Jump to content

Main menu

Main menu

move to sidebar hide

Navigation

- Main page
- Contents
- Current events
- Random article
- About Wikipedia
- Contact us

Contribute

- Help
- Learn to edit
- Community portal
- Recent changes
- Upload file
- Special pages



WIKIPEDIA 25

25 years of the free encyclopedia

Search

Search

Appearance

- Donate
- Create account

- Log in

Personal tools

- Donate
- Create account
- Log in

Contents

move to sidebar hide

- (Top)
- 1 History
- 2 Specification
- 3 Applications
- 4 See also
- 5 References
- 6 Further reading

Toggle the table of contents

Comma-separated values

31 languages

- العربية
- Català
- Čeština
- Cymraeg
- Deutsch
- Esperanto
- Español
- Euskara
- فارسی
- Suomi
- Français
- עברית
- Bahasa Indonesia
- Italiano
- 日本語
- 한국어
- Lëtzebuergesch
- Lietuvių
- Nederlands
- Norsk bokmål
- Polski
- Português
- Русский
- Simple English
- Slovenčina

Comma-separated values

`.csv`
`text/csv`[1]
public.comma-separated-values-text[2]
public.delimited-values-text[2]
multi-platform, serial data streams
database information organized as field separated lists
RFC 4180

Comma-separated values (**CSV**) is a plain text data format for storing tabular data where the fields (values) of a record are separated by a comma as the delimiter and each record is a line (i.e. newline is the delimiter). CSV is commonly used in software that generally deals with tabular data such as a database or a spreadsheet.[3] Benefits cited for using CSV include simplicity of use and human readability.[4] CSV is a form of delimiter-separated values. A **CSV file** is a file that contains CSV-formatted data.

CSV is not limited to a particular character encoding[1] but should be and is commonly used with UTF-8, particularly because it does not provide a way to indicate the character encoding.

History

[edit]

The CSV format predates personal computers by more than a decade. The IBM Fortran (level H extended) compiler under OS/360 supported list-directed ("free-form") input/output, with commas between values, in 1972.[5] List-directed input/output was defined in FORTRAN 77, approved in 1978. List-directed input used commas or spaces for delimiters, so unquoted character strings could not contain commas or spaces.[6]

The term "comma-separated value" and the "CSV" abbreviation were in use by 1983.[7] The manual for the Osborne Executive computer, which bundled the SuperCalc spreadsheet, documents the CSV quoting convention that allows strings to contain embedded commas.[8]

Comma-separated value lists are easier to type (for example into punched cards) than fixed-column-aligned data, and they were less prone to producing incorrect results if a value was punched one column off from its intended location.

Comma separated files are used for the interchange of database information between machines of two different architectures. The plain-text character of CSV files largely avoids incompatibilities

such as byte-order and word size. The files are largely human-readable, so it is easier to deal with them in the absence of perfect documentation or communication.[9]

The main standardization initiative—transforming "*de facto* fuzzy definition" into a more precise and *de jure* one—was in 2005, with

RFC 4180, defining CSV as a MIME Content Type.[10] Later, in 2013, some of RFC 4180's deficiencies were tackled by a W3C recommendation.[11]

In 2014 IETF published

RFC 7111 describing the application of URI fragments to CSV documents. RFC 7111 specifies how row, column, and cell ranges can be selected from a CSV document using position indexes.[12]

In 2015 W3C, in an attempt to enhance CSV with formal semantics, publicized the first *drafts of recommendations* for CSV metadata standards, which began as *recommendations* in December of the same year.[13]

Specification

[edit]

CSV can be informally described as plain text data consisting of one record per line, where each line has the same sequence of fields separated by a comma.[1][14][15] For a simple example:

```
id,name,email
1,John,john.doe@example.com
2,Jane,janey72@test.org
```

The format is more formally described in the 2005 technical standard

RFC 4180 which codifies the CSV format and defines the MIME type `text/csv` for the handling of text-based fields. Among its requirements:

