Proposal for an Oriya Script Root Zone Label Generation Ruleset (LGR)

LGR Version:	3.0
Date:	2018-05-04
Document version:	2.0
Authors:	Neo-Brahmi Generation Panel [NBGP]

1   General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Root Zone Level Generation Rules for the Oriya script. The XML format is under construction. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

"Proposed-LGR-Orya-20180504.xml" (under progress)

Labels for testing can be found in the accompanying text document:

Labels-orya-version-20180504.txt" (Not included yet)

2   Script for which the LGR is proposed

ISO 15924 Code: Orya

ISO 15924 Key N°: 327

ISO 15924 English Name: Oriya (Odia)

Latin transliteration of native script name: oḍiā

Native name of the script: ଓଡ଼ିଆ

Maximal Starting Repertoire (MSR) version: MSR-2

3   Background on Script and Principal Languages using it

Odia (known in Unicode as Oriya) is an Eastern Indic language spoken by about 40 million people mainly in the Indian state of Odisha (Orissa), and also in parts of West Bengal, Jharkhand, Chhattisgarh and Andhra Pradesh. Oriya (Odia) is one of the many official languages of India. It is the official language of Odisha, and the second official language of Jharkhand. Eminent Linguists like John Beames, G. A. Grierson, L.S.S. O'Malley, Suniti Kumar Chatterjee, S. N. Rajaguru, John Boulton and others consider Odia as one of the most ancient languages of India. In Indic family of languages, Oriya (Odia) is closest to Sanskrit and least influenced by foreign languages. Only these two Indic languages (viz. Sanskrit and Odia) have got classical tag due to their rich, uninfluenced and long literary history. According to National Mission for Manuscripts, after Sanskrit (11,66,743), Odia (2,13,088) has the largest number of documented manuscripts in the India.

Odia waspreviouslyspelt as Oriya, and Odisha as Orissa,. However,Odia and Odisha are now the preferred names officially in English as they are closer to their native names: ଓଡ଼ିଆ (oḍiā) [ɔɖiaː] and ଓଡ଼ିଶା (oḍiśā) [ɔɖisaː].

With reference to [Wikipedia](https://en.wikipedia.org/wiki/Odia_language) Oriya (Odia) language is also used by minority populations of the neighboring states of Jharkhand, West Bengal, Chhattisgarh and Andhra Pradesh. The region has been known at different stages of history as Kalinga, Udra, Utkala or Koshala. Odisha was a vast empire in ancient and medieval times, extending from the Ganges in the north to the Godavari in the south. During British rule, however, Odisha lost its political identity and formed parts of the Bengal and Madras Presidencies. The present state of Odisha was formed in 1936. The modern Oriya (Odia) language is formed mostly from Pali words with significant Sanskrit influence. About 28% of modern Oriya (Odia) words have Adivasi origins, and about 2% have Hindustani (Hindi/Urdu), Persian, or Arabic origins. The earliest written texts in the language are about thousand years old.The first Oriya (Odia) newspaper was Utkala Deepika first published on 4 August 1866.

Oriya (Odia) is the only Indo-European language of India other than Sanskrit and the sixth Indian language that has been conferred classical language status and forms the basis of Odissidance and Odissi music.[1]

The region has been known at different stages of history as Kalinga, Udra, Utkala or Koshala. The earliest written texts in the language are about thousand years old. Oriya scriptseems to be a variant of Devanāgarī, the main difference being the absence of the shirorekha or the line

---

[1]The variety of Oriya dialets etc. is reviewed in Appendix B

above the character and also its more rounded shapes. Since initially it was used for commercial ends, it has been referred to as śarāphi (banker's) or mahājani (trader's) script.

Used to write: Oriya (Odia), and also used a number of the other languages spoken in Odisha such as Munda, Santala, Kui, Ho etc, and Sanskrit

3.1   The Evolution of the Script

The Oriya (Odia) script developed from the Kalinga script, one of the many descendents of the Brahmi script of ancient India. (Rajaguru, S.N., Odia Lipira Kramabikash, Odia Sahitya Akademi, page 2). The earliest known inscription in the Oriya (Odia) language, in the Kalinga script, dates from 1051. It descends from a Odra-Magadhi Prakrit similar to ArdhaMagadhi, prevalent in eastern India over 1,500 years ago.

The curved appearance of the Oriya script is a result of the practice of writing on palm leaves, which have a tendency to tear if you use too many straight lines.

