

# Task Force on Technical Approaches to Email Archives: DPC Briefing Day

Chris Prom, University of Illinois at Urbana-Champaign  
Kate Murray, Library of Congress

July 6, 2017  
The National Archives, Kew

# Roadmap

About us: Chris and Kate's background with email projects

Task Force Background and Workplan

Review Email as a Technical Problem

- Researcher Needs

- What is Email?

- Email Lifecycle Approach (understanding workflows and repository needs)

Review Technical Challenges

Describe Some Tools and Approaches We Are Assessing

Discuss Identified Gaps and Recommendations

Next Steps and Questions

## About Us: Chris Prom

Trained as Historian (British Social and Labor History)

Archivist and Andrew S. G. Turyn Professor, University of Illinois

Archival Connections Project, [www.archivalconnections.org](http://www.archivalconnections.org)

Publications Editor, Society of American Archivists

2011 DPC Tech Watch Report *Preserving Email*

Project Partner EPADD Project (Stanford University)

# About Us: Kate Murray

NARA's Transfer Guidance (2014)

Individual messages: XML and MSG

Aggregated messages: PST and MBOX

Sustainability of Digital Formats website (2014-current):

<https://www.loc.gov/preservation/digital/formats/index.html>

NDSA Email Interest Group (2014-2015)

Archiving Email Symposium (2015):

<http://www.digitalpreservation.gov/meetings/archivingemailsymposium.html>

Harvard EAST Workshop (2016)

LC email collections starting to come in

Sponsorship

THE  
ANDREW W.

**MELLON**  
FOUNDATION



Digital**Preservation**Coalition

# Charge

- (a) reexamine and assess current efforts to preserve email;
- (b) articulate a conceptual and technical framework in which these efforts can operate not as competing solutions, but as elements of an interoperable toolkit to be applied as needed;
- (c) construct a working agenda for the community to construct this technical framework, adjust existing tools to work within this framework, and begin to fill in missing elements.

<http://www.emailarchivestaskforce.org>

# Membership

## Executive Committee

Kate Murray (co-chair), Library of Congress

Christopher Prom (co-chair), University of Illinois at U-C

Fran Baker, University of Manchester

Matthew Connelly, Columbia University

Wendy Gogel, Harvard Library Preservation Services

---

Hillel Arnold, Rockefeller Archive Center

Courtney Cain, University of Illinois at U-C

Euan Cochrane, Yale University Library

Kevin DeVorse, National Archives and Records Administration

Glynn Edwards, Stanford University Libraries

Riccardo Ferrante, Smithsonian Institution Archives

William Kilbride, Digital Preservation Coalition

Jessica Meyerson, The University of Texas at Austin

Erin O'Meara, University of Arizona Libraries

Michael Shallcross, University of Michigan

Joel Simpson, Artefactual Systems

Richard Whitt, Google

Julian Zbogar-Smith, Microsoft

# Friends and Consultation Pathways

Council of State Archivists/NAGARA (Webinars)

National Historical Publications and Records  
Commission (NHPRC)

Museums and Web (April 22 in person)

North Carolina Department of Natural and Cultural  
Resources

Coalition for Networked Information (CNI, April  
2017)

Digital Preservation Coalition: UK Briefing Days  
(July and ??)





# Timeline

Date	Task	Status
October 2016	Launch meeting: Identify problem	Completed
November 2016 through April 2017	Working groups meet by phone/email to collaboratively draft text and formulate provisional findings and recommendations	Completed
April 2017	Stakeholder consultations including CNI, COSA (April 18), NAGARA, Museums and Web	Completed
June 30 2017	Draft report released for invited comment	Completed
July 6 2017	DPC Briefing Day to get UK input on early drafts	Right now!
July 2017 through August 2017	Working groups incorporate feedback and continue drafting; Co-chairs and assistant edit draft consultation report	In process
August through October 2017	Open feedback solicited and incorporated into the report	To come
September 2017	Executive Committee Meeting	To come
December 2017	Full task Force meeting in New York to conclude work and approve report	To come
January 2018	Task Force Report Published, DPC Briefing Day	To come

**Task Force on Technical Approaches to Email Archives  
Consultation Report Draft**

**June 30, 2017**

This is a report draft for review, not intended for wide distribution or citation.

Please submit comments on this draft using the form located at  
<https://goo.gl/forms/BTvajRDhnYQs7AKo2>