- A line is terminated per MS-DOS-style: carriage return and line feed (CR/LF) sequence
- A line terminator is optional for the last line
- The data can start with a header record but with no way to test whether the first line is, in fact, a header, care is required when importing
- Each record should contain the same number of fields
- A field containing a comma, double quote or line terminator character should be enclosed in double quotes
- Any field may be enclosed in double quotes
- If a field is enclosed in double quotes, then a double quote embedded in the field must be represented by a sequence of two double quotes

A more complex example, with some of the fields enclosed in double quotes, and fields containing special characters (double quotes, line terminators, commas):

```
Year,Make,Model,Description,Price
1997,Ford,E350,"ac, abs, moon",3000.00
1999,Chevy,"Venture ""Extended Edition""","",4900.00
1999,Chevy,"Venture ""Extended Edition, Very Large""","",5000.00
1996,Jeep,Grand Cherokee,"MUST SELL!
air, moon roof, loaded",4799.00
```

This example illustrates that a CSV cannot be parsed by naïvely splitting the data by line terminators into lines, and then each line by commas. The above data, when correctly parsed, can be represented as this table:

Year	Make	Model	Description	Price
1997	Ford	E350	ac, abs, moon	3000.00
1999	Chevy	Venture “Extended Edition”		4900.00
1999	Chevy	Venture “Extended Edition, Very Large”		5000.00
1996	Jeep	Grand Cherokee	MUST SELL! air, moon roof, loaded	4799.00

Common challenges with CSV include:

- Programs may not support line terminator characters within a field even when properly quoted
- Programs may confuse a header line with data or interpret the first data line as a header
- Double quotes in a field may not be parsed correctly

In 2011, Open Knowledge Foundation (OKF) and various partners created a data protocols working group, which later evolved into the Frictionless Data initiative. One of the main formats they released was the Tabular Data Package. Tabular Data package was heavily based on CSV, using it as the main data transport format and adding basic type and schema metadata. (CSV lacks any type information to distinguish the string 1 from the number 1.)[16] The Frictionless Data Initiative has also provided a standard CSV Dialect Description Format for describing different dialects of CSV, for example specifying the field separator or quoting rules.[17]

In 2013, the W3C “CSV on the Web” working group began to specify technologies providing higher interoperability for web applications using CSV or similar formats.[18] The working group completed its work in February 2016 and is officially closed in March 2016 with the release of a set of documents and W3C recommendations[19] for modeling “Tabular Data”,[13] and enhancing CSV with metadata and semantics. While the well-formedness of CSV data can readily be checked, testing validity and canonical form is less well developed, relative to more precise data formats, such as XML and SQL, which offer richer types and rules-based validation.[20]

Applications

[edit]

CSV is commonly used for data exchange and is widely supported by data-oriented applications. It is often used to move tabular data between programs that natively operate on incompatible data – often in formats that are proprietary or undocumented.[1][21][22] A common scenario is moving data from a database to a spreadsheet which, in general, use completely different formats. Most database systems can export as CSV and most spreadsheet programs can import CSV-formatted data, leveraging CSV as an intermediate format. Every major ecommerce platform provides support for exporting data as a CSV file.[23]

CSV is also used for storing data. Common data science tools such as Pandas include the option to export data to CSV for long-term storage.[24] Benefits of CSV for data storage include the simplicity of CSV makes parsing and creating CSV files easy to implement and fast compared to other data formats, human readability making editing or fixing data simpler,[25] and high compressibility

leading to smaller data files.[26] Alternatively, CSV does not support more complex data relations and makes no distinction between null and empty values, and in applications where these features are needed other formats are preferred.

More than 200 local, regional, and national data portals, such as those of the UK government and the European Commission, use CSV files with standardized data catalogs.[27]

Some applications use CSV as a data interchange format to enhance its interoperability, exporting and importing CSV. Others use CSV as an internal format. CSV is supported by almost all spreadsheets and database management systems.

