

Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: 4 September 2025

C. Bormann  
Universität Bremen TZI  
H. Birkholz  
Fraunhofer SIT  
3 March 2025

Using CDDL for CSVs  
draft-bormann-cbor-cddl-csv-06

## Abstract

The Concise Data Definition Language (CDDL), standardized in RFC 8610, is defined to provide data models for data shaped like JSON or CBOR.

Another representation format that is quite popular is the CSV (Comma-Separated Values) file as defined by RFC 4180.

The present document shows a way how to use CDDL to provide a data model for CSV files.

## Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on 4 September 2025.

## Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights

and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

1. Introduction . . . . .	2
1.1. Terminology . . . . .	2
2. CSV generic data model . . . . .	2
3. Examples . . . . .	4
4. IANA Considerations . . . . .	5
5. Security considerations . . . . .	5
6. References . . . . .	5
6.1. Normative References . . . . .	5
6.2. Informative References . . . . .	6
Appendix A. Example: ietf-system.sid represented in CSV . . . .	7
Acknowledgements . . . . .	9
Authors' Addresses . . . . .	9

## 1. Introduction

The Concise Data Definition Language (CDDL), standardized in [RFC8610], is defined to provide data models for data shaped like JSON or CBOR.

Another representation format that is quite popular is the CSV file as defined by [RFC4180].

The present document shows how to use CDDL to provide a data model for CSV files.

### 1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [BCP14] (RFC2119) (RFC8174) when, and only when, they appear in all capitals, as shown here.

This specification uses terminology from [RFC8610].

## 2. CSV generic data model

The CSV format is defined in [RFC4180]. The generic data model for the data in a CSV file can be described in CDDL as:

```
csv = [?header, *record]
header = [+header-field]
record = [+field]
header-field = text
field = text
```

Note that the elements of this data model describe the interpretation of the data after processing and removal of lexical structure such as newlines, commas, escape characters, and quotation marks.

For the purposes of a specific application, the data model level structure of each field may be described in a more elaborate way, e.g., as a number. A recently approved CDDL extension, [I-D.ietf-cbor-cddl-more-control], provides some CDDL control operators that could be used to express the transformation between the text string in the CSV field and the number that this text string represents at the application data model level; this could be explored in future revisions of this specification. For now, the usage of anything but "text" for a field therefore **MUST** be accompanied by an instruction how to perform the translation. As a preferred choice, the JSON representation of the data model item, if it exists, **MAY** be chosen by that instruction.

Since the CSV media type text/csv defaults to using the US-ASCII character set (i.e., [STD80]; see Section 3 of [RFC4180]), many uses of CSV will need to specify the media type parameter charset. (Note that CDDL can describe text information that is in UTF-8 form, which includes US-ASCII as that is a subset of UTF-8. If a different form that is not a subset of UTF-8 is really still needed, some rules for conversion will need to be defined by the application.)

The media type parameter header **MAY** be used to indicate the presence or absence of a header line; if it is not given, the grammar **MUST NOT** be ambiguous about the presence of a header (i.e., it **MUST** be either mandatory or absent).

Note that the ABNF [STD68] in [RFC4180] does not quite handle the case that charset is not us-ascii. For the purposes of the present specification, the ABNF is understood to allow all characters from the charset except %x22 and %x2C in TEXTDATA. For the purposes of the present specification, the ABNF rule CRLF is read as:

```
CRLF = [CR] LF
```

as is hinted in Section 3 of [RFC4180].

### 3. Examples

A simplified CSV form definition of a SID file [RFC9595] might look like this:

```
; header = absent
```

```
SID-File = [meta-record,  
            ?description-record,  
            *dependency-record,  
            *range-record,  
            *item-record]
```

```
meta-record = ["ietf-sid-file",  
              module-name: text,  
              module-revision: empty / text,  
              sid-file-revision: empty / text,  
              sid-file-status: empty / "unpublished" / "published"]
```

```
description-record = ["description",  
                     description: empty / text]
```

```
dependency-record = ["dependency",  
                    module-name: text,  
                    module-revision: text]
```

```
range-record = ["range",  
               entry-point: uint,  
               size: uint]
```

```
item-record = [; "item", -- useful to elide for bulk of file  
              sid: uint  
              (  
                namespace: "module" / "identity" / "feature"  
                identifier: yang-identifier  
              //  
                namespace: "data"  
                identifier: schema-node-path  
              )  
              status: empty / "stable" / "unstable" / "obsolete"]
```

```
yang-identifier = text .abnf ("yang-identifier" .det id-abnf)  
schema-node-path = text .abnf ("schema-node-path" .det id-abnf)  
id-abnf = '  
  schema-node-path = "/" QID *( "/" OQID)  
  yang-identifier = ID  
  QID = ID ":" ID  
  OQID = ID [ ":" ID]
```

```
ID = I *C
I = "_" / %x41-5a / %x61-7a
C = I / %x30-39 / "-" / "."
,
```

```
empty = ""
```

