



**Smithsonian**  
*Libraries and Archives*

# **DArcMail**

---

*Digital Archiving of Email*

## **Users Guide**

## TABLE OF CONTENTS

<b>ABOUT DARCMAIL .....</b>	<b>3</b>
<i>Issues of scale</i> .....	3
<i>Open Source</i> .....	3
USES FOR DARCMAIL .....	4
<i>Sample Workflows</i> .....	5
IMPORTANT NOTES.....	6
<b>THE DARCMAIL DATABASE .....</b>	<b>6</b>
<b>LAUNCHING DARCMAIL.....</b>	<b>7</b>
USING DARCMAIL .....	7
<i>Open DArCMail</i> .....	7
<i>Logging into the DArCMail database</i> .....	7
NAVIGATING THE DARCMAIL WINDOW .....	8
LOADING DATA INTO THE DARCMAIL DATABASE .....	8
<b>BROWSING DATA IN THE DARCMAIL DATABASE.....</b>	<b>10</b>
SEARCHING MESSAGES.....	16
SEARCHING FOR ADDRESSES OR NAMES.....	22
DELETING DATA FROM THE DARCMAIL DATABASE .....	24
EXPORTING DATA FROM THE DARCMAIL DATABASE .....	25
<b>INSTALLING DARCMAIL .....</b>	<b>26</b>
PREREQUISITES .....	26
<b>DARCMAILXML .....</b>	<b>30</b>
RUNNING DARCMAILXML .....	30
XML FILE NAMES .....	31
CSV OUTPUT.....	31
<b>CMDDARCMAILXML.....</b>	<b>32</b>
RUNNING CMDDARCMAILXML .....	32

## About DArcMail

The **DArcMail** application builds upon the functional core of the Collaborative Electronic Records Preservation (CERP)<sup>1</sup> Parser for the preservation of email messages and attachments. DArcMail delivers a broader range of features and functions and does so on the Python platform. In addition to preservation functionality, DArcMail supports appraisal, processing, and providing access to email accounts or large groups of related email messages. DArcMail's design shrinks processing time and enables it to handle large email accounts on the three most popular operating systems: Windows, Mac OS, and Linux.

Organizations that are only interested in preserving email accounts to Email Account<sup>2</sup> compliant XML may prefer to use **DArcMailXML** (graphical user interface) or **CmdDArcMailXML** (command line user interface).

### *Issues of scale*

In digital preservation circles today, the effective capacity of a tool or a workflow is sometimes described as handling “large amounts” of email. Without actual numbers, a capacity to handle “large amounts” of email is meaningless and does not help the digital preservation practitioner or administrator select a tool suited for their own context.

For clarity's sake, DArcMail is designed to scale, handling anything from a single email message to an account containing over a quarter of a million messages with ease. The Smithsonian Libraries and Archives regularly accessions the email correspondence of key Smithsonian staff, acquiring their whole email account at the time when such a staffperson leaves the Institution. More often than not, Smithsonian museum directors, senior curators and senior scientists may have more than 100,000 emails in their account when acquired.

### *Open Source*

**DArcMail** and **DArcMailXML** are available as open-source software under the MIT license (<http://opensource.org/licenses/MIT>). Its two platform components, Python and SQLite<sup>3</sup>, are also open source, continuing in the spirit of the CERP and the Smithsonian Libraries and Archives' commitment to sharing digital preservation resources. The

---

<sup>1</sup> <https://siarchives.si.edu/oldsite/cerp/index.htm>

<sup>2</sup> Email Account XML Schema (EAXS) [https://raw.githubusercontent.com/StateArchivesOfNorthCarolina/tomes-eaxs/master/versions/1/eaxs\\_schema\\_v1.xsd](https://raw.githubusercontent.com/StateArchivesOfNorthCarolina/tomes-eaxs/master/versions/1/eaxs_schema_v1.xsd)

<sup>3</sup> <https://www.sqlite.org/about.html>

DArcMail application utilizes Python and SQLite for the full set of functionalities. Both Python and SQLite are well established open source tools available on Windows, Macintosh, and Linux platforms. Previous versions of DArcMail also worked with MySQL.

## Uses for DArcMail

DArcMail is intended primarily for the processing, preservation migration, and subsequent access of email accounts or groups of related emails. It requires that the email collection is in the MBOX format.

- Appraisal

Searching through an email account using DArcMail helps an archivist develop a more detailed knowledge of the account owner's use of email, their correspondents, focal topics, etc. Simple and complex searching is available along with filtering and thread tracking features. Individual messages can be identified for exclusion before finalizing the set of emails to be accessioned. A log of senders and subjects can also be generated should the archivist find this helpful in making their determination.

- Processing

Like its usefulness for appraisal, DArcMail's processing capabilities lie in its ability to identify groups of messages based on a wide variety of fielded criteria. Correspondents, subject and body text content, topics, time frames, email threads, attachments – all of these criteria can be used to examine, filter and analyze the full email account or subsets of its messages. The knowledge gained will facilitate the creation of a detailed finding aid or narrowing the scope of the messages to be formally acquired by the organization.

- Preservation

DArcMail's preservation function transforms the original MBOX-formatted email account into XML as the preservation master format of the accessioned email account. Each MBOX file from an account gets its own corresponding XML file.

The XML format is compliant with the Email Account XML Schema<sup>4</sup> (EAXS) and can be validated by most XML editors. The EAXS-compliant file captures every aspect of the email account: 1) internal folder structure and relationships; 2) email message content and associated transmission headers; and 3) messages and

---

<sup>4</sup> [https://raw.githubusercontent.com/StateArchivesOfNorthCarolina/tomes-eaxs/master/versions/1/eaxs\\_schema\\_v1.xsd](https://raw.githubusercontent.com/StateArchivesOfNorthCarolina/tomes-eaxs/master/versions/1/eaxs_schema_v1.xsd)

their attachments. In addition, DArcMail generates a Message Summary log which contains the fields for Date, To, From, Message ID, Subject, and SHA1 Hash<sup>5</sup>.

- Access

DArcMail can be used to provide access to email collections in a reading room setting or for reference staff. The relevant functionality appropriate to this use is the same searching and filtering capabilities described in the previous Appraisal and Processing subsections.

### *Sample Workflows*

#### Acquisition to Preservation

1. Email account is chosen for appraisal.
2. The original account file is scanned for viruses outside of DArcMail suite.
3. If the email account file format is not MBOX, the format is normalized to MBOX outside of DArcMail.
4. The MBOX-formatted account is imported into DArcMail.
5. The archivist reviews the account's messages in DArcMail and exports the email messages that will be accessioned as a set into an MBOX formatted file.
6. Archivist uses the DArcMailXML component to perform the preservation migration from the accessioned MBOX file to the preservation XML file along with the Message Summary logs and a directory of any associated message attachments.

#### Processing / Access

1. Email account's MBOX file/s is imported into DArcMail.
2. Archivist searches through and reviews messages in DArcMail.
3. If the archivist wishes to provide a subset of messages to a researcher, they will create a filtered set of messages using DArcMail as a derivative MBOX file.

---

<sup>5</sup> A message-specific fixity value using the SHA1 algorithm. <https://en.wikipedia.org/wiki/SHA-1>

### *Important Notes*

**Message content.** The content of any given message may be small or large. With multiple attachments, the total content size for a single message can exceed 15 megabytes. Storing all message content for a large account that may contain more than 100,000 messages in the database may be unwise. The main reason for this is not the number of messages, but the potential size of their parts.

.xml files for email attachments are produced by the DArcMailXml conversion process. For example, a Microsoft PowerPoint (ppt) file that was encoded in base 64 initially would be XML encoded, remaining in base 64.

**Folder directory.** Every email account will have one or more folders, with one MBOX file for each email folder. For a given email folder, the directory in which the MBOX file for that folder is stored is the "folder directory" for that folder. If the MBOX file is located directly under the account directory, then the account directory and the folder directory are the same. If you want, you can place multiple MBOX files in the same directory, although for larger email accounts we recommend having each email folder in a separate subdirectory of the account directory.

## The DArcMail Database

If you only want to convert MBOX files to XML using the DArcMailXml or CmdDArcMailXml programs, then you do not need an SQLite database. If you want to search an account for messages, view message content, and export message subsets, then you need the DArcMail program and database. The DArcMail program will work with SQLite as the database.

