

Network Working Group
Internet Draft
[draft-palme-text-html-issues-00.txt](#)
Category-to-be: None
Expires June 1996

Jacob Palme
Stockholm University/KTH
Sweden
December 1995

Issues on sending HTML documents via MIME e-mail

Status of this Memo

This document is an Internet-Draft. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as ``work in progress.''

To learn the current status of any Internet-Draft, please check the ``1id-abstracts.txt'' listing contained in the Internet-Drafts Shadow Directories on ftp.is.co.za (Africa), nic.nordu.net (Europe), munnari.oz.au (Pacific Rim), ds.internic.net (US East Coast), or ftp.isi.edu (US West Coast).

This memo provides information for the Internet community. This memo does not specify an Internet standard of any kind, since this document is mainly a compilation of information taken from other RFC-s.. Distribution of this memo is unlimited.

Abstract

This memo discusses some issues raised by [draft-palme-text-html-00.txt](#) "The Text/HTML content type and the Content-Location MIME header or Sending HTML documents via MIME e-mail" and tries to summarize the discussion on this document.

Table of contents

Issue 1: Syntax of embedding URL-s in message headers
Issue 2: Allow hyperlinks outside multipart/related or not?
Issue 3: Is multipart/related to be used at all
Issue 4: Name of the Content-Location header
Issue 5: Should the Base header (defined in [RFC 1808](#)) be renamed to "Content-Base"?
Issue 6: Relative URL-s and the Content-Base header
Issue 7: Relative URL-s referring to body parts within the same message
Issue 8: The message itself as a Base URL
Issue 9: Priority of bases
Issue 10: Ambiguity of "Content-Base"
Issue 11: Giving body parts names to be used in URL-s
Issue 12: How to indicate that relative URL-s refer to body parts?
Issue 13: Combination of Text/html and Multipart/Alternative
Issue 14: What should "start" refer to
Issue 15: Including remotely available objects in a message
Issue 16: The use of mid-s and cid-s
Issue 17: Content-Location or Content-Disposition

Issue 1: Syntax of embedding URL-s in message headers

Several different ietf memos require the embedding of URL-s in message headers:

- (i) The Content-Location as defined in "[draft-palme-text-html-01.txt](#)".
- (ii) The Base or Content-Base as defined in [RFC 1808](#)
- (iii) In definition of the URL access-type in [draft-ietf-mailext-acc-url-01.txt](#)

Obviously we should agree on one common way of encoding URL-s in all cases where URL-s will appear in message headers.

The syntax problems are

- (a) Which characters need encoding, and if so which encoding scheme should be used?
- (b) How to handle line folding when URL-s can be very long, and blanks are not allowed in URL-s.

[draft-ietf-mailext-acc-url-01.txt](#) defines this as follows:

URL-parameter := <"> URL-word *(*LWSP-char URL-word) <">

URL-word := token

; Must not exceed 40 characters in length

The syntax of an actual URL string is given in [RFC 1738](#). URL strings can be of any length and can contain arbitrary character content. This presents problems when URLs are embedded in MIME body part headers that are wrapped according to [RFC 822](#) rules. For this reason they are transformed into a URL-parameter for inclusion in a message/external-body content-type specification as follows:

- (1) A check is made to make sure that all occurrences of SPACE, CTLs, double quotes, backslashes, and 8-bit characters in the URL string are already encoded using the URL encoding scheme specified in [RFC 1738](#). Any unencoded occurrences of these characters must be encoded. Note that the result of this operation is nothing more than a different representation of the original URL.
- (2) The resulting URL string is broken up into substrings of 40 characters or less.
- (3) Each substring is placed in a URL-parameter string as a URL-word, separated by one or more spaces. Note that the enclosing quotes are always required since all URLs contain one or more colons, and colons are tspecial characters [[RFC 1521](#)].

Extraction of the URL string from the URL-parameter is even simpler: The enclosing quotes and any linear whitespace are removed and the remaining material is the URL string.

