



Email Preservation Workflow

Stage	Components	What we do	Method
TRANSFER	Source E-mail	Receive	
	Metadata – Administrative & Descriptive	Document transfer and object metadata; make back up copy	Manual
SIP	<i>At this point we have a fully defined SIP</i>		
	Preservation/Conservation	ASSESS/RISK ANALYSIS: what formats, what media, obsolescence risks, viruses	Manual - Virus detection: AVG, Symantec, etc.; - Attachment extraction: EZDetach Partially automated - Attachment format identification: JHOVE, DROID
	Metadata – Administrative	Formalize assessment results	Manual
		<i>Optional</i> – Process to produce specialized DIP with sensitive data redacted (highly desirable but not intrinsically necessary for AIP)	Manual
	Metadata – Descriptive	Draft narrative finding aid	Manual
	Preservation	Complete preliminary transformation of grouped messages to MBOX format	Manual MessageSave, Aid4Mail
	Account.XML	PARSE (automated); validate	Automated CERP parser; oXygen XML Editor
	Preservation – MBOX Directory Tree (Parser Directory Tree)	Group attachments, bad messages, and message summary files within the existing MBOX directory tree and zip	Automatic from parser Manual zip
	Metadata, Descriptive – Directory tree of subject-sender logs	Group subject-sender logs (index) and zip	Automatic from parser Manual zip
	Metadata – Administrative	Finalize metadata including Preservation Description Information (PDI)	Manual
	METS	Describe AIP in METS schema and generate METS file. Duplicate METS file for loading into DSpace	Manual
	Metadata – Descriptive	Produce completed finding aid	Manual
	XSLT – display stylesheet	Include display template for access to preserved account	Manual
AIP	<i>AIP = all components above</i>		



TRANSFER

Source e-mail
Basic accessioning metadata

SIP

Data Management

Initial processing
Backup
Virus check
Preservation assessment (physical/file/technical via JHOVE, DROID)

Preservation/Conservation

Preliminary transformation – source format to MBOX format
Final transformation (parsed) – MBOX format to XML

Metadata – Initial

Metadata – administrative with preservation assessment
Metadata – descriptive including narrative Finding Aid

AIP Assembly

Source e-mail (.pst, .msg, etc.)
Preservation e-mail
 Parsed messages (XML)
Parser Directory Tree.zip
 Pre-parsed messages (mbox)
 Bad messages
 Message Summary
 Attachments
Subject-sender log.zip (from parser output)
Metadata narrative.zip – complete (administrative & descriptive)
Finding aid narrative.zip
METS.xml
File Name METS.xml (administrative & descriptive metadata encoded in METS)
Display aids
 XML stylesheet

AIP