The diagram belowshows the major stages in the evolution of Oriya attesting its late divergence from Devanāgarī.
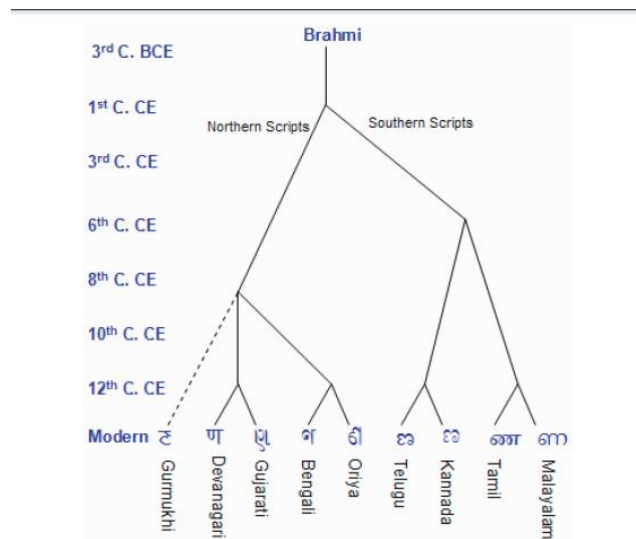

Figure 1: Pictorial depiction of Evolution of Oriya

Oriya (Odia) language literature (Odia: ଓଡ଼ିଆସାହିତ୍ୟ) is the predominant literature of the state of Odisha in India. The actual Oriya script closely resembled Bengali and Assamese scripts, but the one adopted for the printed typesets was completely different, leaning more towards the Tamil script.

3.2   Periods of Odia History

Oriya (Odia) language literature (Odia: ଓଡ଼ିଆସାହିତ୍ୟ)is the predominant literature of the state of Odisha in India.

Historians have divided the history of the Oriya (Odia) language literature into five main stages: Old Oriya (Odia) (8th century to 1300), Early Middle Oriya (Odia) (1300 to 1500), Middle Oriya (Odia) (1500 to 1700), Late Middle Oriya (Odia) (1700 to 1850) and Modern Oriya (Odia) (1850 to present). Further subdivisions, as seen below, can more accurately chart the language's development.

3.2.1   Middle Oriya (Odia) Literature (1500 to 1700)

The next era is more commonly called the Jagannatha Dasa Period and stretches till the year 1700. The period began with the writings of Shri Chaitanya, whose Vaishnava influence brought in a new evolution in Oriya literature. Balarama Dasa, Jagannatha Dasa, Yasovanta, Ananta and Acyutananda were the main exponents in religious works in Oriya. The composers of this period mainly translated, adapted or imitated Sanskrit literature. A few prominent works of this period include the Usabhilasa of SisuSankaraDasa, the Rahasya-manjari of Deva-durlabhaDasa and the Rukmini-bibha of KartikkaDasa. A new form of novels in verse also evolved during the beginning of the 17th century, when Ramachandra Pattanayaka wrote Haravali. Other poets, like Madhusudana, Bhima, Dhivara, Sadasiva and SisuIsvara-dasa, composed another form of poems called Kavyas (long poems based on themes from Puranas). The language used by them was plain and simple Oriya.

3.2.2   Late Middle Oriya Literature (1700 to 1850)

From the turn of the 18th century, verbally tricky Oriya became the order of the day. Verbal jugglery, obscenity and eroticism became the trend of the period, the most notable poet being Upendra Bhanja. Many other poets tried to imitate him, but none could fit into his shoes, with the exception of Bhima-Bhoi and ArakshitaDasa. Family chronicles in prose and literature related to religious festivals and rituals also covered a large portion of this period. The first Oriya printing typeset was cast in 1836, by the Christian missionaries.

3.2.3   Modern Oriya Literature (1850 till present day)

In this period, three great poets and prose writers, Rai Bahadur Radhanatha Ray, Madhusudana Rao and Phakiramohana Senapati, settled in Orissa and made Oriya their own. They brought in a modern outlook and spirit into Oriya literature. Around the same time, the modern drama took birth in the works of Rama Sankara Ray, beginning with Kanci-Kaveri. 20th century writers

in Oriya include Nanda-kisora Bal, Gangadhara Mehera, Chintamani Mahanti and Kuntala-KumariSabatUtkala-bharati, besides NiladriDasa and GopabandhuDasa. The most notable novelists till date are UmesaSarakara, DivyasimhaPanigrahi, Gopala Praharaja and KalindiCharanaPanigrahi. Sachi KantaRauta Ray is the great introducer of the ultra-modern style in modern Oriya poetry. Others who took up this form were Godavarisa Mahapatra, DrMayadharaManasimha, Nityananda Mahapatra and KunjabihariDasa. Prabhasa Chandra Satpati is known for his translations of some western classics, apart from UdayanathaShadangi, Sunanda Kara and Surendranatha Dwivedi. Criticism, Essays and History Criticism, essays and history also became major lines of writing in the Oriya language. The renowned writers in this field were Professor Girija Shankar Ray, Pandit Vinayaka Misra, Professor Gauri Kumara Brahma, JagabandhuSimha and Hare Krushna Mahatab. Oriya literature mirrors the industrious, peaceful and artistic image of the Oriya people, who have offered and gifted much to the Indian civilization, in the field of art and literature.

