

Network Working Group  
Internet Draft  
[draft-ietf-mhtml-spec-01.txt](#)  
Category-to-be: Proposed standard  
Expires: January 1997

Jacob Palme  
Stockholm University/KTH  
Alexander Hopmann  
ResNova Software, Inc.  
July 1996

## MIME E-mail Encapsulation of Aggregate HTML Documents (MHTML)

### Status of this Document

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).

### Abstract

Although HTML [[RFC 1866](#)] was designed within the context of MIME, more than the specification of HTML as defined in [RFC 1866](#) is needed for two electronic mail user agents to be able to interoperate using HTML as a document format. These issues include the naming of objects that are normally referred to by URIs, and the means of aggregating objects that go together. This document describes a set of guidelines that will allow conforming mail user agents to be able to send, deliver and display these HTML objects. In addition it is hoped that these techniques will also apply to the wider category of URI-enabled objects. In order to do this, the document specifies the MIME content-headers "Content-Location" and "Content-Base".



Table of Contents

- [1. Introduction](#)**
- [2. Terminology](#)**
  - 2.1 Conformance requirement terminology**
  - 2.2 Other terminology**
- [4. The Content-Location and Content-Base MIME Content Headers](#)**
  - 4.1 MIME content headers**
  - 4.2 The Content-Base header**
  - 4.3 The Content-Location Header**
  - 4.4 Encoding of URIs in e-mail headers**
- [5. Base URIs for resolution of relative URIs](#)**
- [6. Sending HTML documents without linked objects](#)**
- [7. Use of the Content-Type: Multipart/related](#)**
- [8. Format of Links to Other Body Parts](#)**
  - 8.1 General principle**
  - 8.2 Use of the Content-Location header**
  - 8.3 Use of the Content-ID header and CID URLs**
- [9 Examples](#)**
  - 9.1 Example of a HTML body without included linked objects**
  - 9.3 Example with relative URIs to an embedded GIF picture**
  - 9.4 Example using CID URL and Content-ID header to an embedded GIF picture**
- [10. Content-Disposition header](#)**
- [11. Encoding Considerations for HTML bodies](#)**
  - 11.1 Character set issues**
  - 11.2 Line break characters**
- [12. Security Considerations](#)**
- [13. Acknowledgments](#)**
- [14. References](#)**
- [15. Author's Address](#)**

Mailing List Information

Further discussion on this document should be done through the mailing list [MHTML@SEGATE.SUNET.SE](mailto:MHTML@SEGATE.SUNET.SE).

To subscribe to this list, send a message to  
[LISTSERV@SEGATE.SUNET.SE](mailto:LISTSERV@SEGATE.SUNET.SE)  
which contains the text  
SUB MHTML <your name (not your e-mail address)>

Archives of this list are available by anonymous ftp from

[FTP://SEGATE.SUNET.SE/lists/mHTML/](ftp://SEGATE.SUNET.SE/lists/mHTML/)

The archives are also available by e-mail. Send a message to [LISTSERV@SEGATE.SUNET.SE](mailto:LISTSERV@SEGATE.SUNET.SE) with the text "INDEX MHTML" to get a list of the archive files, and then a new message "GET <file name>" to retrieve the archive files.

Comments on less important details may also be sent to the editor, Jacob  
Palme <jpalme@dsv.su.se>.

More information may also be available at URL:  
[HTTP://www.dsv.su.se/~jpalme/ietf/jp-ietf-home.HTML](http://www.dsv.su.se/~jpalme/ietf/jp-ietf-home.HTML)>

Palme-Hopmann            Do not implement based on this draft            [Page 2]

## **1. Introduction**

The HTML format is a very common format for documents in the Internet, and there is an obvious need to be able to send documents in this format in e-mail [RFC821=SMTP, [RFC822](#)]. The "text/html" media type is defined in [RFC 1866](#) [[HTML2](#)]. This document gives additional specifications on how to use the text/html media type as a Content-Type in MIME [RFC 1521=MIME1] e-mail messages. HTML documents commonly include links to other objects and resources, either embedded or directly accessible through hypertext links. When mailing a HTML document, it is often desirable to also mail all of the additional resources that are referenced in it; those elements are necessary for the complete interpretation of the HTML.

