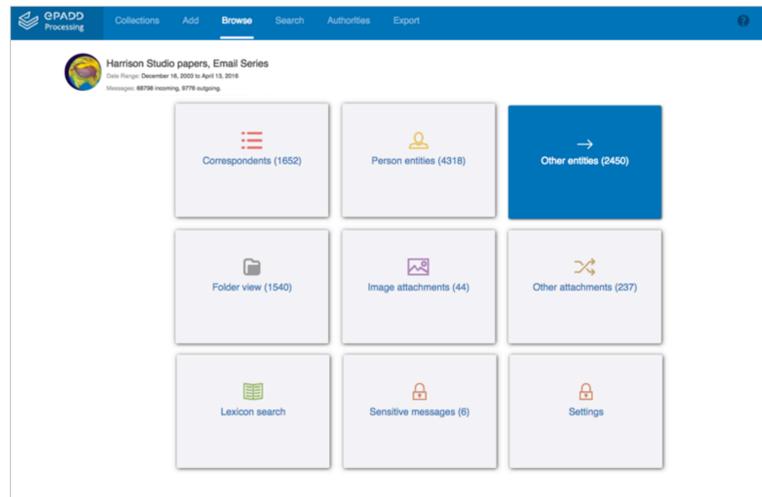


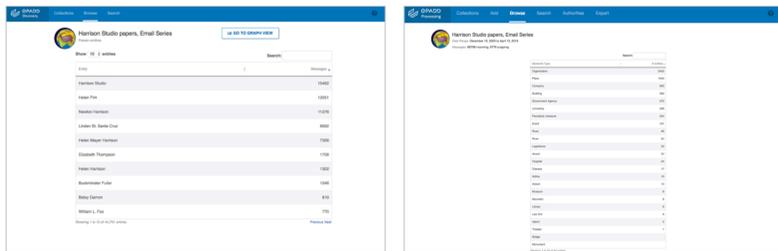
Introduction

ePADD is software that enables the computational analysis of email using named entity recognition and other natural language processing algorithms. It was created to aid archival repositories and other cultural memory insitutions in the appraisal, processing, discovery, and delivery of historically and culturally significant email. It can also be used by journalists, private and family historians, and other individuals seeking to search, browse, analyze, screen, and share email.

Browsing



Named Entity Recognition / Fine-Grained Named Entity Type Browsing: ePADD uses a custom named entity recognizer/classifier that recognizes categories of entities bootstrapped from DBpedia. These include persons, organizations, locations, government entities, political parties, companies, universities, diseases, and awards. ePADD learns from these categories and is also able to recognize likely entities it has not come across before.



Name Resolution / Correspondent Browsing: ePADD resolves names and email addresses associated with a single correspondent, improving browsing and visualization. All decisions can be manually overridden using a dedicated interface. Mailing lists can similarly be tagged and optionally consolidated using this functionality. Resolved correspondent names can be browsed and graphed alphabetically or by volume of messages exchanged with the email account holder.

Image Attachment Browsing: ePADD includes an image attachment browser that supports filtering by file size, file extension, or date range. The results are presented on a scrollable image wall. Selecting any of the resulting images offers the user an option to navigate directly to the related message, to view the image in context. This interface can also be used to screen for sensitive personal images or financial information.

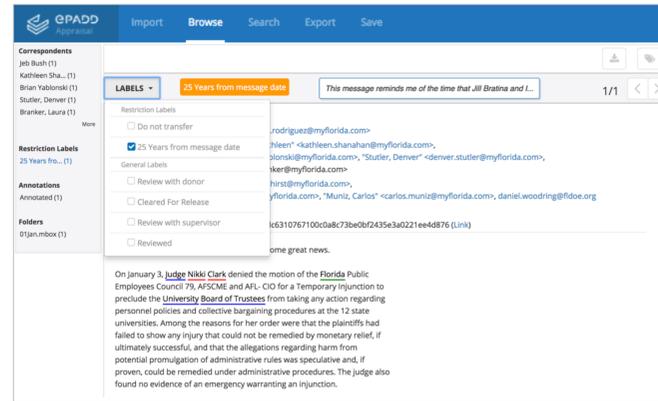
Search

Advanced Search: ePADD includes an advanced search interface enabling sophisticated search queries across text, correspondents, entities, attachments, and more. For instance, users can perform a search for messages containing entities from the disease entity category, or terms from the sensitive lexicon, and further limit this search by mandating that the search should exclude results from a mailing list. In this way a user can create a narrow search for potentially sensitive information to embargo for a specific period of time, or to not share with a repository or other individuals.

Multi-Entity Search: ePADD includes a multi-entity search to aid in comparative entity analysis between the archive and any other textual corpus. Matching entities are highlighted and link to message results.

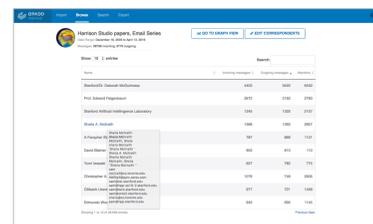
Screening

Bulk Actions and Annotation: ePADD allows the user to apply labels (including restriction periods, processing actions, or more general descriptive labels) and annotations to sets of messages meeting user-defined criteria, including all messages associated with a given correspondent, all messages from a given date range, all messages containing certain keywords or named entities in the subject or message fields, or some combination of the above.

