

**Common Format and MIME Type for CSV Files**  
**draft-shafranovich-mime-csv-05.txt**

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

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>.

This Internet-Draft will expire on October 7, 2005.

Copyright Notice

Copyright (C) The Internet Society (2005).

Abstract

This document documents the format used for Comma-Separated Values (CSV) files and registers the associated MIME type "text/csv".

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Definition of the CSV format . . . . .	<a href="#">3</a>
<a href="#">3.</a>	MIME Type Registration of text/csv . . . . .	<a href="#">5</a>
<a href="#">4.</a>	IANA Considerations . . . . .	<a href="#">6</a>
<a href="#">5.</a>	Security Considerations . . . . .	<a href="#">7</a>
<a href="#">6.</a>	Acknowledgments . . . . .	<a href="#">7</a>
<a href="#">7.</a>	References . . . . .	<a href="#">7</a>
<a href="#">7.1</a>	Normative References . . . . .	<a href="#">7</a>
<a href="#">7.2</a>	Informative References . . . . .	<a href="#">7</a>
	Author's Address . . . . .	<a href="#">8</a>
<a href="#">A.</a>	Status of This Document [To Be Removed Upon Publication] . . .	<a href="#">8</a>
<a href="#">A.1</a>	Discussion Venue . . . . .	<a href="#">8</a>
<a href="#">A.2</a>	Document Repository . . . . .	<a href="#">8</a>
<a href="#">A.3</a>	Document History . . . . .	<a href="#">8</a>
	Intellectual Property and Copyright Statements . . . . .	<a href="#">10</a>



## **1. Introduction**

The comma separated values format (CSV) has been used for exchanging and converting data between various spreadsheet programs for quite some time. Surprisingly, while this format is very common it has never been formally documented. Additionally, while the IANA MIME registration tree includes a registration for "text/tab-separated-values" type, no MIME types have ever been registered with IANA for CSV. At the same time, various programs and operating systems have begun to use different MIME types for this format, many of which vary from system to system. This document seeks to document the format of comma separated values (CSV) files and to formally register the "text/csv" MIME type for CSV in accordance with [RFC 2048 \[1\]](#).

## **2. Definition of the CSV format**

While there are various specifications and implementations for the CSV format (for ex. [\[4\]](#), [\[5\]](#), [\[6\]](#) and [\[7\]](#)), no formal specification exists which causes a wide variety of interpretations for CSV files. This section seeks to document the format that seems to be followed by most implementations:

1. Each record is located on a separate line delimited by a line break (CRLF). For example:

```
aaa,bbb,ccc CRLF
zzz,yyy,xxx CRLF
```

2. The last record in the file may or may not have an ending line break. For example:

```
aaa,bbb,ccc CRLF
zzz,yyy,xxx
```

3. There maybe an optional header line appearing as the first line of the file with the same format as normal record lines. This header will contain names corresponding to the fields in the file and should contain the same number of fields as the records in the rest of the file (the presence or absence of the header line should be indicated via the optional "header" parameter of this MIME type). For example:

```
field_name,field_name,field_name CRLF
aaa,bbb,ccc CRLF
zzz,yyy,xxx CRLF
```



4. Within the header and each record there may be one or more fields, separated by commas. Each line should contain the same number of fields throughout the file. The last field in the record may not be followed by a comma. For example:

```
aaa,bbb,ccc
```

5. Each field may or may not be enclosed in double quotes (however some programs such as Microsoft Excel do not use double quotes at all). If fields are not enclosed with double quotes, then double quotes may not appear inside the fields. For example:

```
"aaa","bbb","ccc" CRLF  
zzz,yyy,xxx
```

6. Field containing line breaks (CRLF), double quotes and commas should be enclosed in double-quotes. For example:

```
"aaa","b CRLF  
bb","ccc" CRLF  
zzz,yyy,xxx
```

