Proposals For Kannada, Oriya and Telugu Scripts' Root Zone Label Generation Rules

Publication Date:	31 October 2018
Prepared By:	IDN Program, ICANN Org

Public Comment Proceeding	
Open Date:	8 August 2018
Close Date:	20 September 2018 Extended to 8 October 2018
Staff Report Due Date:	4 October 2018 Extended to 29 October 2018

Important Information Links

[Announcement](#)
[Public Comment Proceeding](#)
[View Comments Submitted](#)

Staff Contact:	Sarmad Hussain	**Email:**	sarmad.hussain@icann.org

Section I: General Overview and Next Steps

The Neo-Brahmi Script Generation Panel (NBGP) was formed by nine communities that use scripts derived from the Brahmi script. NBGP is developing Root Zone Label Generation Rules (LGR) for Bengali, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu scripts. The GP has published the proposals for eight LGRs from these nine scripts in three sets, releasing proposals for the scripts which share cross-script variant code points together to the extent possible. The first set has undergone public comment and this second set included the following proposals: (1) *Proposal for the Kannada Script Label Generation Rules for the Root Zone,* (2) *Proposal for the Oriya Script Label Generation Rules for the Root Zone,* and (3) *Proposal for the Telugu Script Label Generation Rules for the Root Zone.* As per the [LGR Procedure](#), these proposals were posted for public comments to allow those who have not participated in the NBGP to make their views known. Based on the feedback, the NBGP will finalize each proposal for evaluation and integration into the Label Generation Rules for the Root Zone.

Section II: Contributors

At the time this report was prepared, a total of thirteen (13) community submissions had been posted on the forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.

Organizations and Groups:

Name	Submitted by	Initials
Integration Panel	Asmus Freytag	IP
Myanmar Script Generation Panel	Thin Zar Phyo	MMGP
O Foundation	Prateek Pattanaik Nasim Ali Shitikantha Dash Shreekant Kedia Sailesh Patnaik Jnanaranjan Sahu Chinmayee Mishra Mrutyunjaya Kar Subhashish Panigrahi	OFDN

Individuals:

Name	Affiliation (if provided)	Initials
Shantinath Shirahatti		SS
Vikas Hegde		VH

Ganesh Chandra Singh		GCS
Nistha Ranjan Dash		NRD
Amresh Mishra		AM
Liang Hai		LH
Subhashish Panigrahi		SP
Santosh Kumar Mohanty	Central Institute of Indian Languages	SKM
Anindita Sahoo	Indian Institute of Technology Madras, Chennai India	AS
J. P. Das		JPD

Section III: Summary of Comments

General Disclaimer: This section intends to summarize broadly and comprehensively the comments submitted to the public comment proceeding but does not address every specific position stated by each contributor. The preparer recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

IP reviewed the Kannada and Telugu LGR proposals along with the Sinhala LGR Proposal and makes the following comments:

IP1. IP encourages NBGP to engage with the Sinhala GP to review whether the letter RA/ RAYANNA case rises to the level of a variant code point, given that the Sinhala GP considers no cross-script variant with Kannada and Telugu. All LGR proposals should be updated appropriately as per the mutual decision of the two GPs.

MMGP reviewed the the NBGP LGRs and makes the following comments:

MMGP1. MMGP notes that in the Myanmar LGR proposal, the following sets are being considered as variant code points.

- Set1: ဝ (U+101D) MYANMAR LETTER WA and ଠ (U+0B20) ORIYA LETTER TTHA.
- Set2: ေ (U+1031) MYANMAR VOWEL SIGN E and େ (U+0B47) ORIYA VOWEL SIGN E.

MMGP draws attention of NBGP to consider these code point variants in the Oriya LGR proposal.

OFDN reviewed the Oriya proposal and makes following comments:

OFDN1. Categorizing Odia scripts as a variant of Devanagari script is incorrect, as it existed 1500 years ago in Eastern India, round the time of Sanskrit and Tamil. It also undermines the struggle by Odisha state in India formed in 1936 on linguistic identify.

OFDN2. The proposal should reference primary and secondary sources, and avoid using Wikipedia as a primary source, as latter may have inconsistencies.

OFDN3. In section 3.4, OFDN notes that "ୱ" (U+0B35) is a non-existent character in Odia alphabets, based on various authoritative sources.

OFDN4. In section 3.8, OFDN disagrees with allowing Nukta to be added to "କ" (U+0B15), "ଖ" (U+0B16), "ଗ" (U+0B17), "ଚ" (U+0B1A), "ଜ" (U+0B1C), and "ଫ" (U+0B2B).

OFDN5. OFDN notes that "ଡ଼" and "ଢ଼" are supposed to be treated as characters rather than variations of "ଡ" and "ଢ" as the usage cases for the former two are more than the latter.

SS suggests moving forward to make Kannada domain names available.