[RFC 1808](#) uses the following definition:

```
base-header = "Base" ":" "<URL:" absoluteURL ">"
```

where "Base" is case-insensitive and any whitespace (including that used for line folding) inside the angle brackets is ignored. For example, the header field

Which characters need encoding? Obviously any eight-bit characters in the URL must be encoded. But must ":" and "/" be encoded? Or is it enough to require "<" before and after the URL? Should "<" or "<" and ">" be used to surround the URL string?

Issue 2: Allow hyperlinks outside multipart/related or not?

Issue specification: Should a text/html be allowed to contain hyperlinks to any other part of the same message, or only to other

parts within the same multipart/related?

Palme

[Page 3]

Opinion A: The multipart/related header tells the mailer that "here comes some body parts which are to be treated together in a special way", and as a consequence that a text/html should only be allowed to refer to other body parts which are within this multipart/related group of body parts.

Opinion B: A text/html body part should be allowed to contain hyperlinks to any other body part in this message (or, if CID or MID is used, any body part in any other message).

Arguments for opinion A is that this makes it simpler for the mail receiving agent: When it gets a multipart/related it knows that the body parts within it are to be treated in a special way (usually stored as files, and the start object turned over to a Web browser as a helper application).

The majority seems to be for opinion A.

Issue 3: Is multipart/related to be used at all

Some people in the discussions have proposed that just plain multipart/mixed could be used instead of multipart/related for a set of objects with hyperlinks between them.

The rough consensus seems to be however that a multipart/related should be used.

Issue 4: Name of the Content-Location header

Opinion A: Its name should be Content-Location

Opinion B: Its name should be only Location or only URL

The rough consensus seems to be that its name should be Content-Location, since this is required by MIME. MIME requires that all Content headers begin with the string "Content-".

Issue 5: Should the Base header (defined in [RFC 1808](#)) be renamed to "Content-Base"?

Based on the discussion about the Content-Location header, it seems as if the next revision of [RFC 1808](#) should rename the Base header into Content-Base.

Issue 6: Relative URL-s and the Content-Base header

Issue specification: Under which circumstances should relative URL-s be allowed in text/html body parts, and how should such relative URL-s be resolved?

Palme

[Page 4]

Relative URL-s should only be allowed if their base is known.
The base can be made known in either of two ways:

- (a) There is a BASE element in the HTML document which resolves the relative URL into a non-relative URL.
- (b) There is a Content-Location of the Text/HTML which can then serve as the base.
- (c) There is a Content-Base header (as defined in [RFC 1808](#)), giving the base to be used.

Issue 7: Relative URL-s referring to body parts within the same message

The base for relative URL-s can either be an external base (for example an HTTP base) in which the relative URL-s are resolved according to the scheme for the base URL, or the base can be the multipart/related set of objects within the MIME message.

Issue 8: The message itself as a Base URL

When the Text/HTML uses "cid" URL-s, these might be relative to the message itself. A "Content-Base: CID:://." header might be used to indicate this. Someone suggested that the relative URL-s would then be "../cid:xxx@foo.org" instead of just "cid:xxx@foo.org".

Question: Does this mean that Content-ID-s need not be globally unique? If that is what it means, I am very much against it.

Or is it just a way of indicating that this message contains hyperlinks of the "CID" scheme, and that these hyperlinks refer to objects in the current message, using CID URL-s?

Issue 9: Priority of bases

Bases for relative URL-s in Text/HTML bodies may be defined in three ways:

- (a) There is a BASE element in the HTML document which resolves the relative URL into a non-relative URL.
- (b) There is a Content-Location of the Text/HTML which can then serve as the base.
- (c) There is a Content-Base header (as defined in [RFC 1808](#)), giving

the base to be used.

Question: Suppose more than one of these three methods are used in the same message, then which of them should be used by the recipient?

Palme

[Page 5]

Suggested: Priority as listed above, if more than one Base is specified, BASE elements should be used in preference of Content-Location (since this is the way HTML normally works) and Content-Location should be used in preference of Content-Base (is this the way HTTP works?? when HTTP uses the Base/Content-Base header??)

Issue 10: Ambiguity of "Content-Base"

Some people have pointed out in the discussion that "Content-Base" is ambiguous in a message, since it might either refer to the situation as seen by the sender or as seen by the recipient.