## Launching DArcMail

### Using DArcMail

***Note: Installation directions are at the end of this document***

#### *Open DArcMail*

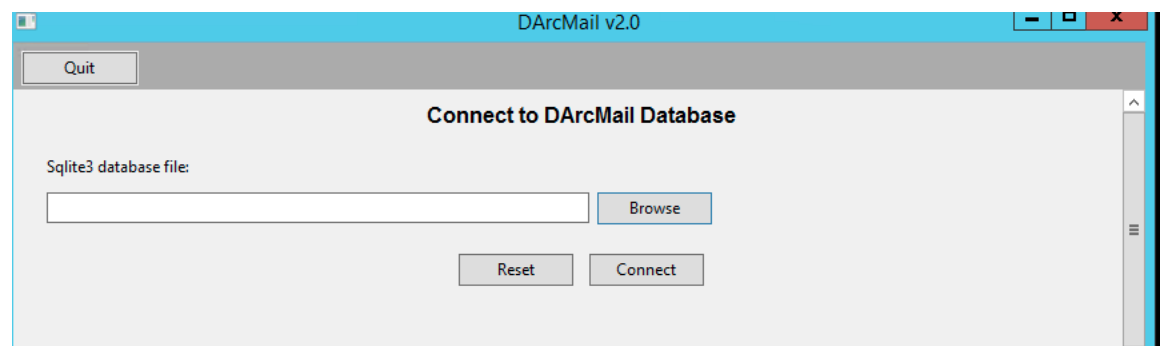
The main DArcMail program is the file DArcMail.py. Create a shortcut for this file and place it on your desktop. If the .py extension has been associated with the Python interpreter (on Windows) or with the Terminal application (on Mac OS X), then you can double-click the file to start the program. If this has not been done, do so following the operating system specific instructions available from Microsoft or Apple. Otherwise, you can start the program from a Windows command prompt window or from a Mac OS X Terminal window.

#### *Logging into the DArcMail database*

The connection screen is the first screen that appears when you start the DArcMail program. When using for the first time, browse to a location (desktop or other) and create a name for the SQLite file. Then select connect. No username or password is required.

An email account is stored in one SQLite database file, and a single SQLite database file can contain only one email account, which can be multiple MBOX files.

An existing SQLite DArcMail file can be used to hold data from another account and can be reloaded at a later time. It is recommended each account have its own SQLite file though since only one account can be searched at a time.



## Navigating the DArcMail Window

After you have successfully logged into the DArcMail database, then the program will show buttons at the top of the window for quitting, loading, deleting, exporting, and browsing data.

## Loading Data into the DArcMail database

The "Load DArcMail Data" screen asks for the following information:

DArcMail v2.0

Quit Load Delete Export Browse

### Load DArcMail Data

Account Name:

Account Directory:  
 Browse

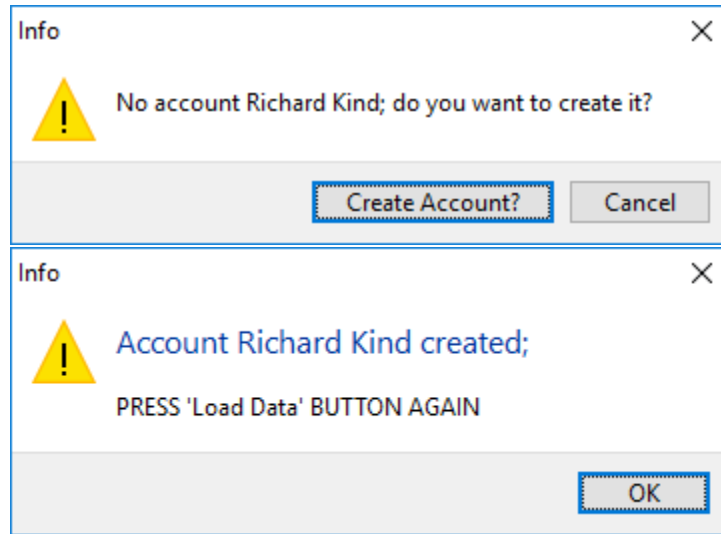
Select one or all folders: ALL FOLDERS ▾

Store attachments externally?  
☐ Yes ☒ No

Reset Load Data

**Account Name.** This is the required DArcMail name of the email account. This can be the email address that is associated with the email account (e.g., "john.smith@jamestowncolony.com") or any other format (e.g., "John Smith", "Smith, J; Jamestown", etc.). If the email account name does not already exist in the DArcMail database, it will be created in the loading process.



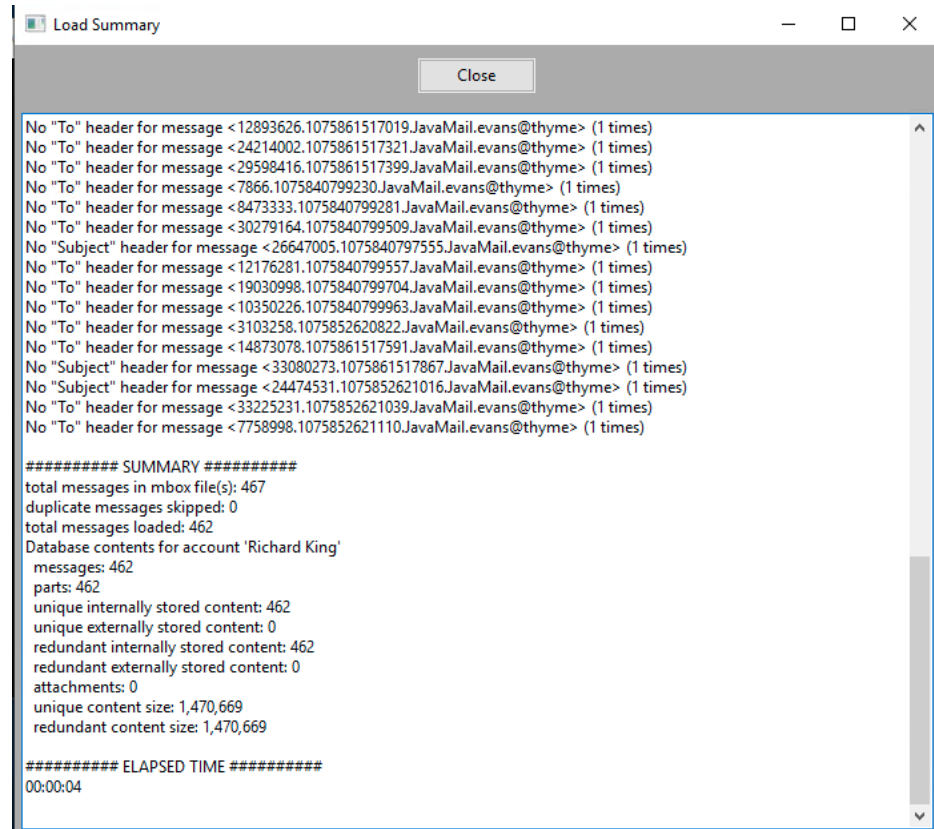


**Account Directory.** This is the directory where the MBOX file/s is located. This must already exist and must contain one or more MBOX-format files with the .mbox file extension.

**Select one or all folders.** This is the option to process all MBOX files or just a specific one, e.g., FY2018 Staff Memos.mbox if you only want the FY2018 Staff Memos to be processed.

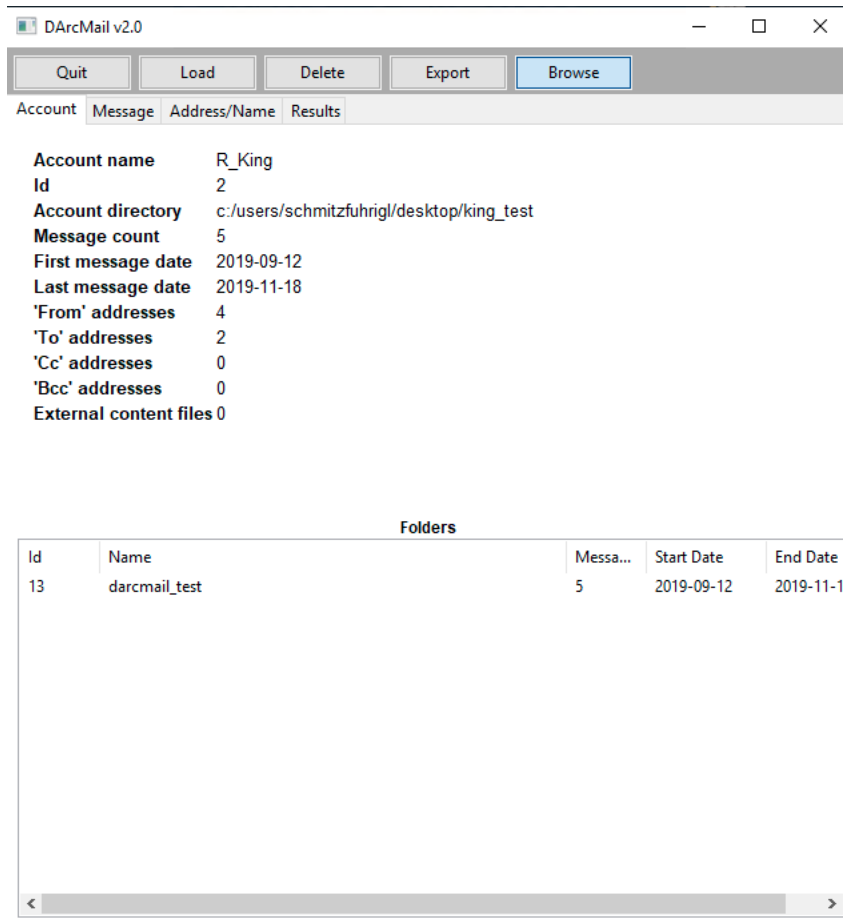
**Store attachments separately.** If the option is set to no, data will be stored within the one SQLite file. If the account is large, the SQLite file will be large. If the option is set to yes, attachments will be stored separately in folders as RAW files.

