सत्यमेव जयते
Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

Draft Policy Document

for

INTERNATIONALIZED DOMAIN NAMES

Language: BANGLA

RECORD OF CHANGES

***A - ADDED M - MODIFIED D - DELETED**

VERSION NUMBER	DATE (DD/MM/YY)	PAGES AFFECTED	A* M D	TITLE OR BRIEF DESCRIPTION	COMPLIANCE VERSION OF MAIN POLICY DOCUMENT
1.0	20/11/09	Whole Document	M	Language Specific Policy Document for BANGLA	1.5
1.1	22/11/10	Page No 11, 21	A	Restriction rule and ccTLD added	1.6
1.2	05/04/11	Whole document	A	Vowel-sequence modified.	1.7
1.3	18/07/2013	Whole document	A,M	Modified as per IDNA 2008, Restriction rules added and modified.	

Table of Contents

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

1.1 Declaration of Variables:

Dash → Hyphen -

Digit → Indo-Arabic digits [0-9]

C → Consonant

V → Vowel

M → Matra

D → Anusvara

B → Chandrabindu

X → Visarga

H → Halant/ Hasant /Virama

N → Nukta

Y → Avagraha

Z → Khanda Ta

1.2 ABNF Operators

Sr. No.	Operator	Function
1	"|"	Alternative
2	"[ ]"	Optional
3	"*"	Variable Repetition
4	"( )"	Sequence Group

In what follows, the Vowel Sequence and the Consonant Sequence pertinent to Bangla are given. To facilitate understanding, equivalents in Devanagari are provided.

1.3 The Vowel Sequence[1]

A vowel sequence is made up of a single vowel. It may be followed but not necessarily (optionally) by an Anusvara (D), Chandrabindu (B) or a Visarga (X). The number of D, B or X which can follow a V in Bangla may not be restricted to one.

The possibility of a Visarga or Anusvara following a Chandrabindu exists in Bangla. Vowel can optionally be followed by a combination of Halant/ Hasant/ Virama [H], Consonant [C] to form a Ya-phalaa. "*Ya-phalaa is a presentation form of U+09AF য bengali letter ya. Represented by the sequence < U+09CD ্ bengali sign virama, U+09AF য bengali letter ya>, ya-phalaa has a special form. When combined with U+09BE া bengali vowel sign aa, it is used for transcribing [æ] as in the "a" in the English word "bat."* [2]

The vowel sequence in Bangla is therefore.

V[D|B|X|BD|BX|HCM[D|B|X|BD|BX]]

A Vowel-sequence admits the following combinations:

1. A Single Vowel
Examples:

V	অ	অ

2. A Vowel can optionally be followed by Anusvara [D] or Chandrabindu [B] or Visarga [X] or Chandrabindu+Anusvara [BD] or Chandrabindu+Visarga [BX] or combination of Halant (also known as Hasant/Virama) [H] followed by Consonant [C] followed by Matra [M].
Examples:

VD	অং	अं

[1] Combinations where Anusvara / Visarga followed by Chandrabindu, though permitted in the language, currently are not rendered because of the underlying shaping engine and the dialog needs to be opened with OS vendors so that problem can be solved in next implementations.

2 Refer chapter 9: http://www.unicode.org/versions/Unicode5.0.0/ch09.pdf

VB	अँ	अँ
VX	अः	अः
VBD	अँं	अँं
VBX	अँः	अँः
VHCM	अ्या (अ+्+य+ा)	

2.1. A VHCM sequence can optionally be followed by Anusvara [D] or Chandrabindu [B] or Visarga [X] or Chandrabindu+Anusvara[BD] or Chandrabindu+Visarga[BX].
Examples:

VHCMD	अ्यां
VHCMB	अ्याँ
VHCMX	अ्याः
VHCMBD	अ्याँं
VHCMBX	अ्याँः

1.4 The Consonant Sequence[3]

A Consonant Sequence admits the following combinations:
1. A single consonant (C)
Example:

C	क	क

2. A Consonant optionally followed by dependent vowel sign / Matra [M] or Anusvara [D] or Chandrabindu [B] or Visarga[X] or Halant (also known as Hasant/Virama) [H] or Chandrabindu+Anusvara [BD] or Chandrabindu+Visarga [BX]

C[M|D|B|X|H|BD|BX]

Example:

[3] Combinations where Anusvara / Visarga followed by Chandrabindu, though permitted in the language, currently are not rendered because of the underlying shaping engine and the dialog needs to be opened with OS vendors so that problem can be solved in next implementations.

CM	कि	कि	
CD	कं	कं	
CB	कँ	कँ	
CX	कः	कः	
CH	क्	क्	(Pure Consonant)
CBD	कँं	कँं	
CBX	कँः	कँः	

2.a. A CM sequence can be optionally followed by D, B, X, BD or BX.