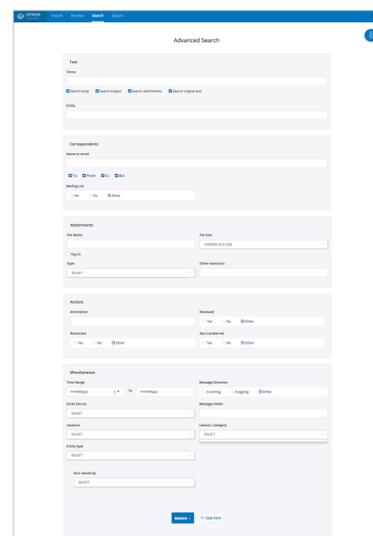


Regular Expression Search: ePADD includes a customizable regular expression search, enabling the user to quickly search a collection for sensitive information such as social security numbers or credit card numbers, or any other expressions.

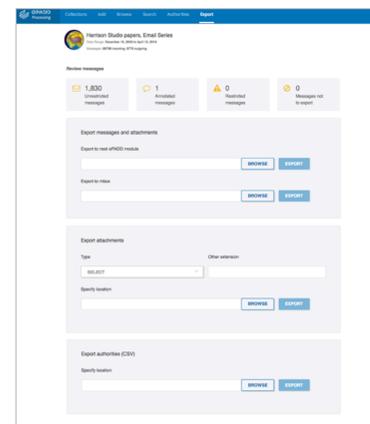
Lexicon Search: ePADD includes tiered thematic keyword searches geared towards broad analysis of a variety of email collections, including lexicons to identify categories of sensitive correspondence. These lexicons can be edited and tuned, or the user can create all new lexicons to suit their research goals.



Name Resolution



Advanced Search



Export Options

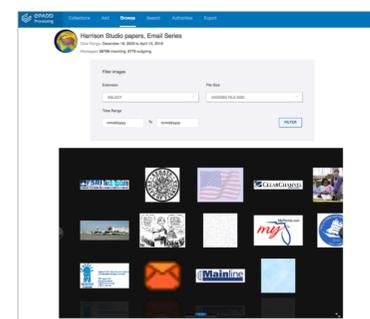
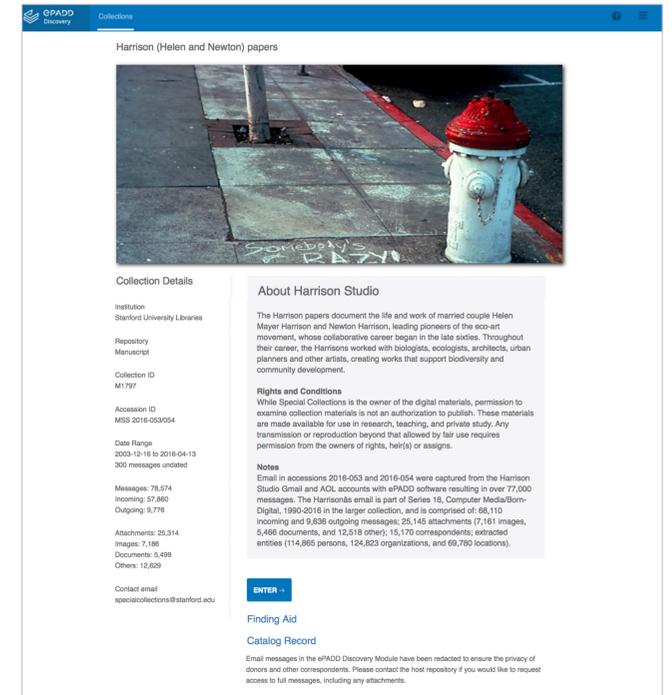


Image Attachment Browsing

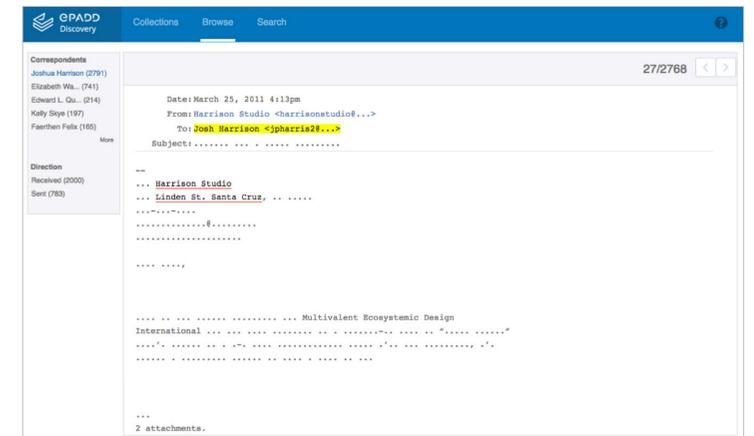
Additional Functionality: ePADD also supports message deduplication, account and folder-level browsing, mailing list identification, graphical visualization, an interface for assigning authorized headings to correspondents and fine-grained named entities, and extensive user-defined export options.

Access

Full Access to Messages and Attachments: ePADD provides a Delivery module that provides full access to all messages and attachments that have been screened and approved by the creator and the insitution.



Redacted View of Messages: ePADD provides an optional online Discovery module that redacts all content other than message dates, correspondents (local-part of email address), and named entities.



An example of ePADD Discovery can be found at epadd.stanford.edu

About the Project

ePADD development is managed by Stanford University's Department of Special Collections & University Archives, part of Stanford University Libraries, in collaboration with partners at Harvard University, the Metropolitan New York Library Council (METRO), University of Illinois at Urbana-Champaign, and University of California, Irvine.



Funding for current ePADD development is provided through an Institute of Museum & Library Studies (IMLS) National Leadership Grant (NLG) for Libraries, which supports projects that address challenges faced by the library and archive fields and that have the potential to advance practice in those fields. Development for the initial 2015 release of ePADD was primarily funded by the National Historical Publications and Records Commission (NHPRC).