This does not seem to me to be any problem. A Content-Base should of course have a scheme. If the scheme is for example "HTTP", then this is a base for HTTP retrieval, if the scheme is "LOCAL-FILE", then this is a base for retrieval of local files in the recipients mailbox (probably files created by saving other body parts of the same message in files).

Issue 11: Giving body parts names to be used in URL-s

If the text/html can contain hyperlinks referring to other body parts, then we need a way to give names to these body parts.

Choice A: Use the file names in "Content-Disposition: inline/filename=" headers in the body parts.

Choice B: Use the Content-ID of the body parts.

Discussion: The advantage with using file names is that most Web browsers are already capable of interpreting relative URL-s which refer to file names. In fact, most Web browsers, when asked to display a file, will assume that relative URL-s within that file refer to other files in the same folder as the file to be displayed. Thus, use of file names means that existing Web browser can be made to display the text/html object if the mailer just saves the various parts of the multipart/related into files in a common folder and then turns the start object over to the Web browser.

The use of Content-ID could be allowed as an alternative, but the use of file names seems to be the easiest choice.

The syntax of these file names should be the subset of file name syntaxes for most platforms, which is eight characters, followed by an extension with a period and three more characters. The characters should only be Latin letters and digits, and the first character should be a letter.

Issue 12: How to indicate that relative URL-s refer to body parts?

[draft-palme-text-html-00.txt](#) proposed a new parameter "linking" to the "Content-Type: Text/HTML" header, with the values "external", "filename", "location" and "cid" to indicate various ways of interpreting URL-s in the Text/HTML body. I was not aware, at that time, of the proposal for the "Base/Content-Base" header in [RFC 1808](#). When the base for relative URL-s are the file names in the Content-Disposition of the referred to objects, then this should in some way be shown in the Content-Base header.

I suggest the following syntax:

Content-Base: "LOCAL-FILE://." where "LOCAL-FILE" is taken from RFC 1521, and "://." is taken from [RFC 1808](#). (Check that I have correctly understood what [RFC 1808](#) means with "://".)

Issue 13: Combination of Text/html and Multipart/Alternative

When a Text/html is sent, many recipients will not be capable of displaying the html text, at least not directly, since their mailers do not support Text/html. There is therefore a need to use Multipart/Alternative. This can however be done in many ways.

Choice a:

The construct shown by the following example was proposed in "draft-palme-text-html-00.txt":

```
Content-Type: Multipart/related; boundary="boundary-example-1";
             type=Text/HTML; start=content-id-example@example.host
```

```
--boundary-example 1
Content-Type: MULTIPART/ALTERNATIVE
Boundary: boundary-example-2
```

```
--boundary-example-2
Content-Type: Text/plain
```

```
... plain text version of the document for recipients
whose mailers cannot handle Text/HTML ...
```

```
--boundary-example-2
Content-Type: Text/HTML
Content-ID: content-id-example@example.host
```

```
... text of the HTML document ...
```

```
--boundary-example-2--  
--boundary-example-1  
Content-Type: Image/GIF
```

```
... a body part, to which the HTML document has a link ...  
--boundary-example-1--
```

An abbreviated form of this, just as a notation within this issue document, is:

```
Multipart/related; type=Text/HTML; start=foo@bar
  Multipart/alternative
    Text/plain
    Text/HTML (contains hyperlink to the Image/GIF object)
      Content-ID: start=foo@bar
  Image/GIF
```

Choice b:

Same as Choice a, but use Multipart/mixed instead of Multipart/related, see issue 2 above.

Choice c:

```
Multipart/alternative
  Multipart/mixed
    Text/Plain
    Image/GIF
  Multipart/Related; type=Text/HTML; start=foo@bar
    Text/HTML
      Content-ID: start=foo@bar
    Message/External-body; access-type=Content-ID
      (pointing to the Image/GIF object)
```

Choice d:

```
Multipart/alternative
  Multipart/mixed
    Text/Plain
    Image/GIF
  Multipart/Related; type=Text/HTML; start=foo@bar
    Text/HTML
      Content-ID: start=foo@bar
    Image/GIF
```

Choice e:

```
Multipart/related; type=Text/HTML; start=foo@bar
  Image/GIF
  Multipart/alternative
    Multipart/mixed
      Text/plain
      message/external-body; access-type=cid:
        (pointer to the image/GIF)
    Text/HTML (contains hyperlink to the Image/GIF object)
      Content-ID: start=foo@bar
```


