

**Usenet Best Practice**  
<[draft-ietf-usefor-useage-00.txt](#)>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC 2026](#).

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

Abstract

This Draft is intended to become a "Best Current Practice" RFC. Its purpose is to set out how software should behave and conventions which users should observe, in order that Netnews in general, and Usenet in particular, should provide the most effective service to its users.

[Remarks enclosed in square brackets and aligned with the left margin, such as this one, are not part of this draft, but are editorial notes to explain matters amongst ourselves, or to point out alternatives, or to assist the RFC Editor.]

[In this draft, references to [NNTP] are to be replaced by [RFC 977](#), or else by references to the RFC arising from the series of drafts draft-ietf-nntpext-base-\*.txt, in the event that such RFC has been accepted at the time this document is published.]

Table of Contents

<a href="#">1.</a>	Introduction .....	<a href="#">4</a>
<a href="#">1.1.</a>	Basic Concepts .....	<a href="#">4</a>

<a href="#">1.2.</a>	Objectives .....	<a href="#">4</a>
<a href="#">2.</a>	Definitions, Notations and Conventions .....	<a href="#">5</a>

2.1.	Definitions .....	5
2.2.	Textual Notations .....	5
3.	The Well-Behaved User Agent .....	6
3.1.	The Well-Behaved Posting Agent .....	6
3.1.1.	Construction of Headers .....	7
3.1.1.1.	Date .....	8
3.1.1.2.	From .....	8
3.1.1.3.	Message-ID .....	8
3.1.1.4.	Subject .....	9
3.1.1.5.	Newsgroups .....	9
3.1.1.6.	Reply-To .....	10
3.1.1.7.	Organization .....	10
3.1.1.8.	Distribution .....	10
3.1.1.9.	Followup-To .....	10
3.1.1.10.	User-Agent .....	10
3.1.2.	Construction of Bodies .....	10
3.1.2.1.	Signatures .....	11
3.1.2.2.	Usage of MIME .....	11
3.1.2.3.	Content-Transfer-Encoding .....	13
3.2.	The Well-Behaved Followup Agent .....	14
3.2.1.	Construction of Headers .....	15
3.2.1.1.	Subject .....	15
3.2.1.1.1.	Examples .....	16
3.2.1.2.	Newsgroups .....	16
3.2.1.3.	Mail-Copies-To .....	16
3.2.1.4.	Posted-And-Mailed .....	17
3.2.1.5.	References .....	17
3.2.2.	Construction of Bodies .....	17
3.2.2.1.	Quoting and Attributions .....	17
3.2.2.2.	Signatures .....	19
3.2.2.3.	Usage of MIME .....	19
3.3.	The Well-Behaved Reading Agent .....	19
3.3.1.	Interpretation of Headers .....	19
3.3.1.1.	Presentation of Articles .....	19
3.3.1.2.	Summary .....	21
3.3.2.	Interpretation of Bodies .....	21
3.3.2.1.	Usage of MIME .....	22
3.4.	The Well-Behaved Reply Agent .....	23
3.5.	User Interfaces .....	23
4.	The Well-Behaved Injecting Agent .....	23
4.1.	Construction of Headers .....	24
4.1.1.	Sender .....	24
4.1.2.	Organization .....	24
4.1.3.	User-Agent .....	24
4.1.4.	Injection-Info .....	25
5.	The Well-Behaved Relaying Agent .....	25
5.1.	The Path Header .....	26
5.1.1.	Suggested Verification Methods .....	26
6.	The Well-Behaved Serving Agent .....	26

<a href="#">6.1.</a>	<a href="#">Control Messages .....</a>	<a href="#">27</a>
<a href="#">6.1.1.</a>	<a href="#">The 'newgroup' and 'mvgroup' Control Messages .....</a>	<a href="#">27</a>
<a href="#">6.1.2.</a>	<a href="#">Cancel Messages .....</a>	<a href="#">27</a>
<a href="#">7.</a>	<a href="#">The Well-Behaved Hierarchy Administrator .....</a>	<a href="#">28</a>
<a href="#">7.1.</a>	<a href="#">Control Messages .....</a>	<a href="#">28</a>

<a href="#">7.2.</a>	Naming of Newsgroups .....	<a href="#">28</a>
<a href="#">7.3.</a>	Format of Bodies .....	<a href="#">30</a>
<a href="#">7.4.</a>	Promulgation .....	<a href="#">30</a>
<a href="#">8.</a>	The Well-Behaved Moderator .....	<a href="#">31</a>
<a href="#">9.</a>	The Well-Behaved Poster .....	<a href="#">31</a>
<a href="#">9.1.</a>	Construction of Headers .....	<a href="#">31</a>
<a href="#">9.1.1.</a>	From .....	<a href="#">31</a>
<a href="#">9.1.2.</a>	Summary .....	<a href="#">31</a>
<a href="#">9.1.3.</a>	Expires .....	<a href="#">32</a>
<a href="#">9.2.</a>	Construction of Bodies .....	<a href="#">32</a>
<a href="#">10.</a>	References .....	<a href="#">32</a>
<a href="#">11.</a>	Acknowledgements .....	<a href="#">34</a>
<a href="#">12.</a>	Contact Address .....	<a href="#">34</a>
<a href="#">Appendix A</a>	- Notices .....	<a href="#">34</a>



## **1. Introduction**

### **1.1. Basic Concepts**

"Netnews" is a set of protocols for generating, storing and retrieving news "articles" (which resemble email messages) and for exchanging them amongst a readership which is potentially widely distributed. It is organized around "newsgroups", with the expectation that each reader will be able to see all articles posted to each newsgroup in which he participates. These protocols are defined in [[USEFOR](#)].

"Usenet" is a particular worldwide open network based upon the Netnews protocols, with the newsgroups being organized into recognized "hierarchies". Anybody can join (it is simply necessary to negotiate an exchange of articles with one or more other participating hosts).

Usenet "belongs" to those who administer the hosts of which it is comprised. There is no Cabal with overall authority to direct what is to be allowed. Nevertheless, there do exist agencies within Usenet that have authority to establish policies and to perform administrative functions, but such authority derives solely from the consent of those sites which choose to recognize it (and who can decline to exchange articles with sites which choose not to recognize it). Usually, the authority of such an agency is restricted to a particular hierarchy, or group of hierarchies.

A "policy" is a rule intended to facilitate the smooth operation of a network by establishing parameters which restrict behaviour that, whilst technically unexceptionable, would nevertheless contravene some accepted standard of "Good Netkeeping". Since the ultimate beneficiaries of a network are its human readers, who will be less tolerant of poorly designed interfaces than mere computers, articles in breach of established policy can cause considerable annoyance to their recipients.

### **1.2. Objectives**

The purpose of this document is to set out how software should behave and conventions which users should observe, in order that Netnews in general, and Usenet in particular, should provide the most effective service to its users.

[[USEFOR](#)] is a standard, and hence its requirements are mandatory. The requirements set out here are in addition to the requirements set out in [[USEFOR](#)]. Their purpose is to establish "Best Current Practice", and hence they are advisory. Nevertheless, failure to observe them will severely prejudice the good order of Usenet, and cause great

inconvenience to the users of that medium.

NOTE: The extreme irritation caused to other readers by such violations is not to be underestimated; however, enforcement of such rules is more a matter of sensible design or of social

C. H. Lindsey

[Page 4]



pressure (whose effectiveness should not be underestimated, even though it cannot be prescribed).

Many of these requirements are matters of policy which may vary from network to network, from hierarchy to hierarchy within one network, and even between individual newsgroups within one hierarchy. It is assumed, for the purposes of this document, that agencies with varying degrees of authority to establish such policies will exist, and that where they do not, policy will be established by mutual agreement. However, it is NOT the purpose of this document to define how the authority of various agencies to exercise control or oversight of the various parts of Usenet is established (that is itself a matter of policy). For the benefit of networks and hierarchies without such established agencies, and to provide a basis upon which all agencies can build, this present document often provides default policy parameters, usually introducing them by a phrase such as "As a matter of policy ...".

NOTE: The practices recommended here relate only to Netnews and Usenet, and not to any other medium. Nevertheless, it may be that some of them may turn out to be helpful for other media such as mailing lists.

## **2. Definitions, Notations and Conventions**

### **2.1. Definitions**

All the technical terms defined in [[USEFOR](#)] [section 2.1](#) are considered to be defined in this document also.

### **2.2. Textual Notations**

This document contains explanatory NOTES using the following format. These may be skipped by persons interested solely in the content of the specification. The purpose of the notes is to explain why choices were made, to place them in context, or to suggest possible implementation techniques.