Sponsored by



f

## Draft Report for Invited Comments

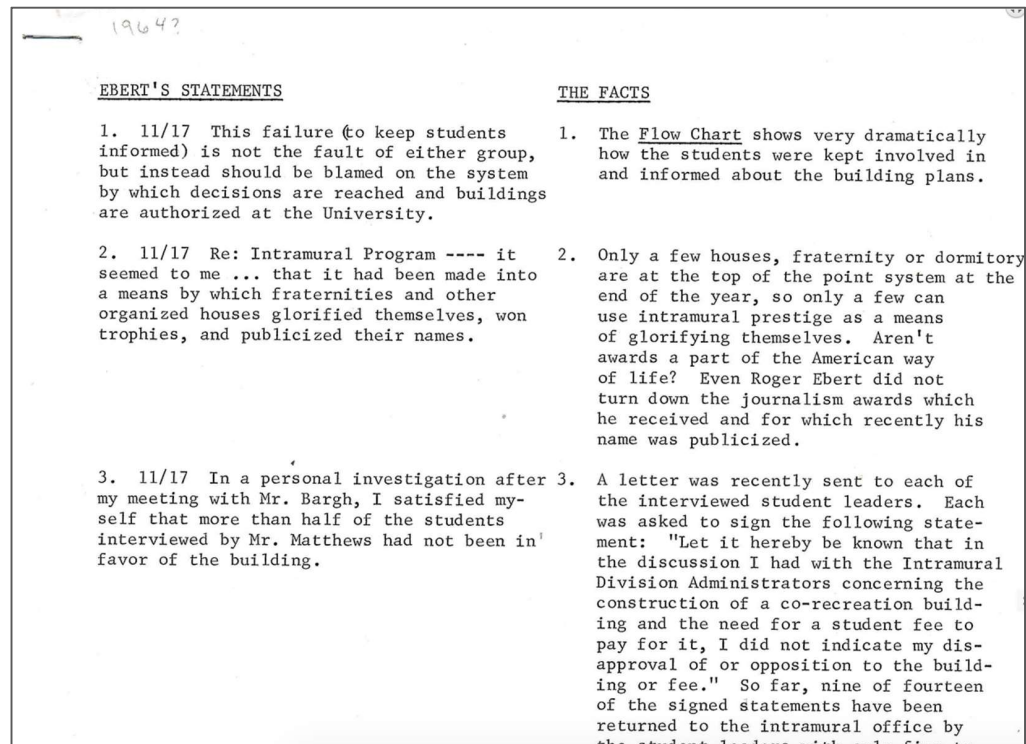
<https://goo.gl/52cpv>

<https://goo.gl/forms/SiC0u5Yx3Y1mtbK62>

# Email as Cultural Documentation

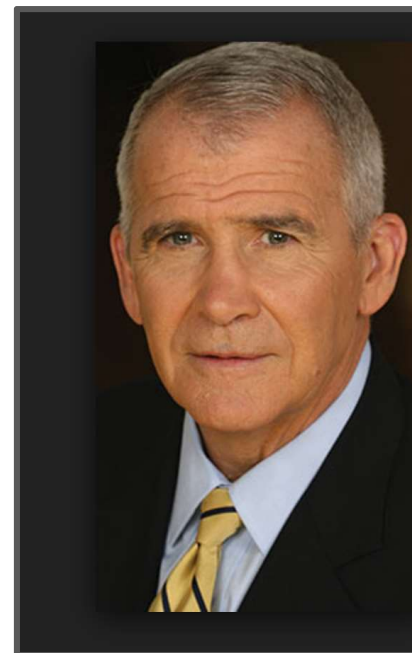
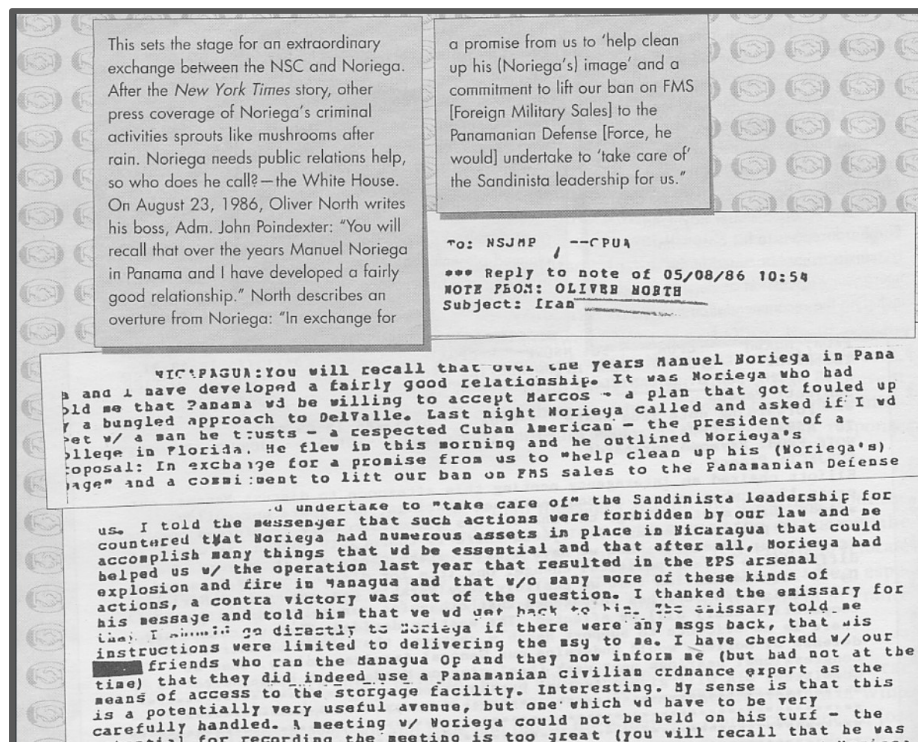


