

IDN Code Points Policy for the .SH Top Level Domain

Purpose:

This document defines the characters that are allowed in the .SH Top Level Domain. Other code points are not allowed unless specified here.

Selection criteria for valid code points:

This code-point document is a “living-document” and as such changes are possible simply by submission to the nic.SH Domain Registry at admin@nic.SH

The nic.SH registry will consider all requests for expanding the allowed Code Points within the .SH Domain Registry on a periodical basis.

The nic.SH registry reserves the right not to allow any proposed Code Point if there is a risk of IDN Homograph Attacks.

Note: The registry system checks all applications for compliance so even if our free encoder (or any other encoder) is used, the XN—encoding will be verified for compliance after the application has been received.

Useful points of reference:

Unicode Code Charts, <http://www.unicode.org/charts/>, equivalent to ISO-10646¹. The code charts are provided as a public service by Unicode, Inc.

The language codes in this document utilise ISO 639.2.

Permitted characters within the .SH domain.

Character	Unicode	Languages	Pre-normalisation Forms
0 Latin digit 0	U+ 0030	LDH	
1 Latin digit 1	U+ 0031	LDH	
2 Latin digit 2	U+ 0032	LDH	
3 Latin digit 3	U+ 0033	LDH	
4 Latin digit 4	U+ 0034	LDH	
5 Latin digit 5	U+ 0035	LDH	
6 Latin digit 6	U+ 0036	LDH	
7 Latin digit 7	U+ 0037	LDH	
8 Latin digit 8	U+ 0038	LDH	
9 Latin digit 9	U+ 0039	LDH	

¹ ISO/IEC, "Information Technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane", ISO/IEC 10646-1:2000, October 2000.

Character		Unicode	Languages	Pre-normalisation Forms
ā	Latin letter A with macron	U+ 0101	Latvian	
a	Latin letter A	U+ 0061	LDH	
á	Latin letter A with acute	U+ 00E1	Castilian (Spanish) Czech Hungarian Icelandic Lule Sami Northern Sami Portugese Slovak UK National Languages	
ă	Latin letter A with breve	U+ 0103	Romanian	
â	Latin letter A with circumflex	U+ 00E2	Albanian French Portugese Romanian UK National Languages	
ä	Latin letter A with diaeresis	U+ 00E4	Bokmål Danish Estonian Finnish German Nynorsk Slovak Swedish UK National Languages	
à	Latin letter A with grave	U+ 00E0	Bokmål Catalan French Nynorsk Portugese UK National Languages	
ą	Latin letter A with ogonek	U+ 0105	Lithuanian Polish	
å	Latin letter A with ring above	U+ 00E5	Bokmål Danish Finnish Lule Sami Northern Sami Nynorsk Southern Sami Swedish	
ã	Latin letter A with tilde	U+ 00E3	Portugese	

Character		Unicode	Languages	Pre-normalisation Forms
æ	Latin letter AE ligature	U+ 00E6	Bokmål Danish French Icelandic Lule Sami Northern Sami Nynorsk Southern Sami	
b	Latin letter B	U+ 0062	LDH	
c	Latin letter C	U+ 0063	LDH	
ć	Latin letter C with acute	U+ 0107	Croatian Polish	
č	Latin letter C with caron	U+ 010D	Croatian Czech Estonian Latvian Lithuanian Northern Sami Slovak Slovene	
ç	Latin letter C with cedilla	U+ 00E7	Albanian Bokmål Catalan French Nynorsk Portugese	
ĉ	Latin letter C with circumflex	U+ 0109	Esperanto	
ċ	Latin letter C with dot above	U+ 010B	Maltese	
d	Latin letter D	U+ 0064	LDH	
ď	Latin letter D with caron	U+ 010F	Czech Slovak	
đ	Latin letter D with stroke	U+ 0111	Croatian Northern Sami	
e	Latin letter E	U+ 0065	LDH	

Character		Unicode	Languages	Pre-normalisation Forms
é	Latin letter E with acute	U+ 00E9	Bokmål Castilian (Spanish) Catalan Czech Danish French Hungarian Icelandic Nynorsk Portugese Slovak Swedish UK National Languages	
ě	Latin letter E with caron	U+ 011B	Czech	
ê	Latin letter E with circumflex	U+ 00EA	Albanian Bokmål French Nynorsk Portugese UK National Languages	
ë	Latin letter E with diæresis	U+ 00EB	Albanian Dutch French UK National Languages	
ė	Latin letter E with dot above	U+ 0117	Lithuanian	
è	Latin letter E with grave	U+ 00E8	Bokmål Catalan French Nynorsk UK National Languages	
ē	Latin letter E with macron	U+ 0113	Latvian	
ę	Latin letter E with ogonek	U+ 0119	Lithuanian Polish	
ŋ	Latin letter Eng	U+ 014B	Lule Sami Northern Sami	
ð	Latin letter Eth	U+ 00F0	Icelandic	
f	Latin letter F	U+ 0066	LDH	
g	Latin letter G	U+ 0067	LDH	
ģ	Latin letter G with cedilla	U+ 0123	Latvian	
ĝ	Latin letter G with circumflex	U+ 011D	Esperanto	