According to [Wikipedia](https://en.wikipedia.org/wiki/Odia_language), https://en.wikipedia.org/wiki/Odia_language the Oriya (Odia) diaspora constitutes a sizeable number of speakers in several countries around the world, pushing the number of Oriya (Odia) speakers globally to 55 million.

Oriya (Odia) language literature (Odia: ଓଡ଼ିଆସାହିତ୍ୟ) is the predominant literature of the state of Odisha in India.

3.3   It has a significant presence in eastern countries such as Bangladesh, Indonesia, mainly carried by the sadhaba, ancient traders from Odisha who carried the language along with the culture during the old-day trading, and in western countries such as the United States, Canada, Australia and England as well. The language has also spread to Burma, Malaysia, Fiji, Sri Lanka and countries of the Middle East.Written Oriya (Odia)

Written Oriya (Odia) (or the standard Oriya (Odia) is used for official purpose. It has elements from different local Oriya (Odia) dialects but it usually avoids words of foreign Origin such as Arabic and Persian. It has also assimilated many tribal words prevalent in Odisha.

3.4   Notable features

Type of writing system: syllabic alphabet in which all consonants have an inherent vowel. Diacritics, which can appear above, below, before or after the consonant they belong to, are used to change the inherent vowel.

When they appear at the beginning of a syllable, vowels are written as independent letters.

When certain consonants occur together, special conjunct shapes are used which combine the essential parts of each letter.

Direction of writing: left to right in horizontal lines.

Used to write: Oriya (Odia), and also used a number of the other languages spoken in Odisha, and Sanskrit

The charts shows the way in which the International Phonetic Alphabet (IPA) represents Oriya (Odia) pronunciations in [Wikipedia articles](#) referred from https://en.wikipedia.org/wiki/Help:IPA/Odia

See Oriya (Odia) phonology for a more thorough discussion of the sounds of Oriya (Odia).

Consonants		
IPA	Oriya (Odia)	English approximation
b	ବ	**b**ig
b^h	ଭ	a**bh**or
d̪[1]	ଦ	**th**at
$d̪^h$[1]	ଧ	
ɖ	ଡ	**d**ice
$ɖ^h$	ଢ	guar**dh**ouse
d͡ʒ	ଜ	**j**unior
$d͡ʒ^h$	ଝ	fri**dge h**andle

g	ଗ	go
gʰ	ଘ	pig**h**eaded
h	ହ	**h**ead
k	କ	s**k**ate
kʰ	ଖ	**K**ate, Stoc**kh**olm
l	ଲ	light (clear L)
l	ଳ	similar to "ld" in old, but with the tongue touching the back of the front teeth
ɭ	ଳ୍ହ	
m	ମ	**m**ight
n	ନ	**n**ot
ɳ	ଣ	American mo**rn**ing
ŋ	ଙ	si**ng**
p	ପ	s**p**ot
pʰ	ଫ	**p**ot
ɾ	ର	Scottish th**r**ee, American la**dd**er
ɽ	ଡ଼	
ɽʰ	ଢ଼	
s	ସ	**s**o
ʈ[1]	ଟ	
ʈʰ[1]	ଠ	**th**ought

t	ଟ	s**t**op
tʰ	O	**t**op
t͡ʃ	ଚ	la**tch**
t͡ʃʰ	ଛ	**ch**oose, whi**ch h**and

Vowels

IPA	Oriya (Odia)	English approximation
aː	◌ା	w**o**man
e	ଏ◌	**a**bility
i	◌ି, ◌ୀ	Engl**i**sh
o	ଏ◌ା	sh**o**rt (UK)
ɔi	ୱ	p**oi**se (UK)
u	◌ୁ ◌ୁ	g**u**ru
u	◌ୁ ◌ୁ	g**u**ru
~	◌ଁ	roughly like **on** (American English), nasalized [ɒ] or [ɑ]. r**en**dez-vous