Choice f:

```
multipart/mixed (Message-ID: message-unique@node.net)
  1: image/gif (Content-ID:<BL8V3T@node..net>
    Content-Disposition: attachment;
    uri=./neat.gif;
    base=file://localhost/anypath/to_here)
  2: multipart/alternative
    text/plain (Content-Disposition: inline;
      including text reference to neat.gif and
      that the GIF is the first part of this MIME
      message)
    text/HTML (Content-disposition: inline; file=me.html;
      embeds URN of
      mid://node..net/message-unique?BL8V3T
      ; or whatever the cid URN syntax is)
```

Issue 14: What should "start" refer to

Which if the following two cases should be used:

```
Multipart/related; type=Text/HTML; start=foo@bar
  Multipart/alternative
    Text/plain
    Text/HTML (contains hyperlink to the Image/GIF object)
      Content-ID: start=foo@bar
  Image/GIF
```

```
Multipart/related; type=Text/HTML; start=foo@bar
  Multipart/alternative
    Content-ID: start=foo@bar
    Text/plain
    Text/HTML (contains hyperlink to the Image/GIF object)
  Image/GIF
```

i.e. should "start" refer to the Text/HTML or to the Multipart/Alternative"??

Issue 15: Including remotely available objects in a message

There are several reasons why a sender of a message, which contains a Text/HTML body part with externally resolvable hyperlinks, might still want to include some or all of these external objects in the message.

Reason i: Because some recipients may have e-mail but not full Internet access.

Reason ii: To make retrieval of the body parts safer and faster for the recipient.

In "[draft-palme-text-html-00.txt](#)" a new header "Content-Location" was proposed for this.

Palme

[Page 9]

The issue has been raised that this should be seen as a "cached" version of the original object, and that a parameter "validity" should maybe be added to indicate the maximum cache time.

Note that this does not mean that the mailer should necessarily put something in the web caches of their web browser. That is a different issue. This is just a way of saying that "if you save this object locally, we recommend a maximum saving time".

Example:

Content-Location: "http://www.jazzie.com/ii/internet/mailnews.html"; LIFN: 1 month.

Question: Has the syntax of such a parameter already been defined in some ietf-draft or RFC? Is LIFN defined in some RFC or internet-draft? If so, can someone refer me to this definition.

Issue 16: The use of mid-s and cid-s

There has been a long discussion in the ietf-types mailing list about how to use mid-s and cid-s, whether cid-s can be qualified by mid-s, whether a cid URL scheme is needed or not etc. I have not understood the whole of this discussion and am not sure whether it should influence the specifications in "[draft-palme-text-html-00.txt](#)" or not. If this discussion requires changes in "[draft-palme-text-html-00.txt](#)", could someone please enlighten me on how this should be done.

Issue 17: Content-disposition inline or attachment

Assume a construct such as this:

Multipart/related; type=Text/HTML; start=foo@bar
Content-Base: "LOCAL-FILE://."

Text/HTML (contains hyperlink to the Image/GIF object)

Content-ID: start=foo@bar

Image/GIF

Content-Disposition: inline/filename=foo.GIF

Should the Content-Disposition above be "inline" or "attachment"?

Discussion: A mailer which does not understand Multipart/related should treat Multipart/related in the same way as Multipart/mixed. From that viewpoint, the Content-Disposition should be "inline" in case the picture is to be shown at the same time as the root text.

A mailer which understands Multipart/related should know that all body parts are to be saved as files, and then turned over to an interpreter for the type of the start object.

"Content-Disposition: attachment" is usually interpreted as "retrieve only if the recipient asks for it" and that is not correct in this case.

A third possible value of "Content-Disposition:" might be "file" which would tell the mailer to store the object as a file.

Issue 18: Content-Location or Content-Disposition

Someone has suggested that instead of

```
Content-Location: "url"
```

we should write

```
Content-Disposition: inline; uri="url".
```

and instead of

```
Content-Base: "base-url"
```

we should write

```
Content-Disposition: inline; base="base-url"
```