NOTE: While such explanatory notes may seem superfluous in principle, they often help the less-than-omniscient reader understand the true intent of the specification in cases where the wording is not entirely clear.

Certain words, when capitalized, are used to define the significance of individual requirements. The key words "MUST", "REQUIRED", "SHOULD", "RECOMMENDED", "MAY" and "OPTIONAL", and any of those words associated with the word "NOT", are to be interpreted as described in [[RFC 2119](#)].

However, as provided in that RFC, the force of these words is lower

here than would have been the case in a standards track document. In particular, violation of a MUST or SHOULD does not necessarily imply a failure of interoperability, but rather that established policy or accepted best practice would be breached, to the detriment of the

C. H. Lindsey

[Page 5]

good order of Usenet.

NOTE: A requirement imposed on a relaying or serving agent regarding some particular article should be understood as applying only if that article is actually accepted for processing (since any agent may always reject any article entirely, for reasons of site policy).

[That NOTE can probably be removed, or severely rewritten, once we have a better idea of the requirements/recommendations we are going to make in this document.]

Wherever the context permits, use of the masculine includes the feminine and use of the singular includes the plural, and vice versa.

Throughout this document we will give various examples. In order to prevent possible conflict with "Real World" entities and people the top level domain ".example" is used in all sample domains and addresses. The hierarchy "example.\*" is also used as a sample hierarchy. Information on the ".example" top level domain is in [RFC 2606].

### **3. The Well-Behaved User Agent**

The term "user agent" comprises posting agents, reading agents and followup agents as defined in [[USEFOR](#)], and also reply agents, by which is meant a user agent that is generating an email, presumably addressed to the poster of an article. Although it is usual for all these functionalities to be included within a single piece of software, it is convenient to discuss them separately here.

This section is addressed primarily to the implementors of user agents. Whilst it is common for such agents to combine the functions of a Netnews User Agent (NUA) and a Mail User Agent (MUA), it needs to be realized that they serve different functions, and adding a few extra features to an MUA is unlikely to result in a good NUA, any more than adding a few extra features to an NUA would result in a good MUA.

#### **3.1. The Well-Behaved Posting Agent**

The implementor of a posting agent SHOULD make it possible for a suitably perseverent poster to generate any article, however absurd, that conforms strictly to [[USEFOR](#)]. On the other hand, it needs to be understood that the difference between a good posting agent and a bad posting agent lies in its ability to encourage the poster to adhere to good standards of "netkeeping", by making it easy to generate articles that will be widely acceptable to the conventions and expectations of the Usenet community, and hard to generate articles outside of those norms. This is largely a matter of choosing

appropriate defaults for various parameters and settings.

Here it should be noted that what is acceptable in Email (which is a one-to-few communication where the author can be expected to be aware of the capabilities and preferences of his correspondents) may not be

C. H. Lindsey

[Page 6]

acceptable in Netnews (which is a one-to-many communication directed at an unseen and unknown audience). Much grief has arisen in the past from poorly designed agents which tried to impose onto Usenet defaults and practices which were perfectly appropriate for Email.

#### **3.1.1. Construction of Headers**

Whilst it SHOULD be possible to insert any legitimate header, not limited to those defined in [[USEFOR](#)] and including experimental headers, there are certain essential headers, namely the Subject-, Newsgroups-, Followup-To- and Reply-To-headers which the poster MUST be able to insert and/or edit (and to do so at any stage during the composition of the article). Note that this specifically includes the possibility of setting the followup to "poster".

Posting agents SHOULD permit the poster to include headers of arbitrary length (and MUST permit at least 79 characters). However, they SHOULD endeavour to keep individual header lines, so far as is possible, within 79 characters (or other established policy limit) by folding them at suitable places (however, the limit of 998 octets ([[USEFOR](#)] 4.5) on any individual header line still applies); but if the poster has manually folded a header within the accepted limits (to achieve some pleasing layout, for example) the posting agent SHOULD respect the poster's intent.

Although header-contents are defined in such a way that folding ([[USEFOR](#)] 4.2.3) can take place between many of the lexical tokens (and even within some of them), folding SHOULD be limited to placing the CRLF at higher-level syntactic breaks, and SHOULD also avoid leaving trailing WSP on the preceding line. For instance, if a header-content is defined as comma-separated values, it is RECOMMENDED that folding occur after the comma separating the values, even if it is allowed elsewhere.

There is a preferred case convention, which posters and posting agents SHOULD use: each hyphen-separated "word" has its initial letter (if any) in uppercase and the rest in lowercase, except that some abbreviations have all letters uppercase (e.g. "Message-ID" and "MIME-Version"). The forms given in the various rules defining headers in [[USEFOR](#)] show the preferred forms (but relaying and reading agents are expected to tolerate articles not obeying this convention).

A comment ([[USEFOR](#)] 4.2.4) is normally used to provide some human readable informational text, except at the end of a mailbox which contains no phrase, as in

```
fred@foo.bar.example (Fred Bloggs)
as opposed to
"Fred Bloggs" <fred@foo.bar.example> .
```

The former is a deprecated, but commonly encountered, usage for indicating the name of the person whose mailbox it is. Posting agents SHOULD NOT now be generating it.

Headers that merely state defaults explicitly (e.g., a Followup-To-header with the same content as the Newsgroups-header, or a MIME Content-Type-header with contents "text/plain; charset=us-ascii"), or state information that reading agents can typically determine easily themselves (e.g. the length of the body in octets) are redundant and posting agents SHOULD NOT include them.

There follow some recommendations specific to particular headers.

#### **3.1.1.1. Date**

It is RECOMMENDED to add a comment, after the date-time, containing the time zone in human-readable form. However, many of the abbreviations commonly used for this purpose are ambiguous, and so the value given by the <zone> is the only definitive form. For example:

Date: Sat, 26 May 2001 11:13:00 -0500 (EST)

#### **3.1.1.2. From**

The mailboxes in the From-content MUST contain syntactically valid email addresses identifying the poster(s). Each such mailbox SHOULD be a working email address, belonging to the poster(s) of the article, or the person or agent on whose behalf the article is posted. When, for whatever reason, a poster does not wish to use a working address, the mailbox concerned SHOULD, to comply with [USEFOR], end in the top level domain ".invalid" [RFC 2606].

NOTE: It is fashionable for posters to disguise their mail addresses to discourage malicious harvesting and for other purposes. Whilst the circumstances which might make this seem desirable are much to be regretted, the practice cannot be regarded as in the best interests of Usenet, and this document does not seek to promote the practice, even though it shows how to do it "correctly". Therefore, it is NOT recommended that implementors should go out of their way to facilitate it.

#### **3.1.1.3. Message-ID**

Posting agents have the option of generating their own message identifiers, or of leaving it to the injecting agent. Recall that it is an absolute requirement of [USEFOR] that message identifiers should be unique with regard to all other Netnews articles or Email messages, past, present or future. However, it would in practice be sufficient to ensure that there were astronomical odds against a duplicated message identifier, and this is usually brought about by using the domain name of the originating site in the id-right of the msg-id, together with the time of composition and other disambiguating material (such as a process number or a serial number)

in the id-left. It is also in order to include additional information of significance to the poster within the id-left, and even to deliberately make a non-unique identifier in cases where the identical message is to be posted by several posters (for example, a



cancel for an article which may also be cancelled by others).  
[Recall that we have two drafts regarding the construction of message identifiers on [www.landfield.com/usefor](http://www.landfield.com/usefor) that were written in the early days of Usefor. Maybe these should be dusted down, published, and referred to here.]

#### **3.1.1.4. Subject**

There is a temptation amongst inventors of new protocols to require particular phrases to be inserted or recognized automatically at particular places within the Subject-header. This temptation is strongly to be resisted.

There are, however, two exceptions to this principle which have become hallowed by longstanding usage:

1. There is an established convention for the Subject-header in a followup to begin with "Re: ", and this SHOULD be supported (see 3.2.1.1).
2. For compatibility with legacy news software, the Subject-content of a control message (i.e. an article that also contains a Control-header) MAY start with the string "cmsg ", and non-control messages SHOULD NOT start with the string "cmsg ". See also [section 6.1](#).

[SHOULD NOT changed from MUST NOT? Do there really still exist servers or other agents that will recognize and act upon "cmsg" in a Subject-header? And if so, maybe that MUST NOT should be moved back into [\[USEFOR\]](#).]

Subject-headers are for humans to read, and the most that user agents should do is to filter them as directed by their human readers. If some enhancement to Netnews requires support within the headers, then the proper procedure is to invent a new header for the purpose, or to adapt an existing header (supposing it had the capability to support such adaptations).