Miriam Blacks Scrapbook,  
Univeristy of Illinois Archives



[archives.library.illinois.edu/ebert/a-student-writer-to-reckon-with/ars-gratia/](http://archives.library.illinois.edu/ebert/a-student-writer-to-reckon-with/ars-gratia/)

# Email and Iran Contra



olovernorth.com



John Poindexter  
historycommons.org

<http://nsarchive.gwu.edu/NSAEBB/NSAEBB113/north07.pdf>

# Email Matters

## Email as a Habitat

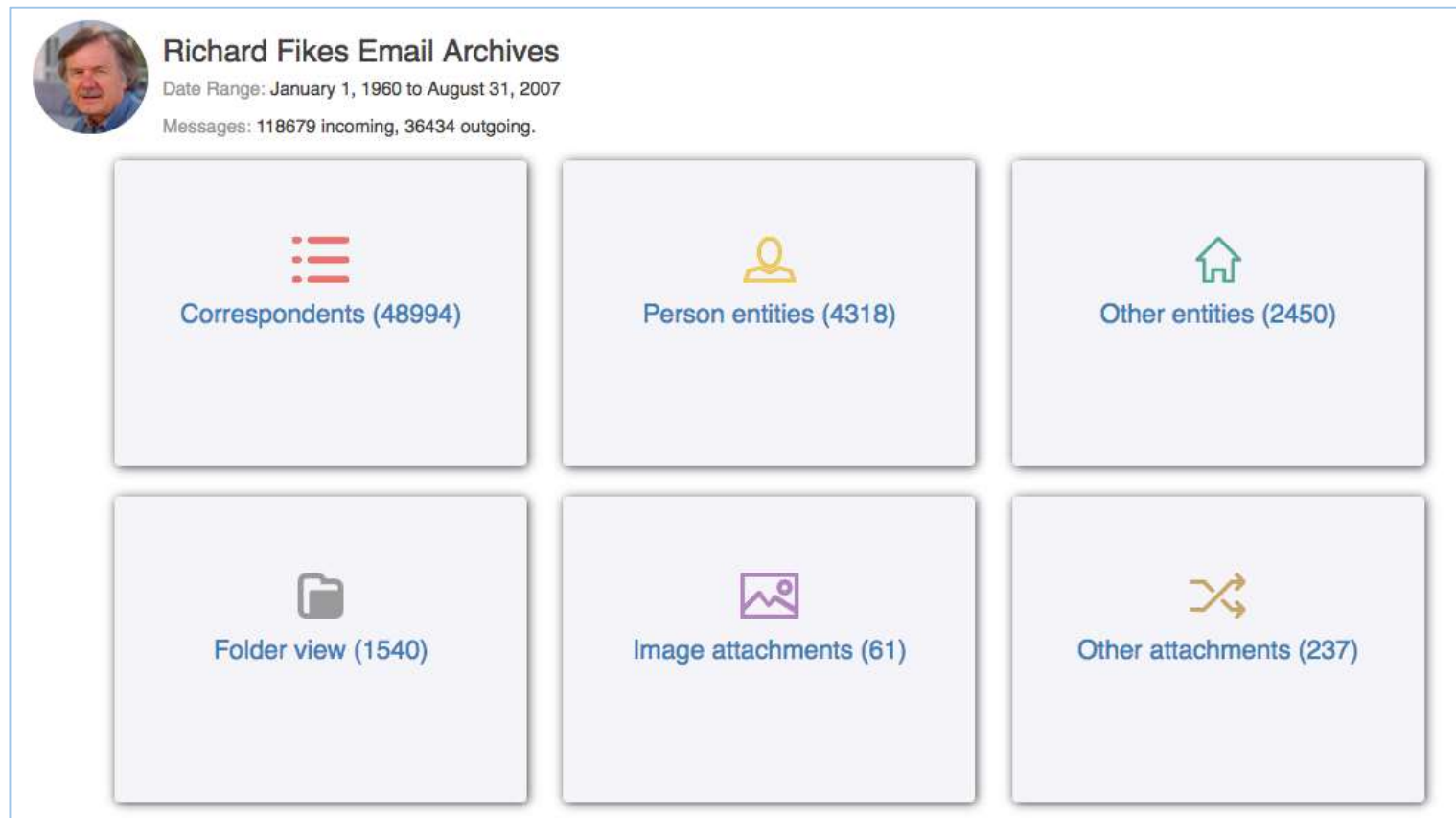
- Email created within an organisation as part of its business, which may be transferred to an in-house archive at the end of its active life.
- Email forming part of archives created by individuals or organisations which are subsequently acquired by libraries or archives (generally referred to as 'collecting institutions') from their original creators.
- Email which is created and preserved by private individuals as part of their own, or their family's, digital archive.
- Email created and curated by informal groups or societies outside of a corporate recordkeeping structure.