Semivowels

IPA	Oriya (Odia)	English approximation
j	ୟ	**y**et

w	ଓ	**w**ine

Reduced vowels		
IPA	Oriya (Odia)	English approximation
ɔ	ଅ	d**o**ctor

Table 1:International Phonetic Alphabet Oriya Pronunciations

3.5   Structured consonants

The structured consonants are classified according to where the tongue touches the palate of the mouth and are classified accordingly into five structured groups. These consonants are shown here with their IAST transcriptions.

	voiceless	voiceless aspirate	voiced	voiced aspirate	nasal
Velars	କ (ka)	ଖ (kha)	ଗ (ga)	ଘ (gha)	ଙ (ṅa)
Palatals	ଚ (ca)	ଛ (cha)	ଜ (ja)	ଝ (jha)	ଞ (ña)
Retroflex	ଟ (ṭa)	ଠ (ṭha)	ଡ (ḍa)	ଢ (ḍha)	ଣ (ṇa)
Dentals	ତ (ta)	ଥ (tha)	ଦ (da)	ଧ (dha)	ନ (na)
Labials	ପ (pa)	ଫ (pha)	ବ (ba)	ଭ (bha)	ମ (ma)

Table 2:structured consonants

3.6   Unstructured consonants

The unstructured consonants are consonants that do not fall into any of the above structures:

ଯ (ja), ୟ (ia), ର (ra), ଲ (la), ଳ (ḷa), ଵ (wa), ଶ (sa), ଷ (sa), ସ (sa), ହ (ha)

3.7   The Implicit Vowel Killer Halant

The character used after a consonant to "strip" it of its inherent vowel. A Halant follows all but the last consonant in every Oriya syllable.

NOTE: A syllable containing halant characters may be shaped with no visible halant signs by using different consonant forms or conjuncts instead.

Halant form of consonants - The form produced by adding the halant (virama) to the nominal shape. The Halant form is used in syllables that have no vowel or as the half form when no distinct shape for the half form exists.

Half form of consonants (pre-base form) - A variant form of consonants which appear to the left of the base consonant, if they do not participate in a ligature. Consonants in their half form precede the ones forming the base glyph. Some Indic scripts, like Devanagari have distinctly shaped half forms for most of the consonants. If nodistinct shape exists, the full form will display with an explicit Virama (same shape as the halant form).

3.8   Matra sign: (Dependent Vowel)

Used to represent a vowel sound that is not inherent to the consonant. Dependent vowels are referred to as "matras" in Sanskrit. They are always depicted in combination with a single consonant, or with a consonant cluster. The greatest variation among different Indian scripts is found in the rules for attaching dependent vowels to base characters. The rules specific to oriya is mentioned in section 6 (variant) and section 7 (WLE Rules).

Following table explains the correlation between Vowel and it's sign (matra).

Correlation between Vowel and its sign (matra)					
Character	Unicode	Name	Character	Unicode	Name
ଅ	U+0B05	ORIYA LETTER A			
ଆ	U+0B06	ORIYA VOWEL LETTER AA	○ା	U+0B3E	ORIYA VOWEL SIGN AA
ଇ	U+0B07	ORIYA VOWEL LETTER I	○ି	U+0B3F	ORIYA VOWEL SIGN I
ଈ	U+0B08	ORIYA VOWEL LETTER II	○ୀ	U+0B40	ORIYA VOWEL SIGN II

ଉ	U+0B09	ORIYA VOWEL LETTER U	◌ୁ	U+0B41	ORIYA VOWEL SIGN U
ଊ	U+0B0A	ORIYA VOWEL LETTER UU	◌ୂ	U+0B42	ORIYA VOWEL SIGN UU
ଋ	U+0B0B	ORIYA VOWEL LETTER VOCALIC R	◌ୃ	U+0B43	ORIYA VOWEL SIGN VOCALIC R
ଏ	U+0B0F	ORIYA VOWEL LETTER E	�ୋ	U+0B47	ORIYA VOWEL SIGN E
ଐ	U+0B10	ORIYA VOWEL LETTER AI	ୈ◌	U+0B48	ORIYA VOWEL SIGN AI
ଓ	U+0B13	ORIYA VOWEL LETTER O	ୋ◌	U+0B4B	ORIYA VOWEL SIGN O
ଔ	U+0B14	ORIYA VOWEL LETTER AU	ୌ◌	U+0B4C	ORIYA VOWEL SIGN AU
ଲ	U+0B32	ORIYA VOWEL LETTER LA	◌ୢ	U+0B62	ORIYA VOWEL SIGN VOCALIC L
ଳ	U+0B33	ORIYA VOWEL LETTER LLA	◌ୣ	U+0B63	ORIYA VOWEL SIGN VOCALIC LL

Table 3: Correlation Between Vowel and It's Sign

This table (Table 3 above) explains Correlation between Vowel and its sign (matra) only. That's why, Candrabindu, Visarga and Anusvara is not part of this table as they do not represent the sign of vowels.

