

# ePADD+ Project

Illinois Email Archiving Symposium

---

**Stephen Abrams**, Harvard University  
**Paul Carlyle**, University of Manchester  
**Sally DeBauche**, Stanford University  
**Tricia Patterson**, Harvard University

# Agenda



- Overview of the ePADD+ Project
- New features
- Get started with ePADD
- ePADD community
- Q&A

# Overview of the ePADD+ Project

---

January 2021 - December 2022

“ Email represents an increasingly important part of the historical record. Preserving and ensuring access to this record are therefore central to the functions and values of archives and archivists. Until we can meet the challenges of email archiving, responsible custody is undermined, accountability is abandoned, and, ultimately, the historical record is imperiled. In short, the problem won't take care of itself, and the time to act is now. ”

# Email preservation



## Applied to email records

- Acquire, as files from email clients/servers or IMAP protocol
- Save original order, e.g., headers, multi-parts, attachments
- Characterize and describe
- Normalize/create derivatives
- Persistent protection and security
- Discovery/delivery

# E@S & ePADD

Functional overlap  
and added value

## Creation and Use



- Create/send/receive
- Daily use of email and attachments

## Appraisal and Selection



- Disposition decisions
- Appraisal
- Donor relations

ePADD

## Acquisition



- Capture or transfer
- Legal stewardship

## Processing



- Arrangement and description
- Restrictions, embargoes, etc.

E@S, ePADD

## Preservation



- Deposit to repository
- Preservation actions

E@S

## Discovery and Access



- Listing in online catalogs
- Availability to researchers

ePADD



Account Holder



Records Manager



Archivist



Technologist



Researcher



Tools such as DarcMail,  
**ePADD, EAS,**  
and TOMES deserve support  
that allows for their  
closer integration and  
alignment over time.



# ePADD+ Project



- Integrate EAS-like long-term preservation function with ePADD's existing support for appraisal, processing, discovery, and delivery
- Collaborative effort by Harvard, Manchester, Stanford, and Amuse Labs
- Funded by University of Illinois (#098468-18280) as part of its EA:BCC regrant initiative, itself supported by the Andrew W. Mellon Foundation (#1905-06768)



# The “e” Team



Stephen Abrams  
Dee Dee Crema  
Jessi Jassal  
Tricia Patterson  
Janet Taylor  
Julie Wetherill  
Andrew Woods



Paul Carlyle  
Jochen Farwer  
Ian Gifford  
Tom Higgins  
Jessica Smith  
Jan Whalen



Sally DeBauche  
Glynn Edwards  
Michael Olson



Ruby Martinez  
Chris Prom



Sudheendra Hangal  
Chinmay Narayan



Matt Hovey



CFRADO

NOW FEATURING

*preservation support*

# New Preservation Features in ePADD

---

ePADD Versions 9.0 and 10.0

# New features: Import

Previous ePADD versions



**After ePADD+ Project**

- Limited header information was captured (To, from, subject, date...)
- Some attachments not handled (e.g. text, html)
- Email text captured but formatting was lost
- No support for multi-part messages

- Complete header information stored
- All attachments types are handled
- Message formatting is captured
- Multi-part messages are handled
- Optional format derivatives
- Optional 'sidecar' files

# New features: Metadata

Previous ePADD versions



**After ePADD+ Project**

- Limited metadata captured
  - Stored in multiple places
- Much richer set of metadata captured:
    - Automatic metadata (e.g. PREMIS events)
    - Manual metadata (additional fields available in GUI)
  - Metadata stored in a single place using PREMIS