An alternative way for sending HTML documents in e-mail is to only send the URL, and let the recipient look up the document using HTTP. That method is described in [[URLBODY](#)] and is not described in this document.

## **2. Terminology**

### **2.1 Conformance requirement terminology**

This specification uses the same words as [RFC 1123](#) [[HOSTS](#)] for defining the significance of each particular requirement. These words are:

- MUST      This word or the adjective "required" means that the item is an absolute requirement of the specification.
- SHOULD    This word or the adjective "recommended" means that there may exist valid reasons in particular circumstances to ignore this item, but the full implications should be understood and the case carefully weighed before choosing a different course.
- MAY        This word or the adjective "optional" means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.

An implementation is not compliant if it fails to satisfy one or more of the MUST requirements for the protocols it implements. An implementation that satisfies all the MUST and all the SHOULD requirements for its protocols is said to be "unconditionally compliant"; one that satisfies all the MUST requirements but not all the SHOULD requirements for its protocols is said to be "conditionally compliant."

Palme-Hopmann

Do not implement based on this draft

[Page 3]

## **2.2 Other terminology**

Most of the terms used in this document are defined in other RFCs.

Absolute URI, AbsoluteURI	See <a href="#">RFC 1808</a> [ <a href="#">RELURL</a> ].
CID	See [ <a href="#">MIDCID</a> ].
Content-Base	See <a href="#">section 4.2</a> below.
Content-ID	See [ <a href="#">MIDCID</a> ].
Content-Location	MIME message or content part header with the URI of the MIME message or content part body, defined in <a href="#">section 4.3</a> below.
CR	See [ <a href="#">RFC822</a> ].
CRLF	See [ <a href="#">RFC822</a> ].
Header	Field in a message or content heading specifying the value of one attribute.
Heading	Part of a message or content before the first CRLFCRLF, containing formatted fields with attributes of the message or content.
HTML	See <a href="#">RFC 1866</a> [ <a href="#">HTML2</a> ].
HTML Aggregate objects	HTML objects together with some or all objects, to which the HTML object contains hyperlinks.
LF	See [ <a href="#">RFC822</a> ].
MIC	Message Integrity Codes, codes use to verify that a message has not been illegally modified.
MIME	See <a href="#">RFC 1521</a> [ <a href="#">MIME1</a> ], [ <a href="#">MIME2</a> ].
MUA	Messaging User Agent.
Relative URI, RelativeURI	See <a href="#">RFC 1866</a> [ <a href="#">HTML2</a> ] and <a href="#">RFC 1808</a> [ <a href="#">RELURL</a> ].
URI, absolute and relative	See <a href="#">RFC 1866</a> [ <a href="#">HTML2</a> ].
URL	See <a href="#">RFC 1738</a> [ <a href="#">URL</a> ].
URL, relative	See [ <a href="#">RELURL</a> ].





### **3. Overview**

An aggregate HTML object is a MIME-encoded message that contains a root document as well as other data that is required in order to represent that document (inline pictures, style sheets, applets, etc.). Aggregate HTML objects can also include additional elements that are linked to the first object. It is important to keep in mind the differing needs of several audiences. Mail sending agents might send aggregate HTML objects as an encoding of normal day-to-day electronic mail. Mail sending agents might also send aggregate HTML objects when a user wishes to mail a particular document from the web to someone else. Finally mail sending agents might send aggregate HTML documents as automatic responders, providing access to WWW resources for non-IP connected clients.

Mail receiving agents also have several differing needs. Some mail receiving agents might be able to receive an aggregate HTML document and display it just as any other text content type would be displayed. Others might have to pass this aggregate HTML document to an HTML browsing program, and provisions need to be made to make this possible.

Finally several other constraints on the problem arise. It is important that it be possible for an HTML document to be signed and for it to be able to be transmitted to a client and displayed with a minimum risk of breaking the message integrity (MIC) check that is part of the signature.

## **4. The Content-Location and Content-Base MIME Content Headers**

### **4.1 MIME content headers**

In order to resolve URI references to other body parts, two MIME content headers are defined, Content-Location and Content-Base. Both these headers can occur in any message or content heading, and will then be valid within this heading and for its content.

In practice, at present only those URIs which are URLs are used, but it is anticipated that other forms of URIs will in the future be used.

