



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

Draft Policy Document For

INTERNATIONALIZED DOMAIN NAMES

Language : BANGLA



समन्वित विभाग
Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

RECORD OF CHANGES

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

VERSION NUMBER	DATE	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



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

Table of Contents

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF).....	4
1.1 Naming of Variables:.....	4
1.2 ABNF Operators.....	4
1.3 The Vowel Sequence.....	5
1.4 Consonant Sequence.....	6
1.5 ABNF Applied to the BANGLA IDN.....	9
2. RESTRICTION RULES.....	10
3. EXAMPLES.....	13
4. LANGUAGE TABLE : BANGLA.....	15
5. NOMENCLATURAL DESCRIPTION TABLE OF BANGLA LANGUAGE TABLE	16
6. VARIANT TABLE.....	20
7. EXPERTS CONSULTED.....	21
8. PROPOSED ccTLD FOR BANGLA.....	22

1. AUGMENTED BACKUS-NAUR FORMALISM (ABNF)

1.1 Naming 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

S. No.	Symbols	Functions
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 use , equivalents in Devanagari are provided.

1.3 The Vowel Sequence*

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 exist 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 ঙ 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.””¹

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 shapes:

1. A Single Vowel

Examples :

V	অ	আ
---	---	---

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 / Hasant / Virama [H], Consonant[C], Matra [M]

Examples :

V[D]	অং	অঁ
V[B]	অঁ	অঁ
V[X]	অঃ	অঃ

* 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 implementation.

1 Refer chapter 9: <http://www.unicode.org/versions/Unicode5.0.0/ch09.pdf>

V[BD]	अँ	अं
V[BX]	अः	अः
V[HCM]	अग (अ + ः + य + ा)	

3. 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:

VHCM[D]	अगं
VHCM[B]	अगँ
VHCM[X]	अगः
VHCM[BD]	अगँं
VHCM[BX]	अगँः

1.4 Consonant Sequence*

A consonant sequence admits the following shapes:

1. A single consonant (C)

Example:

C	क	क
---	---	---

2. A consonant optionally followed by dependent vowel sign[M] or Anusvara[D] or Chandrabindu[B] or visarga[X] or Halant / Hasant / Virama [H] or Chandrabindu+Anusvara[BD] or Chandrabindu+Visarga[BX]
C[M|D|B|X|H|BD|BX]

* 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 implementation.

Example:

C[M]	कि	कि
C[D]	कं	कं
C[B]	कँ	कँ
C[X]	कः	कः
C[H]	क्	क् (Pure Consonant)
C[BD]	कँं	कँं
C[BX]	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:

CM[D]	कीं	कीं
CM[B]	काँ	काँ
CM[X]	वीः	वीः
CM[BD]	काँं	काँं
CM[BX]	काँः	काँः

3. A sequence of consonants (up to 4) joined by Halant / Hasant / Virama
 *3(CH)C

Example:

CHCHCHC क्ष = न् + त् + र् + य न्+त्+र्+य

Subsets

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

Example:

CHC[M]	क्षी	क्की	क्की	क्की
CHC[D]	क्षं	क्कं	क्कं	क्कं
CHC[B]	क्षँ	क्कँ	क्कँ	क्कँ
CHC[X]	क्षः	क्कः	क्कः	क्कः
CHC[BD]	क्षँं	क्कँं	क्कँं	क्कँं
CHC[BX]	क्षँः	क्कँः	क्कँः	क्कँः

3.b. *3(CH)CM may be followed by a D, B, X, BD or BX

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

Example:

CHCM[D]	क्षीं	क्कीं	क्कीं	क्कीं
CHCM[B]	क्षँ	क्कँ	क्कँ	क्कँ
CHCM[X]	क्षः	क्कः	क्कः	क्कः
CHCM[BD]	क्षँं	क्कँं	क्कँं	क्कँं
CHCM[BX]	क्षँः	क्कँः	क्कँः	क्कँः

4. A single Khanda Ta(Z)

Example:

Z

৯

5. A Khanda Ta can be preceded by a consonant and Halant / Hasant / Virama

[CH]Z

Example:

CHZ

ৰ ্ ৯ ৫

The final canonical structure of the consonant sequence in IDN can 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. So a valid IDN label 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))

2. RESTRICTION RULES

The 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 the case of Bangla the following rules apply:

1. H |M |D|B|X cannot occur in the beginning of an Internationalized Domain Name

Example:

্ক	্ক
িক	িক
ংক	ংক
ঁক	ঁক
ঃক	ঃক

As can be seen they will result automatically in a “golu” marking an invalid character. This is an intrinsic property of the Indic syllable and is quasi automatically applied wherever supported by the OS.