# Carl Woese



Evolutionary Microbiologist at the  
University of Illinois  
Revolutionary methods to  
understand characteristics of  
organisms--conservation of  
ribosomal DNA fingerprints  
Discovered the third domain of  
life: Archaea

# Email is both Record and Cultural Documentation





# Email is a Many-Splendored Thing

The term “email” is used to describe:

- An individual electronic mail message

- All of the messages ever sent and received by an account

- The act of composing, sending or reading messages

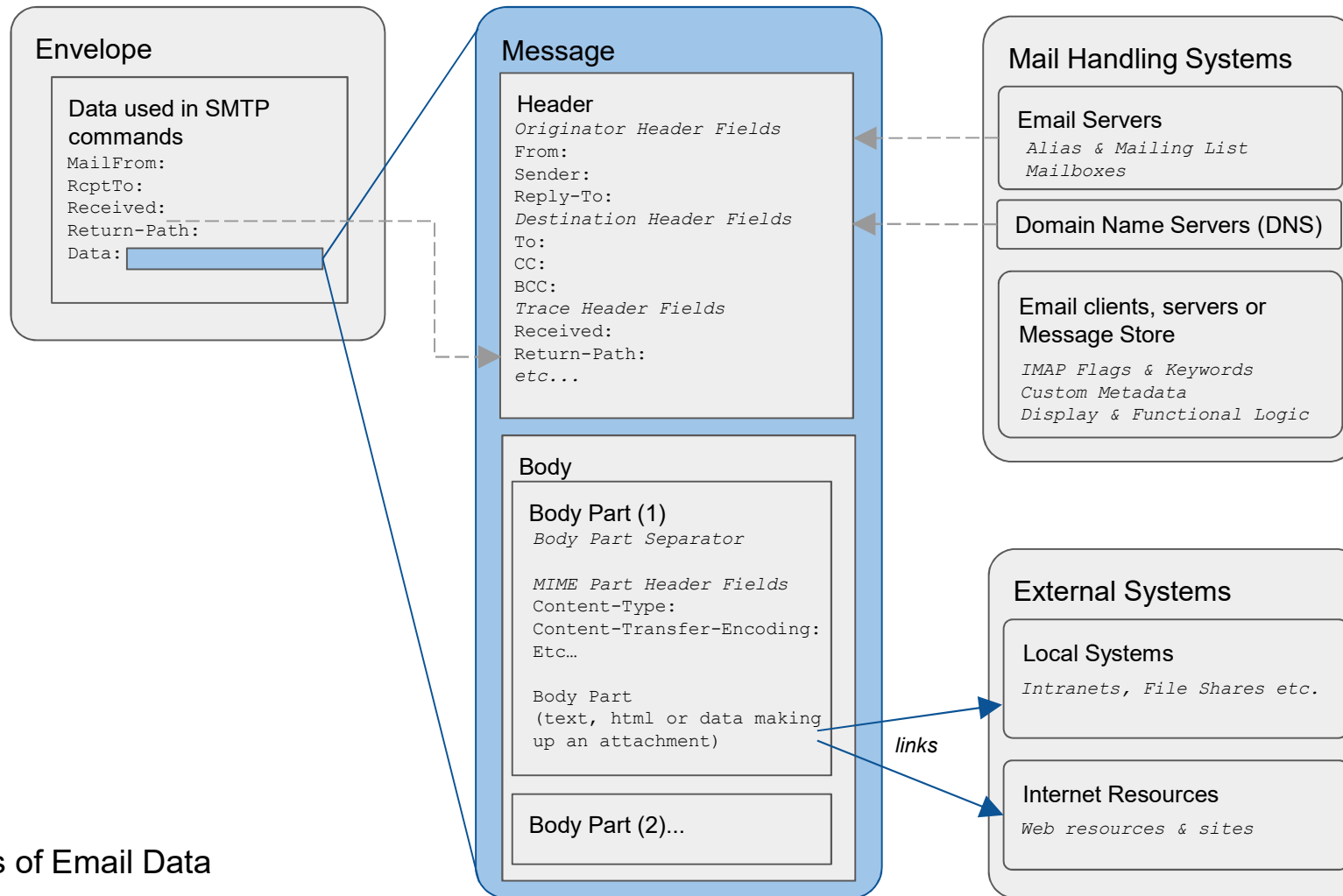
- Applications and systems that are used to transport, receive, and manage messages

Personal information manager (PIM) tools include email, internet chat, calendars, contacts and task all in one application