[MISSING: Discussion of Candrabindu, Visarga and Anusvara]

4   Overall Development Process and Methodology

Under the Neo-Brahmi Generation Panel, there are many different scripts belonging to separate Unicode blocks. Each of these scripts will be assigned a separate LGR; however Neo-Brahmi GP will ensure that the fundamental philosophy behind building those LGRs are all in sync with all other Brahmi derived scripts. This is the Oriya (Odia) LGR, which caters to Oriya (Odia) languages written using Oriya (Odia) belonging to EGIDS scale 1 to 4.

4.1   Guiding Principles

The NBGP adopts following broad principles for selection of code-points in the code-point repertoire across the board for all the scripts within its ambit.

4.1.1   Inclusion principles:

4.1.1.1   *Modern usage:*
Every character proposed should be in the everyday usage of a particular linguistic community. The characters which have been encoded in the Unicode for transcription purposes only or for archival purposes will not be considered for inclusion in the code-point repertoire.

4.1.1.2.Unambiguous use:

Every character proposed should have unambiguous understanding among the linguistic about its usage in the language.

4.1.2   Exclusion principles:

The main exclusion principle is that of External Limits on Scope. These comprise of protocols or standards which are pre-requisites to the Label Generation Rules etc. All further principles are in fact subsumed under these limitations but have been spelt out separately for the sake of clarity.

4.1.2.1   *External Limits on Scope:*
The code point repertoire for root zone being a very special case, up the ladder in the protocol hierarchies, the canvas of available characters for selection as a part of the Root Zone code point repertoire is already constrained by various protocol layers beneath it. Following three main protocols/standards act as successive filters:

i. The Unicode Chart:

Out of all the characters that are needed by the given script, if the character in question is not encoded in Unicode, it cannot be incorporated in the code point repertoire. Such cases are quite rare, given the elaborate and exhaustive character inclusion efforts made by Unicode consortium.

ii. IDNA Protocol:

Unicode being the character encoding standard for providing the maximum possible representation of a given script/language, it has encoded as far as possible all the possible characters needed by the script. However, the Domain name being a specialized case, it is governed by an additional protocol known as IDNA (Internationalized Domain Names in Applications). The IDNA protocol excludes some characters out of Unicode repertoire from being part of the domain names.

Example: Oriya script frequently uses "ଡ" (0B21), "ଢ" (0B22) as well as their respective allophones "ଡ଼", and "ଢ଼". In Oriya (Odia) script, these differ in use of nukta. Thus "ଡ଼" and "ଢ଼" as distinct letters are not allowed but their decomposed form i.e. "ଡ", "ଢ" followed by Oriya (Odia) (Oriya) sign nukta (0B3C) can be used. Similarly for allophones of other consonants like "ଚ" (0B1A), କ (0B15), ଖ (0B16), ଗ (0B17), ଚ (0B1A), ଜ (0B1C), ଫ (0B2B) nukta can be used.

IDNA Protocol also excludes invisible characters Zero Width Non-Joiner (U+200C) and Zero Width Joiner (U+200D), as they require a CONTEXTJ rule. These are required in certain cases where a typical visual shape of an akshar is desired.

iii. Maximal Starting Repertoire:

The Root-zone LGR being a repertoire of the characters which are going to be used for creation of the root zone TLDs, which in turn are an even more specialized case of domain names, the ROOT LGR procedure introduces additional exclusions on IDNA allowed set of characters.

Example: Oriya Sign Avagraha "ଽ" (U+0B3D) even if allowed by IDNA protocol, is not permitted in the Root Zone Repertoire as per the [MSR].

To sum up, the restrictions start off with admitting only such characters as are part of the code-block of the given script/language. This is further narrowed down by the IDNA Protocol and finally an additional filter in the form of Maximal Starting Repertoire restricts the character set associated with the given language even more.

4.1.2.2   No Fraction Marks:

The TLDs being identifiers, fraction markers present in Brahmi based languages such as given below will not be included.

Fraction signs

0B72	꠱	ORIYA FRACTION ONE QUARTER
0B73	꠲	ORIYA FRACTION ONE HALF
0B74	꠳	ORIYA FRACTION THREE QUARTERS
0B75	⁄	ORIYA FRACTION ONE SIXTEENTH
0B76	꠴	ORIYA FRACTION ONE EIGHTH
0B77	꠵	ORIYA FRACTION THREE SIXTEENTHS

Figure 2: Fraction Marks in Oriya

4.1.2.3   No Symbols and Abbreviations:

Abbreviations, weights and measures and other such iconic characters like Isshar"∀" (U+0B70) will not be included.