7. If double-quotes are used to enclosed fields, then a double-quote appearing inside a field must be escaped by preceding it with another double quote. For example:

```
"aaa","b""bb","ccc"
```

The ABNF grammar [2] appears as follows:

```
file = [header CRLF] record *(CRLF record) [CRLF]
```

```
header = name *(COMMA name)
```

```
record = field *(COMMA field)
```

```
name = field
```

```
field = (escaped / non-escaped)
```

```
escaped = DQUOTE *(TEXTDATA / COMMA / CR / LF / 2DQUOTE) DQUOTE
```

```
non-escaped = *TEXTDATA
```

```
COMMA = %x2C
```

```
CR = %x0D ;as per section 6.1 of RFC 2234 [2]
```



DQUOTE = %x22 ;as per [section 6.1 of RFC 2234](#) [2]

LF = %x0A ;as per [section 6.1 of RFC 2234](#) [2]

CRLF = CR LF ;as per [section 6.1 of RFC 2234](#) [2]

TEXTDATA = %x20-21 / %x23-2B / %x2D-7E

### **3. MIME Type Registration of text/csv**

This section provides the media-type registration application (as per [RFC 2048](#) [1], which will be submitted to IANA after IESG approval of this document.

To: ietf-types@iana.org

Subject: Registration of MIME media type text/csv

MIME media type name: text

MIME subtype name: csv

Required parameters: none

Optional parameters: charset, header

Common usage of CSV is US-ASCII, but other character sets as defined by IANA for the "text" tree may be used in conjunction with the "charset" parameter.

The "header" parameter indicates the presence or absence of the header line. Valid values are "present" or "absent". Implementators choosing not to use this parameter must make their own decisions as to whether the header line is present or absent.

Encoding considerations:

As per [section 4.1.1. of RFC 2046](#) [3], this media type uses CRLF to denote line breaks. However, implementors should be aware that some implementations may use other values.

Security considerations:

CSV files contain passive text data which should not pose any risks. However, it is possible in theory that malicious binary data maybe included in order to exploit potential buffer overruns in the program processing CSV data. Additionally, private data





maybe shared via this format (which of course applies to any text data).

#### Interoperability considerations:

Due to lack of a single specification there are considerable differences among different implementations. Implementors should "be conservative in what you do, be liberal in what you accept from others" ([RFC 793](#) [8]) when processing CSV files. An attempt at a common definition can be found in [Section 2](#).

Implementations deciding not to use the optional "header" parameter must make their own decision as to whether the header is absent or present.

#### Published specification:

While numerous private specifications exist for various programs and systems, there is no single "master" specification for this format. An attempt at a common definition can be found in [Section 2](#).

#### Applications which use this media type:

Spreadsheet programs and various data conversion utilities

#### Additional information:

Magic number(s): none

File extension(s): CSV

Macintosh File Type Code(s): TEXT

#### Person & email address to contact for further information:

Yakov Shafranovich <[ietf@shaftek.org](mailto:ietf@shaftek.org)>

Intended usage: COMMON

Author/Change controller: IESG

## **[4.](#) IANA Considerations**

After IESG approval, IANA is expected to register the MIME type "text/csv" using the application provided in [Section 3](#) of this document.



## **5. Security Considerations**

See discussion above

## **6. Acknowledgments**

The author would like to thank Dave Crocker, Martin Duerst, Joel M. Halpern, Clyde Ingram, Graham Klyne, Bruce Lilly, Chris Lilley and members of the IESG for their helpful suggestions. A special word of thanks to Dave for helping with the ABNF grammar.

The author would also like to thank Henrik Lefkowitz, Marshall Rose and the folks at [xml.resource.org](http://xml.resource.org) for providing many of the tools used for preparing RFCs and Internet drafts.

## **7. References**

### **7.1 Normative References**

- [1] Freed, N., Klensin, J., and J. Postel, "Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures", [BCP 13](#), [RFC 2048](#), November 1996.
- [2] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", [RFC 2234](#), November 1997.
- [3] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", [RFC 2046](#), November 1996.