Once loaded, a summary will be displayed. It displays issues it encounters, as well number of messages processed, duplicate messages that were found, and the number of attachments. If you want to go back and re-upload the account with different settings, you must first delete the existing account in DArCMail.



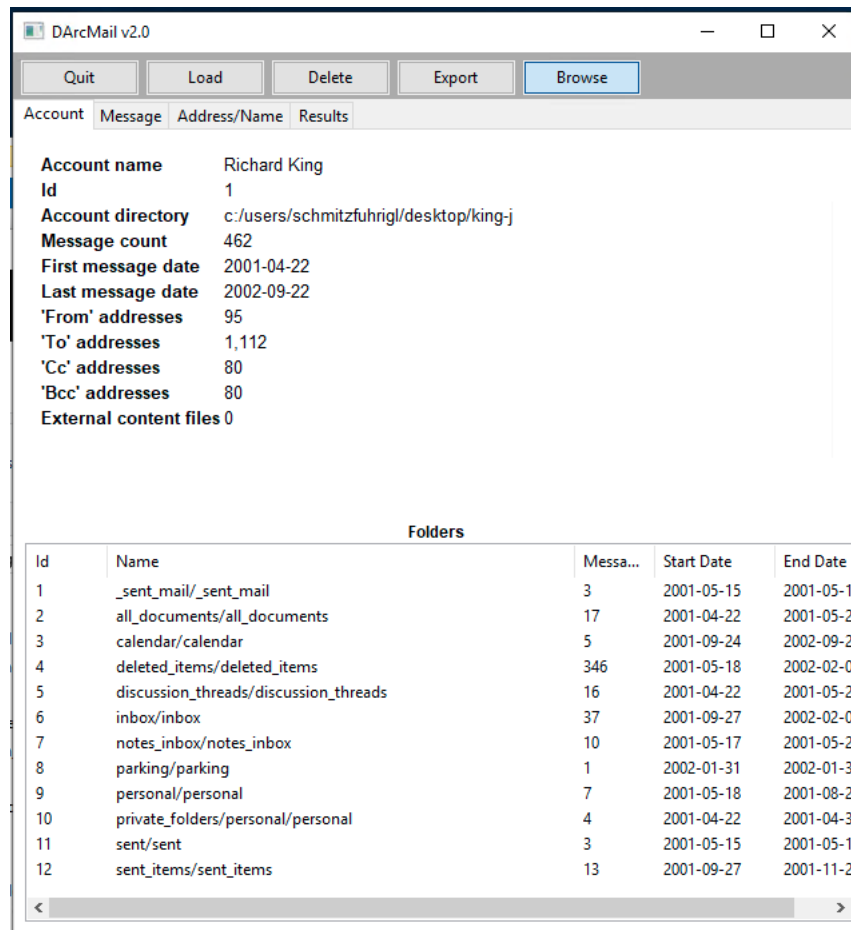
## Browsing data in the DArcMail database

Select the Browse tab to search the account. You can search for accounts, messages, email addresses, and email names. The result of a search is an account list, a message list, or an email address/email name list. Selecting one row of a list will result in an information page for the individual account, message, or email address. Lists and information pages appear under the Results tab. Note that only one account/collection can be searched at a time.



There is a 1,000 results limit on the number of items that can be displayed in a list; if an email address list or message list exceeds 1,000, only the first 1,000 addresses or messages will be shown. If the result list is longer than 30 items, additional tabs will be created for each additional group of 30. This is done in order to preserve fast response times.

When you select the Browse button on the main button bar, you will see four tabs immediately below the main bar that are Account, Message, Address/Name, and Results:



The Account tab provides information about the account, including number of messages, date range, and folder names. At the bottom is shown a further breakdown of information for each individual folder, including the number of messages and start/end dates for each email folder.

