

# Preserving Emails

**Maureen Pennock**  
**Digital Curator, The British Library**

Email, as a technology, is now 40 years old



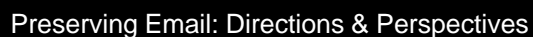




After 40 years and that many millions of messages, you'd think we'd have a solution in place.

# Well we have some solutions:

- David Bearman proposal (1994)
- Digital Preservation Testbed Email preservation recommendations & email/xml plugin (2003)
- XENA Email normalisation (2005?)
- DCC Curation Manual instalment (2006)
- Significant properties of emails (2008)
- CERP Email Preservation Parser (2008)
- PEDALS Email extractor (2010)
- DPC Technology Watch (2011...)



And emails are being archived...

# 1: ENRON



- FERC set
  - Around 1.5 million emails
  - 92% of staff emails
  - Plain text
  - PDFs generated on demand
  - Includes attachments
- Cleaned set
  - Around 500,000 emails
  - 150 users
  - Data integrity problems
  - No attachments
  - Some deletion/redaction



## 2: Warren Spector Collection

- Video Game archive @ University of Texas, Austin
  - Over 56,000 emails
  - Over 6,000 attachments, 50 formats
  - \*.pst files downloaded into DSpace (May 2009)
  - Viewed using Outlook 2003 and other 'appropriate' software
  - XML schema and examples to be worked on.
- <http://bit.ly/pBrDqU>

## 3: Sarah's Inbox

[←](#) [→](#) [C](#) sarahsinbox.com

---

A project of **Sunlight Foundation** ([Who is the Sunlight Foundation?](#))


Join Our Mailing List  Your Email  Your Zip

---

[Home](#)

**Sarah Palin** | [Sign out](#)

---



**Sarah's Inbox**  
A project of the Sunlight Foundation

---

**Inbox**

Sent Mail

Drafts

All Mail

Starred (all)

Starred (by you)

**Sample Searches**

flippinbelieveit

"bridge to nowhere"

"Tina Fey"

"horrible people are bringing you down"

"Alaska natural gas pipeline"

"barack obama"

crud

"first dude"

"who leaked it?"

"What a dumbass he is"

"who's going to trim my hair?"

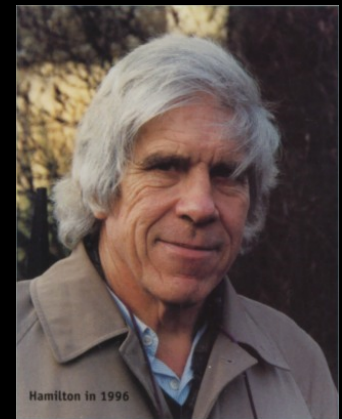
pizza

★	From	Subject	Date
★	Gov. Sarah Palin, Kelly C Goode (GOV), KelyC (GOIIO)	Keller asked to resign	07/01/2009
★	tpalini@mccain08hq.com	PERS and SBS investment funds	10/06/2008
★	'Palin, Gov. Sarah Palin, Kris Perry	Governor's Press Release	09/30/2008
★	Gov. Sarah Palin, Palin	Governor's Press Release	09/30/2008
★	Brad Fluetsch, Gov. Sarah Palin	Give Alaska Housing Finance \$\$\$\$	09/26/2008
★	Gov. Sarah Palin, Kris Perry , Michael A Nizich (GOV), Palin	FW: Ninth Circuit Chistochina decision - CONFIDENTIAL ATTORNEY-CLIENT PRIVILEGED	09/23/2008
★	Gov. Sarah Palin	Legal and lawsuits	09/22/2008
★	Beth Leschper, Bristol Palin, Gov. Sarah Palin, mbkrdk@starband.net	Useful info for Alaska	09/20/2008
★	Gov. Sarah Palin	from Dr. Laurie Roth www.therothshow.com	09/19/2008
★	Gov. Sarah Palin	BBQ at my house manana with the Strange Boys playing!	09/18/2008
★	Gov. Sarah Palin (2)	baby gift from Maya Wrap	09/18/2008
★	Todd Palin	New email address not working	09/18/2008
★		FW: Tesoro Iron Dog, Flat Stanley had the scoop	09/18/2008
★	Gov. Sarah Palin, Kris Perry , Michael A Nizich (GOV), rej (2)	A Difficult Decision	09/17/2008
★	Gov. Sarah Palin	Delivery Notification : Delivery has failed	09/17/2008
★	Beth Leschper, Bristol Palin, mbkrdk@starband.net	Spread the word	09/17/2008
★	Beth Leschper, Bristol Palin, Chuck Heath, mbkrdk@starband.net	shame	09/17/2008
★	Gov. Sarah Palin, Lynne Smith, Michael A Nizich (GOV)	FW: Call from Congressman Young 's Office / Fort Yukon fuel issue?	09/17/2008
★	Gov. Sarah Palin (4)	Delivery Notification: Delivery has failed	09/17/2008
★	Cora Crome, Gov. Sarah Palin, Michael A Nizich (GOV), St...	Exxon Mobile	09/17/2008

