ePADD Shared Discovery Module Website Collection Contributor Guide

February 2021

ePADD is a software application that supports archival processes around the appraisal, ingest, processing, discovery, and delivery of email archives. It is open source and licensed under an Apache Public License, v2.0. The latest information about the software can be found here.

This guide documents the workflows and activities related to **ePADD's** shared Discovery Module. For information related to ePADD's other modules, please refer to the main **ePADD** user guide.

ePADD development is managed by Stanford University's Department of Special Collections & University Archives, in collaboration with partners at Harvard University and the University of Manchester.

Funding for **ePADD** development has been provided by grants from the Andrew W. Mellon Foundation, the Institute of Museum & Library Studies (IMLS) National Leadership Grant (NLG), the Libraries and the National Historical Publications and Records Commission (NHPRC), and Stanford Libraries.

Document Revisions

Date	Document Changes
2/1/2021	Initial Draft

Table of Contents

1   Overview

This user guide will assist ePADD users in contributing their processed collection metadata to the shared ePADD Discovery Module. The module provides tools for repositories to remotely share a redacted view of email archives with users through a public web server discovery environment, accessible at epadd.org.

ePADD is comprised of four modules – Appraisal, Processing, Discovery, and Delivery – which are designed to facilitate the process of working with email archives. This guide is specific to the shared ePADD Discovery Module. For more information on using ePADD's three other modules, please refer to the main [user guide](#).

The ePADD Discovery Module is designed to run under a web-server, and allow remote users to search redacted header info and extracted entities while limiting full-text access to the materials. There are two ways to make ePADD collection metadata available online using the Discovery Module: (1) by mounting a Discovery Module on a local web server at your institution or (2) by contributing the collection metadata to the shared ePADD Discovery Module hosted by Stanford University Libraries.

This guide covers the process for contributing email collection metadata to the shared ePADD Discovery Module hosted by Stanford University Libraries. For more information about mounting a Discovery Module on a local web server at your institution, please refer to the [user guide](#).

2   Getting Started

This section will get you started working with the shared ePADD Discovery Module after you have completed processing an email collection, which typically includes appraisal of materials, annotation of messages, and identification and flagging sensitive information. For more information on processing email collections with ePADD, please refer to Sections 3 and 4 in the [ePADD user guide](#).

2.1   Technical Overview

This section covers technical matters related to the shared ePADD Discovery Module, including details about the Stanford University Libraries ePADD Discovery servers and an explanation of the data contained in the ePADD Discovery package.

2.1.1   The Stanford University Libraries ePADD Discovery Servers

Stanford University Libraries maintains two servers to support the shared ePADD Discovery Module. One server is used for testing and development purposes and the other is the production server for the ePADD Discovery website.

Both servers comply with Stanford University's data security standards for servers.

For questions about technical specifications of the ePADD Discovery servers, please contact epadd_project@stanford.edu

2.1.2   ePADD Discovery Metadata

To contribute ePADD collection metadata to the shared ePADD Discovery Module, institutions will need to transfer a copy of the epadd-discovery folder exported from the Processing Module. For more information on the process of exporting a processed collection from the Processing Module, please refer to section 4.12 of the main ePADD user guide.

Below is a detailed description of the metadata exported by ePADD from the Processing Module. This is the metadata that contributing institutions will transfer to Stanford University Libraries for display in the shared ePADD Discovery Module.
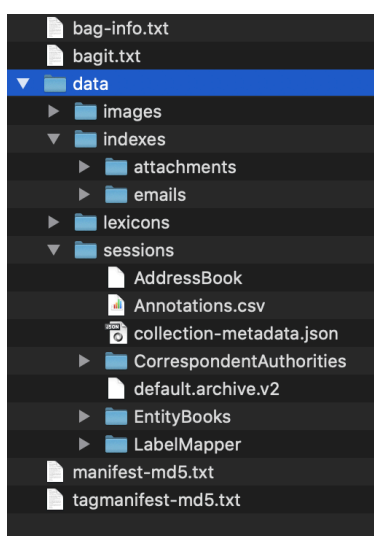
*note: the following directories will be deprecated in the next version of the software:

user\epadd-discovery\ePADD Archive of X-Discovery\data\indexes\attachments

user\epadd-discovery\ePADD Archive of X-Discovery\data\lexicons

user\epadd-discovery\ePADD Archive of X-Discovery\

bag-info.txt	Jan 4, 2021 at 11:33 AM	51 bytes	Plain Text
bagit.txt	Jan 4, 2021 at 11:33 AM	54 bytes	Plain Text
▶ data	Today at 6:59 AM	--	Folder
manifest-md5.txt	Jan 4, 2021 at 11:33 AM	5 KB	Plain Text
tagmanifest-md5.txt	Jan 4, 2021 at 11:33 AM	142 bytes	Plain Text