This CDDL data model assumes that the text strings representing the numbers entry-point, size, and sid are converted to uint. (Note that, due to the way YANG-JSON [RFC7951] defines the representation of uint64 data items, these actually are text strings in JSON, which in CSV is indistinguishable from numbers. However, the CDDL model for the CSV files will be more useful if it takes into account typical CSV applications that automatically convert integer-like text strings into numbers.)

The result of representing in CSV the sid file ietf-system.sid (as defined in Appendix A of [RFC9595]) is shown in Appendix A.

#### 4. IANA Considerations

This document makes no requests of IANA.

#### 5. Security considerations

The security considerations of [RFC8610] and [RFC4180] apply.

#### 6. References

##### 6.1. Normative References

- [BCP14] Best Current Practice 14,  
<<https://www.rfc-editor.org/info/bcp14>>.  
At the time of writing, this BCP comprises the following:
- Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC4180] Shafranovich, Y., "Common Format and MIME Type for Comma-Separated Values (CSV) Files", RFC 4180, DOI 10.17487/RFC4180, October 2005, <<https://www.rfc-editor.org/rfc/rfc4180>>.

- [RFC8610] Birkholz, H., Vigano, C., and C. Bormann, "Concise Data Definition Language (CDDL): A Notational Convention to Express Concise Binary Object Representation (CBOR) and JSON Data Structures", RFC 8610, DOI 10.17487/RFC8610, June 2019, <<https://www.rfc-editor.org/rfc/rfc8610>>.
- [STD68] Internet Standard 68,  
<<https://www.rfc-editor.org/info/std68>>.  
At the time of writing, this STD comprises the following:
- Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234,  
DOI 10.17487/RFC5234, January 2008,  
<<https://www.rfc-editor.org/info/rfc5234>>.

## 6.2. Informative References

- [I-D.ietf-cbor-cddl-more-control]  
Bormann, C., "Concise Data Definition Language (CDDL): Additional Control Operators for the Conversion and Processing of Text", Work in Progress, Internet-Draft, draft-ietf-cbor-cddl-more-control-08, 9 January 2025, <<https://datatracker.ietf.org/doc/html/draft-ietf-cbor-cddl-more-control-08>>.
- [RFC7951] Lhotka, L., "JSON Encoding of Data Modeled with YANG", RFC 7951, DOI 10.17487/RFC7951, August 2016, <<https://www.rfc-editor.org/rfc/rfc7951>>.
- [RFC8792] Watsen, K., Auerswald, E., Farrel, A., and Q. Wu, "Handling Long Lines in Content of Internet-Drafts and RFCs", RFC 8792, DOI 10.17487/RFC8792, June 2020, <<https://www.rfc-editor.org/rfc/rfc8792>>.
- [RFC9595] Veillette, M., Ed., Pelov, A., Ed., Petrov, I., Ed., Bormann, C., and M. Richardson, "YANG Schema Item Identifier (YANG SID)", RFC 9595, DOI 10.17487/RFC9595, July 2024, <<https://www.rfc-editor.org/rfc/rfc9595>>.
- [STD80] Internet Standard 80,  
<<https://www.rfc-editor.org/info/std80>>.  
At the time of writing, this STD comprises the following:
- Cerf, V., "ASCII format for network interchange", STD 80, RFC 20, DOI 10.17487/RFC0020, October 1969,  
<<https://www.rfc-editor.org/info/rfc20>>.

## Appendix A. Example: ietf-system.sid represented in CSV

This appendix shows the CSV file that is automatically generated from Appendix A of [RFC9595]. (Note that plaintext-based RFCs are limited to 72 columns; therefore five long lines in the CSV file have been folded as defined in [RFC8792].)

