txt> (Draft update at <https://datatracker.ietf.org/doc/rfc5322/>)

Risk and Prioritization Analysis

- Low Risk**
 - Moderate Risk**
 - High Risk**
- 30 Numeric Risk Rating**
30 Numeric Prioritization Rating

Proposed Preservation Plan

- Retain** file format in its existing format.
- Transform** file to a new format.
Selected Format:
- Procure/develop tools** to preserve, manage and provide access to records of this type in their existing form.

- Procure/develop tools** to transform the format to the preferred normalized form.
- Provide Additional Information** so that the record type remains understandable/usable over time.
- Explore Additional Options**

Preferred Processing and Transformation Tool(s)

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

Since EML files are created to comply with the industry RFC 822 standard (<https://tools.ietf.org/html/rfc822>), they can be viewed with most email clients, servers and applications. Besides the Microsoft Outlook Express, EML files can be opened using most email clients, such as Microsoft Outlook, Microsoft Entourage, Mozilla Thunderbird, Apple Mail, and IncrediMail. Since EML files are plaintext and formatted much like MHT (MIME HTML) files, they can also be opened directly in the Internet Explorer, Mozilla Firefox and Opera, by first changing the file extension from '.eml' to '.mht'. It is also possible to view EML files using Notepad or any other text editor.

- [CoolUtils Mail Viewer](#)
- [PSTViewerPro](#)
- [DArcMail](#)
- [aid4mail](#) (proprietary license, non-US company, MS-Windows)

Preferred Viewer/Access Software

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

Since EML files are created to comply with the industry RFC 822 standard (<https://tools.ietf.org/html/rfc822>), they can be viewed with most email clients, servers and applications. Besides the Microsoft Outlook Express, EML files can be opened using most email clients, such as Microsoft Outlook, Microsoft Entourage, Mozilla Thunderbird, Apple Mail, and IncrediMail. Since EML files are plaintext and formatted much like MHT (MIME HTML) files, they can also be opened directly in the Internet Explorer, Mozilla Firefox and Opera, by first changing the file extension from '.eml' to '.mht'. It is also possible to view EML files using Notepad or any other text editor.

- [CoolUtils Mail Viewer](#)

Microsoft Outlook Item

NARA Format ID: NF00280

Extension(s):

- msg

Format Documentation

IMF

Internet Message Format (IMF) is the top-level standardized ASCII-based syntax required by SMTP for all email message bit streams used by a message transfer agent, sometimes referred to as a mail transfer agent or MTA, when moving messages between computers. IMF is standardized by RFC 5322. EML is an extension of the IMF standard.

Specification: <https://www.ietf.org/rfc/rfc5322.txt>

<https://www.loc.gov/preservation/digital/formats/fdd/fdd000379.shtml>

Risk and Prioritization Analysis

Low Risk

Moderate Risk

High Risk

33 Numeric Risk Rating

31 Numeric Prioritization Rating

Proposed Preservation Plan

Retain file format in its existing format.

Transform file to a new format.

Selected Format: EML

Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

Provide Additional Information so that the record type remains understandable/usable over time.

Explore Additional Options

Preferred Processing and Transformation Tool(s)

- [CoolUtils Mail Viewer](#)
- [PSTViewerPro](#)
- [DArcMail](#)
- [aid4mail](#) (proprietary license, non-US company, MS-Windows)
- It is also possible to view MSG files using Notepad or any other text editor.

Preferred Viewer/Access Software

- [CoolUtils Mail Viewer](#)
- It is also possible to view MSG files using Notepad or any other text editor.

MBOX Email Format

NARA Format ID: NF00247

Extension(s):

- mbox

Documentation

IMF

Internet Message Format (IMF) is the top-level standardized ASCII-based syntax required by SMTP for all email message bit streams used by a message transfer agent, sometimes referred to as a mail transfer agent or MTA, when moving messages between computers. IMF is standardized by RFC 5322. MBOX is an extension of the IMF standard.