This is the main directory for the data that ePADD exports to the Discovery Module. It is all contained within the "data" directory and includes bagit information.

user\epadd-discovery\ePADD Archive of X-Discovery\data\

▶ images	Sep 4, 2020 at 4:20 PM	--	Folder
▶ indexes	Today at 6:59 AM	--	Folder
▶ lexicons	Jan 4, 2021 at 11:31 AM	--	Folder
▶ sessions	Today at 6:59 AM	--	Folder

There are four directories within the main directory - "images," "indexes," "lexicons," and "sessions."

user\epadd-discovery\ePADD Archive of X-Discovery\data\images\

▼ images	Sep 4, 2020 at 4:20 PM	--	Folder
bannerImage	Sep 4, 2020 at 4:18 PM	722 KB	TextEdit
landingPhoto	Sep 4, 2020 at 4:19 PM	2.5 MB	TextEdit
profilePhoto	Sep 4, 2020 at 4:20 PM	119 KB	TextEdit
▶ indexes	Today at 6:59 AM	--	Folder
▶ lexicons	Jan 4, 2021 at 11:31 AM	--	Folder
▶ sessions	Today at 6:59 AM	--	Folder

The images directory contains the banner image, landing page image, and profile image selected by the archivist to display on the collection information page in the Discovery Module. This directory does not contain any images from the email collection itself.

user\epadd-discovery\ePADD Archive of X-Discovery\data\indexes\

▶ 📁 images	Sep 4, 2020 at 4:20 PM	--	Folder
▼ 📁 indexes	Today at 6:59 AM	--	Folder
▼ 📁 attachments	Jan 4, 2021 at 11:32 AM	--	Folder
📄 segments_1	Jan 4, 2021 at 11:32 AM	69 bytes	TextEdit
📄 write.lock	Jan 4, 2021 at 11:32 AM	Zero bytes	Document
▼ 📁 emails	Jan 4, 2021 at 11:32 AM	--	Folder
📄 _0.cfe	Jan 4, 2021 at 11:32 AM	331 bytes	Document
📄 _0.cfs	Jan 4, 2021 at 11:32 AM	42.4 MB	Document
📄 _0.si	Jan 4, 2021 at 11:32 AM	378 bytes	Document
📄 segments_1	Jan 4, 2021 at 11:32 AM	133 bytes	TextEdit
📄 write.lock	Jan 4, 2021 at 11:31 AM	Zero bytes	Document
▶ 📁 lexicons	Jan 4, 2021 at 11:31 AM	--	Folder
▶ 📁 sessions	Today at 6:59 AM	--	Folder

This directory contains data used by Lucene to index and search entities in the Discovery Module

user\epadd-discovery\ePADD Archive of X-Discovery\data\lexicons\

▶ 📁 images	Sep 4, 2020 at 4:20 PM	--	Folder
▶ 📁 indexes	Today at 6:59 AM	--	Folder
▼ 📁 lexicons	Jan 4, 2021 at 11:31 AM	--	Folder
📄 Persona.academic.adminis...nsitive.duke.english.lex.txt	Jan 4, 2021 at 11:31 AM	988 bytes	Plain Text
📄 Persona.author.princeton.english.lex.txt	Jan 4, 2021 at 11:31 AM	417 bytes	Plain Text
📄 Persona.composer.NYPL.english.lex.txt	Jan 4, 2021 at 11:31 AM	1 KB	Plain Text
📄 Persona.environmental.arti...cts.stanford.english.lex.txt	Jan 4, 2021 at 11:31 AM	6 KB	Plain Text
📄 Persona.faculty.UCI.english.lex.txt	Jan 4, 2021 at 11:31 AM	372 bytes	Plain Text
📄 Persona.journalist.activist....d.travel.UCB.english.lex.txt	Jan 4, 2021 at 11:31 AM	306 bytes	Plain Text
📄 Persona.microbiologist.UIUC.english.lex.txt	Jan 4, 2021 at 11:31 AM	504 bytes	Plain Text
📄 Persona.writer.theater.NYPL.english.lex.txt	Jan 4, 2021 at 11:31 AM	739 bytes	Plain Text
📄 regex.english.lex.txt	Jan 4, 2021 at 11:31 AM	128 bytes	Plain Text
📄 sensitive-cca-french.french.lex.txt	Jan 4, 2021 at 11:31 AM	2 KB	Plain Text
📄 sensitive-ccca-german.german.lex.txt	Jan 4, 2021 at 11:31 AM	22 KB	Plain Text
📄 sensitive.english.lex.txt	Jan 4, 2021 at 11:31 AM	7 KB	Plain Text
📄 sentiments.english.lex.txt	Jan 4, 2021 at 11:31 AM	7 KB	Plain Text
▶ 📁 sessions	Today at 6:59 AM	--	Folder