#### **3.1.1.5. Newsgroups**

There are restrictions on the length of components of newsgroup-names, and on the newsgroup-names themselves, as described more fully in 7.2. Posting and injecting agents MAY attempt to enforce them but, because of the possibility that hierarchy policies or future standards may relax them, it SHOULD be possible for posters to override such checks, and software MUST be so written that they can be disabled altogether.

Posting agents MAY (and followup agents SHOULD) accept articles crossposted to newsgroups which do not exist on their local hosts, though posting agents SHOULD at least alert the poster to the

situation and request confirmation.

#### **3.1.1.6. Reply-To**

In the absence of Reply-To, the reply address(es) is the address(es) in the From-header. For this reason a Reply-To SHOULD NOT be included if it just duplicates the From-header.

NOTE: Use of a Reply-To-header is preferable to including a similar request in the article body, because replying agents can take account of Reply-To automatically.

#### **3.1.1.7. Organization**

Posting agents are discouraged from providing a default value for this header unless it is acceptable to all posters using those agents and unless it contains useful information (including some indication of the poster's physical environment). See [section 4.1.2](#) for an even stronger discouragement for injecting agents.

#### **3.1.1.8. Distribution**

Posting agents SHOULD NOT provide a default Distribution-header without giving the poster an opportunity to override it.

#### **3.1.1.9. Followup-To**

A Followup-To-header SHOULD NOT be included if it just duplicates the Newsgroups-header. At least one of its newsgroup-names SHOULD exist on the posting agent's host (since a well behaved poster ought not to be setting followups to a place that he cannot read). Cf. a similar rule regarding crossposting in [[USEFOR](#)] [section 5.5](#).

#### **3.1.1.10. User-Agent**

Comments in User-Agent-headers should be restricted to information regarding the product named to their left, such as its full name or platform information, and should be concise. Use as an advertising medium (in the mundane sense) is discouraged.

### **3.1.2. Construction of Bodies**

It was the fashion at one time to indicate underlining within body texts using Backspace, in the form of an underscore (US-ASCII 95), a backspace, and a character, repeated for each character that should be underlined. Posting agents MAY support this mechanism, although it is no longer so common for reading agents to process it.

NOTE: using this precise method should ensure that reading agents that cannot display the text underlined will at least display it correctly in an un-underlined form.

The formfeed character (US-ASCII 12) (which is sometimes referred to as the "spoiler character") MAY be used (see 3.3.2 for its effect on reading agents).

In plain-text articles (those with no MIME headers, or those with a MIME Content-Type of "text/plain") posting agents SHOULD endeavour to keep the length of body lines within some reasonable limit. The size of this limit is a matter of policy, the default being to keep within 79 characters at most, and preferably within 72 characters (to allow room for quoting in followups).

NOTE: That policy limit (e.g. 72 or 79) should be expressed as a number of characters (as they will be displayed by a reading agent) rather than as the number of octets used to encode them. For use on occasions where established policy prescribes different line lengths (this usually arises in groups where the charset for the language used is best represented using double width characters) the preferred line length SHOULD be a configurable option. In addition, posting agents MUST permit the poster to create individual lines longer than the default or configured length if he so insists (which may require the cessation of any automatic generation of flowed lines [[RFC 3676](#)] on a temporary basis).

#### **3.1.2.1. Signatures**

A "personal signature" is a short closing text automatically added to the end of articles by posting agents, identifying the poster and giving his network addresses, etc. Whenever a poster or posting agent appends such a signature to an article, it MUST be preceded with a delimiter line containing (only) two hyphens (US-ASCII 45) followed by one SP (US-ASCII 32). The signature is considered to extend from the last occurrence of that delimiter up to the end of the article (or up to the end of the part in the case of a multipart MIME body). Posting agents SHOULD provide a facility to enable the poster to add such signatures, and SHOULD discourage (at least with a warning) signatures of excessive length (4 lines is a commonly accepted limit).

#### **3.1.2.2. Usage of MIME**

When the Content-Type is "text/plain", the recommendations and limits on line lengths set out above SHOULD be observed.

Posting agents MAY use the "format=flowed" parameter of "text/plain" (and also the "DelSp=yes" if appropriate) defined in [[RFC 3676](#)] so as to allow suitably equipped reading agents to reformat flowed paragraphs to suit the width of their display areas. However, it must be understood that many reading agents do not support that feature, and therefore the physical length of all lines SHOULD be restricted to the default preferred length of 72 characters, rather than the 78 recommended in [[RFC 3676](#)]. However, single words longer than that length (and this specifically applies to URIs [[RFC 2396](#)]) MUST NEVER

be split across more than one physical line.

Other forms of text, such as "text/html" SHOULD NOT be used except in groups where established policy or custom so allows (7.3). However, where they are so used then, for the benefit of readers who see it

only in its transmitted form, the material SHOULD be "pretty-printed" (for example by restricting its line length as above and by keeping sequences which control its layout or style separate from the meaningful text).

Likewise, Content-Types requiring special processing for their display, notably the "binary" Content-Types "image", "audio" and "video" (including also material encoded by the "uuencode" protocol), together with most "application" types, SHOULD NOT be used except in groups where established policy or custom so allows (7.3). Exceptionally, those application types defined in [[RFC 1847](#)] and [RFC 3156] for use within "multipart/signed" articles, and the type "application/pgp-keys" (or other similar types containing digital certificates) may be used freely.

The Content-Type "message/partial" is not recommended for textual articles because the Content-Type, and in particular the charset, of the complete article cannot be determined by examination of the second and subsequent parts, and hence (except when they are written in pure US-ASCII) it is not possible to read them as separate articles (as by a reader who wanted to "browse ahead" to see whether it was worth his while to read the whole set). Moreover, for full compliance with [[RFC 2046](#)] it would be necessary to use the "quoted-printable" encoding to ensure the material was 7bit-safe. In any case, breaking such long texts into several parts is usually unnecessary, since modern transport agents should have no difficulty in handling articles of arbitrary length.

On the other hand, "message/partial" may be useful for binaries of excessive length, since reading of the individual parts on their own is not required and they would likely already be encoded in a manner that was 7bit-safe.

The Content-Type "message/rfc822" SHOULD be used where complete news articles or email messages are to be included within another article ([[USEFOR](#)] 6.21.2).

The Content-Type "message/external-body" could be appropriate for texts which it would be uneconomic (in view of the likely readership) to distribute to the entire network.

The Content-Types "multipart/mixed", "multipart/parallel" and "multipart/signed" may be used freely in news articles. However, except where policy or custom so allows, the Content-Type: "multipart/alternative" SHOULD NOT be used, on account of the extra bandwidth consumed and the difficulty of quoting in followups.

The Content-Type: "multipart/digest" is commended for any article composed of multiple messages more conveniently viewed as separate

entities, thus enabling reading agents to move rapidly between them. The "boundary" should be composed of 28 hyphens (US-ASCII 45) (which makes each boundary delimiter 30 hyphens, or 32 for the final one) so as to enable reading agents which currently support the digest usage described in [[RFC 1153](#)] to continue to operate correctly.



NOTE: The various recommendations given above regarding the usage of particular Content-Types apply also within the individual parts of these multipart.

A multipart is preceded and followed by some spare text (a preamble before the first boundary and an epilogue after the last one). It is clear from [\[RFC 2046\]](#) that these texts are not to be considered part of the official message and SHOULD NOT be displayed by reading agents. It is useful for the preamble to contain words such as "This is a multipart message in MIME format" for the benefit of older reading agents that do not support MIME, but the epilogue SHOULD be empty and, in particular, it SHOULD NOT be used to hold the signature (3.1.2.1), as is sometimes done.

### **3.1.2.3. Content-Transfer-Encoding**

The normal expectation ([\[USEFOR\]](#) 6.21.3) is that the Content-Transfer-Encoding will be "8bit (or maybe "7bit" if the charset allows it). Other Content-Transfer-Encodings SHOULD NOT be used unless there are pressing reasons to do so.

The following are examples of such situations where a Content-Transfer-Encoding of other than "8bit" may be necessary.

1. The content type implies that the content is (or may be) "8bit-unsafe"; i.e. it may contain octets equivalent to the US-ASCII characters CR or LF (other than in the combination CRLF) or NUL. In that case one of the Content-Transfer-Encodings "base64" or "quoted-printable" MUST be used, and reading agents MUST be able to handle both of them.

NOTE: If a future extension to the MIME standards were to provide a more compact encoding of binary suited to transport over an 8bit channel, it could be considered as an alternative to base64 once it had gained widespread acceptance.

2. It is often the case that "application" Content-Types are textual in nature, and intelligible to humans as well as to machines, and where this state can be recognized by the posting agent (either through knowledge of the particular application type or by testing) the material SHOULD NOT be treated as 8bit-unsafe; this has the added benefit, where the posting agent uses other than CRLF for line endings internally, of automatically ensuring that line endings are processed correctly during transport.