Specification: <https://www.ietf.org/rfc/rfc5322.txt>

MBOX (Email folders)

MBOX (sometimes known as Berkeley format) is a generic term for a family of related file formats used for storing collections of electronic mail messages. The four formats - mboxo, mboxrd, mboxcl, and mboxcl2 – originate from different versions of Unix. Each mbox file represents a set of email messages that are ordered sequentially and grouped into a 'folder'. Email messages are stored in their source format, e.g. plain text may be stored as ASCII or Unicode, binary data is stored as Base64-encoded text.

Specification: <https://tools.ietf.org/rfc/rfc4155.txt>

Risk and Prioritization Analysis

Low Risk

Moderate Risk

High Risk

34 Numeric Risk Rating

34 Numeric Prioritization Rating

Proposed Preservation Plan

Retain file format in its existing format.

Transform file to a new format.

Selected Format: EML

Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

Provide Additional Information so that the record type remains understandable/usable over time.

Explore Additional Options

Justification: MBOX can be retained as an acceptable preservation format, but it is preferable to transform boxes into individual eml message files for access purposes.

Preferred Processing and Transformation Tool(s)

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

- [CoolUtilsMBOXViewer](#)
- [Emailchemy](#)
- [DArcMail](#)
- [aid4mail](#) (proprietary license, non-US company, MS-Windows)

Preferred Viewer/Access Software

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

- [CoolUtilsMBOXViewer](#)

Microsoft Outlook Personal Folder

NARA Format ID: NF00281

Extension(s):

- pst

Documentation

IMF

Internet Message Format (IMF) is the top-level standardized ASCII-based syntax required by SMTP for all email message bit streams used by a message transfer agent, sometimes referred to as a mail transfer agent or MTA, when moving messages between computers. IMF is standardized by RFC 5322.

Specification: <https://www.ietf.org/rfc/rfc5322.txt>

PST (Email folders)

PST, or Outlook Personal Folders File Format, represent a hierarchy of email folder(s) which contains email messages, which can contain attachments. Information about Folder objects, Message objects, and Attachment objects are stored in properties, which collectively contain all of the information about the particular item.

Specification: [https://msdn.microsoft.com/en-us/library/ff385210\(v=office.12\).aspx](https://msdn.microsoft.com/en-us/library/ff385210(v=office.12).aspx)

Risk and Prioritization Analysis

Supply the Risk Level and Numeric Rating and the Prioritization Numeric Rating as generated with the Format Risk and Prioritization Matrix for this file format.

Low Risk

Moderate Risk

High Risk

13 Numeric Risk Rating

13 Numeric Prioritization Rating

Proposed Preservation Plan

Retain file format in its existing format.

Transform file to a new format.

Selected Format: EML

Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

- Provide Additional Information** so that the record type remains understandable/usable over time.
- Explore Additional Options**

Justification: PST is not an open email standard; files in the PST format should be transformed into EML, which is a documented, open, plain-text standard that can be accessed through a wide range of tools.

Preferred Processing and Transformation Tool(s)

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

- [CoolUtils Outlook Viewer](#)
- [Emailchemy](#)
- [PSTViewerPro](#)
- [DArcMail](#)
- [aid4mail](#) (proprietary license, non-US company, MS-Windows)

Preferred Viewer/Access Software

- [PSTViewerPro](#)
- [CoolUtils Outlook Viewer](#)

Microsoft Outlook Personal Folders ANSI 1997-2002

NARA Format ID: NF00282

Extension(s):

- pst

Documentation

IMF

Internet Message Format (IMF) is the top-level standardized ASCII-based syntax required by SMTP for all email message bit streams used by a message transfer agent, sometimes referred to as a mail transfer agent or MTA, when moving messages between computers. IMF is standardized by RFC 5322.

Specification: <https://www.ietf.org/rfc/rfc5322.txt>

PST (Email folders)

PST, or Outlook Personal Folders File Format, represent a hierarchy of email folder(s) which contains email messages, which can contain attachments. Information about Folder objects, Message objects, and Attachment objects are stored in properties, which collectively contain all of the information about the particular item.

Specification: [https://msdn.microsoft.com/en-us/library/ff385210\(v=office.12\).aspx](https://msdn.microsoft.com/en-us/library/ff385210(v=office.12).aspx)

Risk and Prioritization Analysis

Supply the Risk Level and Numeric Rating and the Prioritization Numeric Rating as generated with the Format Risk and Prioritization Matrix for this file format.