4.1.2.4   No Rare and Obsolete Characters:

There are characters which have been added to Unicode to accommodate rare forms especially like Oriya LETTER VOCALIC RR"ୠ" (U+0B60) and Oriya LETTER VOCALIC LL"ୡ" (U+0B61)as well as their Matra forms "ୢ" (U+0B62) and "ୣ" (U+0B63). All such characters will not be included. This is in compliance with the Conservatism principle as laid down in the Root Zone LGR procedure.

5   Repertoire

This section explains about the section of [MSR] applicable to the Oriya script on which Oriya code point repertoire is based on.

Section 5.1 details the code-point repertoire that the Neo-Brahmi Generation Panel [NBGP] proposes to be included in the Oriya LGR.

5.1   5.1 Oriya section of Maximal Starting Repertoire [MSR] Version 2

Color convention:

All characters that are included in the [MSR]- Yellow background

PVALID in IDNA2008 but excluded from the [MSR]- Pinkish background

Not PVALID in IDNA2008, or are ineligible for the root zone (digits, hyphen) - White background

Figure 3: Oriya Code Page From MSR 3

5.2   Code Point Repertoire:

For each of the code points, language references have been given in the last column titled "Reference". For the entire coverage of Oriya (Odia) code points, references of the same have been given. The examples have been chosen for referencing, they together cover all the code-points required for Oriya (Odia) Language that NBGP has considered.

Sr.No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	Ref.
1	0B01	ଁ	ORIYA SIGN CANDRABINDU	2-Oriya	Candrabindu	[1], [101] , [102] , [103], [104], [105]
2	0B02	ଂ	ORIYA SIGN ANUSVARA	2-Oriya	Anusvara	[1], [101] , [102] , [103], [104], [105]
3	0B03	ଃ	ORIYA SIGN VISARGA	2-Oriya	Visarga	[1], [101] , [102] , [103], [104], [105]
4	0B05	ଅ	ORIYA LETTER A	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
5	0B06	ଆ	ORIYA LETTER AA	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
6	0B07	ଇ	ORIYA LETTER I	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
7	0B08	ଈ	ORIYA LETTER II	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
8	0B09	ଉ	ORIYA LETTER U	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]

9	0B0A	ଊ	ORIYA LETTER UU	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
10	0B0B	ଋ	ORIYA LETTER VOCALIC R	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
11	0B0C	ଌ	ORIYA LETTER VOCALIC L	2-Oriya	Consonant	[1], [101] , [102] , [103], [104], [105]
12	0B0F	ଏ	ORIYA LETTER E	2-Oriya	Vowel	[1], [101] , [102] , [103], [104], [105]
13	0B10	ଐ	ORIYA LETTER AI	2-Oriya	Vowel	[1],[101] ,[102] , [103], [104], [105]
14	0B13	ଓ	ORIYA LETTER O	2-Oriya	Vowel	[1],[101] ,[102] , [103], [104], [105]
15	0B14	ଔ	ORIYA LETTER AU	2-Oriya	Vowel	[1],[101] ,[102] , [103], [104], [105]
16	0B15	କ	ORIYA LETTER KA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
17	0B16	ଖ	ORIYA LETTER KHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
18	0B17	ଗ	ORIYA LETTER GA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
19	0B18	ଘ	ORIYA LETTER GHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]

20	0B19	ଙ	ORIYA LETTER NGA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
21	0B1A	ଚ	ORIYA LETTER CA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
22	0B1B	ଛ	ORIYA LETTER CHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
23	0B1C	ଜ	ORIYA LETTER JA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
24	0B1D	ଝ	ORIYA LETTER JHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
25	0B1E	ଞ	ORIYA LETTER NYA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
26	0B1F	ଟ	ORIYA LETTER TTA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
27	0B20	ଠ	ORIYA LETTER TTHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
28	0B21	ଡ	ORIYA LETTER DDA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
29	0B22	ଢ	ORIYA LETTER DDHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
30	0B23	ଣ	ORIYA LETTER NNA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
31	0B24	ତ	ORIYA	2-Oriya	Consonant	[1],[101] ,[102] ,