(CM)[D|B|X|BD|BX]

Example:

CMD	कीं	कीं
CMB	काँ	काँ
CMX	वीः	वीः
CMBD	काँं	काँं
CMBX	काँः	काँः

3. A sequence of consonants (up to 4) joined by Halant (also known as Hasant/Virama).

*3(CH)C

Example:

CHC	ष्ठ → ष् + ठ	ष्+ठ
CHCHC	ष्ट्र →ष् + ट् + र	ष्+ट्+र
CHCHCHC	ष्ट्र्य → ष् + ट् + र् + य	ष्+ट्+र्+य

Subsets:

While considering its subsets, as a representative example, we will consider the combination CHC only, however the same is equally applicable to CHCHC and CHCHCHC.

3.a. The combination may be followed by M, D, B, X, BD or BX.
Example:

CHCM	क्की → क ् क ी	क्की → क ् क ी
CHCD	क्कं → क ् क ं	क्कं → क ् क ं
CHCB	क्कँ → क ् क ँ	क्कँ → क ् क ँ
CHCX	क्कः → क ् क ः	क्कः → क ् क ः
CHCBD	क्कँं → क ् क ँ ं	क्कँं → क ् क ँ ं
CHCBX	क्कँः → क ् क ँ ः	क्कँः → क ् क ँ ः

3.b. *3(CH)CM may further be followed by a D, B, X, BD or BX
(CHCM)[D|B|X|BD|BX]
Example:

CHCMD	क्कीं → क ् क ी ं	क्कीं → क ् क ी ं
CHCMB	क्काँ → क ् क ा ँ	क्काँ → क ् क ा ँ
CHCMX	क्कीः → क ् क ी ः	क्कीः → क ् क ी ः
CHCMBD	क्काँं → क ् क ा ँ ं	क्काँं → क ् क ा ँ ं
CHCMBX	क्काँः → क ् क ा ँ ः	क्काँः → क ् क ा ँ ः

4. A single Khanda Ta (Z).
Example:

Z ৎ

5. A Khanda Ta can be preceded by a consonant and Halant (also known as Hasant/Virama)

[CH]Z

Example:

CHZ র ্ ৎ ৎ

The final canonical structure of the consonant sequence can thus be defined in ABNF as:

*3(C[N]H)C[N][H|D|B|X|BD|BX|M[D|B|X|BD|BX]] | [CH]Z

1.5 ABNF Applied to the Bangla IDN

The formalism can be applied to create/validate IDN labels in Bangla. So a valid IDN label in Bangla can be defined as follows.

Vowel-sequence →

V [D | B | X | BD | BX | HCM[D|B|X|BD|BX]]

Consonant-sequence →

*3(C[N]H)C[N][H|D|B|X|BD|BX|M[D|B|X|BD|BX]] | [CH]Z

Sequence → Consonant-sequence[Y] | Vowel-sequence[Y]

IDN-label → (sequence | digit) * ([dash] (sequence |digit))

Additional Examples putting more light on Bangla ABNF:

Below are some of the examples which will help a casual reader understand some of the rules ABNF puts in place. These are just given for reference purposes and are not meant to be comprehensive.

1. H, M, D, B or X cannot occur in the beginning of a Bangla IDN.
Example:

◌্ক	◌্ক
◌িক	◌িক
◌ংক	◌ंक
◌ঁক	◌ँक
◌ঃক	◌ःक

As can be seen such combination will result automatically in a “golu” marking it as an invalid formation. This is an intrinsic property of the Indian language syllable and is quasi automatically applied wherever supported by the OS.

2. H is not permitted after V, D, B, X, M, Digit or Dash.
Example:

অ্	अ्
কং◌্	कं◌्
কঁ◌্	कँ◌्
কঃ◌্	कः◌्
কি্	कि्
1◌্	1◌्
-◌্	-◌्

3. Number of D, B or X permitted after Consonant or Vowel or a Matra is restricted to one thus following combinations are invalidated.
Example:

क१०१	कं०
कँ०ँ	कँ०ँ
क१०१	कः०ः
की१०१	कीं०
काँ०ँ	काँ०ँ
की१०१	कीः०ः
अ१०१	अं०
अँ०ँ	अँ०ँ
अ१०१	अः०ः

4. Number of M permitted after Consonant is restricted to one
Example:

कीी	कीी

5. M is not permitted after V
Example:

ओा	ईा

6. The combinations of Anusvara+Visarga as well as Visarga+Anusvara are not permissible
Example:

क१०१	कं०ः
क१०१	कः०

2. RESTRICTION RULES

The Augmented Backus Naur Formalism (ABNF) is generic in nature and when applied to a specific language/script, certain restriction rules apply. In other words, in a given language some of the Formalism structures do not necessarily apply. To take care of such cases restriction rules are set in place. These restrictions will help to fine-tune the ABNF.