DArcMail v2.0

Quit Load Delete Export Browse

Account Message Address/Name Results

Sort Order ☒ Oldest first ☐ Newest first

Message status ☒ Any ☐ Selected ☐ Unselected

Global Id

Date From

Date To

Folder ALL FOLDERS ▾

Subject Line Memo

From Line

To Line

Cc Line

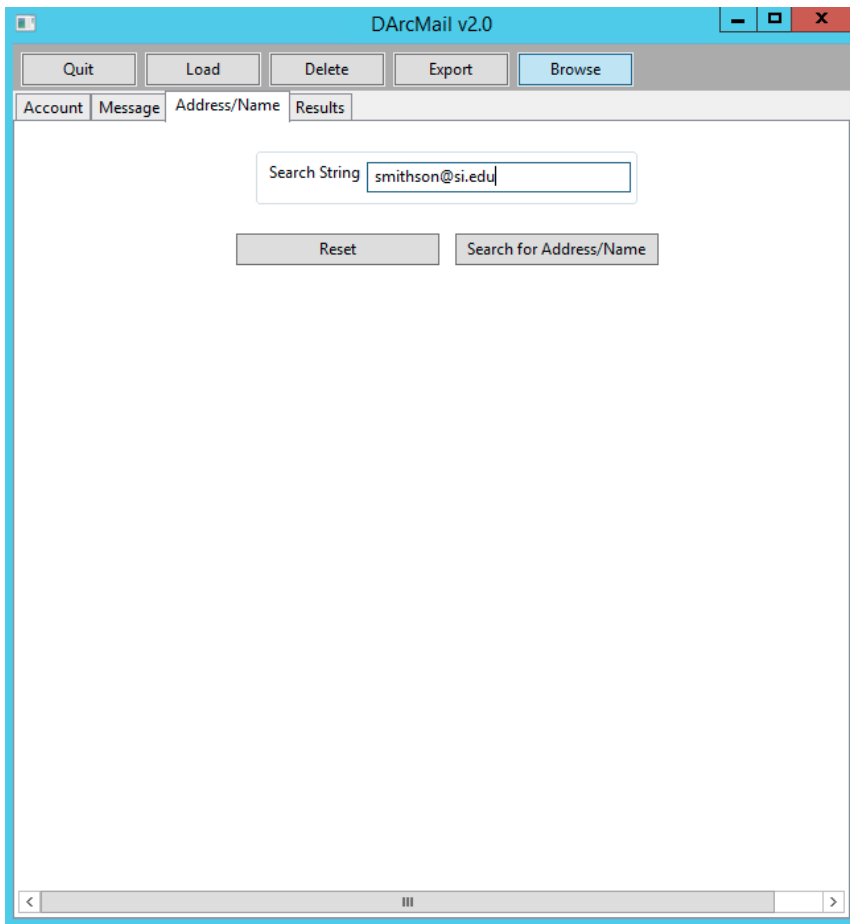
Attachment Name

Body Text

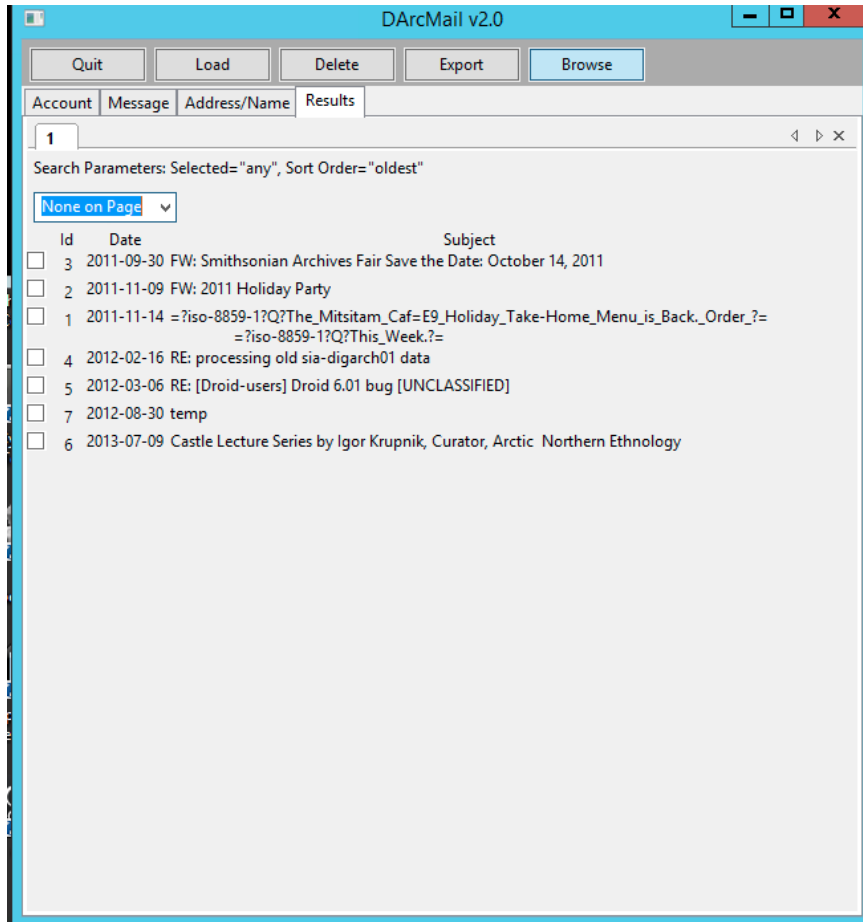
Search body text: ☒ text/plain ☐ text/html

Reset Search for Messages

The remaining tabs are Message



Address/Name tab



The Results display is what the search found in a list and then the message itself. A list or information page remains under the Results tab until you close it with the X in the upper right. If you click on the Id number of a message, a new tab will open with the message itself (see Message Information Page on Page 18)."

## Searching Messages

You can search for messages using any combination of the search fields in the search form. If you supply values for two or more of the search boxes, then the search must match all of the values you supply. This behaves like a Boolean “AND” search query.

All text fields are automatically wild-carded at the beginning and end.

The "Date From" is wild-carded as follows: A year only entry "2014" is interpreted as "2014-01-01." A year-month entry "2014-03" is interpreted as "2014-03-01".

The "Date To" field is wild-carded as follows: A year only entry "2014" is interpreted as "2014-12-31." A year-month entry "2014-03" is interpreted as "2014-03-31".

"From," "To," and "Cc" fields will search both email addresses and email names.

Text searches are not case sensitive.

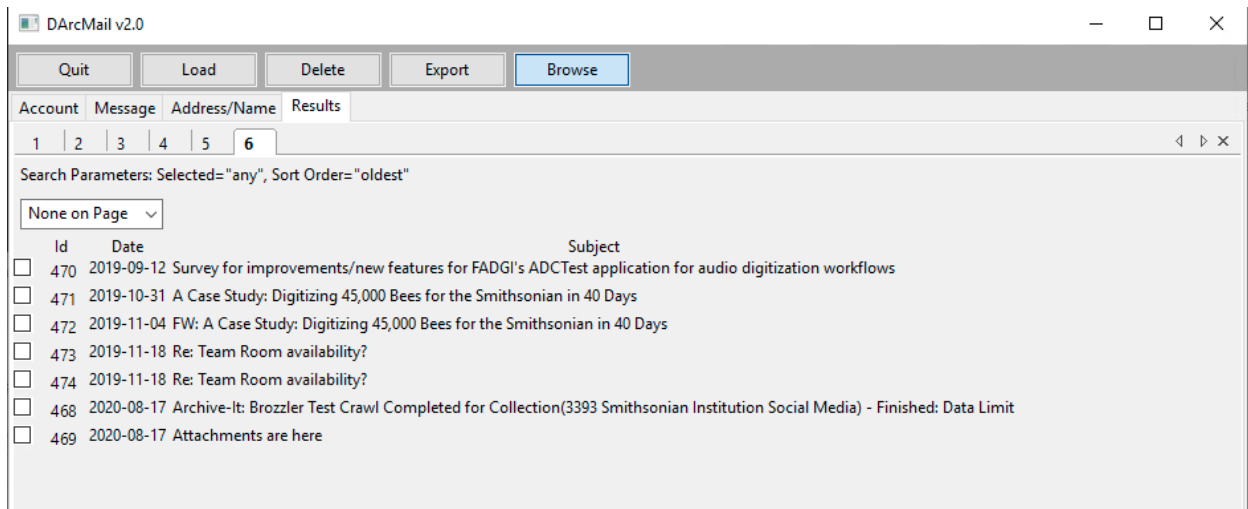
You can search attachment names even if the attachment itself is stored externally. As a text field that is automatically wild-carded, if you supply the string ".jpg" in the "Attachment Name" field, the search will find all messages with at least one part that is an attachment having ".jpg" in the attachment name. Note that quotes are not required for searching.



If you supply a search string for the "Body Text" field, only messages with Content-Type text/plain and text/html will be searched, and only content that is stored internally will be searched.

Message content is stored (and therefore searched) in its original transfer encoding. Content-Type "text/html" is usually encoded as "quoted-printable"; this may affect your search results. In addition, if you specify a search of "text/html" content, your search may match html tags which are not in fact visible in the rendered html. A message part with Content-Type "text/html" is usually accompanied by a part containing the same information with Content-Type "text/plain." For these reasons, you will probably prefer to restrict your search to "text/plain."

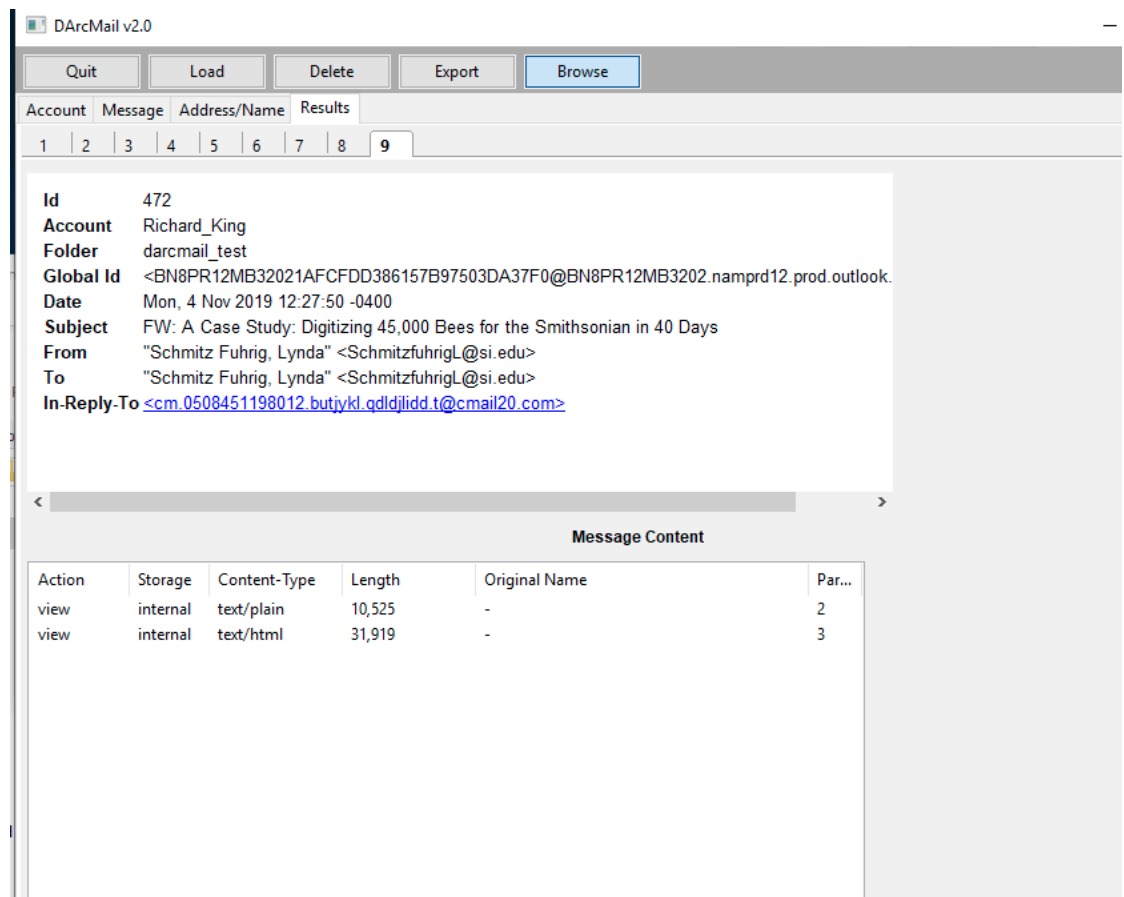
If you do not supply any search string, the search will display all messages. It sorts them by oldest date.



