iPRES2021 Proposal Template Instructions

This is the DOC/DOT template for Paper abstracts, Panels, Poster, Demonstration, Workshop and Tutorial proposals for iPRES2021. This template is loosely based on the IEEE template. If you are writing a full paper, please use any of the papertemplate options instead.

Open the .dotx file to use this template in Word / LibreOffice / OpenOffice.

- This document uses the **Open Sans** font family.
    - When you're on Windows, this font should be included in this template document.
    - On UNIX systems you have to make sure the system has this font installed. You can download the TTF here.
- Remove this preamble text and replace the body text with your intended output.
- Make sure to use the **Styles Pane** (List: styles in current document) or the **Styles Gallery** to format your document elements (titles, author info, headers, references)
- The author info is organized in a table.
    - If your submission has less than three authors, remove the unnecessary columns to space out the authors info evenly over the page.
    - If your submission has more than three authors, duplicate the three rows below the existing one to add extra information
    - In the case where there are 4 authors, remove one column to end up with 4 x 3 cells).
- Make sure to 'Capitalize Each Word' in your headers
- Before exporting: make sure to fill out the Document Properties (*File* > *Properties*)
- When exporting: make sure to
    - embed the font [how-to for Windows]
    - create an accessible PDF [how-to for Office]

17th International Conference on Digital Preservation

iPRES 2021, Beijing,China.

DOI: 10.1145/nnnnnnn.nnnnnnn

EMAIL ARCHIVING:

A Collaborative Journey

Chris Prom

University of Illinois – Urbana Champaign

United States

prom@illinois.edu

Ruby Martinez

University of Illinois – Urbana-Champaign

United States

rubylm2@illinois.edu

Stephen Abrams

Harvard University

United States

Stephen_abrams@harvard.edu

Gregory Wiedeman

University of Albany, SUNY

United States

gwiedeman@albany.edu

Matt Teichman

University of Chicago

United States

teichman@uchicago.edu

Abstract – This panel examines the current and future developments in email archiving. Such developments include 1) the Email Archiving: Building Capacity and Community (a regrant program), 2) Integrating Preservation Functionality into ePADD, 3) Attachment Converter: Preserving the Context of Electronic Correspondence, 4) Mailbag: A Stable Package for Email with Multiple Formats, and 5) Developments in EA-PDF. In addition, the five panelists will consider how their projects' collaborative nature contributes to the broader community's efforts to build flexible, efficient services to meet email archiving challenges.

Keywords – Email Archiving, Best Practice, Community, Collaboration.

Conference Topics – Building the Capacity & Capability; Enhancing the Collaboration.

I. INTRODUCTION

Over the past few years, email preservation has become increasingly feasible—though not yet routine or common—part of archives and digital preservation work. The email archiving community's collaborative efforts simultaneously sustain and build the capacity to preserve email across a wider range of institutions. As email archiving is still an emerging discipline, many practitioners have been working to meet the outstanding need for archives, librarians, and museums to adopt easy-to-implement practices to capture, preserve, render, and distribute email that has continuing value (Task Force on Technical Approaches of Email Archives, 2018: 82-83).

Addressing the challenges of email archiving, toolkit functionality and interoperability, and comprehensive archival frameworks, these presenters trace the course of archiving email along a continuum from development to a community of practice. The presenters show how the combined effort and outcomes of these projects will make significant progress in the adoption, productivity, and efficacy of email archiving.

II. EMAIL ARCHIVING: BUILDING CAPACITY AND COMMUNITY

The Email Archiving: Building Capacity and Community (EA:BCC) re-grant program, led by the University of Illinois at Urbana-Champaign Library, is a four-year program seeking to build a broad community of institutions that preserve email as part of their research collections. In its first round, the EA:BCC awarded grant funding to five institutions deeply involved in innovative email archiving activity: Harvard University; University of Albany, SUNY; Council of State

Archivists, Inc.; Columbia University; and the University of Chicago Library. In addition, all programs seek to engage institutions with the tools that currently exist by developing workflows, interoperable systems, metadata pathways, and archival package structures, then sharing the results in forums that will help others build a similar capacity to preserve email.

