

**COMMENTS**  
**ON**  
**Maximal Starting Repertoire – MSR-1**  
**Overview and Rationale<sup>1</sup>**

---

<sup>1</sup> <http://www.icann.org/en/resources/idn/variant-tlds/msr-overview-27feb14-en.pdf>

## Table of Contents

OVERVIEW.....	3
GENERAL COMMENTS: .....	3
SPECIFIC COMMENTS:.....	4
Page 7: .....	4
Page 7: .....	4
Page 8 et seq.....	4
Page 16 and 20: .....	4
Page 17: .....	4
Page 20: .....	6
Page 21: .....	6
AUTHORS.....	6

## OVERVIEW

The authors of the MSR Overview need to be complimented on having undertaken a herculean task and resolving it with great clarity. However certain niggling comments and remarks need to be pointed out.

Our comments have been divided into two parts:

- a. GENERAL COMMENTS: pertinent to the document as a whole and which suggest modifications and changes
- b. SPECIFIC COMMENTS: pertinent to specific points and issues raised within the document

## GENERAL COMMENTS:

1. Since the issue was raised during the presentation of MSR in Singapore meeting (49<sup>th</sup> ICANN Public Meeting - IDN Root Zone LGR Public Workshop - Integration & Generation Panels), and since Ethnologue data was used to determine which language is in active use; it is suggested that the concept of EGIDS as used by Ethnologue (<https://www.ethnologue.com/about/language-status>) could be incorporated within the document as a short note. This would ensure clarity.
2. After careful perusal of the sections dealing with Indian scripts (cf. msr-non-cjk-27feb14-en.pdf) it is assumed that the character set will be referred to the Neo-Brahmi Generation Panel for vetting. Two types of points have been noted:
  - a. Characters actively used have been deprecated.  
E.g. U+0901- DEVANAGARI SIGN CANDRABINDU.  
Justification for the same is provided in the section dealing with specific comments cf. infra.
  - b. Characters which have been included by virtue of their being PVALID are not necessarily part of the Neo-Brahmi family of languages and need to be excluded from the possible LGR or indicated that they should not be used.
3. Certain statements need to be corrected.

A case in point is pertinent to the use of Candrabindu. The document states as under (Point 5.6):

*Up to Unicode 6.3, these scripts using a CANDRABINDU would have to use 0901 DEVANAGARI SIGN CANDRABINDU instead. As a consequence, their addition would disunify the encoding of that character and change its use.<sup>2</sup>*

The scripts Kannada, Malayalam were not using Devanagari Sign Candrabindu in their text prior to this introduction. As per our knowledge, it was nowhere flagged as requirement either from Unicode or from the language community as well.

Also, in case of Telugu, the script already had a Candrabindu character encoded as (U+0C01 - TELUGU SIGN CANDRABINDU, also known as arasunna). The newly introduced Candrabindu in Telugu is actually a TELUGU SIGN COMBINING CANDRABINDU which is not same as the traditional Candrabindu.

And, anyway using Devanagari Candrabindu with these scripts would not have been possible since the resulting display would show that the combination is illegal and further would not be acceptable to a native user of the script/language:

Below is an example for reference:

కొ - TELUGU LETTER KA + DEVANAGARI SIGN CANDRABINDU

---

<sup>2</sup> Cf. p. 17 of the document

## SPECIFIC COMMENTS:

In each case the page and the comment are provided:

### Page 7:

*Likewise, the scripts of India, Bangladesh, and Sri Lanka follow comparable rules of consonant---vowel combination in rendering.)*

**Comment:** Not all scripts of India follow comparable rules of consonant vowel. This is applicable only to scripts based on Brahmi.

### Page 7:

*Confusability*

**Comment:** Although the section on Confusability is clear an example would help.

### Page 8 et seq.

Use of the word **Bengali** to indicate the script.

**Comment:** It has been noted that **Bengali** is used to indicate the script. However, as per the recent update on the Unicode code charts page, the script is depicted as “Bengali and Assamese” Reproducing below the snapshot for ready reference<sup>3</sup>.



Thus, we feel, the use “Bengali and Assamese” instead of Bengali would be more appropriate since this apparently is the new nomenclature adopted by Unicode 6.3. It is requested that the IP may need to address this point in appropriate text.

### Page 16 and 20:

*Sindhi ampersand and postposition*

**Comment:** It is unfortunate that Sindhi ampersand and postposition have been treated as Punctuations whereas in fact these are the characters that coincide with a given word: *and* and *postposition in*.