Spreadsheets including Apple Numbers, LibreOffice Calc, and Apache OpenOffice Calc. support reading CSV files. Microsoft Excel also supports a dialect of CSV with restrictions in comparison to other spreadsheet software (e.g., as of 2019[update]

Excel still cannot export CSV files in the commonly used UTF-8 character encoding, and separator is not enforced to be the comma). LibreOffice Calc CSV importer is actually a more generic delimited text importer, supporting multiple separators at the same time as well as field trimming.

Various relational databases support saving query results to a CSV file. PostgreSQL provides the `COPY` command, which allows for both saving and loading data to and from a file. `COPY (SELECT * FROM articles) TO '/home/wikipedia/file.csv' (FORMAT csv)` saves the content of a table `articles` to a file called `/home/wikipedia/file.csv`.[28] Some relational databases, when using standard SQL, offer *foreign-data wrapper* (FDW). For example, PostgreSQL offers the `CREATE FOREIGN TABLE`[29] and `CREATE EXTENSION file_fdw`[30] commands to configure any variant of CSV. Databases like Apache Hive offer the option to express CSV or .csv.gz as an internal table format.

Programs that work with CSV may have limits on the maximum number of rows CSV files can have. Examples include Microsoft Excel (1,048,576 rows), Apple Numbers (1,000,000 rows), Google Sheets (10,000,000 cells), and OpenOffice and LibreOffice (1,048,576 rows).[31]

See also

[edit]

- Comparison of data-serialization formats
- Delimiter collision – Formatting challenge caused by data that contains a delimiter
- Flat-file database – Database stored as flat data
- Tab-separated values – Text format for tabular data using a tab between fields

References

[edit]

1. 1 2 3 4 Shafranovich, Y. (October 2005). *Common Format and MIME Type for CSV Files*. IETF. p. 1. doi:10.17487/RFC4180.

   RFC 4180.
2. 1 2 "commaSeparatedText". *Apple Developer Documentation: Uniform Type Identifiers*. Apple Inc. Archived from the original on 2023-05-22. Retrieved 2023-05-22.
3. ↑ "Import or export text (.txt or .csv) files". *Microsoft Support*. Retrieved 2023-08-16.
4. ↑ "What is a CSV file: A comprehensive guide". *flatfile.com*. Retrieved 2024-10-28.
5. ↑ *IBM FORTRAN Program Products for OS and the CMS Component of VM/370 General Information* (PDF) (first ed.), July 1972, p. 17, GC28-6884-0, archived (PDF) from the original on March 4,

2016, retrieved February 5, 2016, For users familiar with the predecessor FORTRAN IV G and H processors, these are the major new language capabilities
6. ↑ "List-Directed I/O", *Fortran 77 Language Reference*, Oracle, archived from the original on 2021-02-26, retrieved 2012-10-26
7. ↑ "SuperCalc², spreadsheet package for IBM, CP/M". Retrieved December 11, 2017.
8. ↑ "Comma-Separated-Value Format File Structure". 1983. Retrieved December 11, 2017.
9. ↑ "CSV, Comma Separated Values (RFC 4180)". *Library of Congress*. Retrieved September 22, 2025.
10. ↑ *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. IETF. doi:10.17487/RFC4180.

    RFC 4180. Retrieved December 22, 2020.
11. ↑ See sparql11-results-csv-tsv, the first W3C recommendation scoped in CSV and filling some of RFC 4180's deficiencies.
12. ↑ *URI Fragment Identifiers for the text/csv Media Type*. IETF. doi:10.17487/RFC7111.