The syntax for these headers is, using the syntax definition tools from [\[RFC822\]](#):

```
content-location ::= "Content-Location:" ( absoluteURI | relativeURI
)
```

```
content-base ::= "Content-Base:" absoluteURI
```

where URI is at present (June 1996) restricted to the syntax for URLs as defined in [RFC 1738](#) [\[URL\]](#).

These two headers are valid only for exactly the content heading or message heading where they occurs and its text. They are thus not valid for the parts inside multipart headings, and are thus meaningless in multipart headings.

Palme-Hopmann            Do not implement based on this draft            [Page 5]

These two headers may occur both inside and outside of a Multipart/Related part.

#### **[4.2 The Content-Base header](#)**

The Content-Base gives a base for relative URIs occurring in other heading fields and in content which do not have any BASE element in its HTML code. Its value MUST be an absolute URI.

Example showing which Content-Base is valid where:

```
Content-Type: Multipart/related; boundary="boundary-example-1";
              type=Text/HTML; start=foo2*foo3@bar2.net
; A Content-Base header cannot be placed here, since this is a
; multipart MIME object.
```

```
--boundary-example-1
```

Part 1:

```
Content-Type: Text/HTML; charset=US-ASCII
```

```
Content-ID: foo2*foo3@bar2.net
```

```
Content-Location: "http://www.ietf.cnir.reston.va.us/images/foo1.bar1"
```

```
;
```

```
; This Content-Location must contain an absolute URI, since no base
; is valid here.
```

```
--boundary-example-1
```

Part 2:

```
Content-Type: Text/HTML; charset=US-ASCII
```

```
Content-ID: foo4*foo5@bar2.net
```

```
Content-Location: "foo1.bar1" ; The Content-Base below applies to
                           ; this relative URI
```

```
Content-Base: "http://www.ietf.cnri.reston.va.us/images/"
```

```
--boundary-example-1--
```

#### **[4.3 The Content-Location Header](#)**

The Content-Location header specifies the URI that corresponds to the content of the body part in whose heading the header is placed. Its value CAN be an absolute or relative URI. Any URI or URL scheme may be used, but use of non-standardized URI or URL schemes might entail some risk that recipients cannot handle them correctly.

The Content-Location header can be used to indicate that the data sent under this heading is also retrievable, in identical format, through normal use of this URI. If used for this purpose, it must contain an absolute URI or be resolvable, through a Content-Base header, into an absolute URI. In this case, the information sent in the message can be

seen as a cached version of the original data.

Palme-Hopmann

Do not implement based on this draft

[Page 6]

The header can also be used for data which is not available to some or all recipients of the message, for example if the header refers to an object which is only retrievable using this URI in a restricted domain, such as within a company-internal web space. The header can even contain a fictitious URI and need in that case not be globally unique.

Example:

```
Content-Type: Multipart/related; boundary="boundary-example-1";
              type=Text/HTML
```

```
--boundary-example-1
```

Part 1:

```
Content-Type: Text/HTML; charset=US-ASCII
```

```
... .. <IMG SRC="fiction1/fiction2"> ... ..
```

```
--boundary-example-1
```

Part 2:

```
Content-Type: Text/HTML; charset=US-ASCII
```

```
Content-Location: "fiction1/fiction2"
```

```
--boundary-example-1--
```

#### **4.4 Encoding of URIs in e-mail headers**

Since MIME header fields have a limited length and URIs can get quite long, these lines may have to be folded. If such folding is done, the algorithm defined in [[URLBODY](#)] [section 3.1](#) should be employed.

### **5. Base URIs for resolution of relative URIs**

Relative URIs inside contents of MIME body parts are resolved relative to a base URI. In order to determine this base URI, the first-listed method in the following list applies.

- (a) There is a base specification inside the MIME body part containing the link which resolves relative URIs into absolute URIs. For example, HTML provides the BASE element for this.
- (b) There is a Content-Base header (as defined in [section 4.2](#)), specifying the base to be used.
- (c) There is a Content-Location header in the heading of the body part which can then serve as the base in the same way as the request URI can serve as a base for relative URIs within a file

retrieved via HTTP [[HTTP](#)].