### Page 17:

*5.6 Code Points for Which the Encoding or Usage May be Unstable*

---

<sup>3</sup> <http://www.unicode.org/charts/>

*The Integration Panel considered allowing the Generation Panel or Panels for Indic scripts to come to a consistent treatment of CANDRABINDU characters across the scripts; unfortunately, this would not be possible with the pending additions being outside the Unicode 6.3 cutoff for the MSR. However, because all indications are that the CANDRABINDU characters are not urgently needed for modern usage, the Integration Panel removed these characters from their respective script repertoires in the MSR.*

- U+0901 DEVANAGARI SIGN
- U+0981 BENGALI SIGN CANDRABINDU
- U+0A01 GURMUKHI SIGN ADAK BINDI
- U+0A81 GUJARATI SIGN CANDRABINDU
- U+0B01 ORIYA SIGN CANDRABINDU

**Comment:**

In the case of Devanagari, Candrabindu is required extensively and hence should not be excluded from MSR. This is shown in the table below:

LANGUAGE	CANDRABINDU USED	COMMENT	ATTESTATION
Hindi	YES	Used to mark Nasality. However the use is restricted to cases where there is no Vowel Sign (Matra) above the shirorekha. Minimal pairs exist such as हंस (swan) vs. हँस (laugh). Recognized as a valid character in grammars as well as by the Central Hindi Directorate	Official document of the Central Hindi Directorate and attested in print and text-books of Hindi
Marathi	YES	Used to show the combination of the loan vowel + the nasal marker. e.g. बॅक फाँट. Rarely used with full vowels. However attestations do exist.	Attested in the Government Resolution issued by the Government of Maharashtra mandating the correct way to write Marathi.
Nepali	YES	Like in Hindi, used to mark Nasality. However unlike Hindi, the Chandrabindu is used uniformly even over vowel signs placed over the Shirorekha.	Attested by Nepal Sahitya Akademi.
Konkani	YES	Used to show the combination of the loan vowel + the nasal marker. e.g. बॅक फाँट. Rarely used with full vowels. However attestations do exist.	Attested by Konkani Akademi and seen in grammars and print.
Maithili	YES	Used to mark Nasality. However the use is restricted to cases where there is no Vowel Sign (Matra) above the shirorekha.	Attested by Maithili Academy
Dogri	NO	Not used	
Boro	YES	To mark nasality. Used in the same manner as Hindi: only when there is no vowel sign above the Shirorekha.	Attested by Boro Sahitya Akademi
Sanskrit	YES	Not used in Traditional Sanskrit. However was used in latter-day Sanskrit texts.	
Ol Chiki / Santali	YES	Corresponding to the Mu Tudag, there is a Nasalization marker. Chandrabindu seems to be used in Santali written in Devanagari script.	
Sindhi	NO	Not used	
Kashmiri	YES	YES. To mark nasality. Used in the same manner as Hindi: only when there is no vowel sign above the Shirorekha.	

In the case of Bengali and Assamese, Candrabindu is similarly used actively and hence should not be excluded from MSR.

Gurmukhi sign Adak Bindi is not a candrabindu per se. It is a combination of two nasal markers Addak (U+0A71) and Bindi (U+A02). It is used very sparingly and may be continued to be excluded from the MSR.

In the case of Gujarati Candrabindu is very sparingly used mainly to show loan words

e.g. ૐ: face

The usage of Candrabindu in this script is slowly spreading though. It may be included in the MSR, an appropriate decision may be taken by the Generation Panel.

In the case of Oriya, as in the case of Devanagari, Candrabindu is actively used and should not be excluded from the MSR. The word for mother ମା is in fact the first word figuring in Oriya primers and hence should not be deprecated.

## Page 20:

### ***5.8. IDNA 2008 Gaps and Side effects***

- *U+0950 DEVANAGARI OM*
- *U+0BDO TAMIL OM*
- *U+0ADO GUJARATI OM*

#### **Comment:**

OM should be grouped under ***5.9 Code points used for Religious Purposes***

To the list should be added:

U+0A74 : ੐ Gurmukhi Ek Onkar

## Page 21:

It is suggested that a separate section to be provided for resurgent scripts i.e. Scripts which have been revived by a Linguistic Community e.g. Ol Ciki, Meetei Mayek. Hebrew is a classic example of such resurgence and acceptance both in writing and speaking.

## **AUTHORS**

This document is an aggregation of comments made by all the authors listed in Alphabetical order:

Akshat Joshi: C-DAC. GIST

Mahesh Kulkarni: ASSOCIATE DIRECTOR & HEAD C-DAC. GIST

Neha Gupta: C-DAC. GIST

Nishit Jain: C-DAC. GIST

Raimond Doctor: CONSULTANT, C-DAC. GIST