

## MailConverter 2.1: Digests And Pseudo-Digests

Digests, ie collections of messages, come in three general flavors:

1. Those which follow the official rfc digest structure. These are true digests and can be parsed unambiguously back into their component messages.
2. Those which don't follow the official structure but which have simple, standard separators between messages (a line of dashes or a line of equal signs). Using a heuristic, MailConverter can do a reasonably good job with these. But it's important to note that this format is a pseudo-digest, not a true one, and carries with it the possibility of ambiguity, and if that happens, there's nothing MailConverter can do to get around it.
3. Those which have no separators at all between messages. These are the trickiest to convert, and have the most potential for ambiguity. These are also, of course, pseudo-digests. In this case, to keep the conversion within reasonable bounds, a small set of standard headers which typically begin messages (stored in the MailConverter Preferences file) is considered when looking for message boundaries, and also when examining the file initially to determine its format. Those headers are: Path:, Newsgroups:, Date:, From:, To:, Subject:, and Received:. To add to this list, you must use ResEdit (see below).

THERE IS NO WAY IN GENERAL FOR MailConverter TO KNOW WHICH CASE IT'S DEALING WITH! To minimize ambiguity, it depends on some help from the user.

Since type (3) is the messiest, MailConverter will not attempt this unless the No pseudo-digest separators option is on. Turn this option on to convert pseudo-digests with no separators between messages, turn it off otherwise.

MailConverter then proceeds as follows. It looks at the first line, and if it's a mail header, the file is assumed to be either type (2) or type (3), depending on the setting of the No pseudo-digest separators option (note that type 2 considers anything that looks vaguely like a header to be a header, whereas type 3 insists on one of a specified list). If the first line is not a header, it will look through the entire file for some indication that the file is a type (1).

The reason MailConverter considers type (1) last is that true digests will be converted more or less correctly by the type (2) and type (3) converters, but not necessarily vice versa; and because type (1) files, despite being the official standard digest format, don't occur all that often in the real world.

On the other hand, type (1) is clearly preferable when appropriate, since it's unambiguous. So if you have a file of type (1), it's a good idea to force type (1) conversion by ensuring that the first line is not a mail header. In Eudora, if you receive a type (1) digest as a message, you can get this to happen by using Save As without headers.

The question that remains is, how can type (1) files be recognized by the user? This is somewhat complicated to explain in general (see rfc934 for specifics), but typically they will contain a line that looks something like this:

----- start of digest (x messages) (RFC 934 encapsulation) -----

where x is the number of messages in the digest.

Adding headers with ResEdit for the no-separators case:

Open System Folder:Preferences:MailConverter Preferences. Then open the STR# resources, resource# 130. To add a new string, click on the last number in the window, then select Insert New Field(s) from the Resource menu, then tab in to the newly created empty box, then type in the field exactly as it would appear in a file, eg "Sender:" (don't include the quotes, of course). Do this as many times as you want for as many headers as you want. Then save the file (cmd-S). Be sure to close it before running MailConverter, and conversely, be sure not to open it in ResEdit while MailConverter is running: very bad things will happen if the settings file is open simultaneously in ResEdit and MailConverter...