## **6. Sending HTML documents without linked objects**

If an HTML document is sent without other objects, to which it is linked, it MAY be sent as a Text/HTML body part by itself. In this case, Multipart/related need not be used.

Such a document may either not include any links, or contain links which the recipient resolves via ordinary net look up, or contain links which the recipient cannot resolve.

Inclusion of links which the recipient has to look up through the net may not work for some recipients, since all e-mail recipients do not have full internet connectivity. Also, such links may work for the sender but not for the recipient, for example when the link refers to an URL within a company-internal network not accessible from outside the company.

Note that documents with links that the recipient cannot resolve MAY be sent, although this is discouraged. For example, two persons developing a new HTML page may exchange incomplete versions.

## **7. Use of the Content-Type: Multipart/related**

The use of URI references creates some additional issues for aggregate HTML objects. Normal URI references can of course be used, however it is likely that many user agents may not be able to retrieve those objects referred to. This document provides a means for these additional objects to be transmitted with the HTML and for the links between these objects to be properly resolved.

If a message contains one or more Text/HTML body parts and also contains as separate body parts, data, to which hyperlinks (as defined in RFC [1866 \[HTML2\]](#)) in the Text/HTML body parts refers, then this set of objects SHOULD be sent within a Multipart/Related body part as defined in [\[REL\]](#).

The root of the Multipart/related SHOULD be of the Content-Type: Text/HTML. Use of the Content-Type: Multipart/Alternative, one of whose parts is of Content-Type: Text/HTML, is also allowed, but implementors are warned that many mail programs treat Multipart/Alternative as if it had been Multipart/Mixed (even though MIME [\[MIME1\]](#) requires support for Multipart/Alternative).

If the root is not the first body part within the Multipart/related, its Content-ID MUST be given in a start parameter to the Content-Type: Multipart/Related header.

When presenting the root body part to the user, the additional body parts within the Multipart/related can be used:

(a) For those recipients who only have e-mail but not full Internet access.

Palme-Hopmann

Do not implement based on this draft

[Page 8]



- (b) For those recipients who for other reasons, such as firewalls or the use of company-internal links, cannot retrieve the linked body parts through the net.

Note that this means that you can, via e-mail, send HTML which includes URIs which the recipient cannot resolve via HTTP or other connectivity-requiring URIs.

- (c) For items which are not available on the web.

- (d) For any recipient to speed up access.

The type parameter of the Content-Type: Multipart/related MUST be the same as the Content-Type of its root.

When a sending MUA sends objects which were retrieved from the WWW, it SHOULD maintain their WWW URIs. It SHOULD not transform these URIs into some other URI form prior to transmitting them. This will allow the receiving MUA to both verify MICs included with the email message, as well as verify the documents against their WWW counterpoints.

The Text/HTML body MAY contain links to MIME body parts outside of the Multipart/Related or in other messages, but such usage is discouraged. Implementors are warned that many receiving mailers may not be able to resolve such links.

Within such a Multipart/related, ALL different parts MUST have different Content-Location or Content-ID values.

## **8. Format of Links to Other Body Parts**

### **8.1 General principle**

A Text/HTML body part may contain hyperlinks to objects which are included as other body parts in the same message and within the same multipart/related content. Often such linked objects are meant to be displayed inline to the reader of the main document. HTML version 2.0 [RFC 1866=HTML2] has only one way of specifying hyperlinks to such inline embedded content, the IMG tag. New tags with this property are however proposed in the ongoing development of HTML (example: applet, frame).

In order to send such messages, there is a need to indicate which other body parts are referred to by the links in the Text/HTML body parts. This is done in the following way: For each distinct URI in the Text/HTML document, which refers to data which is sent in the same MIME message, there SHOULD be a separate body part within the multipart/related part of the message containing this data. Each such body part SHOULD contain a Content-Location header (see [section 8.2](#)) or

a Content-ID header (see [section 8.3](#)).

An e-mail system which claims conformance to this standard MUST support receipt of Multipart/related (as defined in [section 7](#)) with links between body parts using both the Content-Location (as defined in [section 8.2](#)) and the Content-ID method (as defined in [section 8.3](#)).

