EABCC

Building Capacity & Community

The University of Illinois Library is pleased to announce that three new institutions were recommended for funding under the *Email Archives: Building Capacity and Community* (EA:BCC) program. In the second round of grants, the University of Maryland, the University of North Carolina at Chapel Hill and the 92nd Street Y were among those who received funding.

The EA:BCC is a multi-year initiative that brings together, supports, and funds a growing network of institutions and professionals working to develop critical solutions for preserving email. Through two rounds of funding and nearly $650,000 allocated across eight institutions, the EA:BCC program continues to expand the email archiving network and provide resources and solutions for continuous development.

In this issue of the newsletter, you can read about the new programs that have been recommended for funding and get updates on the institutions that were funded in the first round.

ROUND TWO AWARDEES

University of Maryland ($56,949.96)

The University of Maryland will support construction of an online environment in which we will observe "digitally curious" scholars accessing reference email collections using a novel, context-sensitive discovery prototype: EMail CONtextualisation DIScovery Tool (EMCONDIST). These user engagements will occur both online and in-person. In addition to observing these interactions, users will be surveyed about their experiences. These activities will be transcribed, summarized, analyzed, and reported to the relevant user communities and to the email archive community to improve the discovery and accessibility of email archives.

University of North Carolina at Chapel Hill ($87,716.81)

Focusing on important email-related curation use cases, the University of Carolina Chapel Hill aims to enhance software development through the Review, Appraisal and Triage of Mail (RATOM) project. The output of the software is designed to facilitate a wide range of curation activities, including review for sensitivity, appraisal and response to open records requests. With its suite of powerful tools, the RATOM Functional, Interoperability and Reuse Extensions (RATOM FIRE) project will allow easier export of email messages as individual (EML) files; capturing more detailed preservation metadata; and expand the public application programming interface (API) of the RATOM software library to facilitate easier integration into other tools.

92nd Street Y ($100,000.00)

A current count puts the 92Y's Poetry Center's email archive at almost 3 million messages. A collaboration between the Poetry Center and the digital-archive team at Stanford University Libraries, this project will apply ePADD to the assessment and preservation of the email archive with the goal of developing a processing and accessibility model that other cultural centers might learn from and adopt for their own internal and external purposes.

ROUND ONE AWARDEE UPDATES

UNIVERSITY OF ALBANY, SUNY

Mailbag, a functional specification based on the BagIt specification, was developed at the University of Albany, SUNY. A crucial aspect of Mailbag is its ability to preserve many formats for email archives in an organized and practical manner. Email will be captured rapidly and saved in a stable format for subsequent processing and access by archivists.

As part of this project, the team also developed mailbagit, a command-line application for packaging email outputs into mailbags that follow the mailbag specification, while also creating many derivative formats such as PDFs, WARCs, and EMLs. On their project website, you can find out how to install and use mailbagit.

HARVARD UNIVERSITY

ePADD, an open source email archiving software package that already provides archival appraisal, processing, discovery, and delivery, will be updated to include long-term email preservation features. With Integrating Preservation Functionality into ePADD (ePADD+), the functionality of ePADD will be enhanced by Harvard and its project partners, the University of Manchester and Stanford University.

During the first phase of ePADD+, the Manchester development team was successful in meeting the requirements for the new preservation export bag from ePADD. As the project nears completion, the team will be improving code documentation and the ePADD User Guide, as well as putting together training and outreach materials, such as pre-recorded demonstrations of the new features. Join their mailing list to keep updated.

UNIVERSITY OF CHICAGO, LIBRARY

While it is generally recognized that archival practice requires the preservation of electronic documents in their original form, email attachments are typically preserved as separate files and are in a large range of proprietary formats. The Attachment Converter: Preserving the Context of Electronic Correspondence from the University of Chicago, Library.

With the Attachment Converter, the University of Chicago Library will develop software to convert email attachments in all formats into the recommended preservation formats. In the next steps, UChicago will begin testing its email library collections to implement a GUI. To facilitate this work, they have been exploring the concept of using a configuration file in order to make it easier to create output files that are not much different from the original source emails.

COUNCIL OF STATE ARCHIVISTS, INC.

In establishing the groundwork for a nationwide approach across state and territorial archives, the CoSA Prepare project from the Council of State Archivists sets the stage for digital preservation strategies to include email preservation. This multistate initiative to coordinate the preservation of government email will see the development of a suite of resources that will help states determine the most appropriate approach for email processing.

In September 2021, CoSA published their Needs Assessment Survey Analysis & Report that provides insights into the current trends and resources that can be the most beneficial. In addition, CoSA developed resources for determining an email appraisal approach and developing government email preservation policies that can be helpful to those who are just getting started. Those resources can be found here.

COLUMBIA UNIVERSITY

Email is ubiquitous, and yet it's a challenge to preserve and make available for research. Columbia University's "Creating Email Archives from PDFs: The Covid-19 Corpus", addresses both needs by creating an open-source tool that takes emails embedded in PDFs as input and generates an MBOX file format. The draft documentation and prepare codebase for dissemination is publicly available here.

At this point, more than half of the records from the Documenting Covid-19 corpus have been processed, including the extraction of metadata from email fields (from, to, date, etc.) and generated names of persons, places, organizations, and topic words. Currently, Columbia is working extensively to address personally identifiable information (PII) discovered in the processing of the Covid-19 corpus.

STAY TUNED

In our following newsletter, we will provide the results from a regional survey that assesses the state of adoption in Illinois. Currently, little is known about the extent to which institutions are using email archiving software, or what needs they would like to address in future software or program development. We anticipate that the outcomes of this study will be helpful to EA:BCC grantees and others in the community, increasing the number and variety of institutions that archive email.

LEARN MORE

HTTPS://EMAILARCHIVESGRANT.LIBRARY.ILLINOIS.EDU/