This directory contains .txt files containing lexicon terms (example below).

```
Personally Identifiable Information:"social security"|ssn|"credit card"|"driver's
license"|id|visa|"master card"|"american express"|passport|"bank account number"|"credit
card number"|"discover card"|"medical id"|"bank of america"|CapitalOne|"Bank of the
West"|"Wells Fargo"|"Credit Union"|DOB|"date of birth"|pcard|"procurement card"

Employment:"performance appraisal"|"performance evaluation"|"performance review"|
promotion|"annual review"|confidential|tenure|salary|recommendation|"reference letter"|
grievance|fired|"let go"|terminated|demotion|layoff|"letter of support"| "search
committee"|nomination|grievance|harassment|complaint|harass|"sexual harassment"|"sex
discrimination"|"gender discrimination"|discrimination|concealment|toxicant|
drunk|"unwarranted advances"|appointed|"promotion and tenure"|"appointment and
promotion"|"tenure appeal"|"tenure denied"|"tenure denial"|"denied tenure"|"contract
renewal"|compensation|stipend|salary|compensation|stipend|"background check"|"job
offer"|"offer letter"

Student:grade|"education record"|transcript|"student number"|"student id"|expel|
probation|"academic probation"|cheat|"academic fraud"|plagiarize|"underage drinking"|
underage|"underage drinking"

Health:sick|ill|illness|sickness|medical|surgery|endoscopy|hospital|unwell||depressed|
injure|hurt|injury|injured|injurious|doctor|surgeon|surgical|clinic|lethal|suicide|
laceration|trauma|traumatic|schizophrenic|disfigure|radiation|medicine|wound|foetus|fetus|
fetal|injure|disease|infection|vomit|puke|puking|seasick|carsick|nausea|nauseous|
nauseating|nauseated|pallid|wan|miscarriage|"heart attack"|postpartum|hemorrhage|fracture|
casualty|concussion|biopsy|malignant|malignancy|pancreas|pancreatic|gastric|abdominal|
thyroid|cervical|cardiovascular|colorectal|ovarian|gastrointestinal|lung|lungs|thoracic|
oesophagus|alzheimer|parkinson|pediatrician|orthopedic|cardiologist|urologist|oncologist|
gynecologist|dermatologist|neurologist|anesthesiologist|anesthetist|geriatric|"spinal
cord"|"insurance policy"|hospital|doctor|"medical record"|prescription|medication|
```

user\epadd-discovery\ePADD Archive of X-Discovery\data\sessions\

▼ 📁 sessions	Today at 6:59 AM	--	Folder
AddressBook	Jan 4, 2021 at 11:33 AM	259 KB	TextEdit
Annotations.csv	Jan 4, 2021 at 11:33 AM	21 bytes	CSV Document
collection-metadata.json	Jan 4, 2021 at 11:33 AM	5 KB	JSON File
▼ 📁 CorrespondentAuthorities	Jan 4, 2021 at 11:33 AM	--	Folder
CandidateCorrespondentAuthorities.csv	Jan 4, 2021 at 11:33 AM	17 KB	CSV Document
ConfirmedCorrespondentAuthorities.csv	Jan 4, 2021 at 11:33 AM	731 bytes	CSV Document
default.archive.v2	Jan 4, 2021 at 11:33 AM	3.9 MB	Document
▶ 📁 EntityBooks	Today at 7:00 AM	--	Folder
▼ 📁 LabelMapper	Jan 4, 2021 at 11:33 AM	--	Folder
docidmap.csv	Jan 4, 2021 at 11:33 AM	18 bytes	CSV Document
label-info.json	Jan 4, 2021 at 11:33 AM	423 bytes	JSON File

This directory contains data that ePADD manages including the address book, user created annotations, user created collection metadata, and user created/applied labels.