VH agrees with the Kannada LGR proposal. VH suggests that exclusion of ZWNJ and ZWJ may need to be reviewed in future.

GCS agrees with the Oriyda LGR proposal and makes the comment that the usage of Nukta with various consonants addresses support of various other tribal languages using the Oriya script, such as Munda, Kui, Ho, Juanga, Oram, Kishan, Gond, Koya, etc**.**

NRD supports the inclusion of bilabial plosive Ba (ବ), labiodental Va (ୱ) and Wa (ଵ) in the Oriya LGR proposal.

AM supports the Oriya LGR proposal and affirms good coverage of different domains in Oriya language by the code points shortlisted.

LH reviewed the Oriya proposal and makes the following comments:

LH1. LH suggests that the following points or sections should be revised.

(1) The accurate transliteration of the script name should be used throughout the document.

(2) In section 3 and 3.1, the term "Oriya" is not only known in Unicode. The Oriya script is not a variant of Devanagari script and the relationship with Bangla should also be discussed.

(3) In section 3.4, the IPA table should be revised or removed.

(4) In section 3.6, the accurate transliteration should be used.

(5) In section 3.7, the Pre-base form is not relevant to the Odia discussion.

(6) In section 3.10, either phonetic transcription or transliteration should be used. It could also be disregarded to avoid confusion.

(7) In section 3.11, the example saṁkhyā shows rendering failure.

LH2. LH raises the following questions and discussion points.

(1) In section 3.8, the usage of Nukta is not only limited to the Perso-Arabic words.

(2) In section 3.12: Why is vocalic rr excluded but vocalic l and vocalic ll are included?

(3) In section 5.2 and 5.3.2, does not U+0B35 ORIYA LETTER VA fall into "4.1.2.4 No Rare and Obsolete Characters"? Why is U+0B57 ORIYA AU LENGTH MARK excluded but U+0B56 ORIYA AI LENGTH MARK is included?

LH3. LH disagrees with the no-inscript variant statement. LH suggests that Odia has a serious problem of confusable code points because of multiple ways of encoding the signs of ba and ya. The proposal should capture this case as variant code points. LH proposes the disposition to be "allocatable"

LH4. LH suggests the pattern representation of Oriya labels as:
`C[N][M][B|D|X] | V[B|D|X] | C[N]H`

SP makes following comments in response to LH comments on Oriya proposal:

SP1. SP agrees largely with LH comments and suggests referring to primary sources instead of Wikipedia.

SP2. In section 3.4, SP notes that the "ଵ"(U+0B35) is a borrowed character from Sanskrit which is not generally used.

SP3. In section 3.8, SP disagrees with adding Nukta to "କ" (U+0B15),"ଖ" (U+0B16), "ଗ" (U+0B17), "ଚ" (U+0B1A), "ଜ" (U+0B1C), and "ଫ" (U+0B2B) as there is no reference to support it. However, per LH's point, SP supports "ଡ଼" and "ଢ଼" in Oriya script. The very name of the language "ଓଡ଼ିଆ" and the geographical place — the state of Odisha (ଓଡ଼ିଶା) contains these characters. "ଡ଼" and "ଢ଼" are supposed to be treated as characters rather than variations of "ଡ" and "ଢ" as the usage case for the former two are more than the latter.

LH reviewed the Telugu proposal and makes the following comments:

LH5. LH suggests that the following points or sections should be revised.

(1) The accurate transliteration of the script name should be used throughout the document.

(2) In section 3.5, the number of dependent signs should be 15.

(3) In section 3.5.1, Vocalic I should be categorized with vocalic rr and vocalic ll. Transliteration of vocalic ll is wrong. In transcribing character sequences, distinction between phonetic sequences and written forms should be clarified, e.g. in R1 for inherent vowel. In section 3.5.3, it should be clarified if arasunna is used today or not.

(4) In section 4.2, more clarification is needed regarding if ZWNJ can be used in domain names and if not, why it is not allowed and what is the impact on domain names, as section 4.1 indicates ZWNJ is needed.

(5) In section 5, is it appropriate to exclude U+0C58 tsa and U+0C59 dza? In section 5.2, U+0C44 TELUGU VOWEL SIGN VOCALIC RR should be excluded.

(6) In section 5.3, the description on various signs needs to be clarified per the feedback. Also, clarification is needed on why the phonological variants are not permitted.

(7) In section 6, it should be noted that there are no in-script variant code points due to the WLE rules.

(8) In section 6.1, the disposition should not be indicated in the table as they are not variant code points. Vowel sign o and oo need to be discussed separately as they have a different behavior. Also, ౚ may also need to be discussed in this section, even if it is excluded.

(9) Section 6.2 provides an inappropriate restriction for reasons provided. Further, in section 6.4.1, the analysis should be revised. Confusable standalone letters do not necessarily mean confusable contextual forms. Has the contextual behavior analyzed? Table 16 is a duplication of Table 10.

