

PRESERVING EMAIL ATTACHMENTS WITH ATTACHMENT CONVERTER

Matt Teichman

*Digital Library
Development Center
University of Chicago Library
USA
teichman@uchicago.edu
<https://orcid.org/0000-0002-9637-6613>*

Ashley Gosselar

*Hanna Holborn Gray Special
Collections Research Center
University of Chicago Library
USA
agosselar@uchicago.edu
<https://orcid.org/0009-0000-4172-7618>*

Obi Obetta

*Digital Library
Development Center
University of Chicago Library
USA
obettasteph@uchicago.edu*

Nishchay Karle

*Digital Library
Development Center
University of Chicago Library
USA
nishchaykarle@uchicago.edu*

Keywords – email, MIME, attachment, digital preservation, digital formats

Conference Topics – Sustainability: Real and Imagined ; From Theory to Practice

I. WORKSHOP CONTENT

Email increasingly forms a core portion of many archives, and while tools are available for basic preservation of email messages, less consideration has been given to archival preservation and control of email attachments. It is essential to future research that these important records are preserved in the context of their associated messages.

With the generous support of an “Email Archives: Building Capacity and Community” grant from the University of Illinois, the University of Chicago Digital Library Development Center has created a free and open-source tool called Attachment Converter. Attachment Converter utilizes file format conversion utilities already installed on a user’s system to batch-convert common email attachments to formats recommended for archival preservation and access, retaining the connection between the migrated attachments and their associated emails.

In this workshop, iPres attendees will learn the basic backend anatomy of an email, how to convert email from PST to MBOX, and inspect and analyze the contents for common attachments. We will demonstrate how to migrate attachments to preservation formats using Attachment Converter and discuss how archivists without UNIX/Linux experience can collaborate with their institution’s IT staff to implement the tool.

II. WORKSHOP AGENDA

The workshop will have two parts. First, we will explore the basic technical components of email and share techniques for working with them. The full email specification is too complicated for a 90-minute workshop, but we will provide everything that an archivist needs to know for the purpose of working with email attachments. The goal is for all attendees to walk away with the ability to convert email from Outlook PST to MBOX format, open the MBOX in a text editor, and inspect the raw contents of the mailbox, regardless of their technical background. Second, we will teach attendees how to set up and use Attachment Converter.

The workshop is designed to be interactive and attendees are encouraged to follow along on their computers using a sample mailbox that will be provided to them. The sample emails will contain attachments in file formats archivists may commonly wish to preserve: JPEG, GIF, PCX, PDF, DOC, DOCX, RTF, XLS, and XLSX. We will show participants how to open a mailbox in a text editor and identify where the attachments are in an email, how to guess the file format of an attachment, and also explain why that has to be a guess. We will then demonstrate how to convert all the attachments in a single email to archivally stable formats using Attachment Converter. Following that, we will show how to batch-convert attachments in an entire mailbox.

Finally, we will explain how to set Attachment Converter up to use a utility of the user's choice to convert attachments in their emails. This process requires more technical expertise---so we won't go into full detail---but we will explain that feature at a high level and stress that any archivist working somewhere with an IT staff that knows UNIX/Linux system administration should be able to make use of this feature.

III. LEARNING OBJECTIVES

At the end of this workshop participants will be able to:

- Understand the structure of an email;
- Convert email from Outlook PST to MBOX format;
- Open and analyze MBOX email in a text editor;
- Install and run Attachment Converter on a sample MBOX (provided);
- Communicate with IT staff about configuring Attachment Converter for local needs.

IV. WORKSHOP REQUIREMENTS

This will be a 90-minute virtual workshop. Students will need the following:

- A computer (Mac or Windows) and internet connection, and the ability to download files from Box;
- A text editor (Notepad or Text Edit);
- Optional advanced: If students want to follow along, they will need admin rights on their

computer, and will need to install Homebrew (Mac) or WSL Ubuntu (Windows). Installation guides will be provided beforehand via Box.

Attachment Converter will be demonstrated in a UNIX environment. Knowledge of how to use UNIX is not required for this workshop, but basic knowledge of command line would be useful for understanding the demonstration.