**About**

This site parses the emails sent and received by Sarah Palin while she was governor of Alaska and presents them in a more familiar interface.

# 4: Email Preservation at the British Library



# BL: Enhanced Curation





# Wendy Cope email archive

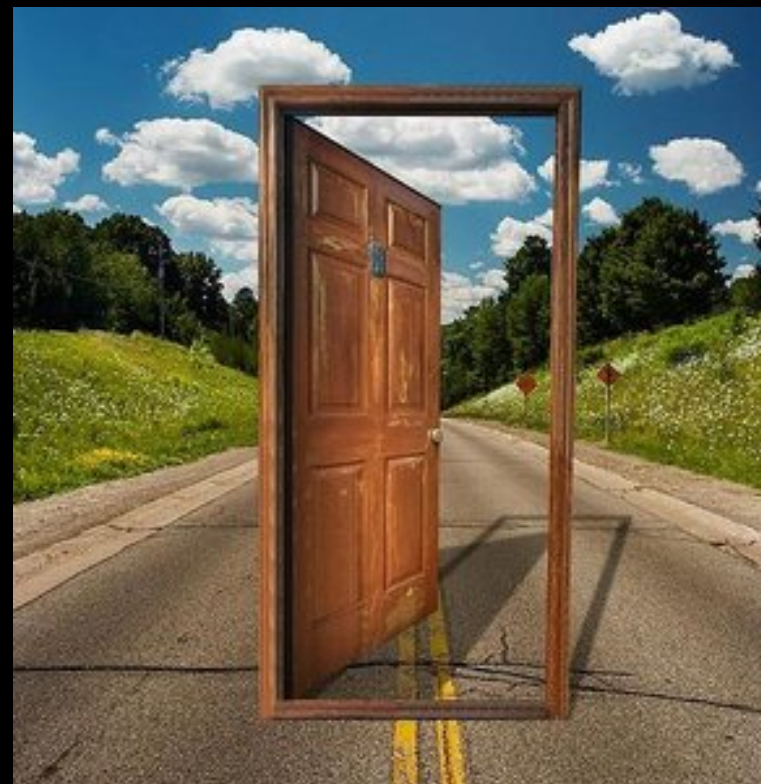
- Yahoo email account
  - Some IMAP, some POP3
- Captured via owners PC
  - Captured inbox in entirety
  - Includes attachments
  - I'm not sure about address book?
- Mix of content: some personal, some professional
  - Important to understand context of emails