			LETTER TA			[103], [104], [105]
32	0B25	ଥ	ORIYA LETTER THA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
33	0B26	ଦ	ORIYA LETTER DA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
34	0B27	ଧ	ORIYA LETTER DHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
35	0B28	ନ	ORIYA LETTER NA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
36	0B2A	ପ	ORIYA LETTER PA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
37	0B2B	ଫ	ORIYA LETTER PHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
38	0B2C	ବ	ORIYA LETTER BA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
39	0B2D	ଭ	ORIYA LETTER BHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
40	0B2E	ମ	ORIYA LETTER MA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
41	0B2F	ଯ	ORIYA LETTER YA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
42	0B30	ର	ORIYA LETTER RA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]

43	0B32	ଲ	ORIYA LETTER LA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
44	0B33	ଳ	ORIYA LETTER LLA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
45	0B35	ଵ	ORIYA LETTER VA	2-Oriya	Consonant	[4],[101] ,[102] , [103], [104], [105]
46	0B36	ଶ	ORIYA LETTER SHA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
47	0B37	ଷ	ORIYA LETTER SSA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
48	0B38	ସ	ORIYA LETTER SA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
49	0B39	ହ	ORIYA LETTER HA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
50	0B3C	଼	ORIYA SIGN NUKTA	2-Oriya	Nukta	[1],[101] ,[102] , [103], [104], [105]
51	0B3E	ା	ORIYA VOWEL SIGN AA	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
52	0B3F	ି	ORIYA VOWEL SIGN I	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
53	0B40	ୀ	ORIYA VOWEL SIGN II	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]

54	0B41	ଁ	ORIYA VOWEL SIGN U	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
55	0B42	ଁ	ORIYA VOWEL SIGN UU	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
56	0B43	ଁ	ORIYA VOWEL SIGN VOCALIC R	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
57	0B44	ଁ	ORIYA VOWEL SIGN VOCALIC RR	2-Oriya	Matra	[5],[101] ,[102] , [103], [104], [105]
58	0B47	ଋ	ORIYA VOWEL SIGN E	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
59	0B48	ଋ	ORIYA VOWEL SIGN AI	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
60	0B4B	ଋ	ORIYA VOWEL SIGN O	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
61	0B4C	ଋ	ORIYA VOWEL SIGN AU	2-Oriya	Matra	[1],[101] ,[102] , [103], [104], [105]
62	0B4D	ଁ	ORIYA SIGN VIRAMA	2-Oriya	Halant / VIRAMA	[1],[101] ,[102] , [103], [104], [105]

63	0B56	◌ୖ	ORIYA AI LENGTH MARK	2-Oriya	Matra	[2],[101] ,[102] , [103], [104], [105]
64	0B5F	ୟ	ORIYA LETTER YYA	2-Oriya	Consonant	[1],[101] ,[102] , [103], [104], [105]
65	0B71	ୱ	ORIYA LETTER WA	2-Oriya	Consonant	[4],[101] ,[102] , [103], [104], [105]

Table 4: Code Point Repertoire

5.2.1 Codepoint Excluded Variables involved

Sr.No.	Unicode Code Point	Glyph	Character Name	Language with EGIDS	Indic Syllabic Category	Ref.
1	0B57	◌ୗ	ORIYA AU LENTH MARK	2-Oriya	Matra	[1],[101], [102] [103] [104] [105]

Table 5: Code Point Excluded from Repertoire

Since the matra 0B57 (◌ୗ) ORIYA AU LENTH MARK is not in use at present by Oriya Community it is decided by the NBGP to exclude it.

5.2.2 Variables involved

C   →   Consonant

M   →   Matra

V   →   Vowel

B   →   Anusvara

H   →   Halant / Virama

N   →   Nukta

C1   →   {ଡ 0B21, ଢ 0B22, କ 0B15, ଖ 0B16, ଗ 0B17, ଚ 0B1A, ଜ 0B1C, ପ 0B2B}

X   →   Visarga

D   →   Candrabindu

6   Variants

6.1   In-Script Variants

In Oriya script, there are no characters/character sequences which can be created by using the Oriya characters permitted as per the [MSR] and looks identical. There is no In-Script Variants.

6.2   Cross-Script Variants

A cross-script variant, also sometimes referred to as "Whole Label confusable", is the variant case where one label in one script can be composed in such a way that it can resemble another entire label in a different script.

Every individual LGR under NBGP is supposed to provide a set of cross script variants it identifies with all other scripts under NBGP.

NBGP has ensured that not only the individual characters but also most of the akshar variations are taken into consideration during the Cross-script variant analysis of Oriya with all the other scripts under NBGP. It was achieved by sharing a list of most (a word 'most' is used here as all the possible Consonant + Halant + Consonant+…. cases cannot be practically covered. Case of all the Oriya "Consonant + Halant + Consonant" was included in the analysis.) of the akshar combinations with all the other script teams.

