



DDI Alliance Controlled Vocabulary for CharacterSet

CV definition

Standard set of characters upon which many character encodings are based (Wikipedia).

Details

CV short name: CharacterSet

CV name: Character Set

CV notes:

This vocabulary lists some popular character sets but is not exhaustive. If the character set is not mentioned in the list, use 'Other' for this element and enter the specific name into the OtherValue element. You can search for the name of the character set in the list published by IANA (Internet Assigned Character Numbers). Use the "preferred MIME-name" if the IANA list specifies one, otherwise use the format in the Name field - <http://www.iana.org/assignments/character-sets>. The vocabulary was first published by the DDI Alliance. Please see: <https://ddialliance.org/controlled-vocabularies/all>.

Language: English (en)

Version: 1.0.1

Version notes:

To align with a new versioning system, this version has been automatically created by copying the previous version PUBLISHED-1.0.0.

Canonical URI: <urn:ddi:int.ddi.cv.CharacterSet:1.0.1>

Agency: [DDI Alliance](#)

Code list

| Code value | Code descriptive term | Code definition |
|------------|-----------------------|--|
| ASCII | ASCII | The official name is US-ASCII but use the format: ASCII. The ISO code for ASCII is ISO 14962. |
| ISO88591 | ISO-8859-1 | For ISO standards, use format: ISO-n-n in Caption and ISOnn in Code. ISO standards are also known as Latin, for example, ISO-8859-1 as Latin1. |
| ISO88592 | ISO-8859-2 | |
| ISO88593 | ISO-8859-3 | |
| ISO88594 | ISO-8859-4 | |
| ISO88595 | ISO-8859-5 | |
| ISO88596 | ISO-8859-6 | |
| ISO88597 | ISO-8859-7 | |

| Code value | Code descriptive term | Code definition |
|-------------|-----------------------|---|
| ISO88598 | ISO-8859-8 | |
| ISO88599 | ISO-8859-9 | |
| ISO885910 | ISO-8859-10 | |
| ISO885911 | ISO-8859-11 | |
| ISO885913 | ISO-8859-13 | |
| ISO885914 | ISO-8859-14 | |
| ISO885915 | ISO-8859-15 | |
| ISO885916 | ISO-8859-16 | |
| MacOSRoman | Mac OS Roman | |
| UTF8 | UTF-8 | For Unicode Transformation Formats, use format: UTF-n in Caption and UTFn in Code. |
| UTF16 | UTF-16 | |
| UTF32 | UTF-32 | |
| Windows1251 | Windows-1251 | For MS-Windows character sets, use format: Windows-n in Caption and Windowsn in Code. |
| Windows1252 | Windows-1252 | |
| Windows1253 | Windows-1253 | |
| Windows1254 | Windows-1254 | |
| Windows1255 | Windows-1255 | |
| Windows1256 | Windows-1256 | |
| Windows1257 | Windows-1257 | |
| Windows1258 | Windows-1258 | |
| Unspecified | Unspecified | Use if the character set is not known, for example for some proprietary data files. |
| Other | Other | Use if the character set is known, but not found in the list. |

Deprecated codes list

| |
|--|
| |
|--|

Usage

[DDI-L 3.3](#)

Module name: reusable

Element name: [CharacterSet](#)

[DDI-L 3.2](#)

Module name: reusable

Element name: [CharacterSet](#)

[DDI-C 2.5](#)

Element/Attribute name: [@charset](#)

Element/Attribute XPath: /codeBook/fileDscr/fileTxt/fileType/@charset

Copyright and license

Copyright © [DDI Alliance](#) 2020.



This work is licensed under a [Creative Commons Attribution 4.0 International](#).

Citation: DDI Alliance. (2020). Character Set (Version 1.0) [Controlled vocabulary]. CESSDA.

urn:ddi:int.ddi.cv:CharacterSet:1.0. Available from:

<http://vocabularies.cessda.eu:80/urn/urn:ddi:int.ddi.cv:CharacterSet:1.0.1>



DDI Alliance Controlled Vocabulary for CharacterSet

CV definition

Standard set of characters upon which many character encodings are based (Wikipedia).