## Email is...

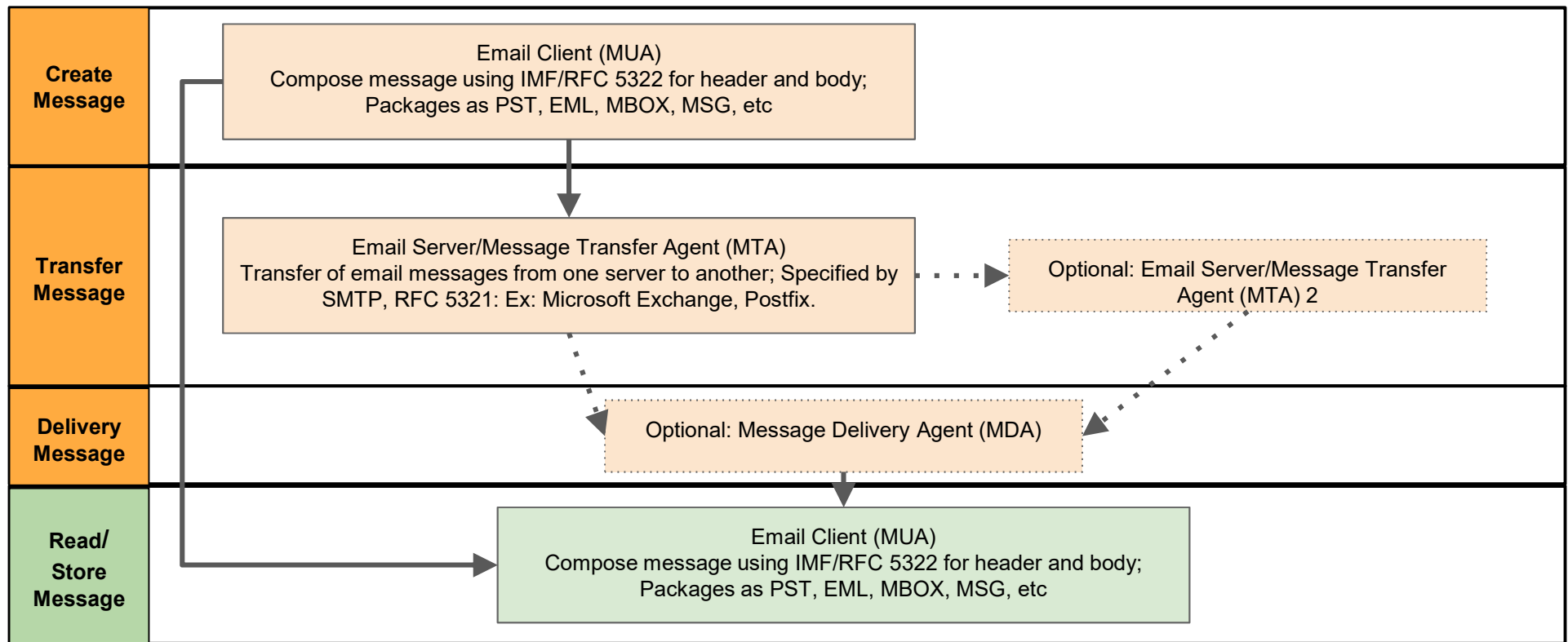
At its heart email is transactional process a sender transmits a message to a recipient.

From a technical perspective, It is useful to think of email as a protocol that defines a series of commands and responses that operate in a manner not unlike a computer programming language and which permits mundane email processes to occur.