If, on the other hand, the posting agent recognizes that the material is not textual, or cannot reasonably determine it to be so, then the material MUST be encoded as for 8bit-unsafe (however, in that case, it is the responsibility of the agent generating the

material to ensure that lines endings, if any, are represented correctly).

NOTE: All the application types defined by [\[USEFOR\]](#), namely "application/news-transmission", "application/news-groupinfo" and "application/news-checkgroups" are textual, and indeed designed for human reading.

3. Although the "text" Content-Types should normally be encoded as 8bit (or 7bit), if the character set specified by the "charset=" parameter can include the 3 disallowed octets, then the material MUST be encoded as for 8bit-unsafe. This is most likely to arise in the case of 16-bit character sets such as UTF-16 ([\[UNICODE 3.2\]](#) or [\[ISO/IEC 10646\]](#)). In addition, where it is known that the material is subsequently to be gatewayed from Netnews to Email ([\[USEFOR\]](#) 8.8), the encoding "quoted-printable" MAY be used (otherwise the gateway might have to re-encode it itself).
4. Some protocols REQUIRE the use of a particular Content-Transfer-Encoding. In particular, the authentication protocol based on OpenPGP defined in [\[RFC 3156\]](#) mandates the use of one of the encodings "quoted-printable" or "base64". Whilst posters might be tempted to risk the use of "8bit" or "7bit" encodings (and indeed the referenced standard recommends that signed messages using those encodings be accepted and interpreted), they should be warned that differences in the treatment of trailing whitespace between OpenPGP [\[RFC 2440\]](#) and earlier versions of PGP may render signatures written with the one unverifiable by the other; and, moreover, Usenet articles are very likely to include trailing whitespace in the form of a personal signature (3.1.2.1).
5. The Content-Type message/partial [\[RFC 2046\]](#) is required to use encoding "7bit" (the encapsulated complete message may itself use encoding "quoted-printable" or "base64", but that information is only conveyed along with the first of the partial parts).

NOTE: Although there would actually be no problem using encoding "8bit" in a pure Netnews (as opposed to Email) environment, this document discourages the use of "message/partial" except for binary material, which will likely be encoded to pass through "7bit" in any case.

It may be necessary to change the Content-Transfer-Encoding at gateways. For example in the case where such an encapsulated news article with the Content-Type "message/rfc822" is to be transported by email and it has Content-Transfer-Encoding "8bit", the Content-Transfer-Encoding may need to be changed, although there may well be no problems in practice if the email transport supports 8BITMIME [\[RFC 2821\]](#).

### [3.2.](#) The Well-Behaved Followup Agent

Usenet is primarily a medium for discussion. The majority of articles that are posted are in fact followups to previous articles, and exceedingly complex threads can develop. Therefore, it is essential that user agents provide facilities for followups that will enable such elongated discussions to proceed smoothly.

### **3.2.1. Construction of Headers**

The requirements on inserting and editing headers already set out in 3.1.1 still apply, and apply in particular to those headers for which the followup agent has set default values.

#### **3.2.1.1. Subject**

The Subject of the followup is, by default, taken from that of the precursor, but users are able to override that default; indeed they are to be encouraged to do so whenever appropriate in order to avoid long threads which have wandered far from the topic with which they originated, but which still adhere to the original Subject.

It has been a long standing practice, both on Usenet and in Email, to prepend the back-reference "Re: " ([[USEFOR](#)] 5.4) to the Subject when preparing a followup, as an indication to the reader that this is a continuation of discussion of an earlier topic rather than the start of a new one. [[USEFOR](#)] does not require this practice, but permits it so long as it is not applied if such a back-reference is already present, and provided no string other than "Re: " is used for the purpose.

However, the practice is not without its difficulties:

1. Although the "Re" (which is an abbreviation for the Latin "In re", meaning "in the matter of", and not an abbreviation of "Reference" as is sometimes erroneously supposed) may be understood by English speakers, and indeed by speakers of most European Languages, its use in a newsgroup where articles were customarily written in Arabic, or Hindi, or Chinese would be less than helpful.
2. It requires extra processing (to ignore it) in some reading agents which choose to consult the Subject-header when deciding the best order in which to present articles to the reader (see 3.3.1.1). This burden has to be weighed against the relatively small benefit of the indication provided directly to readers.
3. Sometimes, followup agents attempt to use translations of "Re: " into other languages, as in "Sv: " and "Antwort: ". But it is not practicable for those reading agents which take some special note of "Re: " also to take note of translations into an indeterminate number of other languages, and for this reason [[USEFOR](#)] makes it clear that such translations SHOULD NOT be used.
4. Even the presence of "Re: " at the start of a Subject may occasionally be misleading, because it might have been deliberately placed there by a poster rather than having been generated automatically by a followup agent.

5. And finally, there are philosophical arguments against features within an unstructured header which imply specific recognition and support within user agents (for reason already explained in 3.1.1.4). Indeed, the only reason why [[USEFOR](#)] permits this

particular exception is on account of its current widespread usage.

For these reasons, this document does not seek to perpetuate this practice, and indeed it might be better if its use were eventually to be phased out. Nevertheless, it is certain that it will continue to happen for some considerable period of time in newgroups where English is the primary language, simply on account of the inertia already behind it. For this reason, [section 3.3.1.1](#) RECOMMENDS stripping away any initial "Re: " when comparing Subjects.

It would be wiser for any followup agents which are able to recognize non-standard back-references such as "Re(2): ", "Sv: ", etc. to refrain from prepending anything further, but other attempts to mend that problem are likely to do more harm than good.

As well as the addition of "Re: ", the Subject-header MAY be refolded (which MAY include collapsing/expanding whitespace to/from a single SP at any point where the folding is changed). However, it MUST NOT (except by deliberate act of the poster) be truncated, extended or changed in any other way that might cause a reading agent to deduce that the subject of a thread had changed.

[Bruce wants users to be requested to confirm that they are happy with the default Subject as provided.]

#### [3.2.1.1.1](#). Examples

In the following examples, please note that only "Re: " has any official status (and hence may be utilized by reading agents). "was: " is a convention used by many English-speaking posters to signal a change in subject matter. Software can always recognize that such changes have occurred from the References-header.

```
Subject: Film at 11
Subject: Re: Film at 11
Subject: Godwin's law considered harmful (was: Film at 11)
Subject: Godwin's law (was: Film at 11)
Subject: Re: Godwin's law (was: Film at 11)
Subject: Re: Godwin's law
```

#### [3.2.1.2](#). Newsgroups

Followup agents SHOULD accept articles crossposted to newsgroups which do not exist on their local hosts (as opposed to posting agents, for which that requirement is only "MAY").

#### [3.2.1.3](#). Mail-Copies-To

If the user attempts to email the poster as well as to followup, in the case where the Mail-Copies-To-header is absent, and even more so

when it is present and there is an explicit "nobody", the followup agent SHOULD issue a warning and ask for confirmation.



NOTE: This header is only relevant when posting followups to Netnews articles, and is to be ignored when sending pure email replies to the poster, which are handled as prescribed under the Reply-To-header.

#### **3.2.1.4. Posted-And-Mailed**

NOTE: In addition to the Posted-And-Mailed-header, some followup agents also include within the body a mention that the article is both posted and mailed, for the benefit of reading agents that do not normally show that header.

#### **3.2.1.5. References**

Followup agents SHOULD trim message identifiers out of a References-header but SHOULD NOT do so until the number of message identifiers exceeds 21, at which time trimming SHOULD be done by removing sufficient identifiers starting with the second from the left so as to bring the total down to 21 (but the first message identifier MUST NOT be trimmed). However, it would be wrong to assume that References-headers containing more than 21 message identifiers will not occur.

#### **3.2.2. Construction of Bodies**

Followup agents SHOULD follow policies already described for posting agents (3.1.2) regarding the length of lines when generating new text. Exceptionally, they SHOULD NOT adjust the length of quoted lines (3.2.2.1) in followups unless they are able to reformat them in a consistent manner.

##### **3.2.2.1. Quoting and Attributions**

It is customary for the body of a followup to commence with an "attribution" referring to the "precursor" and to "quote" any text copied verbatim from the precursor with a suitable prefix. Followup agents MUST facilitate the automatic incorporation of these things, even though they are not mandated by any standard, in a manner consistent with the conventions described below.

These conventions for quotations and attributions describe widely used practices. Since much software will attempt to recognize and act upon them, questions of interoperability can arise, and so the words "MUST", "SHOULD", etc. are here to be understood as more than advisory.