# New features: Metadata

; separated trusted email addresses

RECOMPUTE ADDRESS BOOK

Set Owners C

owners email id

PREMIS Meta

Event

Ingestion

Identifier assignment

Fixity check

Message digest calculation

Quarantine

Unquarantine

Unpacking

Name cleanup

Virus check

Format identification

Validation

Normalization

Transcription

Creation

Other

Ingestion



# New features: Metadata

```
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
epaddPremis.xml
646 <premis:linkingAgentIdentifierType>local</premis:linkingAgentIdentifierType>
647 <premis:linkingAgentIdentifierValue>ePADD - Release 10-alpha - Appraisal - running in Java 11.0.17 - Windows 10</premis:linkingAgentIdent
648 <premis:linkingAgentRole authority="eventRelatedAgentRole" authorityURI="http://id.loc.gov/vocabulary/preservation/eventRelatedAgentRole"
649 </premis:linkingAgentIdentifier>
650 </premis:event>
651 <premis:event>
652 <premis:eventIdentifier>
653 <premis:eventIdentifierType>UUID</premis:eventIdentifierType>
654 <premis:eventIdentifierValue>9d7b091c-64d8-4a2a-b1f2-7e5b4d05f1ed</premis:eventIdentifierValue>
655 </premis:eventIdentifier>
656 <premis:eventType>Other</premis:eventType>
657 <premis:eventDateTime>2023-01-16T19:06:00+0000</premis:eventDateTime>
658 <premis:eventDetailInformation>
659 <premis:eventDetail>Email subset chosen by Kelly Miller</premis:eventDetail>
660 </premis:eventDetailInformation>
661 <premis:eventOutcomeInformation>
662 <premis:eventOutcome></premis:eventOutcome>
663 </premis:eventOutcomeInformation>
664 <premis:linkingAgentIdentifier>
665 <premis:linkingAgentIdentifierType>local</premis:linkingAgentIdentifierType>
666 <premis:linkingAgentIdentifierValue>ePADD - Release 10-alpha - Appraisal - running in Java 11.0.17 - Windows 10</premis:linkingAgentIdent
667 <premis:linkingAgentRole authority="eventRelatedAgentRole" authorityURI="http://id.loc.gov/vocabulary/preservation/eventRelatedAgentRole"
668 </premis:linkingAgentIdentifier>
669 </premis:event>
670 <premis:agent>
671 <premis:agentIdentifier>
672 <premis:agentIdentifierType>local</premis:agentIdentifierType>
673 <premis:agentIdentifierValue>ePADD</premis:agentIdentifierValue>
674 </premis:agentIdentifier>
675 <premis:agentName>ePADD</premis:agentName>
676 <premis:agentType>software</premis:agentType>
677 <premis:agentVersion>Release 10-alpha</premis:agentVersion>
678 </premis:agent>
679 <premis:rights xmlID="AAAAG" version="3.0">
```

# New features: Export

Previous ePADD versions



- Appraisal module export included only post-appraisal data for processing module
- Key preservation info e.g. MBOX structure, duplicates/folder names etc. not included

**After ePADD+ Project**

- 'Preservation ready' Bag structured to include multiple exportable assets including:
  - Full MBOX structure
  - Canonical version of archive
  - Derivative versions
  - Additional sidecar files
  - Metadata



# New features: Emailchemy integration

Weird Kid Software

Industrial-strength email migration, conversion and archival software

Emailchemy - the Email Alchemist [Emailchemy for ePADD](#) [Buy](#) [Blog](#) [Support](#)