Addressbook (example below)

```
├── Archive owner
Don Knuth
don knuth
don knurh
Donald Knuth
Donald knuth
donald knuth
DonaldKnuth
Donald E. Knuth
Donald E.Knuth
Donald E Knuth
Donald Ervin Knuth
Donand Ervin Knuth
donald ervin knuth
Knuth, Don
Knuth, Donald
Don
Don K.
Donald
donald
knuth
Knuth
Professor Donald Knuth
Prof. Knuth
Maggie McLoughlin
maggie mcloughlin
maggie a mcloughlin
Maggie McLaughlin
Mam
mam
```

Collection-metadata.json (example below)


```
3    "repository": "Manuscripts",
4    "collectionTitle": "Robert Creeley email archive",
5    "collectionID": "M0662",
6    "findingAidLink": "https://oac.cdlib.org/findaid/ark:/13030/tf7b69n911/",
7    "catalogRecordLink": "https://searchworks.stanford.edu/view/4084039",
8    "contactEmail": "specialcollections@stanford.edu",
9    "rights": "Property rights reside with the repository. Literary rights reside with the creators of the do
10   "notes": "Email messages in the ePADD Discovery Module have been redacted to ensure the privacy of donors
11   "scopeAndContent": "",
12   "shortTitle": "Robert Creeley",
13   "shortDescription": "American poet and author",
14   "timestamp": 1580925274460,
15   "tz": "America/Los_Angeles",
16   "nDocs": 48919,
17   "nIncomingMessages": 31934,
18   "nOutgoingMessages": 17573,
19   "nHackyDates": 79,
20   "nBlobs": 8698,
21   "nUniqueBlobs": 6363,
22   "nImageBlobs": 4265,
23   "nDocBlobs": 2347,
24   "nOtherBlobs": 2086,
25   "ownerName": "Robert Creeley",
26   "about": "Recognized as a seminal figure of American letters in the second half of the 20th century, Robe
27   "entityCounts": {
28      "0": 22625,
29      "3": 2299,
30      "2": 348,
31      "4": 109,
32      "5": 82,
33      "9": 41,
34      "10": 38,
35      "17": 26,
36      "18": 146,
37      "19": 44,
```

user\epadd-discovery\ePADD Archive of X-Discovery\data\sessions\CorrespondentAuthorities

▼ 🗀 sessions	Today at 6:59 AM	--	Folder
📄 AddressBook	Jan 4, 2021 at 11:33 AM	259 KB	TextEdit
📄 Annotations.csv	Jan 4, 2021 at 11:33 AM	21 bytes	CSV Document
📄 collection-metadata.json	Jan 4, 2021 at 11:33 AM	5 KB	JSON File
▼ 🗀 CorrespondentAuthorities	Jan 4, 2021 at 11:33 AM	--	Folder
📄 CandidateCorrespondentAuthorities.csv	Jan 4, 2021 at 11:33 AM	17 KB	CSV Document
📄 ConfirmedCorrespondentAuthorities.csv	Jan 4, 2021 at 11:33 AM	731 bytes	CSV Document
📄 default.archive.v2	Jan 4, 2021 at 11:33 AM	3.9 MB	Document
▶ 🗀 EntityBooks	Today at 7:00 AM	--	Folder
▶ 🗀 LabelMapper	Jan 4, 2021 at 11:33 AM	--	Folder

This directory contains information about correspondent authorities including Name, FAST ID, VIAF ID, LOC Named ID, Local ID, and date extent in a .csv file (example below).

	A	B	C	D	E	F	G	H	I	J	K
1	Canonical Name	Name	FAST Id	VIAF Id	Wikipedia Id	LoC Subject I	LoC Named Authority File Id	Local Id	Extent	Is manually assigned	
2	don knuth	Knuth, Donald Ervin	1442423	7466303	Donald_Knuth	unknown	n79135509	unknown	1938-	N	
3	graham ronald	Graham, Ronald L.	51917	108818002	Ronald_Graham	unknown	n80033469	unknown	1935-	N	
4	kailath thomas	Kailath, Thomas	20030	73928044	Thomas_Kailath	unknown	n77000709	unknown	unknown	N	
5	e henry lowood	Lowood, Henry	1441610	4967324	unknown	unknown	n79104244	unknown	1953-	N	
6	nils nilsson	Nilsson, Nils J.	1442319	84480189	Nils_John_Nilsson	unknown	n79131975	unknown	1933-	N	
7											
8											