===== NOTE: '\\' line wrapping per RFC 8792 =====

```
ietf-sid-file,ietf-system,2014-08-06,,
description,Example sid file
dependency,ietf-yang-types,2013-07-15
dependency,ietf-inet-types,2013-07-15
dependency,ietf-netconf-acm,2018-02-14
dependency,iana-crypt-hash,2014-08-06
range,1700,100
1700,module,ietf-system,
1701,identity,authentication-method,
1702,identity,local-users,
1703,identity,radius,
1704,identity,radius-authentication-type,
1705,identity,radius-chap,
1706,identity,radius-pap,
1707,feature,authentication,
1708,feature,dns-udp-tcp-port,
1709,feature,local-users,
1710,feature,ntp,
1711,feature,ntp-udp-port,
1712,feature,radius,
1713,feature,radius-authentication,
1714,feature,timezone-name,
1715,data,/ietf-system:set-current-datetime,
1775,data,/ietf-system:set-current-datetime/input,
1776,data,/ietf-system:set-current-datetime/input/current-datetime,
1717,data,/ietf-system:system,
1718,data,/ietf-system:system-restart,
1719,data,/ietf-system:system-shutdown,
1720,data,/ietf-system:system-state,
1721,data,/ietf-system:system-state/clock,
1722,data,/ietf-system:system-state/clock/boot-datetime,
1723,data,/ietf-system:system-state/clock/current-datetime,
1724,data,/ietf-system:system-state/platform,
1725,data,/ietf-system:system-state/platform/machine,
1726,data,/ietf-system:system-state/platform/os-name,
1727,data,/ietf-system:system-state/platform/os-release,
1728,data,/ietf-system:system-state/platform/os-version,
1729,data,/ietf-system:system/authentication,
1730,data,/ietf-system:system/authentication/user,
```

```
1731,data,/ietf-system:system/authentication/user-authentication-\
                                         order,
1732,data,/ietf-system:system/authentication/user/authorized-key,
1733,data,/ietf-system:system/authentication/user/authorized-key/\
                                         algorithm,
1734,data,/ietf-system:system/authentication/user/authorized-key/key\
                                         -data,
1735,data,/ietf-system:system/authentication/user/authorized-key/\
                                         name,
1736,data,/ietf-system:system/authentication/user/name,
1737,data,/ietf-system:system/authentication/user/password,
1738,data,/ietf-system:system/clock,
1739,data,/ietf-system:system/clock/timezone-name,
1740,data,/ietf-system:system/clock/timezone-utc-offset,
1741,data,/ietf-system:system/contact,
1742,data,/ietf-system:system/dns-resolver,
1743,data,/ietf-system:system/dns-resolver/options,
1744,data,/ietf-system:system/dns-resolver/options/attempts,
1745,data,/ietf-system:system/dns-resolver/options/timeout,
1746,data,/ietf-system:system/dns-resolver/search,
1747,data,/ietf-system:system/dns-resolver/server,
1748,data,/ietf-system:system/dns-resolver/server/name,
1749,data,/ietf-system:system/dns-resolver/server/udp-and-tcp,
1750,data,/ietf-system:system/dns-resolver/server/udp-and-tcp/\
                                         address,
1751,data,/ietf-system:system/dns-resolver/server/udp-and-tcp/port,
1752,data,/ietf-system:system/hostname,
1753,data,/ietf-system:system/location,
1754,data,/ietf-system:system/ntp,
1755,data,/ietf-system:system/ntp/enabled,
1756,data,/ietf-system:system/ntp/server,
1757,data,/ietf-system:system/ntp/server/association-type,
1758,data,/ietf-system:system/ntp/server/iburst,
1759,data,/ietf-system:system/ntp/server/name,
1760,data,/ietf-system:system/ntp/server/prefer,
1761,data,/ietf-system:system/ntp/server/udp,
1762,data,/ietf-system:system/ntp/server/udp/address,
1763,data,/ietf-system:system/ntp/server/udp/port,
1764,data,/ietf-system:system/radius,
1765,data,/ietf-system:system/radius/options,
1766,data,/ietf-system:system/radius/options/attempts,
1767,data,/ietf-system:system/radius/options/timeout,
1768,data,/ietf-system:system/radius/server,
1769,data,/ietf-system:system/radius/server/authentication-type,
1770,data,/ietf-system:system/radius/server/name,
1771,data,/ietf-system:system/radius/server/udp,
1772,data,/ietf-system:system/radius/server/udp/address,
1773,data,/ietf-system:system/radius/server/udp/authentication-port,
```



1774,data,/ietf-system:system/radius/server/udp/shared-secret,

#### Acknowledgements

Rob Wilton, unknowingly, made us write this specification. We hope it will be useful. Laurent Toutain inspired the SID CDDL format with an example.

#### Authors' Addresses

Carsten Bormann  
Universitt Bremen TZI  
Postfach 330440  
D-28359 Bremen  
Germany  
Phone: +49-421-218-63921  
Email: cabo@tzi.org

Henk Birkholz  
Fraunhofer SIT  
Rheinstrasse 75  
64295 Darmstadt  
Germany  
Email: henk.birkholz@sit.fraunhofer.de