Results of a search. To view the message, click on the Id number.

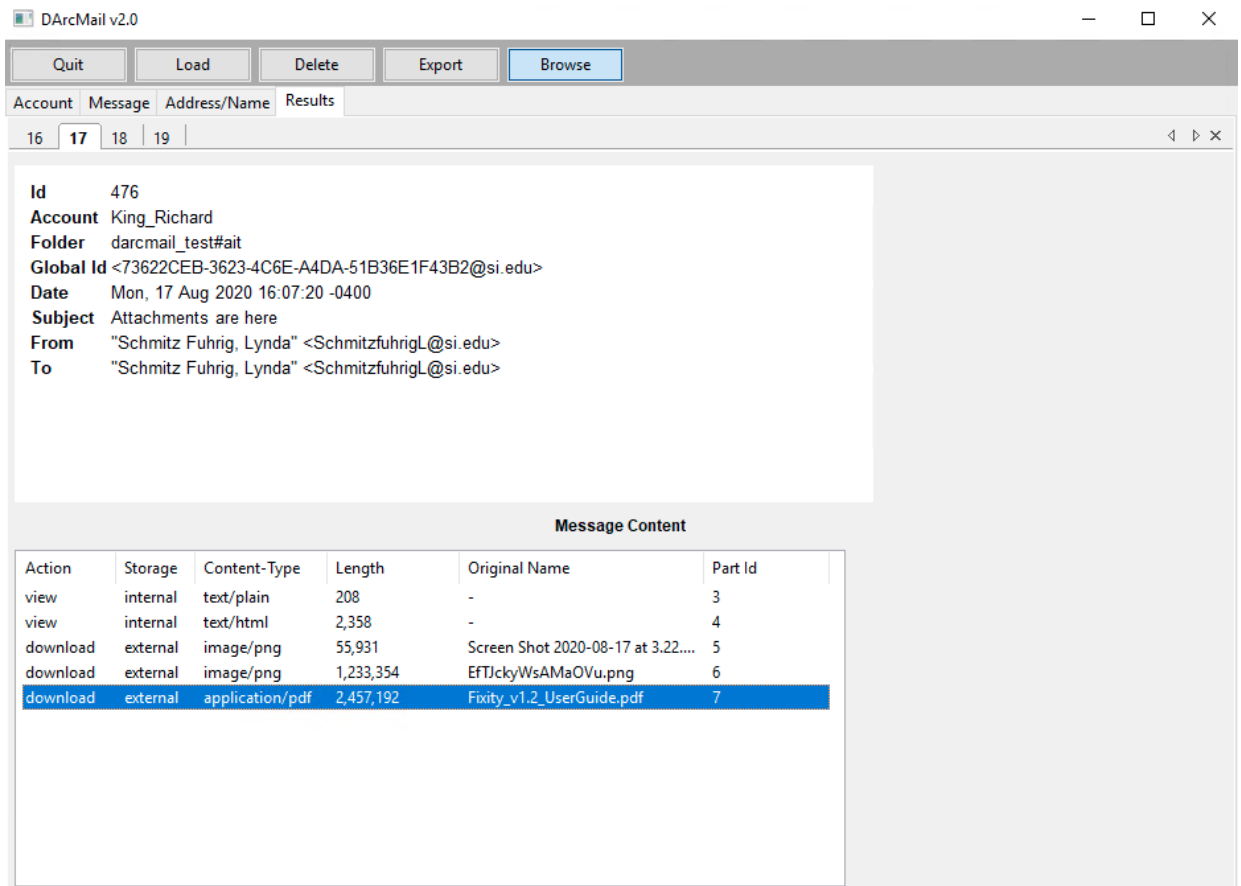
## Message information page

A message information page has two parts:



The top part of the page has attributes of the message, including information from the MBOX header lines "Message-Id:", "Date:", "Subject:", "From:", "To:", and if present, "Cc:", "Bcc:", and "In-Reply-To." If the message has an "In-Reply-To" header and if the replied-to message is in the DArcMail database, then the message ID of the replied-to message is hyperlinked to its message information page.

If the message itself has replies, then the text "Get Replies" (or "Get Reply") is hyperlinked to a list of the replying messages.



### Message Content List

The bottom part of the page is the Message Content list. An MBOX message may have just one body part or it may have multiple body parts (with a Content-Type of multipart/alternative or multipart/mixed or multipart/related or multipart/report, etc). If a message is multipart, then its first part is simply the multipart container whose contents are the other parts. A multipart message part (a "container" part) does not appear in the message content list.

The message content list shows these fields for each non-empty, non-multipart message part:

- Action. This is “view,” “download,” or “show path”.
  - o If the Content-Type of the message part is “text/html” or “text/plain” AND either
    - (a) the Storage is “internal” OR
    - (b) Storage is “external” and the external file is visible from the computer on which the DArcMail application is running, and the external content has NOT been wrapped in XML,
 then the action is “view,” meaning that you can view the content directly in the DArcMail application.
- If the Content-Type is not “text/plain” or “text/html” AND either
  - (a) the Storage is “internal” OR
  - (b) Storage is “external” and the external file is visible from the computer on which the DArcMail application is running,
 then the action is “download.”
- If the Storage is “external” and the external file is NOT visible from the computer on which the DArcMail application is running, then the action is “show path,” meaning that DArcMail will show the full path of the external file in a separate window.

If a message part has a Content-Transfer-Encoding of “base64” or “quoted-printable,” DArcMail will decode the content to its original representation before viewing or downloading. However, if the Storage is external and if the external content has been wrapped in XML, then the content will not be decoded.

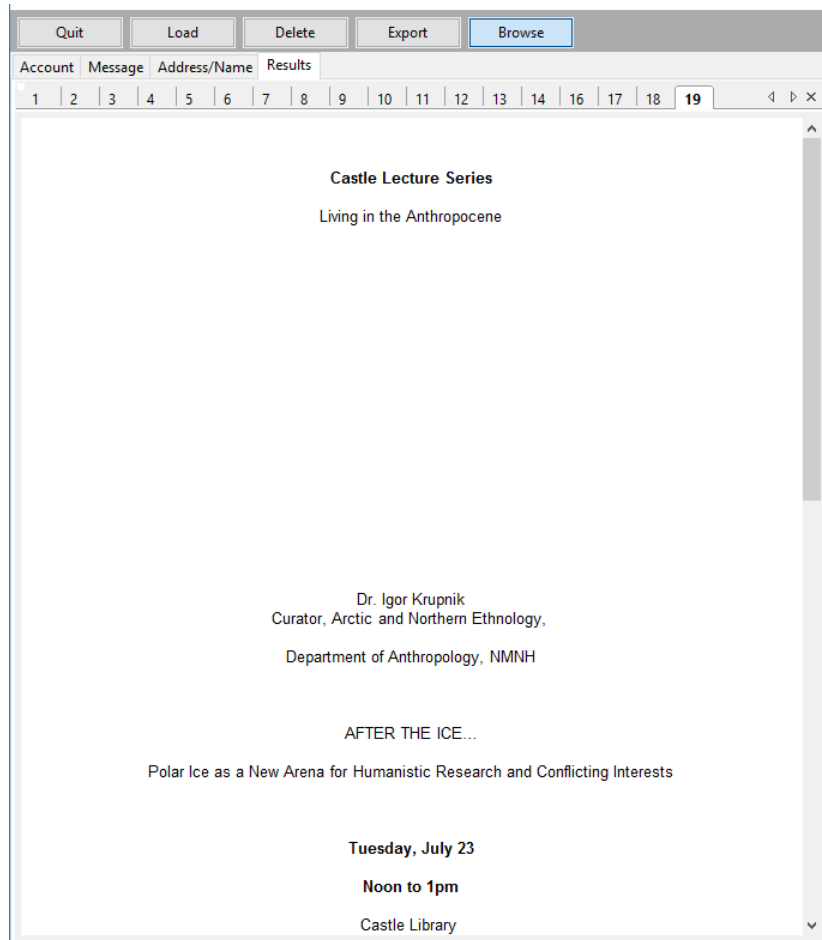
**Content-Type.** This is the MIME type of the message part, as given in the Content-Type header of the message part. If a message part has Content-Type “text/html,” the preceding or following message part may be a “text/plain” rendering of the same information.