user\epadd-discovery\ePADD Archive of X-Discovery\data\sessions\EntityBooks

▼ 📁 sessions	Today at 6:59 AM	--	Folder
📄 AddressBook	Jan 4, 2021 at 11:33 AM	259 KB	TextEdit
📊 Annotations.csv	Jan 4, 2021 at 11:33 AM	21 bytes	CSV Document
🗓 collection-metadata.json	Jan 4, 2021 at 11:33 AM	5 KB	JSON File
▶ 📁 CorrespondentAuthorities	Jan 4, 2021 at 11:33 AM	--	Folder
📄 default.archive.v2	Jan 4, 2021 at 11:33 AM	3.9 MB	Document
▼ 📁 EntityBooks	Today at 7:00 AM	--	Folder
▼ 📁 Airline	Jan 4, 2021 at 11:33 AM	--	Folder
📄 EntityBook	Jan 4, 2021 at 11:33 AM	284 bytes	TextEdit
▶ 📁 Airport	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Award	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Bridge	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Building	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Company	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Disease	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Event	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Government Agency	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Hospital	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Island	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Law firm	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Legislature	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Library	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Monument	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Mountain	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Museum	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Organisation	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Other	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Periodical Literature	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Person	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Place	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 River	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Road	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 Theater	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 University	Jan 4, 2021 at 11:33 AM	--	Folder
▶ 📁 LabelMapper	Jan 4, 2021 at 11:33 AM	--	Folder

This directory contains a sub-directory for each of the entities in ePADD. Within each of those sub-directories is a file containing all of the terms in that specific entity (example below).

```
AIR
--
ALASKA Airlines
--
Aer Lingus
--
Air Canada
--
Air Canada Express
--
Air Canada Rouge
--
Air France
--
American Airlines
--
British Airways
--
KLM CITYHOPPER
--
Norwegian Air Shuttle
--
Olympic Air
--
Tyrolean Airways
--
UNITED EXPRESS
--
United Airlines
--
```

user\epadd-discovery\ePADD Archive of X-Discovery\data\sessions\LabelMapper

▼ 📁 sessions	Today at 6:59 AM	--	Folder
📄 AddressBook	Jan 4, 2021 at 11:33 AM	259 KB	TextEdit
📄 Annotations.csv	Jan 4, 2021 at 11:33 AM	21 bytes	CSV Document
📄 collection-metadata.json	Jan 4, 2021 at 11:33 AM	5 KB	JSON File
▶ 📁 CorrespondentAuthorities	Jan 4, 2021 at 11:33 AM	--	Folder
📄 default.archive.v2	Jan 4, 2021 at 11:33 AM	3.9 MB	Document
▶ 📁 EntityBooks	Today at 7:00 AM	--	Folder
▼ 📁 LabelMapper	Jan 4, 2021 at 11:33 AM	--	Folder
📄 docidmap.csv	Jan 4, 2021 at 11:33 AM	18 bytes	CSV Document
📄 label-info.json	Jan 4, 2021 at 11:33 AM	423 bytes	JSON File

This directory contains information related to user created and/or applied labels including a .json document containing all labels and a .csv containing the mappings between labels and individual messages to which they are applied.

Label-info.json (example below)

```
● label-info.json*  ×
 1 ▽{
 2 ▽    "13": {
 3          "labType": "GENERAL",
 4          "labId": "13",
 5          "labName": "Stanford archives staff",
 6          "description": "",
 7          "isSysLabel": false,
 8          "restrictionType": "OTHER",
 9          "restrictedUntilTime": 0,
10          "restrictedForYears": 0
11      },
12 ▽    "17": {
13          "labType": "GENERAL",
14          "labId": "17",
15          "labName": "Accession2.Reviewed",
16          "description": "This message was reviewed",
17          "isSysLabel": false,
18          "restrictionType": "OTHER",
19          "restrictedUntilTime": 0,
20          "restrictedForYears": 0
21      }
22 }
23
```

3 How to Contribute to the Shared Discovery Module

This section will cover the workflow for contributing email collection metadata to the shared ePADD Discovery Module hosted by Stanford University Libraries.