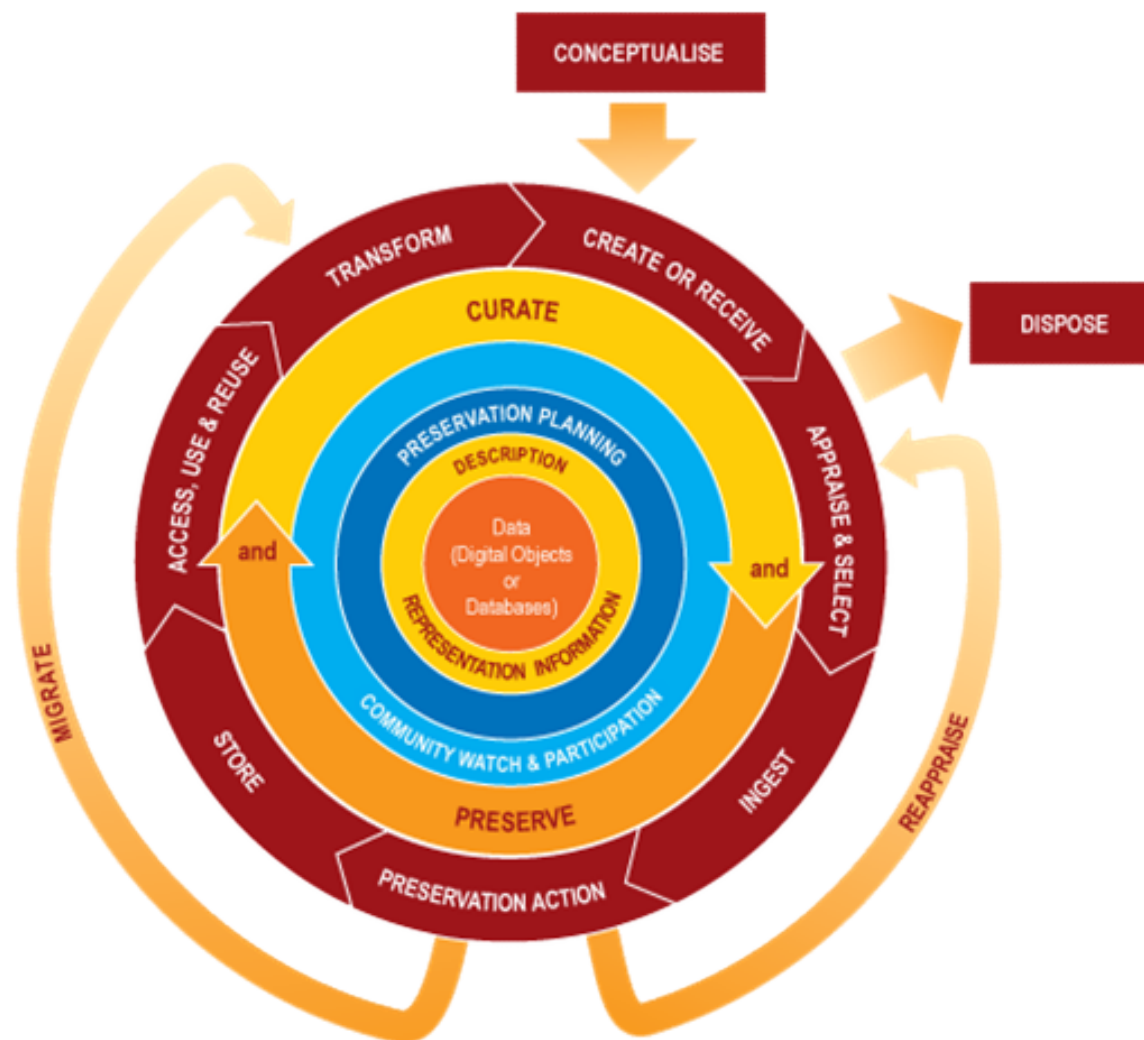
Types of Email Data

# Standardization Across Systems



# DCC Curation Lifecycle Model

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>



# Email Lifecycle



# Technical Challenges

Capture

Ensuring Authenticity

Working at Scale

Addressing Security Concerns

Messages and Formats

Preserving Embedded Attachments  
and linked documents

Tool Interoperability



**Student at Lincoln Hall in Wheelchair, ca.  
1957-1967 Division of Rehabilitation  
Services: Director's Office Photograph File,  
University of Illinois Archives**

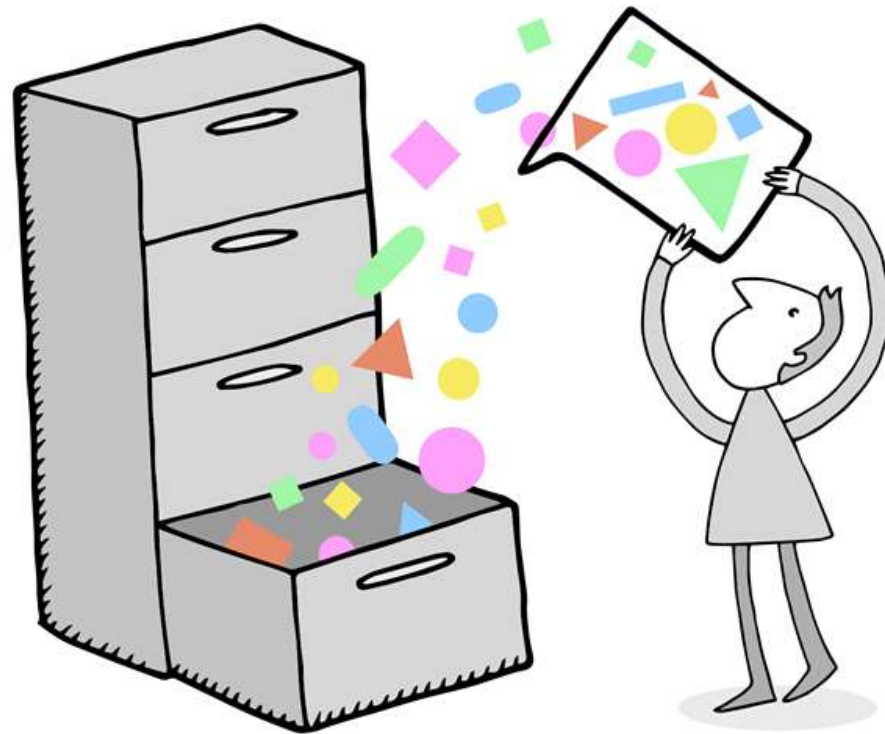
# Capture

Building trust

Aggregating data

Workflows

Chaining tools



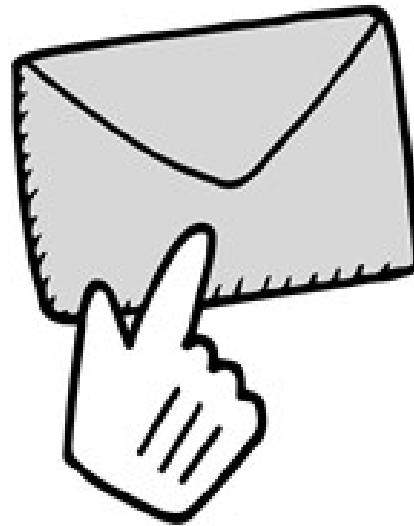
Credit: Digital Preservation Business Case Toolkit