### **7.2 Informative References**

- [4] Repici, J., "HOW-TO: The Comma Separated Value (CSV) File Format", 2004,  
<<http://www.creativyst.com/Doc/Articles/CSV/CSV01.htm>>.
- [5] Edoceo, Inc., "CSV Standard File Format", 2004,  
<<http://www.edoceo.com/utilis/csv-file-format.php>>.
- [6] Rodger, R. and O. Shanaghy, "Documentation for Ricebridge CSV Manager", February 2005,  
<<http://www.ricebridge.com/products/csvman/reference.htm>>.
- [7] Raymond, E., "The Art of Unix Programming, Chapter 5", September 2003,  
<<http://www.catb.org/~esr/writings/taoup/html/ch05s02.html>>.
- [8] Postel, J., "Transmission Control Protocol", STD 7, [RFC 793](#),



September 1981.

#### Author's Address

Yakov Shafranovich  
SolidMatrix Technologies, Inc.

Email: [ietf@shaftek.org](mailto:ietf@shaftek.org)

URI: <http://www.shaftek.org>

## **Appendix A. Status of This Document [To Be Removed Upon Publication]**

### **A.1 Discussion Venue**

Discussion about this document should be directed to the IETF-TYPES mailing list <<http://www.alvestrand.no/mailman/listinfo/ietf-types/>> which is also reachable via <[ietf-types@iana.org](mailto:ietf-types@iana.org)>. Of course, comments directly to the author are always welcome.

### **A.2 Document Repository**

Copies of this and earlier versions including multiple formats can be found at <<http://www.shaftek.org/publications/drafts/mime-csv/>>.

### **A.3 Document History**

Changes from [draft-shafranovich-mime-csv-04](#) to [draft-shafranovich-mime-csv-05](#):

- o Fixes in ABNF grammar: comma excluded from TEXTDATA and added to escaped text, 2\*DQUOTE changed to 2DQUOTE.

Changes from [draft-shafranovich-mime-csv-03](#) to [draft-shafranovich-mime-csv-04](#):

- o Fixed ABNF grammar in response to IESG comments: VCHAR changed to TEXTDATA, DQUOTE excluded from TEXTDATA, spaces are included in TEXTDATA, CRLF taken out since CR and LF are already included.
- o Added clarification that double quotes may not be included inside non-quoted fields.
- o Added clarification that the same number of fields should appear in every line.
- o Added the optional "header" parameter to indicate the presence or absence of a header line.



Changes from [draft-shafranovich-mime-csv-02](#) to [draft-shafranovich-mime-csv-03](#):

- o Changed text to prohibit the last field ending with a comma matching the ABNF grammar
- o The double quote escaping is now set to two double quotes instead of three
- o Moved some of the references between informative and normative sections

Changes from [draft-shafranovich-mime-csv-01](#) to [draft-shafranovich-mime-csv-02](#):

- o Minor errors in ABNF grammar corrected in response to AD comments
- o Minor spelling mistakes corrected

Changes from [draft-shafranovich-mime-csv-00](#) to [draft-shafranovich-mime-csv-01](#):

- o Type "text/comma-separated-values" has been removed
- o The "encoding consideration" paragraph of [Section 3](#) has been changed to allow CRLF only as per [section 4.1.1. of RFC 2046](#) [3]. This has been reflected in the ABNF grammar in [Section 2](#).
- o ABNF grammar in [Section 2](#) has been cleaned up.
- o Acknowledgements and status sections were added.
- o CSV format definition was moved to the normative section of the document





## Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights 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; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat 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 implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at [ietf-ipr@ietf.org](mailto:ietf-ipr@ietf.org).

## Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM 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.

## Copyright Statement

Copyright (C) The Internet Society (2005). This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

## Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

