

## Email Collection

### Description

**Email Collection** refers to the identification and isolation of electronic mail (email) messages that pertain to a specific legal matter in civil litigation cases.

### What gets collected

What is actually being collected during email collections can be one of two things:

1. Files representing the contents of the transmitted email messages themselves (usually in MSG, HTML, EML or RTF format).
2. Container (or store) files that hold the contents and data associated with multiple email messages, usually all of the emails for a specific custodian.

Whether files for individual emails or container files are collected depend mostly on the type of email system being used by the custodian. If the custodian is a user of Microsoft Outlook for instance, then either container files or individual email files may be produced. If the custodian is a user of a webmail service, such as Gmail or Yahoo!, then it is likely only individual email files can be collected.

### How it's done

Software such as [Harvester](#) from [Pinpoint Labs](#) can search the PST store files produced by Microsoft Outlook and Exchange email systems for individual emails containing specific criteria, such as who sent the email, who received it, when these actions occurred and whether the subject, body, or attachments contain specified key words. It can also produce the result to either individual email files or whole, reconstructed container files, known as [PST regeneration](#).

With other email systems, either the whole container file can be copied and sorted through manually, or the individual emails can be manually identified and exported as individual email files.

### What to remember

As with any data being collected, the two concepts to remember are preservation and validation.

Preservation refers to keeping the metadata about the individual messages as well as the metadata contained within each of the messages intact so as to maintain their admissibility. [PST regeneration](#) is especially desirable in this case because it maintains both the email data and the data that linked it to contact data, task list data and other data integrated with these types of email messages.

Validation refers to the policy of insuring, either by hash value comparison (analogous to fingerprints for

data) or bit-wise comparison, that the contents of the copy are the same as the contents of the original.

Software such as [Harvester](#) and [SafeCopy 2](#), both from [Pinpoint Labs](#), have built-in preservation and validation systems to certify that both of these conditions are always met.

**Date**

04/18/2025

**Date Created**

12/06/2010