**Length.** This is the length of the content in bytes.

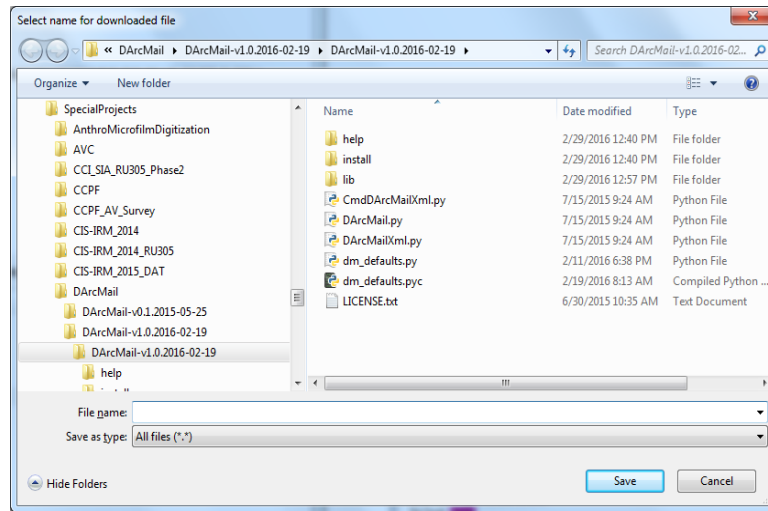
**Original Name.** DArcMail preserves the original names of attachments.

**Part Id.** This is the sequential part number assigned in the original MBOX format input. If there is a numbering gap in the Part Id column, then the missing part was a multipart “container,” or it was a part with zero length.

Click on the desired row in the Message Content window to view the part's contents. When a message part is viewed, it will appear in another numbered tab and display as indicated below.



If the action specified in the message content list is “download,” you will be prompted to select a location.



## Searching for Addresses or Names

An email address is a string like "john.smith@jamestowncolony.com." An email name is a string like "Smith, John" which might appear paired with an email address in a message header, like this:

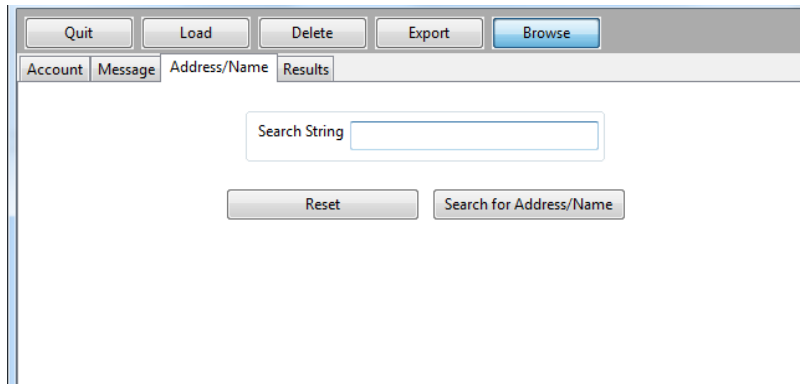
To: Smith, John <john.smith@jamestowncolony.com>

Email names are supplied by contact directories, which may differ from user to user, so it is possible for a unique email address to be associated with multiple email names.

To: Cristoforo Colombo <columbusc@expeditions.es>

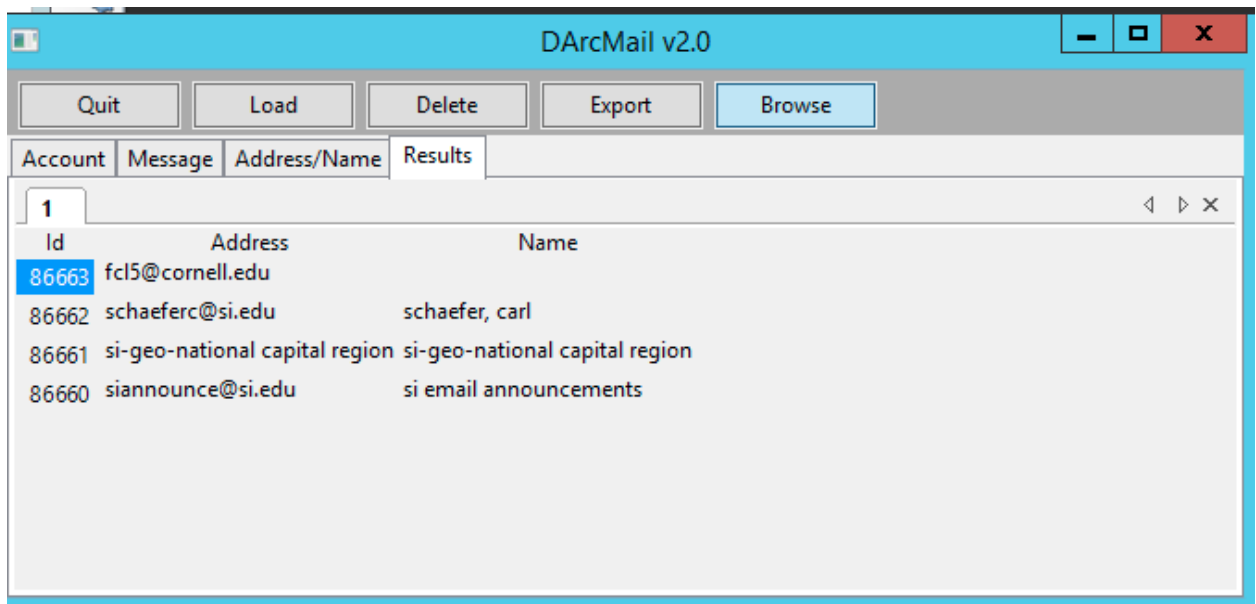
To: Columbus, Christopher <columbusc@expeditions.es>

Note: It is possible that an MBOX export of a message may have an email name that is not paired with an email address.



### Searching for an email address or email name

- You can search on a partial or complete email address or email name.
- The search string you supply will be automatically wild-carded at the beginning and end.
- Searches are not case sensitive.
- If you do not supply any search string, the search will match all email addresses used in the account.
- If a search string matches an email name but does not match an email address, then the search will return a result only if the email name is paired with an email address in at least one message.



### Address information page

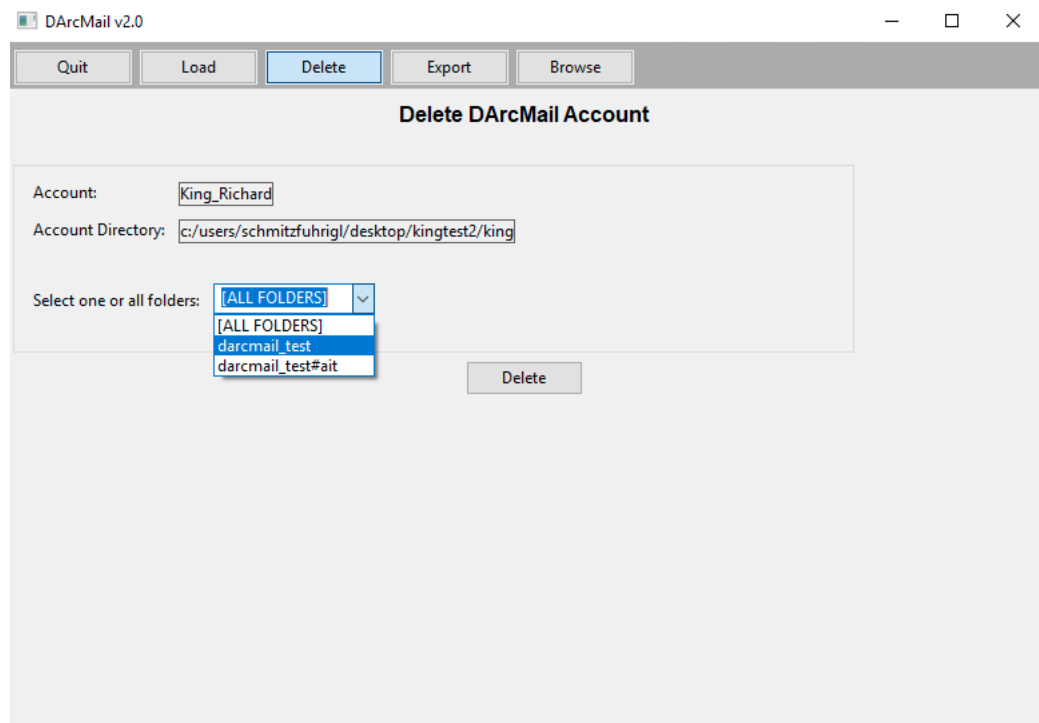
Similar to the message information page, the address information page displays attributes of the email address. Counts of the number of times an email address appears in a From, To, Cc, or Bcc field are linked to corresponding message lists.