In case of Bangla the following rules apply:

1. Khanda ta is NOT allowed in the beginning of an IDN label.

2. CH can come with Khanda Ta in only the case where C is র (09B0).

3. Nukta shall be allowed only after following characters:

   ড (09A1)

   ঢ (09A2)

   য (09AF)

4. Only following combinations with VHCM will be allowed.

   অ্যা → অ (0985) + ্(09CD) + য(09AF) + া(09BE)

   এ্যা → এ (098F) + ্(09CD) + য(09AF) + া(09BE)

5. A consonant sequence that is intended to end with Halant [H] can only be followed by Hyphen, Digit or Avagraha [Y]. Thus following combinations are permissible.

   ক্-

   ক্1

   ক্হ

6. Consecutive Hyphens will not be permitted in a domain name.

7. The number of consecutive identical consonants joined by a Halant within a label shall not exceed two. Thus क्क (ka+halant+ka) is permitted but not क्क्क (ka+halant+ka+halant+ka).

8. A label containing not more than three "akshara", which have got variants shall be permitted. As an example let us consider a, b, c and d as four aksharas in a given label having a', b', c' and d' as variants in which case such a label will be disallowed. (Example of disallowed label - abcd, acdb, cdaba and so on)

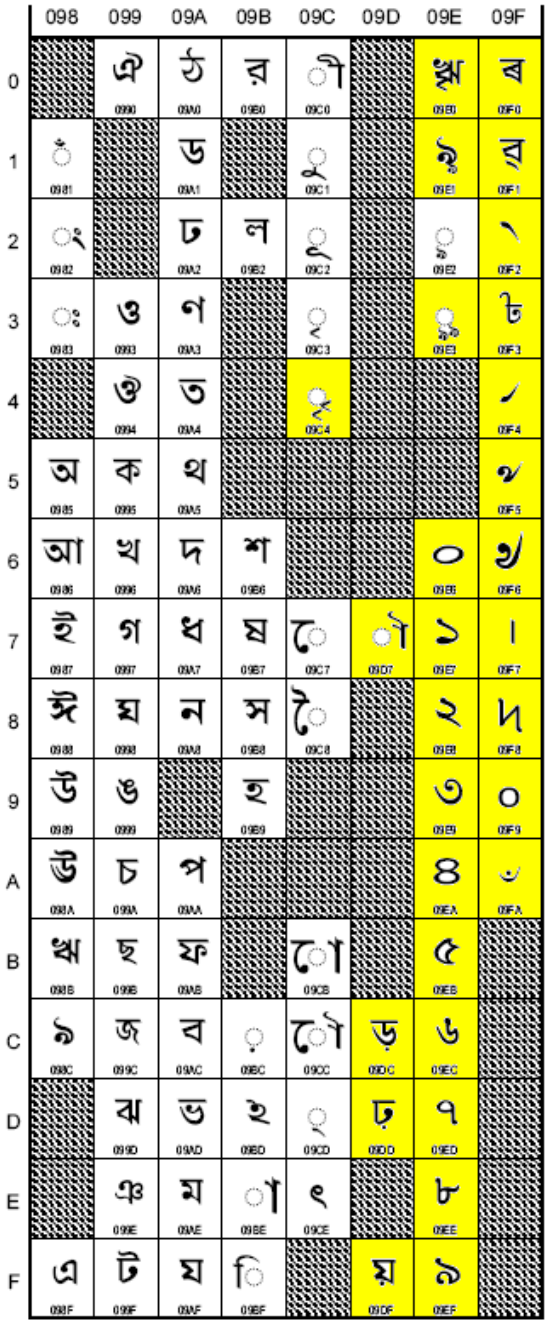
Additional Note:
Wherever a variant is present in a given label, the variants shall be strictly symmetric and non-transitive. Thus given a variant ◌ी and ◌ीं, the variants of a label such as ब्रनी shall be ब्रनीं. ब्रनीँं generated by adding an extra ◌ीं to ◌ी shall not be permitted regardless of its ABNF Validity status. This ensures that over generativity does not take place.

3. EXAMPLES

Combination	Example	Word with combination
C	ক	কল
CN	ড়	বড়
CH	হ্	আল্লাহ্
CM	চা	চাল
CD	রং	বরং
CB	গঁ	গঁদ
CX	নঃ	মনঃ
CMD	রাং	রাংতা
CMB	চাঁ	চাঁদ
CMX	দুঃ	দুঃখ
CMBD	হ্যাঁং	হ্যাঁংচা
CMBX	হ্যাঁঃ	হ্যাঁঃ
CHC	দ্ম	পদ্ম
CHCHC	ন্ত্র	মন্ত্রণা
CHCHCHC	স্ট্র্যা	স্ট্র্যাপ
V	আ	আকাশ
VD	ইং	ইংরেজি

VB	আঁ	আঁক
VX	আঃ	আঃ
VHCM	অ্যা	অ্যাকশন
Z	ৎ	হঠাৎ
CHZ	ৎ	ভর্ৎসনা

4. LANGUAGE TABLE[4]: BANGLA[5]



[4] This language table is based on Unicode Chart for Bengali script provided by the Unicode Consortium
5 Characters marked in yellow are not applicable to the language.