Palme-Hopmann            Do not implement based on this draft            [Page 9]

## **[8.2](#) Use of the Content-Location header**

If there is a Content-Base header, then the recipient MUST employ relative to absolute resolution as defined in [RFC 1808](#) [[RELURL](#)] of URIs in both the HTML markup and the Content-Location header before matching a hyperlink in the HTML markup to a Content-Location header. The same applies if the Content-Location contains an absolute URL, and the HTML markup contains a BASE element so that relative URL-s in the HTML markup can be resolved.

If there is NO Content-Base header, and the Content-Location header contains a relative URL, then NO relative to absolute resolution SHOULD be performed (even if there is a BASE element in the HTML markup), and exact textual match of the relative URL-s in the Content-Location and the HTML markup is performed instead (after removal of LWSP introduced as described in [section 4.4](#) above).

The URI in the Content-Location header need not refer to an object which is actually available globally for retrieval using this URI (after resolution of relative URIs).

## **[8.3](#) Use of the Content-ID header and CID URLs**

When CID (Content-ID) URLs as defined in [RFC 1738](#) [[URL](#)] and [RFC 1873](#) [[MIDCID](#)] is used for links between body parts, the Content-Location statement will normally be replaced by a Content-ID header. Thus, the following two headers are identical in meaning:

```
Content-ID: foo@bar.net
Content-Location: CID: foo@bar.net
```

Note: Content-IDs MUST be globally unique [[MIME1](#)]. It is thus not permitted to make them unique only within this message or within this multipart/related.

## **[9](#) Examples**

### **[9.1](#) Example of a HTML body without included linked objects**

The first example is the simplest form of an HTML email message. This is not an aggregate HTML object, but simply one by itself. This message contains a hyperlink but does not provide the ability to resolve the hyperlink. To resolve the hyperlink the receiving client would need either IP access to the Internet, or an electronic mail web gateway.

```
From: foo1@bar.net
To: foo2@bar.net
Subject: A simple example
Mime-Version: 1.0
```

Content-Type: Text/HTML; charset=US-ASCII

Palme-Hopmann      Do not implement based on this draft      [Page 10]

```
<HTML>
<head></head>
<body>
<h1>Hi there!</h1>
An example of an HTML message.<p>
Try clicking <a href="http://www.resnova.com/">here.</a><p>
</body></HTML>
```

## **[9.2](#) Example with absolute URIs to an embedded GIF picture:**

```
From: foo1@bar.net
To: foo2@bar.net
Subject: A simple example
Mime-Version: 1.0
Content-Type: Multipart/related; boundary="boundary-example-1";
              type=Text/HTML; start=foo3*foo1@bar.net

--boundary-example 1
  Content-Type: Text/HTML; charset=US-ASCII
  Content-ID: foo3*foo1@bar.net

  ... text of the HTML document, which might contain a hyperlink
  to the other body part, for example through a statement such as:
  <IMG SRC="http://www.ietf.cnri.reston.va.us/images/ietflogo.gif"
  ALT="IETF logo">

  --boundary-example-1
  Content-Location:
    "http://www.ietf.cnri.reston.va.us/images/ietflogo.gif"
  Content-Type: IMAGE/GIF
  Content-Transfer-Encoding: BASE64

  R0lGODlhGAGgAPEAAP/////ZRaCgoAAAACH+PUNvcHlyaWdodCAoQykgMTk5
  NSBJRVRGLiBVbmF1dGhvcml6ZWQgZHVwbGljYXRpb24gcHJvaGliaXRlZC4A
  etc...

  --boundary-example-1--
```

## **[9.3](#) Example with relative URIs to an embedded GIF picture**

```
From: foo1@bar.net
To: foo2@bar.net
Subject: A simple example
Mime-Version: 1.0
Content-Base: "http://www.ietf.cnri.reston.va.us"
Content-Type: Multipart/related; boundary="boundary-example-1";
              type=Text/HTML

--boundary-example 1
Content-Type: Text/HTML; charset=ISO-8859-1
```

Content-Transfer-Encoding: QUOTED-PRINTABLE

Palme-Hopmann      Do not implement based on this draft      [Page 11]

... text of the HTML document, which might contain a hyperlink to the other body part, for example through a statement such as:  
<IMG SRC="/images/ietflogo.gif" ALT="IETF logo">

Example of a copyright sign encoded with Quoted-Printable: =A9

Example of a copyright sign mapped onto HTML markup: &#168;

--boundary-example-1

Content-Location: "/images/ietflogo.gif"

Content-Type: IMAGE/GIF

Content-Transfer-Encoding: BASE64

R0lGODlhGAGgAPEAAAP/////ZRACgoAAAACH+PUNvcHlyaWdodCAoQykgMTk5  
NSBJRVRLiBVbmF1dGhvcml6ZWQgZHVwbGljYXRpb24gcHJvaGliaXRlZC4A  
etc...