Character		Unicode	Languages	Pre-normalisation Forms
ġ	Latin letter G with dot above	U+ 0121	Maltese	
h	Latin letter H	U+ 0068	LDH	
ĥ	Latin letter H with circumflex	U+ 0125	Esperanto	
ħ	Latin letter H with stroke	U+ 0127	Maltese	
i	Latin letter I	U+ 0069	LDH	
í	Latin letter I with acute	U+ 00ED	Castilian (Spanish) Catalan Czech Hungarian Icelandic Portugese Slovak UK National Languages	
î	Latin letter I with circumflex	U+ 00EE	Albanian French Romanian UK National Languages	
ï	Latin letter I with diaeresis	U+ 00EF	Catalan French UK National Languages	
ì	Latin letter I with grave	U+ 00EC	UK National Languages	
ī	Latin letter I with macron	U+ 012B	Latvian	
į	Latin letter I with ogonek	U+ 012F	Lithuanian	
j	Latin letter J	U+ 006A	LDH	
ĵ	Latin letter J with circumflex	U+ 0135	Esperanto	
k	Latin letter K	U+ 006B	LDH	
ķ	Latin letter K with cedilla	U+ 0137	Latvian	
l	Latin letter L	U+ 006C	LDH	
ĺ	Latin letter L with acute	U+ 013A	Slovak	
ľ	Latin letter L with caron	U+ 013E	Slovak	
ļ	Latin letter L with cedilla	U+ 013C	Latvian	
ł	Latin letter L with middle dot	U+ 0140	Catalan	
ł̣	Latin letter L with stroke	U+ 0142	Polish	

Character		Unicode	Languages	Pre-normalisation Forms
m	Latin letter M	U+ 006D	LDH	
n	Latin letter N	U+ 006E	LDH	
ń	Latin letter N with acute	U+ 0144	Lule Sami Northern Sami Polish	
ň	Latin letter N with caron	U+ 0148	Czech Slovak	
ņ	Latin letter N with cedilla	U+ 0146	Latvian	
ñ	Latin letter N with tilde	U+ 00F1	Bokmål Castilian (Spanish) Nynorsk	
o	Latin letter O	U+ 006F	LDH	
ó	Latin letter O with acute	U+ 00F3	Bokmål Castilian (Spanish) Catalan Czech Hungarian Icelandic Nynorsk Polish Portugese Slovak UK National Languages	
ô	Latin letter O with circumflex	U+ 00F4	Albanian Bokmål French Nynorsk Portugese Slovak UK National Languages	
ö	Latin letter O with diaeresis	U+ 00F6	Danish Estonian Finnish German Hungarian Icelandic Swedish UK National Languages	
ő	Latin letter O with double acute	U+ 0151	Hungarian	
ò	Latin letter O with grave	U+ 00F2	Bokmål Catalan Nynorsk UK National Languages	

Character		Unicode	Languages	Pre-normalisation Forms
ø	Latin letter O with stroke	U+ 00F8	Bokmål Danish Lule Sami Northern Sami Nynorsk Southern Sami	
õ	Latin letter O with tilde	U+ 00F5	Estonian Portugese	
œ	Latin letter OE ligature	U+ 0153	French	
p	Latin letter P	U+ 0070	LDH	
q	Latin letter Q	U+ 0071	LDH	
r	Latin letter R	U+ 0072	LDH	
ř	Latin letter R with acute	U+ 0155	Slovak	
ř	Latin letter R with caron	U+ 0159	Czech	
ŗ	Latin letter R with cedilla	U+ 0157	Latvian	
s	Latin letter S	U+ 0073	LDH	
ś	Latin letter S with acute	U+ 015B	Polish	
š	Latin letter S with caron	U+ 0161	Croatian Czech Estonian Finnish Latvian Lithuanian Northern Sami Slovak Slovene	
ș	Latin letter S with cedilla	U+ 015F	Romanian	
ŝ	Latin letter S with circumflex	U+ 015D	Esperanto	
t	Latin letter T	U+ 0074	LDH	
ť	Latin letter T with caron	U+ 0165	Czech Slovak	
ț	Latin letter T with cedilla	U+ 0163	Romanian	
ƚ	Latin letter T with stroke	U+ 0167	Northern Sami	
þ	Latin letter Thorn	U+ 00FE	Icelandic	
u	Latin letter U	U+ 0075	LDH	

