Preservation Action Plan: Email
National Archives and Records Administration (NARA)
Plan Date: 20220714
Template: 202105

Email

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.

Significant Properties of Email

Appearance properties are not generally significant to email, with the possible exception of text, which has similar appearance characteristics to textual records. Layout might also be considered in certain circumstances, as specified below. Behavior/User interaction with email does not generate a property that can be preserved or replicated after the fact.

Appearance

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.

Structure

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 a significant property and should be preserved.

Behavior

Name	Definition	Function Description
User Interaction		User interaction with email does not generate a property 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.
Originator/ 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.
Originator/ 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.
Originator/ Domain-literal	The IP address of the source or destination domain.	Establishes the provenance (and as a result supports or contradicts its authenticity) of message by identifying the machine address from which the message originated.

Originator/ Display name (if present)	A plain text indication of the Agent's name.	Establishes the provenance (and as a result supports or contradicts its authenticity) of message by identifying the name of the Agent specified for the mail account.
Originator/ 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 supports or contradicts 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 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 conversation.	Present in child messages on email threads; the fields will contain the contents of the parent's "References:" 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.

Current NARA Transfer Guidance for Email
[Bulletin 2014-04](#)

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

Current NARA Format(s) for Public Access and Reference for Email
Formats for Public Access are those made available online through the National Archives Catalog. Formats for Reference are defined as those made available to researchers upon direct requests for digital copies.

Formats Available for Public Access: Content created or delivered for public access in the Catalog is delivered primarily in the following file formats: PDF (Textual and Image), JPEG (Textual and Image), MP3 (Audio), and MP4 (Audio/Video) and ASCII (Datasets). Other file formats may be present depending on when they were added to the Catalog. Email records may be available via FOIA requests from Presidential Libraries and presented through the National Archives Catalog or FOIA Reading Room.

Format(s) Available for Reference: PDF, native plain text format.