Details

CV short name: CharacterSet

CV name: Character Set

CV notes:

This vocabulary lists some popular character sets but is not exhaustive. If the character set is not mentioned in the list, use 'Other' for this element and enter the specific name into the OtherValue element. You can search for the name of the character set in the list published by IANA (Internet Assigned Character Numbers). Use the "preferred MIME-name" if the IANA list specifies one, otherwise use the format in the Name field - <http://www.iana.org/assignments/character-sets>. The vocabulary was first published by the DDI Alliance. Please see: <https://ddialliance.org/controlled-vocabularies/all>.

Language: English (en)

Version: 1.0.0

Version notes:

Version changes:

Canonical URI: <urn:ddi:int.ddi.cv.CharacterSet:1.0>

Agency: [DDI Alliance](#)

Code list

| Code value | Code descriptive term | Code definition |
|------------|-----------------------|--|
| ASCII | ASCII | The official name is US-ASCII but use the format: ASCII. The ISO code for ASCII is ISO 14962. |
| ISO88591 | ISO-8859-1 | For ISO standards, use format: ISO-n-n in Caption and ISOnn in Code. ISO standards are also known as Latin, for example, ISO-8859-1 as Latin1. |
| ISO88592 | ISO-8859-2 | |
| ISO88593 | ISO-8859-3 | |
| ISO88594 | ISO-8859-4 | |
| ISO88595 | ISO-8859-5 | |
| ISO88596 | ISO-8859-6 | |
| ISO88597 | ISO-8859-7 | |

| Code value | Code descriptive term | Code definition |
|-------------|-----------------------|---|
| ISO88598 | ISO-8859-8 | |
| ISO88599 | ISO-8859-9 | |
| ISO885910 | ISO-8859-10 | |
| ISO885911 | ISO-8859-11 | |
| ISO885913 | ISO-8859-13 | |
| ISO885914 | ISO-8859-14 | |
| ISO885915 | ISO-8859-15 | |
| ISO885916 | ISO-8859-16 | |
| MacOSRoman | Mac OS Roman | |
| UTF8 | UTF-8 | For Unicode Transformation Formats, use format: UTF-n in Caption and UTFn in Code. |
| UTF16 | UTF-16 | |
| UTF32 | UTF-32 | |
| Windows1251 | Windows-1251 | For MS-Windows character sets, use format: Windows-n in Caption and Windowsn in Code. |
| Windows1252 | Windows-1252 | |
| Windows1253 | Windows-1253 | |
| Windows1254 | Windows-1254 | |
| Windows1255 | Windows-1255 | |
| Windows1256 | Windows-1256 | |
| Windows1257 | Windows-1257 | |
| Windows1258 | Windows-1258 | |
| Unspecified | Unspecified | Use if the character set is not known, for example for some proprietary data files. |
| Other | Other | Use if the character set is known, but not found in the list. |

Deprecated codes list

| |
|--|
| |
|--|

Usage

[DDI-L 3.3](#)

Module name: reusable

Element name: [CharacterSet](#)

[DDI-L 3.2](#)

Module name: reusable

Element name: [CharacterSet](#)

[DDI-C 2.5](#)

Element/Attribute name: [@charset](#)

Element/Attribute XPath: /codeBook/fileDscr/fileTxt/fileType/@charset

Copyright and license

Copyright © [DDI Alliance](#) 2020.



This work is licensed under a [Creative Commons Attribution 4.0 International](#).

Citation: DDI Alliance. (2020). Character Set (Version 1.0) [Controlled vocabulary]. CESSDA.

urn:ddi:int.ddi.cv:CharacterSet:1.0. Available from:

<http://vocabularies.CESSDA.eu:80/urn/urn:ddi:int.ddi.cv:CharacterSet:1.0>