3.1 Planning to Contribute Email Collection Metadata

3.1.1 Contacting the ePADD Team

The first step in contributing your email collection metadata to the shared ePADD Discovery Module is to contact the ePADD team at **epadd_project@stanford.edu**. A member of the team will consult with you to make sure that you are ready to contribute your email collection metadata and explain the process for transferring your email collection metadata files to Stanford. Before transferring email collection metadata, contributing institutions should perform thorough review to restrict sensitive data contained in their collections.

3.1.2 Terms of Service

All contributing institutions should thoroughly review Stanford University Libraries' Terms of Service document for the shared ePADD Discovery Module. Contribution of email collection metadata will be considered consent to the Terms of Service.

3.2 Using Github to Track Your Contribution

We use Github issues to track all actions and communications regarding your contribution of email collection metadata to the shared ePADD Discovery Module. Such actions could include: requesting to transfer email collection metadata to Stanford, requesting to remove email collection metadata from the shared Discovery Module either permanently or temporarily, troubleshooting your transfer, or reporting a bug or requesting an enhancement to the ePADD software. If you do not already have a Github account, you will first need to create one. You can find information about creating a free Github account on the [Github documentation site](#).

To initiate the transfer of your email collection metadata, create a new issue on the [ePADD Github repository](#). You can find a step-by-step guide to creating a new issue on Github on the [Github documentation site](#). A member of the ePADD team will respond to the issue to provide information about how to transfer your email collection metadata.

3.3 Transferring Email Collection Metadata

We use Google Drive to transfer email collection metadata files. The epadd-discovery folders are typically fairly small as they only contain text metadata. The ePADD team will provide access to a shared Google Drive folder where the contributing institution can upload their email collection metadata.

It is best practice to maintain backup copies of your epadd-discovery folders locally at your institution.

3.3.1 Transferring Repository Information

In addition to the epadd-discovery folder containing your email collection metadata, you will need to supply the information that will be displayed on your repository "about" page. This information includes details a short statement about your repository, preferred contact information, and a logo.

For guidelines on creating repository information, see Appendix D.

3.3.2 Uploading and Testing Email Collection Metadata

The ePADD team will download your email collection metadata files from Google Drive and upload them to the ePADD Discovery test server. The ePADD team will provide the contributing institution with a link to view their email collection metadata to allow them to make sure that it looks correct.

Once the contributing institution confirms that their collection looks correct, the ePADD team will transfer their collection metadata from the ePADD Discovery test server to the production server at which point it will be displayed on the shared ePADD Discovery Module website.

3.4   Removing Email Collection Metadata from the Shared ePADD Discovery Module

If you need to remove your email collection metadata from the shared ePADD Discovery Module either permanently or temporarily, create a new issue in Github to make your request. Please advise whether you plan on re-submitting your email collection metadata in the future. The ePADD team will delete your email collection metadata from both the test and production servers and confirm this action on your Github issue. If you would like to re-submit your email collection metadata after your remediation, please create a new Github issue to initiate the transfer process.

It is best practice to maintain backup copies of your epadd-discovery folders locally at your institution. Stanford University Libraries will not save copies of your epadd-discovery folders while you perform remediation of your collection metadata.

Appendix A

ePADD Shared Discovery Module Terms of Service

Terms of Service
for
Hosting of ePADD Discovery Metadata
3/18/2021

The ePADD Discovery site hosts automatically extracted collection-level metadata, including the number of email messages in the email archive (incoming/outgoing), inclusive collection dates, number of attachments, browsable correspondent names (with the domain of associated email addresses redacted), and browsable fine-grained entities

(including people, organizations, universities, diseases) extracted from message subjects and bodies by a custom named entity recognizer.

The ePADD Discovery site also hosts user-supplied collection-level metadata added to the ePADD Processing Module, including the name of host institution, name of host repository, collection ID, biographical note, collection summary, scope and contents note, rights and conditions, links to catalog and/or finding aid records, a contact information for the Contributing Institution that is monitored, and three (3) user-supplied images intended to highlight the email creator.

Stanford Libraries makes the ePADD Discovery site services available under the below terms and conditions. In depositing their metadata, the metadata owner (the Contributing Institution) accepts these terms.

1. **Service Availability**
   1.1 Stanford Libraries will make the ePADD Discovery site available to the Contributing Institution 24 hours a day 7 days a week except for scheduled maintenance windows or unexpected outages.
   1.2 Stanford Libraries will notify the Contributing Institution via e-mail at least 24 hours in advance of any planned maintenance.
   1.3 Stanford Libraries will monitor the system for unanticipated outages and will notify the Contributing Institution via email within 72 hours of the detection of a system outage.
   1.4 Stanford Libraries will address unexpected outages during its regular business hours and will update the Contributing Institution at least every 72 hours.