III. Integrating Preservation Functionality into ePADD

The goal of this collaborative project between Harvard University, the University of Manchester, and Stanford University is to integrate long-term email preservation functionality into Stanford's open-source email archiving software program, ePADD, which already supports archival appraisal, processing, discovery, and delivery. The enhanced product will provide the digital archiving community with a tool that supports the email archiving lifecycle more comprehensively. Because the requirements of preservation infrastructure and workflows vary greatly across institutions, the enhanced ePADD will support functions for local customization and extensibility.

There are several expected benefits from this project, the central one being the provision of preservation functionality in an already popular email archiving tool. The wider digital archiving community will be better equipped to make use of ePADD to complete significant curatorial activities not currently possible without utilizing a matrix of separate tools. The collaborative nature of the project will contribute to an international community of engaged users with the new system functionality. Sustainability of the software will be another key outcome and benefit of this project, as responsibility for code maintenance and support is extended beyond Stanford to include project partners Harvard and Manchester and, through affirmative project outreach activities, the wider email archiving community.

IV. Attachment Converter: Preserving the Context of Electronic Correspondence

At the University of Chicago, we frequently come into the possession of email backups from faculty and staff that are of historical interest. As we accession these important documents, we are always looking for ways of safeguarding them against the dangers of technological obsolescence. To meet these challenges, the Hanna Holborn Gray Special Collections Research Center and the Digital Library Development Center are working together to develop a command line tool for archivists called Attachment Converter. What Attachment Converter will do is convert all of the attachments in an email collection to archivally stable formats, so that future researchers will be able to browse the collection using an email client of their choice, viewing and opening the converted attachments in their respective emails alongside the originals. We have begun work on the project as part of a Masters Program in Computer Science research practicum, and through the Email Archives: Building Capacity and Community grant, we will be hiring two interns to develop the initial command line version of the application over the Summer of 2021.

V. Mailbag: A Stable Package for Email with Multiple Formats

Currently there is no single effective preservation format for email. The Mailbag project aims to draft a specification and create an open source tool for preserving email archives using multiple formats, such as MBOX, PDF, and WARC. MBOX or EML files provide structured access for computational use, PDF files preserve the document-like rendering of email well and provide easy dissemination, and web archives preserve the potential interactivity of email HTML and CSS, as well as embedded and linked Web content from external sources. The Mailbag approach is to preserve these multiple formats in a stable and computer-actionable package that will be will be an extension of the Bagit specification. A Mailbag will be a special type of "bag," using filesystem conventions and metadata that will connect different formats of the same email along with information about its source. The project also aims to create a Mailbag tool which will enable archivists to process email shortly after capture through either the command line or a basic graphical user interface (GUI).

VI. Developments in EA-PDF

Email technology does not include a concept of a "native" email presentation; preservation outside the source systems implies some degree of transformation. PDF, on the other hand, is a format that is broadly adopted for presentation and preservation purposes, prevalent in business and industry, and with viewers

installed on the operating systems of most consumer computers and handheld devices.

PDF is a natural target format for email preservation. Existing package structures, such as MBOX, reflect application-specific features, and content cannot be easily rendered outside of an email client environment. Domain-specific tools rely on internal databases and are not preservation solutions. PDF, on the other hand, is a supported file format in most existing preservation repositories and digital libraries. In addition to its familiar page rendering capability, PDF is a highly structured and documented container format supporting dozens of document-specific features and capabilities. PDF technology represents a platform independent free-form database with built in support for XMP (Extensible Metadata Packaging). These qualities explain its broad appeal and implantation, as well as its suitability for packaging metadata together with visual content. Relevant archives user communities, including local, state, and federal archives, as well as museum archives, university archives, and special collections unites, have requested PDF-based archiving options for email and a specification has been released for further development (Specification for Using PDF to Package and Represent Email, 2020).

VII. CONCLUSION

This session will inform attendees of current and future developments in email archiving. Chris Prom, EA:BCC Principal Investigator, and Ruby Martinez, Email Archives Community Fellow, will facilitate a discussion amongst the awardees focusing on how their programs complement each other as part of a broader community effort to build flexible, efficient services to meet specific email archiving needs. All programs seek to engage institutions with the tools that currently exist by developing workflows, interoperable systems, metadata pathways, and archival package structure, then sharing the results in forums that will help others build a similar capacity to preserve email.

REFERENCES

[1] The Future of Email Archives: A Report from the Task Force on Technical Approaches for Email Archives, https://www.clir.org/pubs/reports/pub175/

[2] A Specification for Using PDF to Package and Represent Email, https://www.ideals.illinois.edu/handle/2142/109251