--boundary-example-1--

#### **9.4 Example using CID URL and Content-ID header to an embedded GIF picture**

From: foo1@bar.net

To: foo2@bar.net

Subject: A simple example

Mime-Version: 1.0

Content-Type: Multipart/related; boundary="boundary-example-1";  
type=Text/HTML

--boundary-example 1

Content-Type: Text/HTML; charset=US-ASCII

... text of the HTML document, which might contain a hyperlink to the other body part, for example through a statement such as:  
<IMG SRC="cid:foo4\*foo1@bar.net" ALT="IETF logo">

--boundary-example-1

Content-ID: foo4\*foo1@bar.net

Content-Type: IMAGE/GIF

Content-Transfer-Encoding: BASE64

R0lGODlhGAGgAPEAAAP/////ZRACgoAAAACH+PUNvcHlyaWdodCAoQykgMTk5  
NSBJRVRLiBVbmF1dGhvcml6ZWQgZHVwbGljYXRpb24gcHJvaGliaXRlZC4A  
etc...

--boundary-example-1--

## **10. Content-Disposition header**

Note the specification in [[REL](#)] on the relations between Content-Disposition and Multipart/Related.

Palme-Hopmann

Do not implement based on this draft

[Page 12]



## 11. Encoding Considerations for HTML bodies

### 11.1 Character set issues

A mail user agent that is composing a message using HTML has a choice in how to represent and subsequently encode characters for the transmission of the mail message.

However, there are some differences as to the default character encoding, specified by the MIME "charset" parameter. If this parameter is omitted: When transferred through HTTP, the default is [HTTP]:

content-type: Text/HTML; charset=ISO-8859-1

When transferred via e-mail, the default is [MIME1]:

content-type: Text/HTML; charset=US-ASCII

To avoid confusion, the MIME Content-Type parameter for Text/HTML SHOULD always include a charset value, and not rely on the MIME e-mail default of US-ASCII if no charset value is specified.

When sending HTML via MIME e-mail, three layers of encoding are relevant as shown in Figure 1:

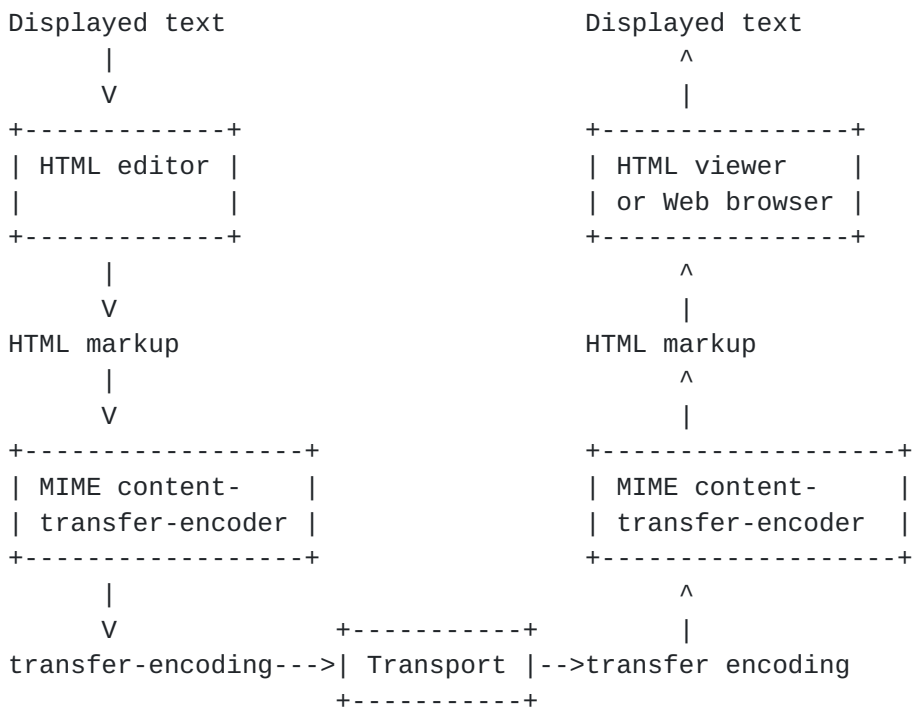


Figure 1

Definitions (see Figure 1):

Displayed text    A visual representation of the intended text.