2. **Stanford Libraries Roles and Responsibilities**
   2.1 Stanford Libraries will share guidelines on how to create a processed email archive using the ePADD Processing Module and how to share collection- and message-level metadata with Stanford Libraries.
   2.2 Stanford Libraries will troubleshoot to support adding Discovery export files (i.e. collection- and message-level metadata associated with archival email at the Contributing Institution) to Stanford Libraries' ePADD Discovery site (currently hosted at epadd.org).
   2.3 Stanford Libraries will remove collection- and message-level metadata from the ePADD Discovery server when requested by Contributing Institutions. Removal of collection- and message-level metadata will occur within 72

business hours of receipt of a request during regular business hours, Monday — Friday 8am-5pm Pacific Time.

2.4 Stanford Libraries may remove collection- and message-level metadata at its discretion, for example if contacted by an individual included in collection with a takedown request. In such cases, Stanford Libraries will contact the Contributing Institution to notify them of the issue.

2.5 When hosting a copy of collection- and message-level metadata for discovery purposes, Stanford Libraries claims no ownership over the contents or intellectual property of the data hosted. Note that the Stanford copy of the metadata is for access and convenience; it may be changed or deleted at any point.

3. Contributing Institution Roles and Responsibilities

3.1 The Contributing Institution is solely responsible for the accuracy and completeness of the data it contributes. The Contributing Institution will ensure that user-supplied collection-level metadata deposited to the ePADD Discovery site is correct and that all metadata is created and transmitted according to guidelines shared in advance by Stanford Libraries.

3.2 The Contributing Institution will ensure that its data files for the ePADD Discovery site are exported from a currently supported version of ePADD software.

3.3 The Contributing Institution will maintain a copy of record for all data locally for all collection- and message-level metadata deposited to the ePADD Discovery site. Stanford Libraries is not responsible for data loss associated with the ePADD Discovery site.

3.4 The Contributing Institution will remediate or update collection- and message-level metadata when necessary.

3.5 The Contributing Institution will monitor the public contact email address associated with its shared email archives.

4. Communication

4.1 All application updates, including metadata specification updates, and outage and maintenance notifications, will be communicated in a timely manner to an email address identified by Contributing Institutions.

4.2 The Contributing Institution will make all service requests, including requests to transfer collection- and message-level metadata to Stanford Libraries, remove collection- and message-level metadata from Stanford Libraries servers via the creation of Github issues on the ePADD Github repository.

5. Miscellaneous

5.1 Stanford Libraries may modify these terms from time to time at its sole discretion. Stanford Libraries will notify the Contributing Institution at least 60 days before any modifications become effective.

5.2 Either party may terminate this arrangement in whole or in part with 30 days' notice. On termination, Stanford Libraries will delete the Contributing Institution's metadata within 30 days.

5.3 Limitation of Liability -

In no event shall Stanford University, Stanford Libraries or any other organization that has participated in hosting Contributing Institution's metadata in the ePADD Discovery site be held liable for any loss, damage, lost profits, lost savings, or any other direct or indirect damages associated with the inaccuracy, incompleteness, or content of the metadata or to the use or inability to use either the metadata or the ePADD services, even if Contributing Institutions or their agents have been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause is in contract, tort, or otherwise. The agents and employees of Contributing Institutions are not authorized to modify warranties or representations, or the disclaimers thereof, or to make additional warranties or representations binding on Stanford University. Accordingly, additional statements, whether written or oral, do not constitute warranties of Stanford University and should not be relied upon as a warranty of Stanford University. Without limiting the foregoing, neither party shall be liable to the other for special, incidental, consequential or punitive damages of any nature, for any reason, including, without limitation, the breach of the agreement or any termination of this agreement, whether such liability is asserted on the basis of contract, tort (including negligence or strict liability) or otherwise, even if the other party has been warned of the possibility of such damages.

5.4. Contributing Institutions will be responsible for addressing any third party queries or claims relating to the third party's discovery or use of the metadata hosted by Stanford University. Stanford University will promptly forward any such claims or queries it receives to the Contributing Institution.