## Deleting data from the DArcMail database

You can delete an entire email account from the SQLite file, or you can delete a single email folder from an account. In the latter case, this removes that folder's information from the database, as if it had never been loaded. It does not modify the original MBOX file content. If you then export this filtered version of the account from DArcMail, the new MBOX file generated will not include the deleted folder.

Choose the Delete button. The Delete DArcMail Account screen asks for the following information:

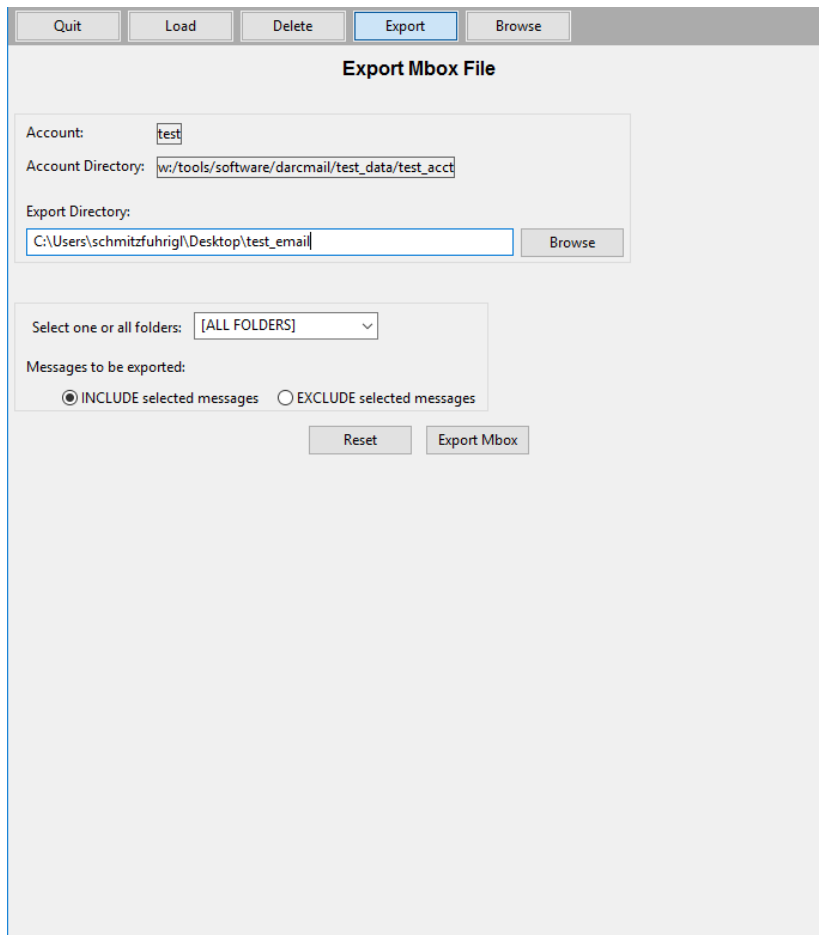
- Name. This is the DArcMail name of the email account and will be auto populated.
- Folder. If no folder is specified, then the program will delete the email account and all its data from the DArcMail database. Only one folder can be deleted as well. For example, if `"/archive/sent_items "` is specified, then the program will delete only message data that was loaded from the MBOX file named `"/archive/sent_items"` and not `"/sent_items."` A Delete Summary window will appear when the action has finished.





## Exporting data from the DArcMail database

To export content from DArcMail, choose the Export button.



The screenshot shows the 'Export Mbox File' dialog box. At the top, there is a toolbar with buttons: 'Quit', 'Load', 'Delete', 'Export' (highlighted in blue), and 'Browse'. The dialog box has a title bar 'Export Mbox File'. Inside, there are several input fields and controls:

- Account:** A text box containing 'test'.
- Account Directory:** A text box containing 'w:/tools/software/darcmail/test\_data/test\_acct'.
- Export Directory:** A text box containing 'C:\Users\schmitzfuhrig\Desktop\test\_email' and a 'Browse' button to its right.
- Select one or all folders:** A dropdown menu showing '[ALL FOLDERS]'.
- Messages to be exported:** Two radio buttons: 'INCLUDE selected messages' (selected) and 'EXCLUDE selected messages'.
- At the bottom right, there are two buttons: 'Reset' and 'Export Mbox'.

Specify the location where the exported MBOX file will go under “Export Directory.” Select [ALL FOLDERS] or choose one of the folders in the email account. Indicate if the selected messages are to be included or excluded from the export. Excluding messages may be more useful when the undesired messages are a small number in proportion to the account as a whole.

## Installing DArcMail

### Prerequisites

DArcMail runs on Python, using SQLite3. wxPython also is needed.

In order to use DArcMail, you must download both Python and the GUI toolkit wxPython. If you are interested in only using the command line version, DArcMailXml, only Python is required. DArcMail will run on Windows, Mac, and Linux (Ubuntu) systems. Installing Python3 (version 3.6 or later) is required. See <https://www.python.org> for the downloadable installer files. DArcMail uses SQLite for its file-based database engine, which is packaged with standard Python distributions.

No special installation for SQLite is required beyond having Python3 installed. Having module wxPython (version 4.0 or later) installed is also required (see <https://wxpython.org>). wxPython can be installed directly from the command line using Python's pip tool, or by downloading directly from the website.

Please refer to the source websites for Python and wxPython for help with specific issues. The following instructions are meant to serve as general guidance for potential Python accessibility issues as they relate to DArcMail.

- Check the system type of your environment (32- or 64-bit). Python offers installation packages for both systems.
- Python offers three installation options: executable installer file (.exe), ZIP file, or web-based installer file. Pick whichever one that best suits your work environment.
- Check your settings when installing Python. To make it easier to install wxPython, we recommend making sure the pip tool is downloaded during installation, and that the checkbox for setting Python to PATH is checked.



- Test out Python first. Once you have installed Python, open the command line and type

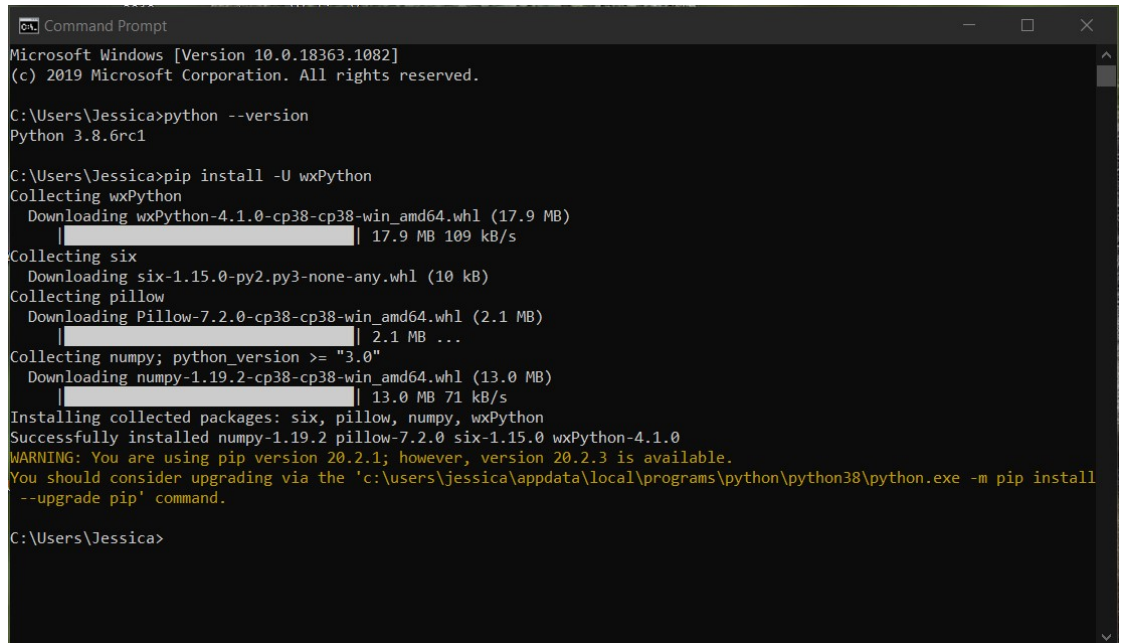
Python --version

The command line should print the version of Python you are running on your computer (e.g., Python 3.8.5).