2. H is not permitted after V, D, B, X, M, digit and dash

Example:

অ্	অ্
কং্	কং্
কঁ্	কঁ্
কঃ্	কঃ্
ক্	ক্



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

1ॢ

1ॢ

-ॢ

-ॢ

3. Number of D,B or X permitted after consonant-sequence or vowel-sequence or M is restricted to one

Example:

कॢं

कंॢ

कँॢ

कँॢ

कःॢः

कःॢः

कीॢं

कींॢ

काँॢ

काँॢ

कीःॢः

कीःॢः

अॢं

अंॢ

अँॢ

अँॢ

अःॢः

अःॢः

4. Number of M permitted after consonant-sequence is restricted to one

Example:

कीी

कीी

5. M is not permitted after V

Example:

ईा

ईा

6. The combination of Anusvara+Visarga as well as Visarga+Anusvara is not permissible

Example:

कंः कंः
कःं कःं

7. Khanda ta is NOT allowed in the beginning of a IDN label.

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

9. Nukta can be only allowed after following characters:

उ (09A1)

ऊ (09A2)

य (09AF)

10. Only following combination with VHCM will be allowed.

अा → अ(0985) + ्(09CD) + य(09AF) + ाँ(09BE)

एा → ए(098F) + ्(09CD) + य(09AF) + ाँ(09BE)

11. A consonant syllable that is intended to end with Halant [H] can only be followed by Hyphen, digit or Avagraha [Y].



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

3. EXAMPLES

Following are the some examples to show BANGLA IDN Label.

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	হাঁং	হাঁংচা
CMBX	হাঁঃ	হাঁঃ
CHC	দ্ম	পদ্ম
CHCHC	ভ্র	মন্ত্রণা



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

CHCHCHC	স্ট্র	স্ট্রাপ
V	আ	আকাশ
VD	ইং	ইংরেজি
VB	আঁ	আঁক
VX	আঃ	আঃ
VHCM	অ্যা	অ্যাকশন
Z	ৎ	হঁতাৎ
CHZ	ৎ	ভৎসনা



सूचना एवं संचार विभाग
Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

4. LANGUAGE TABLE : BANGLA²

0980

Bengali

09FF

	098	099	09A	09B	09C	09D	09E	09F
0		ঐ	ঔ	র	ী		ঋ	ঌ
1	ঐ		ড		ঊ		ঋ	ঌ
2	ং		ঢ	ল	ঋ		ঋ	ঌ
3	ং	ও	ণ		ঊ		ঋ	ঌ
4		ও	ত		ঊ		ঋ	ঌ
5	অ	ক	খ					ঋ
6	আ	খ	দ	শ			ঊ	ঌ
7	ই	গ	ধ	ষ	ে	ী	ঊ	ঌ
8	ঐ	ঘ	ন	স	ৈ		ঊ	ঌ
9	উ	ঙ		হ			ঊ	ঌ
A	উ	চ	প				ঊ	ঌ
B	ঋ	ঋ	ফ		ৌ		ঊ	ঌ
C	ঋ	জ	ব	়	ৌ	ড়	ঊ	ঌ
D		ঝ	ভ	হ	্	ঢ	ঊ	ঌ
E		ঞ	ম	া	ৎ		ঊ	ঌ
F	এ	ট	য	ি		য়	ঊ	ঌ

Bengali

² 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	ঐ	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	ল	BENGALI LETTER L
098B	ঝ	BENGALI LETTER VOCALIC R
Consonants (C)		
0995	ক	BENGALI LETTER KA
0996	খ	BENGALI LETTER KHA



समन्वित वास्तव
Department of Information Technology,
Ministry of Communications and Information Technology,
Government of India, New Delhi

0997	গ	BENGALI LETTER GA
0998	ঘ	BENGALI LETTER GHA
0999	ঙ	BENGALI LETTER NGA
099	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



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

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	হ	BENGALI LETTER HA
09DC	ড়	BENGALI LETTER RRA
09DD	ঢ	BENGALI LETTER RHA
09DF	য়	BENGALI LETTER YYA

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



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

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
KHANDATA (Z)		
09CE	৓	BENGALI LETTER KHANDA TA
AVAGRAHA (Y)		
09BD	৑	BENGALI SIGN AVAGRAHA
NUKTA (N)		
09BC	ৎ	BENGALI SIGN NUKTA



6. VARIANT TABLE

Actual Character / characters	Variant Character / characters
ड + ः 09A1+ 09BC 09DC	ड़ 09DC
ढ + ः 09A2 + 09BC	ढ़ 09DD
य + ः 09AF + 09BC	य़ 09DF
09C3	09C2
ी + ः 09C0 + 0981	ी़ 09C0
ो + ः 09CC + 0981	ो़ 09CC



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

7. EXPERTS CONSULTED

Expertise provided by C-DAC KOLKATA.



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

8. PROPOSED ccTLD FOR BANGLA

India (Bhārat) localized in Bangla - **ভারত**

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