(10) The conclusion of NBGP in section 6.4.2 needs to be supported by stronger arguments in addition to those provided.

LH6. LH suggests reviewing the following rules.

(1) Rule 5 is inappropriate and over-restrictive, per the feedback provided by LH on section 6.2.

(2) Rule 6 is also unnessarily restrictive, as NBGP does not have a right to force the community to abandon preferred spelling conventions.

LH7. LH suggests the pattern representation of Telugu labels as: `C[M][B|X] | V[B|X] | CH`.

LH reviewed the Kannada proposal and makes the following comments:

LH8. LH suggests that the following revisions.

(1) The transliteration of the script name as kannaḍa should be used throughout the document.

(2) In section 5.3 it should not be stated that the characters do not belong to Kannaada because of the following: U+0CBC KANNADA SIGN NUKTA is a Kannada grapheme, but not commonly used. U+0CD5 KANNADA LENGTH MARK and U+0CD6 KANNADA AI LENGTH MARK are technically used as part of a vowel sign character's canonical decompositions, even if not included by IDNA2008 due to NFC requirement.

(3) In section 6.1, it should be noted that there are no in-script variant code points due to the WLE rules.

(4) The ZWJ and ZWNJ discussion is important and should be included in the main text. The rationale for not including them is weak and should be revised.

(5) The proposal should discuss the inconsistent encoding and rendering of <ra, virama, ra> /rra/, as per the details provided by LH.

LH9. LH suggests that the limitation in section 3.4.7.4 and WLE rule 5 are too restrictive and should be reviewed.

LH10. LH suggests the pattern representation of Kannada labels as: `C[M][B|X] | V[B|X] | CH`.

LH reviewed the Oriya proposal and makes the following comments:

(1) In section 2, "oḍiā": Not an accurate transliteration as it doesn't distinguish ଓଡିଆ and ଓଡ଼ିଆ.

(2) In sections 3 and 3.1, LH suggests reviewing the description related to origins of Oriya from Devanagari and its relationship with Bangla script.

(3) LH points some inaccuracies in transcription section 3.4 and suggests Table 1 could be removed. Similarly, LH notes that transliterations in section 3.6 are not accurate and should be reviewed.

(4) In section 3.7, use of "pre-base form" is not relevant for Oriya.

(5) In section 3.8, LH notes that the Nukta use is not just limited to the use of Perso-Arabic style, e.g. on dda and ddha.

(6) In section 3.10, the transcription /ãala/ is incorrect and should be reviewed or removed. In section 3.11, the second example has a rendering issue.

(7) In section 3.12, LH inquires why is vocalic rr excluded but vocalic l and vocalic ll included? It is suggested to be consistent with the discussions in later sections.

(8) In section 5.2 and 5.2.1, LH inquires if U+0B35 ORIYA LETTER VA fall into "4.1.2.4 No Rare and Obsolete Characters"? Why is U+0B57 ORIYA AU LENGTH MARK excluded but U+0B56 ORIYA AI LENGTH MARK is included?

(9) In section 6.1, LH disagrees that Oriya does not have in-script variant code points. LH says that Oriya has a seriously problem of confusables because of multiple ways of encoding the signs of ba and ya. Per the details provided by LH, his problem should be discussed, and these be considered as cases of "allocatable" variant code points. There are also other ambiguities that need to be addressed, as shown in https://en.wikipedia.org/wiki/Odia_alphabet#Ambiguities.

(10) For other reviewers' reference Oriya text can be formed using the scheme: `C[N][M][B|D|X] | V[B|D|X] | C[N]H`

SKM thanks the team for their work and makes the following comments:

SKM1. The IPA and Nukta characters in section 3.4 and 3.8 should be reviewed.

SKM2. Some of the references from Wikipedia should be verified from authentic sources.

AS reviewed the Oriya LGR proposal and makes the following comments:

AS1. Referring to the Oriya LGR proposal, AS notes that "ଡ଼" and "ଢ଼" are distinct consonants, not the allophones of "ଡ" and "ଢ". Further, "ଡ଼" and "ଢ଼" should not be considered as the decomposed form of "ଡ଼" and "ଢ଼".

AS2. AS disagrees with allowing Nukta with the following letters: "କ", "ଖ", "ଗ", "ଚ", "ଜ", and "ଫ".

JPD suggests that letters with Nukta below letters should be rejected even for "ଡ" and "ଢ".

Section IV: Analysis of Comments

General Disclaimer: This section intends to provide an analysis and evaluation of the comments submitted along with explanations regarding the basis for any recommendations provided within the analysis.

These comments are being submitted to the Neo-Brahmi Generation Panel for their consideration and incorporation (as required) in the final version of the proposals.