When the precursor had used the "format=flowed" parameter of text/plain [[RFC 3676](#)], and when the followup agent also supports "format=flowed", flowed paragraphs in the precursor (including any flowed lines within quotations in the precursor) SHOULD be reflowed.

Thus, if all agents supported "format=flowed", no physical line, quoted or not, would ever exceed the default (or policy) limit, except by the deliberate intent of the poster. Where the precursor was not flowed, its lines SHOULD be left alone when quoting, except

that already quoted lines which appeared (from the presence of trailing SP) to have been flowed by one of the precursor's precursors MAY be treated as such.

When a followup agent incorporates the "precursor" as a quotation, it MUST be distinguished from the surrounding text in some way, and SHOULD be so distinguished by prefacing each line of the quoted text (even if it is empty) with the character ">" (or perhaps with "> " in the case of a previously unquoted line). This will result in multiple levels of ">" when quoted content itself contains quoted content, and it will also facilitate the automatic analysis of articles.

NOTE: Whilst posters should edit quoted context to trim it down to the minimum necessary, followup agents SHOULD NOT attempt to enforce this beyond issuing a warning (past attempts to do so have been found to be notably counter-productive).

The followup agent SHOULD also precede the quoted content by an "attribution line" (however, readers are warned not to assume that they are accurate, especially within multiply nested quotations). The following convention for such lines is intended to facilitate their automatic recognition and processing by sophisticated reading agents. The attribution SHOULD contain the name and/or the email address of the precursor's poster, as in

Joe D. Bloggs <jdbloggs@foo.example> wrote:

or

Helmut Schmidt <helmut@bar.example> schrieb:

The attribution MAY contain also a single newsgroup-name (the one from which the followup is being made), the precursor's message identifier and/or the precursor's Date and Time. Any of these that are present, SHOULD precede the name and/or email address. However, the inclusion or not of such fields SHOULD always be under the control of the poster.

To enable this line, and the message identifier and the email address within it, to be recognized (for example to enable suitable reading agents to retrieve the precursor or email its poster by clicking on them), the following conventions SHOULD be observed:

- o The precursor's message identifier SHOULD be enclosed within <...> or <news:...>
- o The precursor's poster's email address SHOULD be enclosed within <...>
- o The various fields may be separated by arbitrary text and they may be folded in the same way as headers, but attributions SHOULD always be terminated by a ":" followed by CRLF.

Further examples:

On comp.foo in <1234@bar.example> on 24 Dec 2001 16:40:20 +0000,  
"Joe D. Bloggs" <jdbloggs@bar.example> wrote:

Am 24. Dez 2002 schrieb Helmut Schmidt <helmut@bar.example>:

### **3.2.2.2. Signatures**

Followup agents, when incorporating quoted text from a precursor, SHOULD NOT include the signature in the quotation.

### **3.2.2.3. Usage of MIME**

Followup agents which quote parts of a precursor SHOULD initially include all parts of the precursor that were displayed inline, as if they were a single part.

## **3.3. The Well-Behaved Reading Agent**

### **3.3.1. Interpretation of Headers**

Reading agents need to be prepared for ancient usages (and even non-compliance) which nevertheless still appear from time to time. In particular, the following is often seen:

```
fred@foo.bar.example (Fred Bloggs)
as opposed to
"Fred Bloggs" <fred@foo.bar.example> .
```

The former is a deprecated, but commonly encountered, usage and reading agents SHOULD take special note of such comments as indicating (e.g. in killfiles) the name of the person whose mailbox it is.

[Reading agents SHOULD make all headers available on user request.]  
[What about headers etc that are unparseable?]

### **3.3.1.1. Presentation of Articles**

[The following text might be better placed in the proposed section concerning requirements for user interfaces, if we decide to go ahead with that section.]

Reading agents SHOULD present the articles in each newsgroup in an order which ensures that the reader never sees a followup or reply to an article unless he has already had an opportunity to read the original. However, this may be easier said than done. Here are some methods commonly used to fulfil this aim; none of them works perfectly.

1. Present the articles in the order they were received at the local serving agent. However, articles propagated via different routes with different delays may well arrive out of order, so this may not be reliable.
2. Sort the articles into order according to their Date-headers. This will usually be better than the first method, but relies on the clock and timezone settings in posting agents being approximately

correct. And although it satisfies the minimal recommendation at the head of this section, it will likely result in totally separate threads of discussion being merged in an unhelpful order.

3. Sort the articles according to their Subject-headers (or group them according to their Subject-headers, with the groups being presented in order of the Date-header). Within a group with the same Subject, sort according to the Date-header. This works tolerably well, but within a long discussion with many divergent subthreads, those subthreads are still merged in an unhelpful order. Moreover, it will occasionally bring together totally unrelated articles that just happen to have the same Subject by chance.
4. Construct a tree in which each article is within a sub-tree headed by each article mentioned in its References-header, and present articles by a depth-first traversal of that tree, sorting the siblings within each branch according to their Date-headers. This method is usually superior to the ones mentioned earlier, but it can go wrong for a number of reasons.
  - a) References-headers are sometimes absent, or incomplete (and are even permitted to be trimmed when they get too long), and earlier articles in the threads may have expired off the local server. Nevertheless, with careful implementation, these problems are mostly surmountable.
  - b) A poster may join an existing discussion (and clearly intend to do so by using the same Subject-header, possibly with a prepended "Re: ") and yet his article might not be created as a followup to any specific precursor and hence would not have a References-header. Hence it would be presented quite apart from the other (sub-)threads of that discussion.
  - c) Conversely, the topic of some sub-thread might have diverged so far from the original topic of discussion that some poster decides to create a totally new Subject for his followup. Nevertheless, that followup, and the whole sub-thread which issues from it, will still be presented in the midst of the other sub-threads of the original discussion.
5. To counter these various deficiencies, various hybrid schemes have been devised which take account of all three headers, References-, Subject- and Date-, and these often succeed in providing a more pleasing presentation to the reader. However, different readers can be pleased in different ways, and so it is often the case that reading agents provide configurable options to choose between several methods.

This document does not single out any particular method as "the best". They are all to be considered acceptable, and implementors are encouraged to experiment accordingly. Nevertheless, it is inevitable that some combination of Subjects and followups will eventually arise

that defeats even the most sophisticated scheme.

It must be noted, however, in the case of those methods which rely on the comparison of Subject-headers, whether to detect equality or for sorting, that there are certain additional precautions that need to



be taken, such as:

- a) [[USEFOR](#)] permits a back-reference "Re: " to be prepended (optionally) to a Subject when creating a followup. Therefore, that back-reference SHOULD be stripped away before performing any comparison of Subjects. On the other hand, "Re:" is the only back-reference permitted, and therefore it is not necessary for translations of "Re: " into other languages to be recognized (even though such translations are sometimes generated by non-compliant followup agents). Likewise, that "Re: " is case-sensitive, although non-compliant agents that generate "RE: " are common enough that it might be wiser to accept that form also.

[The above wording is subject to change according to what is finally said in [[USEFOR](#)].]

- b) It is not unknown for non-compliant followup agents to truncate the Subject-header. Some reading agents therefore truncate the Subject before making any comparison. Sometimes this makes things better; sometimes it makes them worse.
- c) The use of encoded-words ([[RFC 2047](#)]) within Subject-headers can give rise to different ways of encoding the same Subject. Therefore, such encoding SHOULD be undone before any comparison of Subject-headers is made. It cannot even be assumed that the back-reference "Re: " is not within an encoded-word.

[It is possible that this matter will ultimately be addressed in [[USEFOR](#)] rather than here.]

#### **[3.3.1.2.](#) Summary**

Although this header is not widely used, reading agents SHOULD make provision for it to be displayed if present (at least as the default).

#### **[3.3.2.](#) Interpretation of Bodies**

Implementors of reading agents need to be aware of ancient usages (and even non-compliance) which nevertheless still appear from time to time, and SHOULD endeavour to recognize them and display them appropriately.

An example of this is the use of Backspace by posting agents in order to construct composite characters (e.g. by underlining) (3.1.2).

Tab (US-ASCII 9) SHOULD be interpreted as sufficient horizontal white space to reach the next of a set of fixed positions (customarily set at every 8th character). Formfeed (US-ASCII 12) (which is sometimes referred to as the "spoiler character") signifies a point at which the reading agent SHOULD pause and await reader interaction before displaying further text.

Reading agents MUST provide facilities to display the whole of long lines up to the maximum of 998 characters (whether by wrapping or by providing horizontal scroll bars). However, cutting and pasting of

wrapped lines SHOULD copy the original unwrapped line (i.e. all CRLFs not in the original should be discarded).