5. NOMENCLATURAL DESCRIPTION TABLE OF BANGLA LANGUAGE TABLE

CHANDRABINDU (B)		
0981	ঁ	BENGALI SIGN CANDRABINDU
ANUSVARA (D)		
0982	ং	BENGALI SIGN ANUSVARA
VISARGA (X)		
0983	ঃ	BENGALI SIGN VISARGA
INDEPENDENT VOWELS (V)		
0985	অ	BENGALI LETTER A
0986	আ	BENGALI LETTER AA
0987	ই	BENGALI LETTER I
0988	ঈ	BENGALI LETTER II
0989	উ	BENGALI LETTER U
098A	ঊ	BENGALI LETTER UU
098F	এ	BENGALI LETTER E
0990	ঐ	BENGALI LETTER AI
0993	ও	BENGALI LETTER O
0994	ঔ	BENGALI LETTER AU
098C	ঌ	BENGALI LETTER L
098B	ঋ	BENGALI LETTER VOCALIC R
CONSONANTS (C)		
0995	ক	BENGALI LETTER KA
0996	খ	BENGALI LETTER KHA
0997	গ	BENGALI LETTER GA
0998	ঘ	BENGALI LETTER GHA
0999	ঙ	BENGALI LETTER NGA

099A	চ	BENGALI LETTER CA
099B	ছ	BENGALI LETTER CHA
099C	জ	BENGALI LETTER JA
099D	ঝ	BENGALI LETTER JHA
099E	ঞ	BENGALI LETTER NYA
099F	ট	BENGALI LETTER TTA
09A0	ঠ	BENGALI LETTER TTHA
09A1	ড	BENGALI LETTER DDA
09A2	ঢ	BENGALI LETTER DDHA
09A3	ণ	BENGALI LETTER NNA
09A4	ত	BENGALI LETTER TA
09A5	থ	BENGALI LETTER THA
09A6	দ	BENGALI LETTER DA
09A7	ধ	BENGALI LETTER DHA
09A8	ন	BENGALI LETTER NA
09AA	প	BENGALI LETTER PA
09AB	ফ	BENGALI LETTER PHA
09AC	ব	BENGALI LETTER BA
09AD	ভ	BENGALI LETTER BHA
09AE	ম	BENGALI LETTER MA
09AF	য	BENGALI LETTER YA
09B0	র	BENGALI LETTER RA
09B2	ল	BENGALI LETTER LA
09B6	শ	BENGALI LETTER SHA
09B7	ষ	BENGALI LETTER SSA
09B8	স	BENGALI LETTER SA

09B9	ঽ	BENGALI LETTER HA
VOWEL SIGNS (MATRAS) (M)		
09BE	া	BENGALI VOWEL SIGN AA
09BF	ি	BENGALI VOWEL SIGN I
09C0	ী	BENGALI VOWEL SIGN II
09C1	ু	BENGALI VOWEL SIGN U
09C2	ূ	BENGALI VOWEL SIGN UU
09C3	ৃ	BENGALI VOWEL SIGN VOCALIC R
09C7	ে	BENGALI VOWEL SIGN E
09C8	ৈ	BENGALI VOWEL SIGN AI
09CB	ো	BENGALI VOWEL SIGN O
09CC	ৌ	BENGALI VOWEL SIGN AU
09E2	ৢ	BENGALI VOWEL SIGN VOCALIC L
VIRAMA (H)		
09CD	্	BENGALI SIGN VIRAMA
KHANDA TA (Z)		
09CE	ৎ	BENGALI LETTER KHANDA TA
AVAGRAHA (Y)		
09BD	ঽ	BENGALI SIGN AVAGRAHA
NUKTA (N)		
09BC	়	BENGALI SIGN NUKTA

6. VARIANT TABLE FOR BANGLA

VARIANTS	
ৃ 09C3	ূ 09C2
ীঁ 09C0 + 0981	ী 09C0
োঁ 09CC + 0981	ো 09CC

7. EXPERTS CONSULTED

Expertise provided by C-DAC KOLKATA.

8. PROPOSED ccTLD FOR BANGLA

India (Bhārat) localized in Bangla - ভারত

Note: You can send your feedbacks to idn-feedback@cdac.in