Appendix B

ePADD Discovery Module Roles and Responsibilities

This matrix defines the responsibilities of parties involved in maintaining and contributing to the ePADD shared Discovery Module, including Stanford University Libraries (SUL) staff and representatives from contributing institutions. Within SUL there are two sub-groups with different responsibilities, the ePADD Team and the Digital Library Systems and Services (DLSS) group.

Task	SUL - ePADD Team	SUL - DLSS	Contributing Institution
Maintain ePADD Discovery servers, including performing necessary security patches		x	
Monitor ePADD Github issues and project email account for communications	x		x
Facilitate transfer of email collection metadata to SUL	x		
Manage email collection metadata on ePADD Discovery servers including uploading and deleting metadata	x		
Test ePADD Discovery Module functionality	x		
Communicate information about server outages to contributing institutions	x	x	
Backup and recovery of ePADD Discovery servers		x	
Monitor ePADD application	x		
Upgrade ePADD Discovery standalone application when new versions are released	x		
Provide ePADD collection metadata to SUL			x

![ePADD logo]

Update ePADD collection metadata when necessary			**x**
Evaluate takedown requests			**x**

Appendix C

ePADD Collection Metadata Guide

This table provides suggested guidelines for ePADD collection metadata. This metadata is input by the processing archivist in the Processing Module.

Field	Definition/Specifications	Example
Landing Page image	The image that appears on the ePADD discovery site for the collection The image should have a 4:3 aspect ratio	
Institution	High level organization/institution	Stanford University
Repository	Sub-repository or library	Special Collections
Collection Title	Personal or organizational name followed by the phrase "email archive"	Robert Creeley email archive
Short Title	Personal or organizational name	Robert Creeley
Collection ID	ID for the collection as assigned by the institution	M0062
Finding Aid Link	URI to the online finding aid	http://www.oac.cdlib.org/findaid/ark:/13030/kt2k4035s1/
Catalog Record Link	URI to the online catalog record	https://searchworks.stanford.edu/view/4

		085328
Contact Email Address	Address where you would like questions/comments from users to be sent to	specialcollections@stanford.edu
Short Description	Text that appears on the collection card to describe the collection In the style of a Wikipedia disambiguation description of a person Max characters - approx. 75	Stanford professor emeritus, computer scientist, and mathematician American poet and author
About	General description of the creator of the collection, similar to a Biographical or Historical Note	Recognized as a seminal figure of American letters in the second half of the 20th century, Robert White Creeley was born in Arlington, Massachusetts, on May 21, 1926, attended the Holderness School and then Harvard College. He received degrees from The Black Mountain College (B.A., 1956) and the University of New Mexico (M.A., 1960).
Rights and Conditions	Information about the copyright status of the collection and conditions of use	Property rights reside with the repository. Literary rights reside with the creators of the documents or their heirs. To obtain permission to publish or reproduce, please contact the Public Services Librarian of the Dept. of Special Collections.
Notes	General note that can be used for any purpose designated by the institution	Email messages in the ePADD Discovery Module have been redacted to ensure the privacy of donors and other correspondents. Please contact the host repository if you would like to request access to full messages, including any attachments.
Scope and Content	Information about the nature of the materials and activities reflected in the collection	This email collection spans messages from January 1999 through January 2019. The messages reflect Knuth's

		work in the study and teaching of computer programming, computer systems, mathematics, and organ performance.

Appendix D

Repository Metadata Guide

Field	Definition/Specifications	Example
Repository Logo	Institution or Repository Logo image file	
Repository Name (optional)	Name of sub-repository within the institution	Department of Special Collections
Institution Name	Name of parent institution	Stanford University
Library Name & Location	Physical location of repository reading room.	Green Library, Stanford, CA, 94305-6004
Contact Phone (optional)	Preferred contact phone number for reference inquiries	(650) 725-1161
Contact Fax (optional)	Preferred contact fax number for reference inquiries	(650) 723-8690
Contact Email (optional)	Preferred contact email for reference inquiries	specialcollections@stanford.edu
Repository/Institution Website	URL for repository or institution website	https://library.stanford.edu/spc/manuscripts-division
Repository/Institution About Information	Text describing repository or institution. Maximum of 500 characters.	The Collection Services division in the Department of Special Collections at

		Stanford University arranges, describes, preserves, and makes available documents and born-digital materials acquired by subject specialists across disciplines in the Stanford Libraries. These are records of enduring historic value primarily acquired to support the research needs of the undergraduates, graduate students, faculty and other scholars at Stanford University and beyond.