<http://wiki.dpconline.org/>



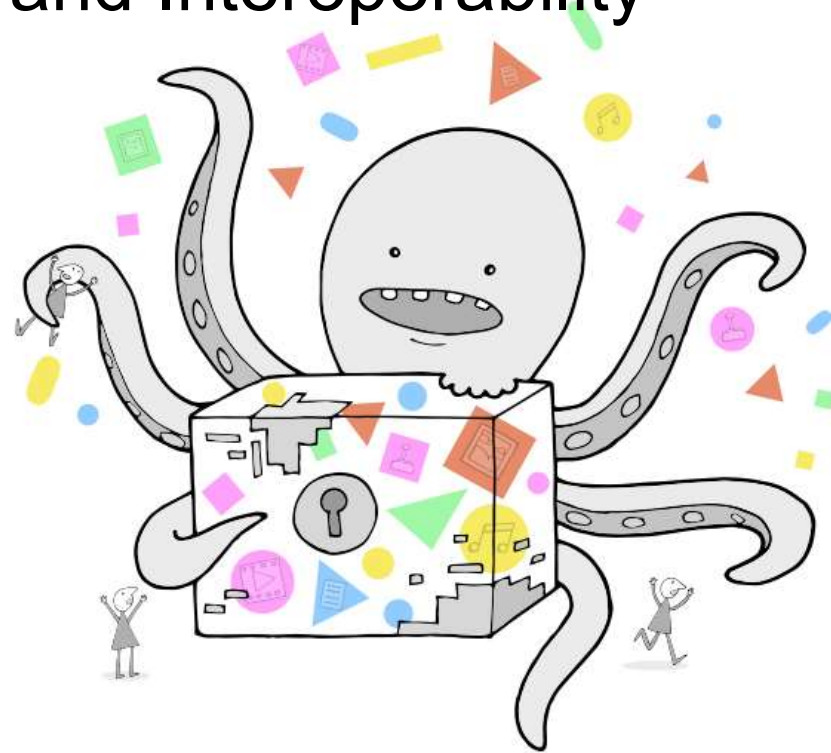
# Capture and Authenticity

- Current Capture Methods rely on stored copies
- Envelope, Header Fields, and Message Data



# Accessioning, Processing and Interoperability

- Significant properties?
- Semi-standardized workflows?
- API development?
- User experience?



Digital Preservation Business Case Toolkit

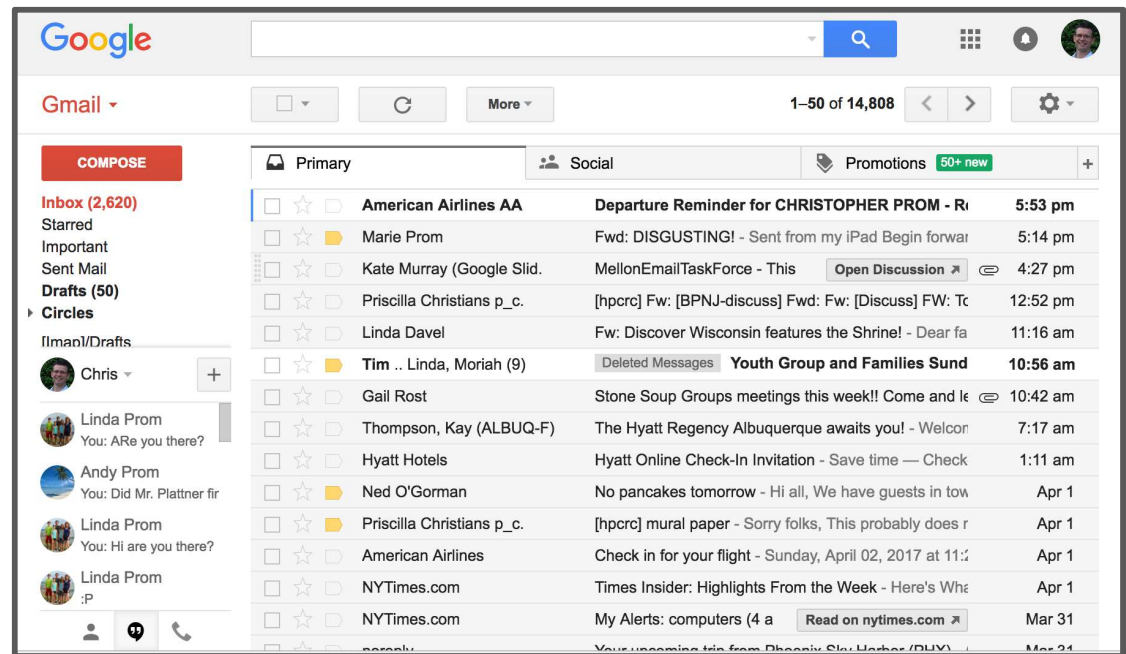
<http://wiki.dpconline.org/>

# Ensuring Authenticity

Ingest methods matter

Key functions and  
metadata stored in  
systems, not outputs

Preserving the chain of  
actions that  
archivists take



# Attachments/ Linked Content

**Matthew J. Connelly**

To: Christopher John Prom Cc: Murray, K  
Re: Mellon Email Task Force - checking in