Emailchemy for ePADD



for



# New features: Emailchemy integration

Previous ePADD versions



- Conversion to mbox format required outside of ePADD application

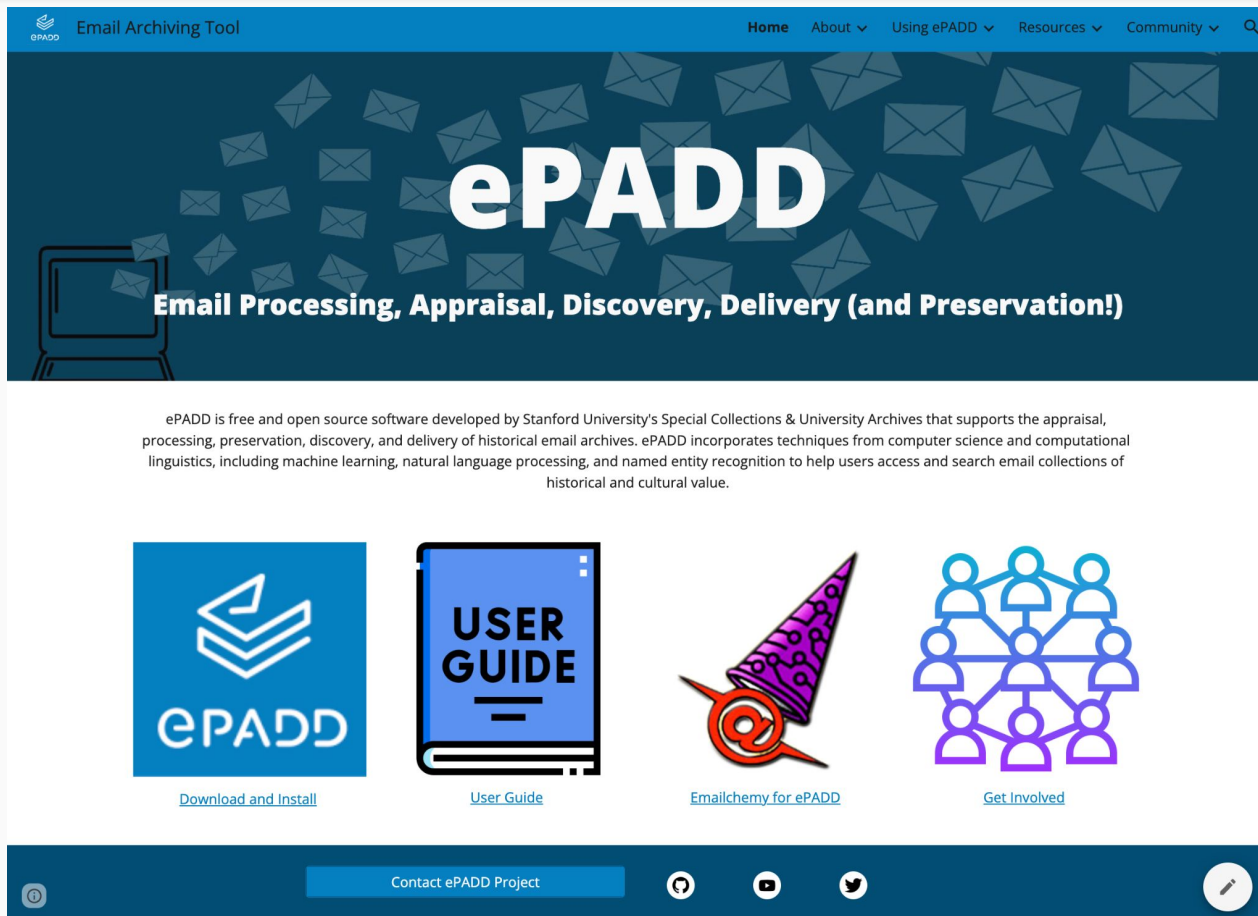
**After ePADD+ Project**

- On purchase of an Emailchemy licence and creation of dongle
- Conversion to mbox from multiple email formats can take now place within ePADD
- Conversion of appraised or processed files to EML format is also possible on export

# Getting started

---

Installation, user guide, and workflows



The screenshot shows the homepage of the ePADD project. At the top, there is a blue navigation bar with the text "Email Archiving Tool" on the left and a menu on the right containing "Home", "About", "Using ePADD", "Resources", and "Community", along with a search icon. The main header area features a dark blue background with a pattern of white envelope icons. The word "ePADD" is prominently displayed in large white letters, with the tagline "Email Processing, Appraisal, Discovery, Delivery (and Preservation!)" below it. A central paragraph describes the software as free and open source, developed by Stanford University's Special Collections & University Archives, and lists its capabilities in processing, preservation, discovery, and delivery of historical email archives. Below the text are four icons: the ePADD logo, a book icon labeled "USER GUIDE", a purple party hat with an @ symbol, and a group of people icon. Each icon has a corresponding text link below it: "Download and Install", "User Guide", "Emailchemy for ePADD", and "Get Involved". The footer is a dark blue bar with a "Contact ePADD Project" button, social media icons for GitHub, YouTube, and Twitter, and a pencil icon for editing.


Email Archiving Tool


Home About Using ePADD Resources Community


# ePADD


**Email Processing, Appraisal, Discovery, Delivery (and Preservation!)**

ePADD is free and open source software developed by Stanford University's Special Collections & University Archives that supports the appraisal, processing, preservation, discovery, and delivery of historical email archives. ePADD incorporates techniques from computer science and computational linguistics, including machine learning, natural language processing, and named entity recognition to help users access and search email collections of historical and cultural value.

 [Download and Install](#)