#### **3.3.2.1. Usage of MIME**

Even though this document, or applicable policy, may discourage the use of some Content-Types, all reading agents SHOULD make some realistic attempt to display at least all text types (especially where the Content-Disposition is "inline", even if all that can be done is to exhibit any formatting information as received (thus allowing a suitably knowledgeable reader to interpret it manually).

The same applies to unrecognized charsets. It is not expected that reading agents will necessarily be able to present characters in all possible character sets (for example, a reading agent might be able to present only the ISO-8859-1 (Latin 1) characters [ISO 8859]), but where unpresentable characters arise they SHOULD be presented in some escaped notation, e.g. octal or hexadecimal (rather than as some single distinctive glyph or by exhibiting a warning).

Reading agents MAY interpret image, audio and video Content-Types inline, but few in fact do so (and the use of such Content-Types is anyway deprecated in the absence of established policy to the contrary - see 3.1.2.2). Likewise, reading agents MAY interpret "application" types (and SHOULD at least display those types which are inherently textual in nature). However, there are security risks inherent in some application types, and even in "text/html" ([[USEFOR](#) 9.2.2). Even requiring the reader to click on some icon before proceeding with the application has proven notoriously ineffective against malicious attacks. The only safe alternative is to execute the application within a protected environment, or "sandbox", outside of which its side effects cannot occur.

Of the multipart Content-Types, reading agents MUST handle correctly at least "multipart/mixed" and "multipart/alternative". Other multipart types that are not implemented directly MUST be treated as "multipart/mixed". It is a regular practice for some Usenet articles to consist of digests of other messages or informative documents (usually known as "FAQ"s). These take the form of digests, as defined in [[RFC 1153](#)] or of the MIME Content-Type "multipart/digest". Reading agents SHOULD recognize both of these formats and enable the individual digest items to be presented separately, as if they were separate articles.

Reading agents SHOULD honour any Content-Disposition-header that is provided (in particular, they SHOULD display any part of a multipart for which the disposition is "inline", possibly distinguished from adjacent parts by some suitable separator). In the absence of such a header, the body of an article or any part of a multipart with

Content-Type "text" SHOULD be displayed inline.

### **3.4. The Well-Behaved Reply Agent**

First and foremost, a reply agent is an Email agent, and therefore its primary responsibility is to generate messages that are compliant with [[RFC 2822](#)] and other applicable Email standards and conventions.

When a reply is to be emailed to the poster of an article, the reply agent **MUST** initially create a To-header from the Reply-To- or From-header, as appropriate, of the precursor.

NOTE: A distinction is to be made between when a reply is emailed to the poster of an article, and when such a reply is also posted during the course of generating a followup; in the latter case (but not the former) it is expected that any Mail-Copies-To header will have been observed.

Note also that use of the Posted-And-Mailed header is appropriate whenever a message is both posted and emailed, whether or not this is done during the course of a formal followup.

Since addresses ending in ".invalid" are undeliverable, reply agents **SHOULD** warn any user attempting to reply to them and **SHOULD NOT**, in any case, attempt to deliver to them (since that would be pointless anyway).

### **3.5. User Interfaces**

[At this point we need to consider whether to add a section regarding the user interfaces to NUAs (commands/menus and the like). There is much in the GNKSA of that nature which we might choose to adopt. Indeed, the next step should be a careful comparison of what is contained in the GNKSA and what has been said here, since there are undoubtedly cases where requirements are less strict than those put forward in the GNKSA, and vice versa. Such a comparison might suggest some further changes and features to be considered for this draft.]

## **4. The Well-Behaved Injecting Agent**

The injecting agent bears a responsibility towards the rest of the network for ensuring both that the articles it injects are compliant with [[USEFOR](#)], and that they conform with the general expectations of the rest of the network as to what constitutes "proper behaviour". [[USEFOR](#)] therefore imposes a duty on it to check articles for compliance rather thoroughly, but also a general duty to be responsive to complaints concerning the behaviour of those who are permitted to post through it.

An injecting agent **MAY** take account of the policies of any newsgroups or hierarchies that the article is posted to (though it would be

unreasonable to expect it to be aware of the policies and idiosyncrasies of all the hierarchies that it might encounter).

As part of their responsibility for the actions of their posters, injecting agents MAY cancel articles which they have previously injected ([USEFOR] 7.3).

[That paragraph will move back to USEFOR if the rules governing who may issue cancels are moved back.]

#### **4.1. Construction of Headers**

According to [USEFOR], an injecting agent MAY add other headers not already provided by the poster, but SHOULD NOT alter, delete, or reorder any existing header. However, the addition of non-mandatory headers by the injecting agent may alter the posting agent's preferred presentation of information.

Insofar as the injecting agent needs to add headers not present in the proto-article (whether mandatory headers or otherwise), it MUST also behave as a well-behaved posting agent (3.1) with regard to those headers, including the insertion of appropriate folding so as to keep line lengths within the accepted limits.

##### **4.1.1. Sender**

The generation of the Sender-header is to be regarded as the responsibility of the posting agent. Although adding this header by injecting agents is not forbidden by [USEFOR] (though overwriting an existing one is), and although some agents indeed do so, this practice SHOULD be phased out. Exposing a sender's mailbox has privacy implications; where the main or only purpose for doing so is as tracing information, it is preferable to use instead one of the options provided for the Injection-Info-header.

##### **4.1.2. Organization**

The general discouragement from providing a default value for this header (3.1.1.7) applies even more to injecting agents. Where all the posters using a given injecting agent belong to a single organization, including the name of that organization as the default might well be reasonable. But if the injecting agent is merely providing a service to the general public, providing the name of the service provider as the default organization is mere advertising, and makes no allowance for the possibility that subscribers to the service who do not provide an Organization-header of their own might prefer not to have one at all.

##### **4.1.3. User-Agent**

There is provision in [USEFOR] for injecting agents to include (or augment if already present) a User-Agent-header to identify the software that they use but, again, use as an advertising medium (in the mundane sense) is discouraged (cf. 3.1.1.10).





#### **4.1.4. Injection-Info**

Various headers such as NNTP-Posting-[Host, Date, etc.] (which actually had nothing to do with NNTP) and X-trace have not been standardized in [USEFOR], but have instead been incorporated in the new Injection-Info-header (whose syntax incorporates more room for future extension). Use of those headers SHOULD therefore be phased out.

The purpose of the various parameters of the Injection-Info-header is to enable the injecting agent to make assertions about the origin of the article, in fulfilment of its responsibilities towards the rest of the network. These assertions can then be utilized as follows:

1. To enable the administrator of the injecting agent to respond to complaints and queries concerning the article. For this purpose, the parameters included SHOULD be sufficient to enable the administrator to identify its true origin (which parameters are best suited to this purpose will vary with the nature of the injecting site and of its relationship to the posters who use it - there is no benefit in including parameters which contribute nothing to this aim). An administrator MAY, with those parameters where the syntax so allows, use cryptic notations interpretable only by himself if he considers it appropriate to protect the privacy of that origin.
2. To enable relaying, serving and reading agents to recognize articles from origins which they might wish to reject, divert, or otherwise handle specially, for reasons of site policy.
3. To enable the timely identification of spews of articles arising from a common origin.

NOTE: Administrators of injecting agents can choose which selection of the various parameters best enables them to fulfil their responsibilities. Some of these parameters identify the source of the article explicitly whereas others do so indirectly, thus affording more privacy to posters who value their anonymity, but also making harder the tracking of malicious disruption of the network, especially so if the administrators choose not to cooperate. There is thus a balance to be struck between the needs of privacy on the one hand and the good order of Usenet on the other, and administrators need to be aware of this when formulating their policies.

#### **5. The Well-Behaved Relaying Agent**

[USEFOR] establishes as a basic principle that relaying agents are not to alter articles in any way during transmission (except for

those headers explicitly defined to be "variant"). This applies even if the article is perceived not to be conformant with [[USEFOR](#)]; in such a case it MUST either be passed on as it stands, or else it should be discarded altogether. In this way, it will be ensured that all copies of a given article, wherever they appear throughout

Usenet, will be identical.

In particular, [USEFOR] requires serving and relaying agents to accept any syntactically correct newsgroup-name in Newsgroups-headers, even if it would violate one or more of the policy restrictions set out in [section 7.2](#); i.e. the injecting agent is the last place for such checks to be made (3.1.1.5).

## **[5.1.](#) The Path Header**

It is important to be able to determine where a given article was injected into Usenet and the route it took to reach each site at which it appears. Both the Path- and Injection-Info-headers have an important part to play in this. [USEFOR] therefore imposes a strong obligation on relaying agents to verify where articles reached them from and to record this information in the Path-header. It is important that these new requirements in [USEFOR] be adopted by all injecting and relaying agents at the earliest opportunity.