New contact info found in this email: Mat

Hi Chris and Kate:

Here is a link to the draft of what I have so far for the report:

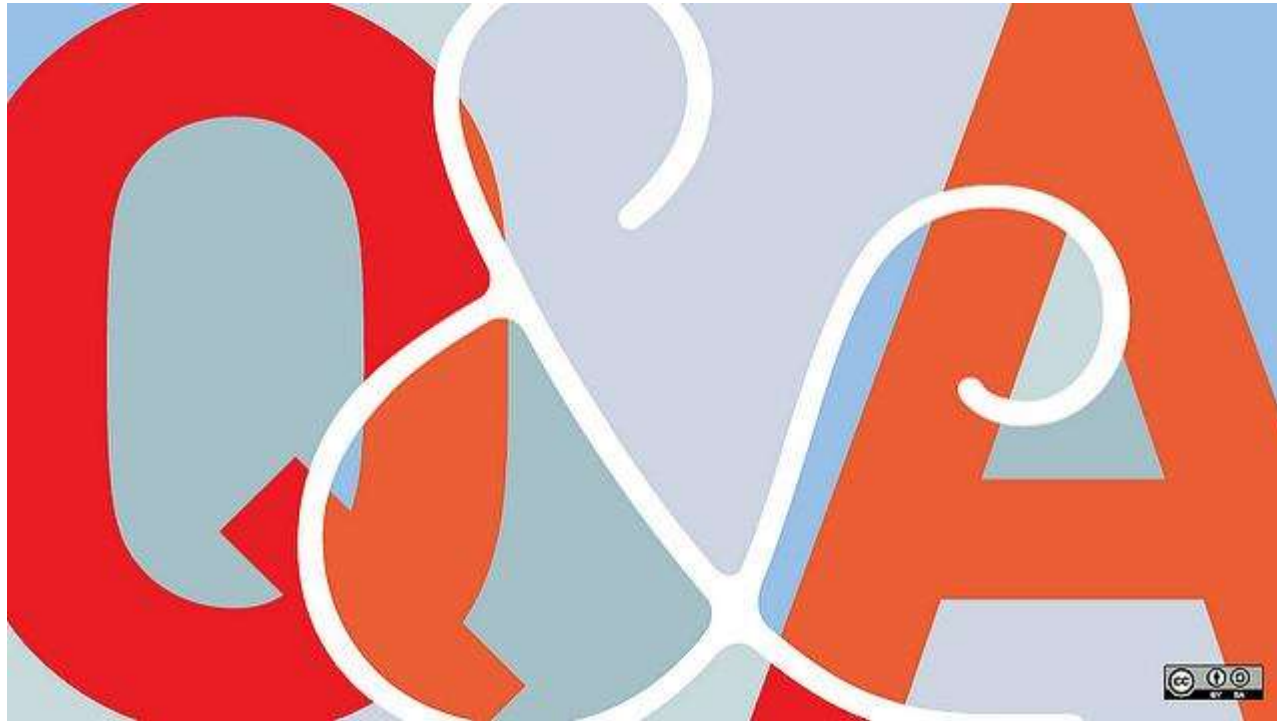
[https://docs.google.com/document/d/1xrA1RmFq57sGgjsfd0HUibMgda\\_m8q9-H5vnNCxpVjA/edit?usp=sharing](https://docs.google.com/document/d/1xrA1RmFq57sGgjsfd0HUibMgda_m8q9-H5vnNCxpVjA/edit?usp=sharing)

It just covers what's outlined below so far, but I'll try to make more progress tomorrow on the rest of it. can you remind me of how many words you would want my section to be?

Best,  
Matt

o Topics:

In November 1986, a senior official at the National Security Council sat down at his White House computer, entered his password, and started deleting e-mail. It seemed worth it at the time, because Lieutenant Colonel Oliver North was trying to cover up a vast conspiracy. He believed they had safely destroyed the evidence. The National Security Adviser, who had gotten a Ph.D. from CalTech, knew his e-mail was being backed up. But Poindexter also knew that the magnetic tapes were regularly recycled and would soon be overwritten. He therefore joined with North in writing what was meant to be the authoritative account of what happened. It depicted an earnest effort to free long-suffering hostages and achieve a diplomatic breakthrough with Iran -- with nary a word about the Contras. They called it a "Historical Chronology."



Credit: <https://www.flickr.com/photos/opensourceway/5556249000>