- Install wxPython using pip. Type in the command line the following text:

pip install -U wxPython

You may receive errors about not having the latest version of wxPython, which can be ignored if wxPython downloads correctly (see screenshot of command line below).



```

C:\Users\Jessica>python --version
Python 3.8.6rc1

C:\Users\Jessica>pip install -U wxPython
Collecting wxPython
  Downloading wxPython-4.1.0-cp38-cp38-win_amd64.whl (17.9 MB)
    |#####| 17.9 MB 109 kB/s
Collecting six
  Downloading six-1.15.0-py2.py3-none-any.whl (10 kB)
Collecting pillow
  Downloading Pillow-7.2.0-cp38-cp38-win_amd64.whl (2.1 MB)
    |#####| 2.1 MB ...
Collecting numpy; python_version >= "3.0"
  Downloading numpy-1.19.2-cp38-cp38-win_amd64.whl (13.0 MB)
    |#####| 13.0 MB 71 kB/s
Installing collected packages: six, pillow, numpy, wxPython
Successfully installed numpy-1.19.2 pillow-7.2.0 six-1.15.0 wxPython-4.1.0
WARNING: You are using pip version 20.2.1; however, version 20.2.3 is available.
You should consider upgrading via the 'c:\users\jessica\appdata\local\programs\python\python38\python.exe -m pip install
--upgrade pip' command.

C:\Users\Jessica>

```

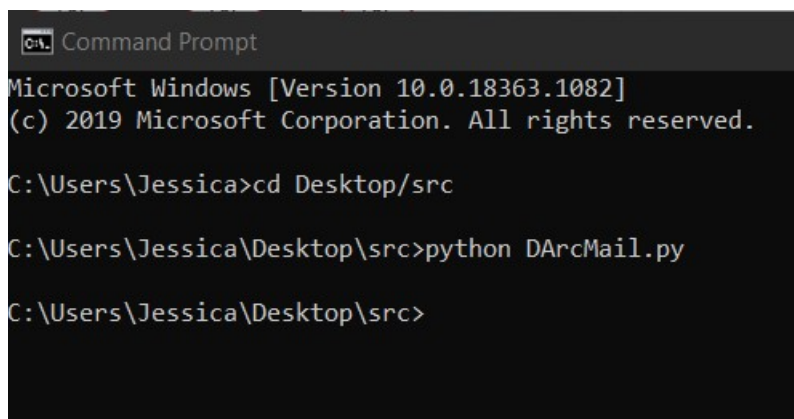
The DArcMail programs are in the src directory of the distribution. They can be copied to, and run from, anywhere on your computer. You can run them from where you have unzipped the distribution package, or you can copy all the src/\*.py files in the distribution to another location.

Wherever the distribution is stored, make sure the command line is pointed to the correct location on your computer in order to access DArcMail.

For example, if the folder src/\*.py is on your Desktop, type the following command into the command line to change the directory to this folder:

```
cd Desktop/src
```

Then type python DArcMail.py to open the application.



```

C:\Users\Jessica>cd Desktop/src

C:\Users\Jessica\Desktop\src>python DArcMail.py

C:\Users\Jessica\Desktop\src>

```

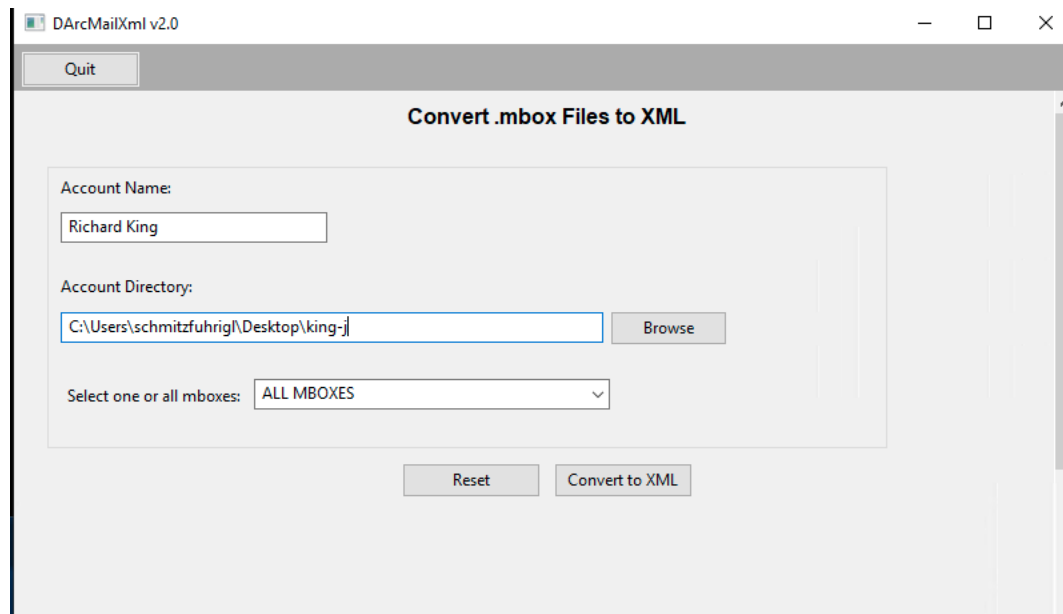
Another option for opening DArcMail is to double-click on the DArcMail file in the src folder.

In a networked environment, the DArcMail application scripts can be run on any computer that can access the SQLite file. In a standalone environment, the SQLite files must be present on the same computer as the DArcMail scripts.

## DArcMailXML

DArcMailXml allows you to convert one or more .mbox files to an XML file, without using a database. The XML follows the preservation EAXS (Email Account XML) schema, co-developed by North Carolina State Archives and Smithsonian Institution Archives. The schema is at

[https://raw.githubusercontent.com/StateArchivesOfNorthCarolina/tomes-eaxs/master/versions/1/eaxs\\_schema\\_v1.xsd](https://raw.githubusercontent.com/StateArchivesOfNorthCarolina/tomes-eaxs/master/versions/1/eaxs_schema_v1.xsd)



### Running DArcMailXml

The main DArcMailXml program is the file DArcMailXml.py. If the .py extension has been associated with the Python interpreter (on Windows) or with the Terminal application (on Mac OS X), then you can double-click the file to start the program. Otherwise, you can start the program from a Windows command prompt window or from a Mac OS X Terminal window.

*Note that issues have been detected when trying to run on a MacOS when there is an Anaconda platform installed.*

When launched, the "Convert MBOX Files to XML" screen asks for the following information:

**Account Name.** This is the name of the email account entered by the user.

**Account Directory.** This is the account directory. This must already exist and must contain one or more MBOX-format files with the .mbox file extension.

Each MBOX file will have a corresponding XML file with attachments stored externally. Each message attachment will be stored in an XML file separate from the main XML file. Each attachment file is located in a subdirectory of the folder directory and is referenced from the main XML file. These subdirectories are used to prevent the creation of directories with thousands of files.

### XML file names

If you convert an entire email account, then the name of the XML file/s will be the name of the MBOX files with the ".xml" file extension. If the MBOX name contains characters that are not permitted in a file name, then each occurrence of an illegal character will be replaced with an underscore. Characters that are replaced with an underscore include any character with a numeric value less than decimal 32 ( ' ') or greater than decimal 127 (~), as well as these characters:

/ ? < > \ : \* | " ^

The XML file/s for an account will be created in the account directory. There will be one XML file for each corresponding MBOX file so one large XML file is not created.

### CSV output

DArcMailXml produces a CSV log for each XML file. A row in the CSV file corresponds to one message of an XML file.

The CSV file has six columns:

From

To

Date

Subject

Message ID

SHA1 Hash

The file is in "comma-separated values" so that it can be loaded into a spreadsheet program like Microsoft Excel or OpenOffice Calc.

*Example: sent items.mbox will have sent items.xml and sent items.csv files.*

If errors are encountered parsing the email messages, the first error message encountered will be noted in another column in the CSV file.

## CmdDArcMailXml

### Running CmdDArcMailXml

CmdDArcMailXml is a command-line version of DArcMailXml. It has the same options and produces the same outputs as DArcMailXml and requires Python.

You can run CmdDArcMailXml from a MacOS X terminal window or from a Windows command prompt window. Command-line options are as follows:

Long Option	Short Option	Kind	Value Type	Meaning	Default
--help	-h			display command options	
--account	-a	Required	string	account name	None
--directory	-d	Required	string	account directory	None
--folder	-f	Optional	string	folder name	None

### CmdDArcMailXML Results

XML

MBOX

CSV (MessageSummary log)

External directory structure (contains all exported attachments as XML files)