# Wendy Cope email archive

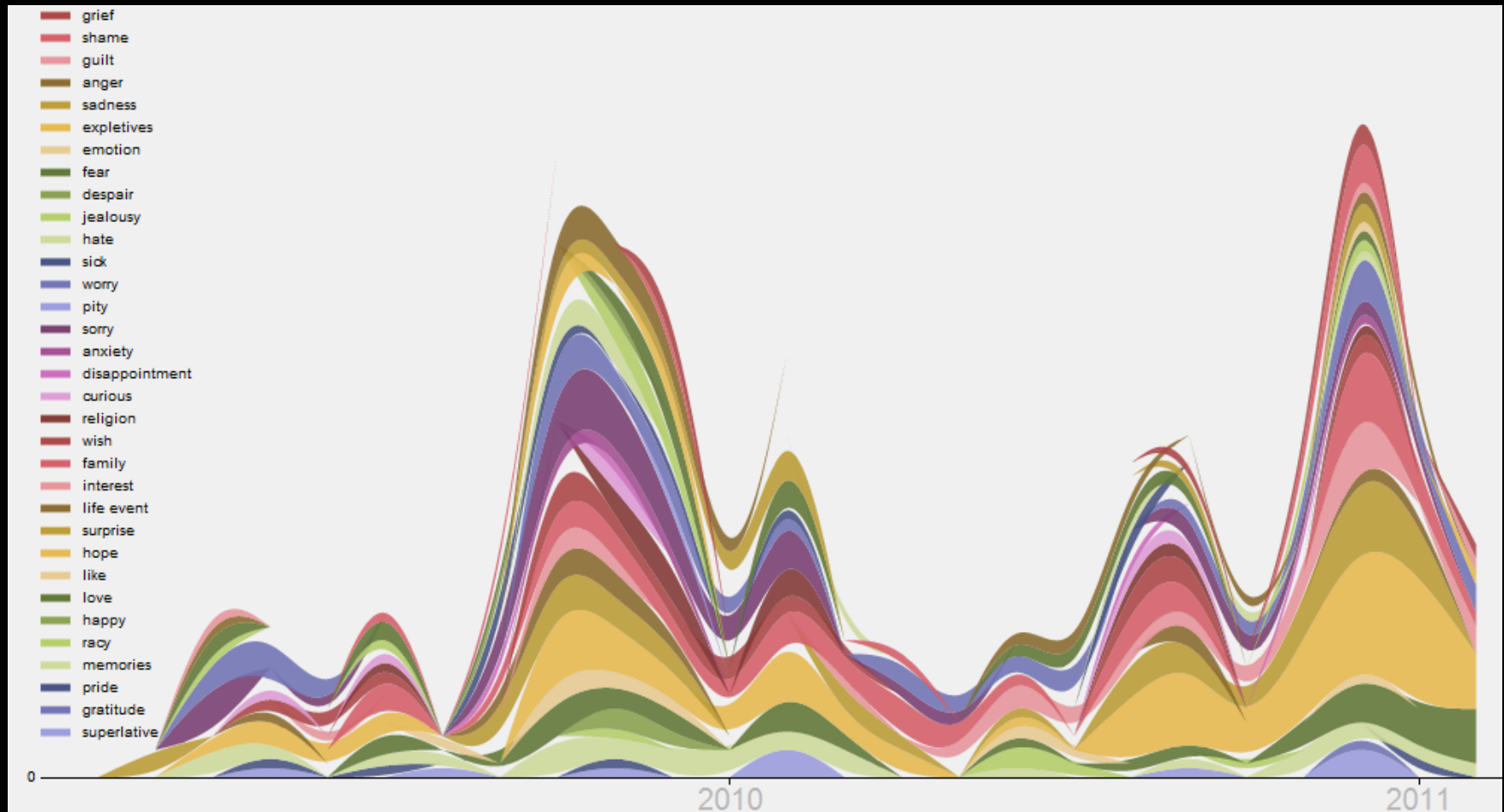
- Experimenting with various tools in eMSS lab:
  - CERP Email Preservation Parser
  - Aid4Mail
  - Emailchemy
- Anticipating some XML & PDF conversion
- Issues around copyright (incl. third party)
  - Access via reading rooms

# What of the future?

- Long-Term Preservation?
- Integrated social media accounts?
- Access?
  - Inbox interface
  - Data mining
  - Visualisations