HTML markup      A sequence of characters formatted according to the  
HTML specification [[HTML2](#)].

Palme-Hopmann      Do not implement based on this draft      [Page 13]

MIME encoding      A sequence of octets physically forwarded via e-mail, may include MIME content-transfer-encoding as specified in [\[MIME1\]](#).

HTML editor        Software used to produce HTML markup.

MIME content-transfer-encoder      Software used to encode and decode non-US-ASCII characters according to the MIME standard.

HTML viewer        Software used to display HTML documents to recipients.

Note: Real implementations need not split functions into different modules as described above. The figure above is a logical model in order to explain how rewriting and transport is done.

If the displayed text contains non-US-ASCII characters, these characters might have to be rewritten if the transport (as is common in e-mail) is set to handle only 7-bit characters.

HTML markup allows some characters at the displayed text level to be represented using either entity references or numeric character references (as defined in [\[HTML2\] section 3.2.1](#)). For example, a "small a, acute accent" may be represented by the entity reference "&acute;" or the numeric character reference "&#255;". Alternatively, the same character might appear directly in the HTML document, but for transmission through MIME 7-bit-systems, the entire HTML document is encoded using a Content-Transfer-Encoding (as defined in [\[MIME1\] section 5](#)).

In sending a message containing non US-ASCII characters, both these rewriting methods MAY be used, and any mixture of them MAY occur when sending the document via e-mail. Receiving mailers (together with the Web browser they may use to display the document) MUST be capable of handling any combinations of these rewriting methods.

The value of the charset attribute of the Content-Type header field should be US-ASCII if and only if the HTML markup contains only US-ASCII characters (even if the displayed text contains non-US-ASCII characters).

Example of non-US-ASCII characters in HTML: See [section 9.3](#) above.

## **[11.2](#) Line break characters**

The MIME standard [\[MIME1\]](#) specifies that line breaks in the MIME encoding (see figure 1) MUST be CRLF. The HTTP standard [\[HTTP\]](#) specifies that line breaks in transported HTML markup (see figure 2) may be either bare CRs, bare LFs or CRLFs. To allow data integrity checks through checksums, MIME encoding of line breaks SHOULD be such that after decoding, the line break representation of the original HTML markup is

returned.

Palme-Hopmann

Do not implement based on this draft

[Page 14]

Note that since the mail content-MD5 is defined to a canonical form with all line breaks converted to CRLF, while the HTTP content-MD5 is defined to apply to the transmitted form. This means that the Content-MD5 HTTP header may not be correct for Text/HTML that is retrieved from a HTTP server and then sent via mail.

## **12. Security Considerations**

Some Security Considerations include the potential to mail someone an object, and claim that it is represented by a particular URI (by giving it a Content-Location: header). There can be no assurance that a WWW request for that same URI would normally result in that same object. It might be unsuitable to cache the data in such a way that the cached data can be used for retrieval of this URI from other messages or message parts than those included in the same message as the Content-Location header. Because of this problem, receiving User Agents SHOULD not cache this data in the same way that data that was retrieved through an HTTP or FTP request might be cached.

URLs, especially File URLs, may in their name contain company-internal information, which may then inadvertently be revealed to recipients of documents containing such URLs.

One way of implementing messages with linked body parts is to handle the linked body parts in a combined mail and WWW proxy server. The mail client is only given the start body part, which it passes to a web browser. This web browser requests the linked parts from the proxy server. If this method is used, and if the combined server is used by more than one user, then methods must be employed to ensure that body parts of a message to one person is not retrievable by another person. Use of passwords (also known as tickets or magic cookies) is one way of achieving this. Note that some caching HTML proxy servers may not distinguish between cached objects from e-mail and HTTP, which may be a security risk.