### **[5.1.1.](#) Suggested Verification Methods**

It is preferable to verify the claimed path-identity against the source than to make routine use of the '?' path-delimiter ([USEFOR] 5.6.1), with consequential wasteful double-entry Path additions.

If the incoming article arrives through some TCP/IP protocol such as NNTP, the IP address of the source will be known, and will likely already have been checked against a list of known FQDNs, IP addresses, or other registered aliases that the receiving site has agreed to peer with.

Since the source host may have several IP addresses, checking the claimed FQDN or IP address against the source IP, or finding a suitable FQDN to report with a '?' path-delimiter, may involve several DNS lookups, following CNAME chains as required. Note that any reverse DNS lookup that is involved needs to be confirmed by a forward one.

If the incoming article arrives through some other protocol, such as UUCP, that protocol MUST include a means of verifying the source site. In UUCP implementations, commonly each incoming connection has a unique login name and password, and that login name (or some alias registered for it) would be expected as the path-identity.

If none of these methods is applicable, relaying agents SHOULD require connecting hosts to identify themselves using some cryptographic authentication mechanism.

[What references should be given here? SASL?]

## **[6.](#) The Well-Behaved Serving Agent**

The principles set out in [section 5](#) regarding not altering articles in any way apply equally to serving agents. The article as stored MUST be identical to the article as injected (variant headers

excepted).

## **6.1. Control Messages**

Serving agents SHOULD deny group control messages ([[USEFOR](#)] 7.2) not issued by the appropriate administrative agencies, and therefore SHOULD take such steps as are reasonably practicable to validate their authenticity, e.g. by checking digital signatures in cases where they are provided.

### **6.1.1. The 'newgroup' and 'mvgroup' Control Messages**

Serving agents SHOULD, insofar as they are conveniently able to, reject all 'newgroup' and 'mvgroup' messages not meeting the policies of the relevant hierarchy.

Since the 'mvgroup' control message was a feature newly introduced by [[USEFOR](#)], the requirements set for it were relatively light, so as to facilitate a rapid deployment within Usenet (treating it as a 'newgroup' message is minimally conformant). Nevertheless, to achieve full benefit, serving agents need to arrange to service requests for access to the old group by providing access to the new. [[USEFOR](#)] states how that MAY be done, but this document goes further; serving agents SHOULD be upgraded to do so at the earliest opportunity.

### **6.1.2. Cancel Messages**

A cancel message may be issued in the following circumstances.

1. The poster of an article (or, more specifically, any entity mentioned in the From-header or the Sender-header, whether or not that entity was the actual poster) is always entitled to issue a cancel message for that article, and serving agents SHOULD honour such requests. Posting agents SHOULD facilitate the issuing of cancel messages by posters fulfilling these criteria.
2. The agent which injected the article onto the network (more specifically, the entity identified by the path-identity in front of the leftmost '%' delimiter in the Path-header or in the Injection-Info-header and, where appropriate, the moderator (more specifically, any entity mentioned in the Approved-header) is always entitled to issue a cancel message for that article, and serving agents SHOULD honour such requests.
3. Other entities MAY be entitled to issue a cancel message for that article, in circumstances where established policy for any hierarchy or group in the Newsgroup-header, or established custom within Usenet, so allows (such policies and customs are not defined by this document). Such cancel messages MUST include an

Approved-header identifying the responsible entity. Serving agents MAY honour such requests, but SHOULD first take steps to verify their appropriateness.

[There was one request to move that back into [\[USEFOR\]](#). Any one else?]

## **[7.](#) The Well-Behaved Hierarchy Administrator**

The term "hierarchy administrator" means any agency responsible for administration of a (sub-)hierarchy (1.1), or in the absence of such an agency, the custom and usage generally accepted for that (sub-)hierarchy, insofar as such can be determined.

### **[7.1.](#) Control Messages**

In those hierarchies where appropriate administrative agencies exist (see 1.1), group control messages SHOULD NOT be issued except as authorized by those agencies, in which case the administrator needs to establish just what person (or other entity) is to be permitted to issue those messages; moreover he should at the same time establish a digital signature key to be used for authenticating them ([\[USEFOR\]](#) 7.1), and finally he SHOULD ensure that this information is widely promulgated for use by serving agents worldwide.

For compatibility with legacy news software, the Subject-content of a control message (i.e. an article that also contains a Control-header) MAY start with the string "cmsg ", and non-control messages SHOULD NOT start with the string "cmsg ".

[SHOULD NOT changed from MUST NOT. Do there really still exist servers or other agents that will recognize and act upon "cmsg" in a Subject-header? And if so, maybe that MUST NOT should be moved back into [\[USEFOR\]](#).]

The newsgroup-name in 'newgroup' control messages (and the second (new-)newsgroup-name in 'mvgroup' control messages) SHOULD conform to whatever policies have been established by the administrator (7.2).

Although, in accordance with [\[RFC 2822\]](#) and [\[USEFOR\]](#), a newsgroups-line (as found in both 'newgroup' and 'checkgroups' messages) could have a maximum length of 998 octets, as a matter of policy a far lower limit, expressed in characters, SHOULD be set. The current convention is to limit its length so that the newsgroup-name, the HTAB(s) (interpreted as 8-character tabs that takes one at least to column 24) and the newsgroup-description (excluding any moderation-flag) fit into 79 characters. This document does not seek to enforce any such rule, but any decision to extend it should be made as a specific decision for the hierarchy. Reading agents SHOULD therefore enable a newsgroups-line of any length to be displayed, e.g. by wrapping it as required.

### **[7.2.](#) Naming of Newsgroups**

Because group control messages can only be issued on the authority of

the responsible agency, it follows that the agency has complete control of the names of the newsgroups to be considered as valid members of that (sub-)hierarchy. Consequently, it needs to establish policies for the format of the newsgroup-names it intends to permit;

C. H. Lindsey

[Page 28]



these policies can be both technical and aesthetic.

[USEFOR] provides by default the following technical restrictions upon which hierarchy administrators can then build, and which SHOULD in any case be applied in hierarchies not subject to such management.

NOTE: These restrictions are intended to reflect existing practice and are intended both to avoid certain technical difficulties and to avoid unnecessary confusion. They may well change over time in the light of future experience.

1. Uppercase letters are forbidden.

NOTE: Traditionally, newsgroup-names have been written in lowercase. However, posting agents SHOULD NOT convert uppercase characters to the corresponding lowercase forms except under the explicit instructions of the poster.

2. A component name is forbidden to consist entirely of digits.

NOTE: This requirement was in [\[RFC 1036\]](#) but nevertheless several such groups have appeared in practice and implementors should be prepared for them. A common implementation technique uses each component as the name of a directory and uses numeric filenames for each article within a group. Such an implementation needs to be careful when this could cause a clash (e.g. between article 123 of group xxx.yyy and the directory for group xxx.yyy.123). Once the latter group exists, the subsequent creation of the former would be precluded for all time.

3. A component is limited to 30 component-graphemes and a newsgroup-name to 66 component-graphemes (counting also the '.'s separating the components).

NOTE: Whilst there is no longer any technical reason to limit the length of a component (formerly, it was limited to 14 octets) nor of a newsgroup-name, it should be noted that these names are also used in the newsgroups-line where another overall policy limit applies (7.1) and, moreover, excessively long names can be exceedingly inconvenient in practical use. The 66 limit on newsgroup-names ensures that a Followup-To-Header with such a name will still fit within 79 characters overall.

In the event that some future extension to [\[USEFOR\]](#) allows internationalized newsgroup-names including non-ASCII characters, there will be further technical issues to be taken into account, including:

4. What non-ASCII punctuations and other symbols are to be allowed.

5. What normalizations need to be observed to overcome multiple ways of constructing glyphs with identical or similar appearance.

6. Restrictions on mixing alphabets within one component of a name (so as to avoid confusion between, for example, Latin A and Greek Alpha, and similar confusions between some Latin and Cyrillic letters - though retaining the restriction on uppercase letters will mitigate these problems somewhat).

Aesthetic reasons for policy limitations are likely to include insistence upon a clear hierarchical structure (the tree of names needs to be neither too broad nor too deep), that the components of newsgroup-names are meaningful in the context of the language(s) expected to be used, that frivolous names are avoided, and that abbreviations are likely to be recognized by the intended readership. [David Wright has a FAQ on hierarchical naming which might give us some help.]