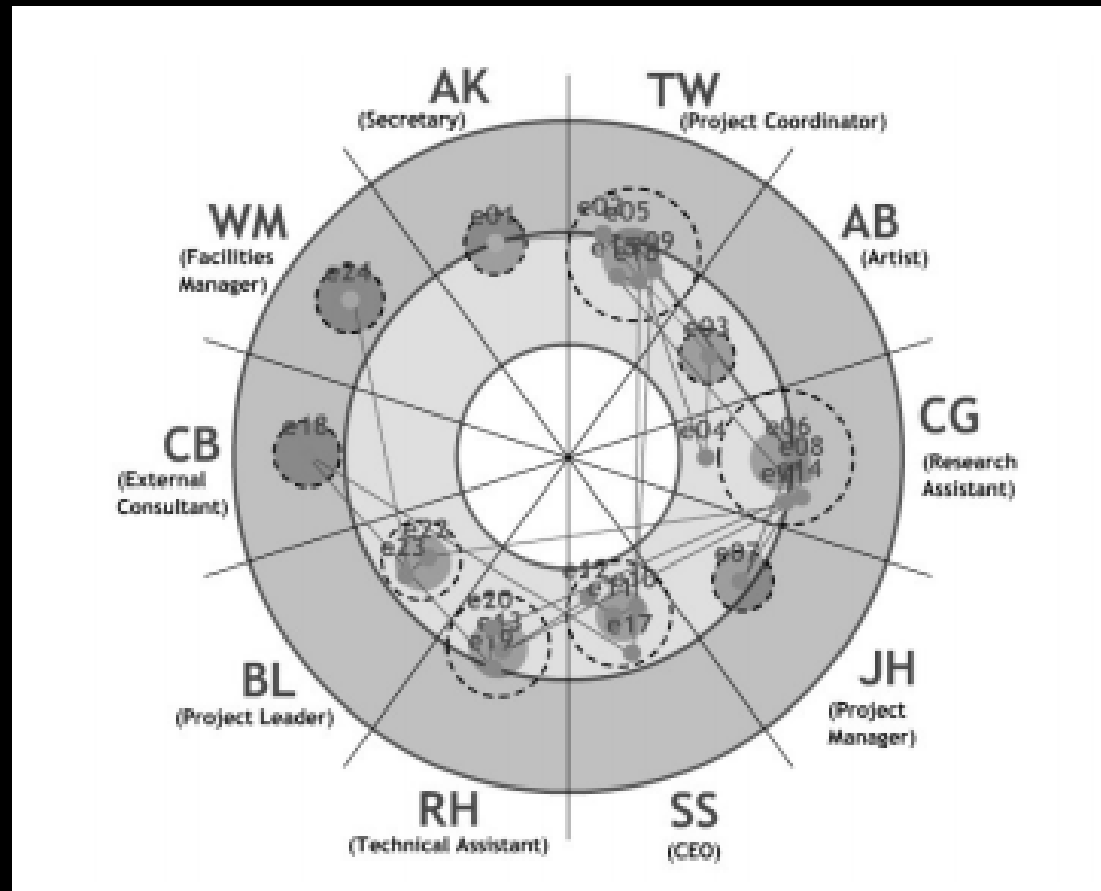


# Inbox Visualisation: Muse





# Inbox Visualisation: Collaborations



# Preserving email?

*‘Just Do It’*

# Image credits

- Brick wall: [http://sv.wikipedia.org/wiki/Fil:Solna\\_Brick\\_wall\\_vilt\\_forband.jpg](http://sv.wikipedia.org/wiki/Fil:Solna_Brick_wall_vilt_forband.jpg)
- Birthday cake:  
<http://jomkepandan.blogspot.com/2011/05/happy-birthday-candle.html>
- 60 seconds infographic:  
<http://www.infographicsarchive.com/tech-and-gadgets/things-that-happen-on-internet-every-sixty-seconds/>
- Wendy Cope panoramic:  
<http://pressandpolicy.bl.uk/Resource-Library/Wendy-Cope-s-study-3d8.aspx>
- Open door:  
<http://maryjaneryan.wordpress.com/2009/06/01/one-door-closes%E2%80%A6/>
- Muse visualisation:  
<http://litbrarian.wordpress.com/2011/03/20/reaching-out/>
- Collaboration visualisation:  
Visualising Collaboration via Email: Finding the Key Players (2006)