Low Risk

Moderate Risk

High Risk

13 Numeric Risk Rating

13 Numeric Prioritization Rating

Proposed Preservation Plan

Retain file format in its existing format.

Transform file to a new format.

Selected Format: EML

Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.

Procure/develop tools to transform the format to the preferred normalized form.

Provide Additional Information so that the record type remains understandable/usable over time.

Explore Additional Options

Justification: PST is not an open email standard; files in the PST format should be transformed into EML, which is a documented, open, plain-text standard that can be accessed through a wide range of tools.

Preferred Processing and Transformation Tool(s)

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

- [CoolUtils Outlook Viewer](#)
- [Emailchemy](#)
- [PSTViewerPro](#)
- [DArcMail](#)
- [aid4mail](#) (proprietary license, non-US company, MS-Windows)

Preferred Viewer/Access Software

- [PSTViewerPro](#)
- [CoolUtils Outlook Viewer](#)

Microsoft Outlook Personal Folders Unicode 2003-2007

NARA Format ID: NF00283

Extension(s):

- pst

Documentation

IMF

Internet Message Format (IMF) is the top-level standardized ASCII-based syntax required by SMTP for all email message bit streams used by a message transfer agent, sometimes referred to as a mail transfer agent or MTA, when moving messages between computers. IMF is standardized by RFC 5322.

Specification: <https://www.ietf.org/rfc/rfc5322.txt>

PST (Email folders)

PST, or Outlook Personal Folders File Format, represent a hierarchy of email folder(s) which contains email messages, which can contain attachments. Information about Folder objects, Message objects, and Attachment objects are stored in properties, which collectively contain all of the information about the particular item.

Specification: [https://msdn.microsoft.com/en-us/library/ff385210\(v=office.12\).aspx](https://msdn.microsoft.com/en-us/library/ff385210(v=office.12).aspx)

Risk and Prioritization Analysis

- Low Risk
 - Moderate Risk
 - High Risk
- 13 Numeric Risk Rating
13 Numeric Prioritization Rating

Proposed Preservation Plan

- Retain file format in its existing format.
- Transform file to a new format.
Selected Format: EML
- Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.
- Procure/develop tools to transform the format to the preferred normalized form.
- Provide Additional Information so that the record type remains understandable/usable over time.
- Explore Additional Options

Justification: PST is not an open email standard; files in the PST format should be transformed into EML, which is a documented, open, plain-text standard that can be accessed through a wide range of tools.

Preferred Processing and Transformation Tool(s)

Note: This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

- [CoolUtils Outlook Viewer](#)
- [Emailchemy](#)
- [PSTViewerPro](#)
- [DArcMail](#)
- [aid4mail](#) (proprietary license, non-US company, MS-Windows)

Preferred Viewer/Access Software

- [PSTViewerPro](#)
- [CoolUtils Outlook Viewer](#)

Stationery for Mac OS X

NARA Format ID: NF00412

Extension(s):

- mailstationary

Documentation

An email stationery file created and used by Mail, an email application bundled with OS X; contains files that make up a stationery design, such as .JPG, .PNG, and .HTML files, along with .STRINGS files and a .PLIST file that describe the design.

Risk and Prioritization Analysis

- Low Risk
- Moderate Risk
- High Risk
- 3** Numeric Risk Rating
- 3** Numeric Prioritization Rating

Proposed Preservation Plan

- Retain file format in its existing format.
- Transform file to a new format.
Selected Format:
- Procure/develop tools to preserve, manage and provide access to records of this type in their existing form.
- Procure/develop tools to transform the format to the preferred normalized form.
- Provide Additional Information so that the record type remains understandable/usable over time.
- Explore Additional Options

Justification: A proprietary add-on to email messages that can only be viewed in Apple Mail. It is unclear if mailstationary files are record or non-record, or if they can be transformed to be viewed with any EML file.

Preferred Processing and Transformation Tool(s)

This section reflects federal and legislative records processing, but is not yet finalized for EOP tools.

- Apple Mail

Preferred Viewer/Access Software

- Apple Mail