### **7.3. Format of Bodies**

Hierarchy administrators MAY declare, as a matter of policy, which languages and charsets are to be considered appropriate within their hierarchies (or within particular groups). Whereas in principle, any character set may be specified in the "charset=" parameter of a Content-Type, readers cannot be expected to possess agents capable of displaying characters not needed for those chosen languages, hence administrators SHOULD choose charsets accordingly and/or limit the planes to be allowed within charsets based on [UNICODE 3.2], such as UTF-8.

The document has already provided (3.1.2,3.2.2) for a default limit on the length of lines (79, or preferably 72) within plain-text articles, and hierarchy administrators MAY change this, as a matter of policy (though there would seem to be little reason to do so except where the intended language and charsets so dictate - e.g. because of a need to use double-width characters).

This document has also limited (3.1.2.2), by default, the Content-types that may be used in articles to "text/plain". Hierarchy administrators MAY relax this, as a matter of policy (by allowing, for example, "text/html", the "binary" types "audio", "image" and "video", and selected "application" types), and they MAY similarly regulate the use of "message/partial".

Hierarchy administrators MAY also impose other restrictions relevant to the nature of their hierarchy, such as limits on the overall size of articles, on the length of signatures, the topics to be discussed (usually set out in a charter for each newsgroup) and the extent of advertising to be permitted.

### **7.4. Promulgation**

The policies established by each hierarchy administrator SHOULD be publicised (in the form of guidelines, FAQs and charters) in suitable \*.announce groups within each hierarchy, and also on suitable web sites (although it should be understood that Usenet exists as a separate entity from the World Wide Web, and it would be wrong to

assume that every Usenet user has web - or even email - access).

NOTE: The promulgation of policies is one thing; the enforcement of policies is quite another. With the exception of newsgroup-names, for which technical controls exist, policy enforcement is a matter of peer pressure (which, when consistently applied, can be remarkably effective), possibly with the aid of the administrators of injecting agents through their ability, and even duty (4), to apply disciplinary pressure to their users.

## **8. The Well-Behaved Moderator**

A moderator MAY inform the poster if an article is accepted, and he SHOULD inform the poster if it is rejected (except where it appears to be a deliberate and malicious attempt to disrupt).

A moderator SHOULD NOT (absent any established and widely promulgated policy to the contrary) remove any newsgroup-name from the Newsgroups-header, nor split an article into two versions with disjoint Newsgroups-headers. These are matters more usually within the prerogative of the poster; moreover splitting can lead to fragmentation of threads.

## **9. The Well-Behaved Poster**

[What you see here is but the tip of a very large iceberg, being the particular advice to posters which has been transported the earlier drafts of [\[USEFOR\]](#). There is much more that could, and probably should, be said.

However, it would first be advisable to study [\[RFC 1855\]](#) and to decide whether we want to adopt and adapt what is already stated there, even to the extent of obsoleting it entirely.]

### **9.1. Construction of Headers**

Posters SHOULD NOT include redundant headers such as Reply-To and Followup-To that merely duplicate the defaults (c.f. 3.1.1.6 and 3.1.1.9).

#### **9.1.1. From**

Whether or not a valid address can subsequently be extracted from an address ending in ".invalid" falls outside the scope of this document but, obviously, posters wishing to disguise their address should not suppose that just adding ".invalid" to it will achieve that effect.

#### **9.1.2. Summary**

The summary should be terse. Posters SHOULD avoid trying to cram

their entire article into the headers; even the simplest query usually benefits from a sentence or two of elaboration and context, and not all reading agents display all headers. On the other hand the summary should give more detail than the Subject.

### **9.1.3. Expires**

An Expires-header should only be used in an article if the requested expiry time is earlier or later than the time typically to be expected for such articles. Local policy for each serving agent will dictate whether and when this header is obeyed and posters SHOULD NOT depend on it being completely followed.

### **9.2. Construction of Bodies**

Posters SHOULD avoid using control characters and escape sequences except for tab (US-ASCII 9), formfeed (US-ASCII 12) and, possibly, backspace (US-ASCII 8), for reasons already explained in [section 3.3.2](#).

NOTE: Backspace was historically used for underlining, done by an underscore (US-ASCII 95), a backspace, and a character, repeated for each character that should be underlined. Posters are warned that underlining is not available on all output devices or supported by all reading agents and is best not relied on for essential meaning.

When preparing followups, posters SHOULD edit quoted context to trim it down to the minimum necessary.

Posters SHOULD observe the policies established for each hierarchy (7.3) or, in the absence of such policies, to the defaults set out in this document, as regards:

- o The languages and charsets to be used;
- o The length of lines;
- o The acceptability of various Content-Types, and especially of "text/html" and the "binary" types;
- o Conventions regarding the advisability of using "message/partial";
- o Limits on the overall size of articles;
- o The topics to discussed in each group, as determined by its charter;
- o The acceptability of advertising.

## **10. References**

[ISO 8859] International Standard - Information Processing - 8-bit Single-Byte Coded Graphic Character Sets. Part 1: Latin alphabet No. 1, ISO 8859-1, 1987. Part 2: Latin alphabet No. 2, ISO 8859-2, 1987. Part 3: Latin alphabet No. 3, ISO 8859-3, 1988. Part 4: Latin alphabet No. 4, ISO 8859-4, 1988. Part 5: Latin/Cyrillic alphabet, ISO 8859-5, 1988. Part 6: Latin/Arabic alphabet, ISO 8859-6, 1987. Part 7: Latin/Greek alphabet, ISO 8859-7, 1987. Part 8: Latin/Hebrew alphabet, ISO 8859-8, 1988.

[ISO/IEC 10646] "International Standard - Information technology -  
Universal Multiple-Octet Coded Character Set (UCS) - Part 1:  
Architecture and Basic Multilingual Plane", ISO/IEC 10646-

C. H. Lindsey

[Page 32]



1:2000, 2000.

- [RFC 1036] M. Horton and R. Adams, "Standard for Interchange of USENET Messages", [RFC 1036](#), December 1987.
- [RFC 1153] F. Wancho, "Digest Message Format", [RFC 1153](#), April 1990.
- [RFC 1847] J. Galvin, S. Murphy, S. Crocker, and N. Freed, "Security Multiparts for MIME: Multipart/Signed and Multipart/Encrypted", [RFC 1847](#), October 1995.
- [RFC 1855] S. Hambridge, "Netiquette Guidelines", [RFC 1855](#), October 1995.
- [RFC 2046] N. Freed and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", [RFC 2046](#), November 1996.
- [RFC 2047] K. Moore, "MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text", [RFC 2047](#), November 1996.
- [RFC 2119] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), March 1997.
- [RFC 2396] T. Berners-Lee, R. Fielding, U.C. Irvine, and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", [RFC 2396](#), August 1998.
- [RFC 2440] J. Callas, L. Donnerhake, H. Finney, and R. Thayer, "OpenPGP Message Format", [RFC 2440](#), November 1998.
- [RFC 2606] D. Eastlake and A. Panitz, "Reserved Top Level DNS Names", [RFC 2606](#), June 1999.
- [RFC 2821] John C. Klensin and Dawn P. Mann, "Simple Mail Transfer Protocol", [RFC 2821](#), April 2001.
- [RFC 2822] P. Resnick, "Internet Message Format", [RFC 2822](#), April 2001.
- [RFC 3156] M. Elkins, D. Del Torto, R. Levien, and T. Roessler, "MIME Security with OpenPGP", [RFC 3156](#), August 2001.
- [RFC 3676] R. Gellens, "The Text/Plain Format and DelSp Parameters", [RFC 3676](#), February 2004.
- [UNICODE 3.2] The Unicode Consortium, "The Unicode Standard - Version 3.2, being an amendment to [UNICODE 3.1]", Unicode Standard Annex #28 <<http://www.unicode.org/unicode/reports/tr28>>, 2002.

[USEFOR] Charles H. Lindsey, "News Article Format", [draft-ietf-usefor-article-format](#)/\*.txt.

## [11.](#) Acknowledgements

## [12.](#) Contact Address

Editor

Charles. H. Lindsey  
5 Clerewood Avenue  
Heald Green  
Cheadle  
Cheshire SK8 3JU  
United Kingdom  
Phone: +44 161 436 6131  
Email: chl@clw.cs.man.ac.uk

[

Working group chairs

Alexey Melnikov <alexey.melnikov-usefor@isode.com>

]

Comments on this draft should preferably be sent to the mailing list of the Usenet Format Working Group at

usenet-format@landfield.com.

This draft expires six months after the date of publication (see Page 1) (i.e. in Nov 2004).

## Appendix A - Notices

### Intellectual Property

The IETF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on the IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in [BCP-11](#). Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementors or users of this specification can be obtained from the IETF Secretariat.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary

rights which may cover technology that may be required to practice this standard. Please address the information to the IETF Executive Director.

C. H. Lindsey

[Page 34]

## Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