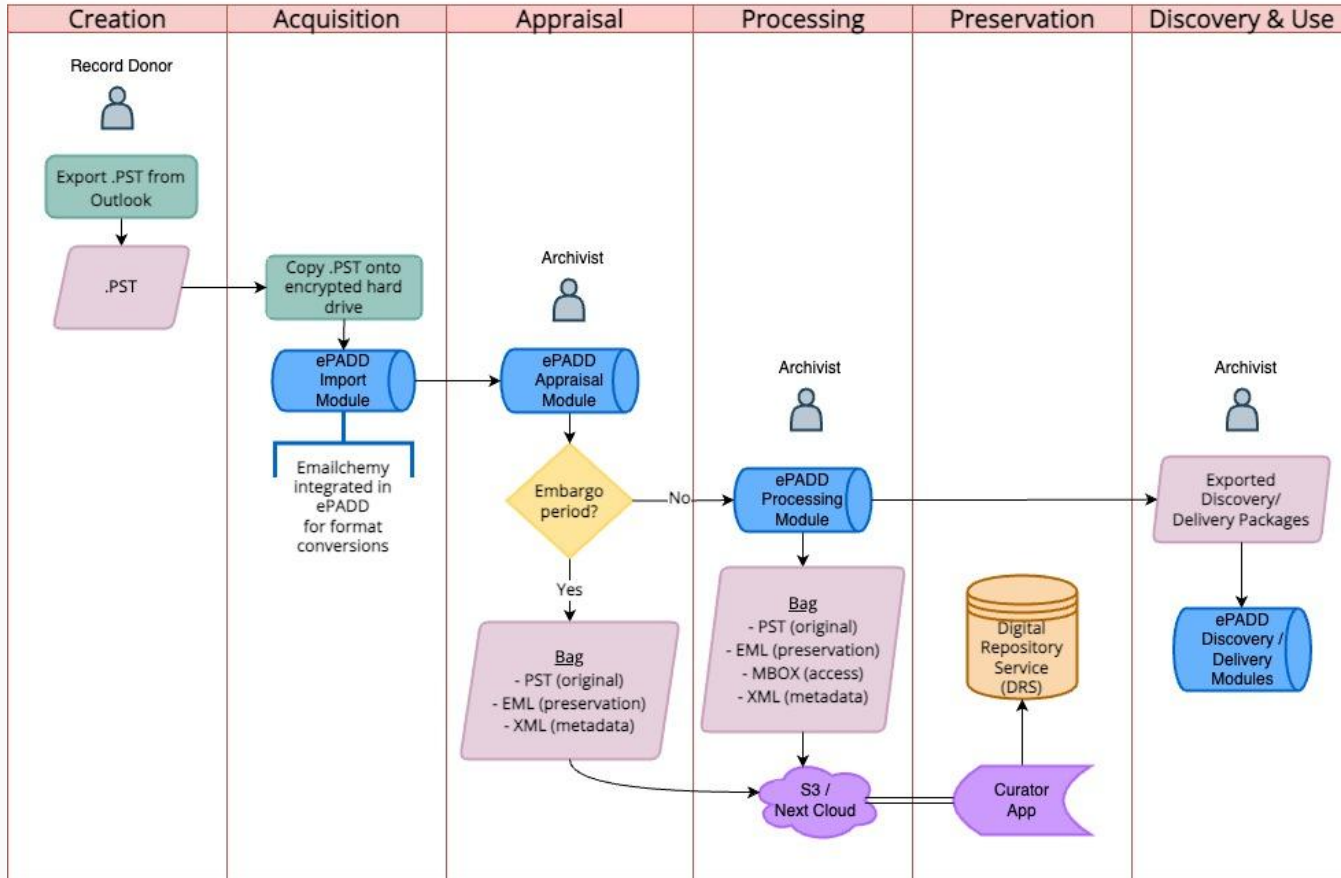
 [User Guide](#)

 [Emailchemy for ePADD](#)

 [Get Involved](#)

Contact ePADD Project

# Sample institutional workflow - Harvard University



# Getting involved

---

ePADD community members unite!

# Steering group

## Join the Steering Group to:

- Advance the ePADD roadmap through collective efforts and strategic grant applications
- Collaborate with other ePADD users on improving documentation and enhancing user guidance
- Share ideas on workflows, projects, and more

## Bi-monthly meetings:

- Tuesday, 11 July 2023, 11am - 12pm ET
- Tuesday, 12 September 2023, 11am - 12pm ET
- Tuesday, 12 December 2023, 11am - 12pm ET

# Code group

## Join the Code Group to:

- Learn how to contribute to the ePADD codebase
- Help address maintenance and security issues with the tool
- Merge your enhancements in a sustainable way
- Improve code documentation and testing

## Monthly open meetings:

- Thursday, 6 July 2023, 10am-11am ET
- Thursday, 14 September 2023, 10am-11am ET
- Thursday, 12 October 2023, 10am-11am ET
- Thursday, 9 November 2023, 10am-11am ET
- Thursday, 7 December 2023, 10am-11am ET



# Join the ePADD community!

## **Stay informed**

- Website - [epaddproject.org](http://epaddproject.org)
- ePADD listserv
- Twitter: @e\_padd

## **Get involved**

- ePADD steering group
- ePADD code group
- Share your workflows & use cases

# Q & A

---

Questions, comments, discussion

# Thank you!

---

**Stephen Abrams**, [stephen\\_abrams@harvard.edu](mailto:stephen_abrams@harvard.edu)

**Paul Carlyle**, [paul.carlyle@manchester.ac.uk](mailto:paul.carlyle@manchester.ac.uk)

**Sally DeBauche**, [debauche@stanford.edu](mailto:debauche@stanford.edu)

**Tricia Patterson**, [tricia\\_patterson@harvard.edu](mailto:tricia_patterson@harvard.edu)