In addition, by allowing people to mail aggregate HTML objects, we are opening the door to other potential security problems that until now were only problems for WWW users. For example, some HTML documents now either themselves contain executable content (JavaScript) or contain links to executable content (The "INSERT" specification, Java). It would be exceedingly dangerous for a receiving User Agent to execute content received through a mail message without careful attention to restrictions on the capabilities of that executable content.

Palme-Hopmann

Do not implement based on this draft

[Page 15]

### **13. Acknowledgments**

Harald T. Alvestrand, Richard Baker, Dave Crocker, Martin J. Duerst, Lewis Geer, Roy Fielding, Al Gilman, Paul Hoffman, Richard W. Jesmajian, Mark K. Joseph, Greg Herlihy, Valdis Kletnieks, Daniel LaLiberte, Ed Levinson, Jay Levitt, Albert Lunde, Larry Masinter, Keith Moore, Gavin Nicol, Pete Resnick, Jon Smirl, Einar Stefferud, Jamie Zawinski and several other people have helped us with preparing this document. I alone take responsibility for any errors which may still be in the document.

### **14. References**

Ref.	Author, title
-----	-----
[CONDISP]	R. Troost, S. Dorner: "Communicating Presentation Information in Internet Messages: The Content-Disposition Header", <a href="#">RFC 1806</a> , June 1995.
[HOSTS]	R. Braden (editor): "Requirements for Internet Hosts -- Application and Support", STD-3, <a href="#">RFC 1123</a> , October 1989.
[HTML2]	T. Berners-Lee, D. Connolly: "Hypertext Markup Language - 2.0", <a href="#">RFC 1866</a> , November 1995.
[HTTP]	T. Berners-Lee, R. Fielding, H. Frystyk: Hypertext Transfer Protocol -- HTTP/1.0. <a href="#">RFC 1945</a> , May 1996.
[MIDCID]	E. Levinson: "Message/External-Body Content-ID Access Type", <a href="#">RFC 1873</a> , December 1995.
[MIME1]	N. Borenstein & N. Freed: "MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies", <a href="#">RFC 1521</a> , Sept 1993.
[MIME2]	N. Borenstein & N. Freed: "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types". <a href="#">draft-ietf-822ext-mime-imt-02.txt</a> , December 1995.
[NEWS]	M.R. Horton, R. Adams: "Standard for interchange of USENET messages", <a href="#">RFC 1036</a> , December 1987.
[REL]	Harald Tveit Alvestrand, Edward Levinson: "The MIME Multipart/Related Content-type", < <a href="#">draft-levinson-multipart-related-00.txt</a> >, January 1995.
[RELURL]	R. Fielding: "Relative Uniform Resource Locators", <a href="#">RFC</a>

[1808](#), June 1995.

Palme-Hopmann

Do not implement based on this draft

[Page 16]



- [RFC822] D. Crocker: "Standard for the format of ARPA Internet text messages." STD 11, [RFC 822](#), August 1982.
- [SMTP] J. Postel: "Simple Mail Transfer Protocol", STD 10, [RFC 821](#), August 1982.
- [URL] T. Berners-Lee, L. Masinter, M. McCahill: "Uniform Resource Locators (URL)", [RFC 1738](#), December 1994.
- [URLBODY] N. Freed and Keith Moore: "Definition of the URL MIME External-Body Access-Type", [draft-ietf-mailext-acc-url-01.txt](#), November 1995.

## **[15. Author's Address](#)**

For contacting the editors, preferably write to Jacob Palme rather than Alex Hopmann.

Jacob Palme  
Stockholm University and KTH  
Electrum 230  
S-164 40 Kista, Sweden

Phone: +46-8-16 16 67  
Fax: +46-8-783 08 29  
E-mail: [jpalme@dsv.su.se](mailto:jpalme@dsv.su.se)

Alex Hopmann  
President  
ResNova Software, Inc.  
**[5011 Argosy Dr. #13](#)**  
Huntington Beach, CA 92649

E-mail: [alex.hopmann@resnova.com](mailto:alex.hopmann@resnova.com)

Working group chairman:

Einar Stefferud <[stef@nma.com](mailto:stef@nma.com)>