    RFC 7111. Retrieved December 22, 2020.
13. 1 2 "Model for Tabular Data and Metadata on the Web". 17 December 2015. Retrieved March 23, 2016. (W3C Recommendation)
14. ↑ "Comma Separated Values (CSV) Standard File Format". Edoceo, Inc. Archived from the original on July 14, 2020. Retrieved June 4, 2014.
15. ↑ * Creativyst (2010), *How To: The Comma Separated Value (CSV) File Format*, creativyst.com, archived from the original on April 4, 2021, retrieved May 24, 2010
16. ↑ "Tabular Data Package". *Frictionless Data Specs*.
17. ↑ "CSV Dialect". *Frictionless Data Specs*.
18. ↑ "CSV on the Web Working Group". W3C CSV WG. 2013. Retrieved 2015-04-22.
19. ↑ "CSV on the Web Repository". *GitHub*. (on GitHub)
20. ↑ "Rules Or Schemas". CsvPath Project. 2024. Retrieved 2025-02-13.
21. ↑ "CSV - Comma Separated Values". Archived from the original on 2021-03-07. Retrieved 2017-12-02.
22. ↑ "CSV Files". Archived from the original on April 30, 2021. Retrieved June 4, 2014.
23. ↑ "CSV Supported Ecommerce Platforms". *RFM Calc*. Retrieved 2025-03-09.
24. ↑ "pandas.DataFrame.to_csv — pandas 2.0.3 documentation". *pandas.pydata.org*. Retrieved 2023-08-16.
25. ↑ "CSV Format: History, Advantages and Why It Is Still Popular". *ByteScout*. 2021-09-15. Retrieved 2023-08-16.
26. ↑ "Comparison of different file formats in Big Data". *www.adaltas.com*. 2020-07-23. Retrieved 2023-08-16.
27. ↑ Mahmud, S M Hasan; Hossin, Md Altab; Jahan, Hosney; Noori, Sheak Rashed Haider; Bhuiyan, Touhid (2018). *CSV-ANNOTATE: Generate annotated tables from CSV file*. 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE. pp. 71–75. doi:10.1109/ICAIBD.2018.8396169. ISBN 978-1-5386-6987-7.
28. ↑ "Documentation: 14: COPY". PostgreSQL. Retrieved 2024-05-12.
29. ↑ "Documentation: 14: F.35. postgres_fdw". PostgreSQL. 2022-02-10. Retrieved 2022-03-04.
30. ↑ "Documentation: 14: F.14. file_fdw". PostgreSQL. 2022-02-10. Retrieved 2022-03-04.
31. ↑ "Understanding CSV and row limits". Archived from the original on September 15, 2025. Retrieved Sep 23, 2026.

Further reading

[edit]

- “IBM DB2 Administration Guide - LOAD, IMPORT, and EXPORT File Formats”. IBM. Archived from the original on 2016-12-13. Retrieved 2016-12-12. (Has file descriptions of delimited ASCII (.DEL) (including comma- and semicolon-separated) and non-delimited ASCII (.ASC) files for data transfer.)

* v * t * e Data exchange formats
* CSV * EDIFACT * JSON * Web Encryption * Web Token * Web Signature * Markdown * Property list * RDF * Rebol * TOML * TOON * XML * Atom * YAML
* AMF * Ascii85 * ASN.1 * SMI * Avro * Base32 * Base64 * Bencode * BSON * UBJSON * Cap'n Proto * CBOR * FlatBuffers * MessagePack * Property list * Protocol Buffers * Thrift * Cyphal DSDL * XDR * uuencode * yEnc
Comparison of data-serialization formats

Retrieved from “https://en.wikipedia.org/w/index.php?title=Comma-separated_values&oldid=1379342910”

Categories:

- Delimiter-separated formats
- Open formats
- Spreadsheet file formats

Hidden categories:

- Articles with short description
- Short description is different from Wikidata
- Articles containing potentially dated statements from 2019
- All articles containing potentially dated statements
- This page was last edited on 9 October 2026, at 10:54 (UTC).
- Page was rendered with Parsoid.
- Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
- Privacy policy
- About Wikipedia
- Disclaimers
- Contact Wikipedia
- Legal & safety contacts
- Code of Conduct
- Developers
- Statistics
- Cookie statement
- Mobile view





Search

Search

Toggle the table of contents

Comma-separated values

31 languages Add topic