Character		Unicode	Languages	Pre-normalisation Forms
ú	Latin letter U with acute	U+ 00FA	Castilian (Spanish) Catalan Czech Hungarian Icelandic Portugese Slovak UK National Languages	
ŭ	Latin letter U with breve	U+ 016D	Esperanto	
û	Latin letter U with circumflex	U+ 00FB	Albanian UK National Languages	
ü	Latin letter U with diaeresis	U+ 00FC	Bokmål Castilian (Spanish) Catalan Danish Estonian French German Hungarian Nynorsk Swedish UK National Languages	
ű	Latin letter U with double acute	U+ 0171	Hungarian	
ù	Latin letter U with grave	U+ 00F9	French UK National Languages	
ū	Latin letter U with macron	U+ 016B	Latvian Lithuanian	
ų	Latin letter U with ogonek	U+ 0173	Lithuanian	
ů	Latin letter U with ring above	U+ 016F	Czech	
v	Latin letter V	U+ 0076	LDH	
w	Latin letter W	U+ 0077	LDH	
ŵ	Latin letter W with circumflex	U+ 0175	UK National Languages	
x	Latin letter X	U+ 0078	LDH	
y	Latin letter Y	U+ 0079	LDH	
ý	Latin letter Y with acute	U+ 00FD	Czech Icelandic Slovak UK National Languages	

Character		Unicode	Languages	Pre-normalisation Forms
ÿ	Latin letter Y with circumflex	U+0177	Albanian UK National Languages	
ÿ	Latin letter Y with diæresis	U+00FF	French UK National Languages	
z	Latin letter Z	U+007A	LDH	
ż	Latin letter Z with acute	U+017A	Polish	
ž	Latin letter Z with caron	U+017E	Croatian Czech Estonian Latvian Lithuanian Northern Sami Slovak Slovene	
ż	Latin letter Z with dot above	U+017C	Maltese Polish	

Additional permitted scripts are:

CJK, as per http://en.wikipedia.org/wiki/CJK_Unified_Ideographs,
EXCLUDING ...

Ideographic Description Characters (2FF0-2FFF)

CJK Symbols and Punctuation (3000-303F)

And also:

Arabic - http://en.wikipedia.org/wiki/Arabic_script

Armenian - http://en.wikipedia.org/wiki/Armenian_alphabet

Bengali - http://en.wikipedia.org/wiki/Bengali_alphabet

Braille - <http://en.wikipedia.org/wiki/Braille>

Burmese - http://en.wikipedia.org/wiki/Burmese_script

Bopomofo - <http://en.wikipedia.org/wiki/Zhuyin>

a phonetic system for transcribing Chinese

Canadian Aboriginal -

http://en.wikipedia.org/wiki/Canadian_Aboriginal_syllabics

Cherokee - http://en.wikipedia.org/wiki/Cherokee_syllabary

Cyrillic - http://en.wikipedia.org/wiki/Cyrillic_script

Devanagari - <http://en.wikipedia.org/wiki/Devanagari>

Ge'ez - http://en.wikipedia.org/wiki/Ge%27ez_alphabet

Georgian - http://en.wikipedia.org/wiki/Georgian_alphabet

Gurmukhi - http://en.wikipedia.org/wiki/Gurmukhi_script

Greek & Coptic - http://en.wikipedia.org/wiki/Greek_alphabet

Gujarati - http://en.wikipedia.org/wiki/Gujarati_alphabet

Hebrew - http://en.wikipedia.org/wiki/Hebrew_alphabet

Kannada - http://en.wikipedia.org/wiki/Kannada_alphabet

Lao - http://en.wikipedia.org/wiki/Lao_script

Malayalam - http://en.wikipedia.org/wiki/Malayalam_script

Mongolian - http://en.wikipedia.org/wiki/Mongolian_script

Oriya - http://en.wikipedia.org/wiki/Oriya_script

Philippine scripts Buhid, Hanunó'o, Tagalog, and Tagbanwa

Tamil - http://en.wikipedia.org/wiki/Tamil_script

Telugu - http://en.wikipedia.org/wiki/Telugu_script

Thai - http://en.wikipedia.org/wiki/Thai_alphabet

Tibetan - http://en.wikipedia.org/wiki/Tibetan_script

Please note: We reserve the right not to process or delegate any domain name, you will not be charged in such circumstances.