Oriya script has a set of possible cross-script variants only with the Malayalam script. Cases listed in Table 6 are of the variants that are proposed to be cross-script variants between Oriya and Malayalam.

It is to be noted that none of the combinations listed in Table 6 are termed to be equivalents of each other semantically or otherwise. They are only grouped based on possible visual confusability.

NBGP has ensured that Oriya and Malayalam LGR teams propose a same set of cross-script variants by meeting face-to-face on many occasions as well as through mail communications. The same set of cross-script variants (with Malayalam) is supposed to be found in the Malayalam LGR documents.

Variant Set	Oriya		Malayalam	
	CP	Glyph	CP	Glyph
1.	0B02	◌ͦ	0D02	◌o
2.	0B03	◌ଃ	0D03	◌ː
3.	0B20	O	0D20	O

Table 6: Variant set between Oriya script and Malayalam script

The cases listed in Appendix B are the confusable code points for reference, but they are not defined as variant code points for this proposal.

7   Whole Label Evaluation Rules (WLE)

This section provides the whole label evaluation rules for text written in Oriya script. The rules have been drafted in such a way that they can be easily translated into the LGR specification.

Below are the symbols used in the WLE rules for each of the "Indic Syllabic Category" as mentioned in Table 7: Code point repertoire

In addition, we have created a few additional symbols related to matras and vowels for the explanation of the rules.

C   →   Consonant

M   →   Matra

V   →   Vowel

B   →   Anusvara

H   →   Halant / Virama

N   →   Nukta

C1   →   {ଡ0B21, ଢ0B22, କ 0B15, ଖ 0B16, ଗ 0B17, ଚ 0B1A, ଜ 0B1C, ଫ 0B2B}

X   →   Visarga

D   →   Candrabindu

Rule1: N(଼) must be preceded only by C1

For example:

ଡ(0B21)+଼(0B3C) = ଡ଼

ଢ(0B22)+଼ (0B3C) = ଢ଼

Rule2: B (ଂ) must be preceded by V, C, N or M

i)   B must be preceded by V (examples: ଅଂଶ,)

ii)   B must be preceded by C, (example: ସଂସାର, ବଂଶ)

iii)   B must be preceded by N (example: ଡଂଗା)

iv)   B must be preceded by M, (examples: ସିଂହ, ମାଂସ, ବିଂଶ, ସୁତରାଂ)

Rule3: X (ଃ) must be preceded by C, V, N or M

i)   X must be preceded by C, (example: ପ୍ରାୟଃ, କ୍ରମଶଃ)

ii)   X must be preceded by N (example: ଡଃ)

iii)   X must be preceded by M, (examples: ଦୁଃଖା, ଦୁଃଖିତ,)

iv)   X must be preceded by V, (examples:ଅଃ, ଆଃ, ଇଃ, ଉଃ) commonly used when writing
sanskrit or when there is religious requirement

Rule4: D (ଁ) must be preceded by V, C, N or M

i)   D must be preceded by V (examples: ପାଇଁ, ଯେଉଁ ,ନିଆଁ)

ii)   D must be preceded by C (example: ମୁହଁ, ପହଁରା, ନୁହଁ)

iii)   D must be preceded by N (example: ଡ଼ାଁଣ)

iv)   D must be preceded by M (examples: ନାହିଁ, ନାଁ, ଗାଁରେ)

Rule5: H (◌୍) must be preceded by C or N

i)   H must be preceded by C, (example: ଠିକ୍, ଭୁଲ୍)

ii)   H must be preceded by N (example: ଢ୍ୟୁତି, ଢ୍ରାଗନ)

Rule6: M must be preceded by C or N

i)   M must be preceded by C (example: ମୁହଁ, ପହଁରା, ନୁହଁ)

ii)   M must be preceded by N (example:ଡ଼ାଇଲମିଟର)

8 Contributors

NBGP Co-chairs: Dr. Udaya Narayan Singh, Mr. Mahesh D Kulkarni and Dr. Ajay Data

Following is the full list of NBGP members with their language expertise.

Name	Language Expertise
Udaya Narayana Singh	Bengali, Maithili, Hindi, English
Ajay Data	Hindi
Mahesh D. Kulkarni	Marathi, Hindi
Anupam Agrawal	Hindi, Bengali
Akshat S. Joshi	Hindi, Marathi

Abhijit Dutta	Bengali, Hindi
Neha Gupta	Hindi
Nishit Jain	Hindi
Prabhakar Pandey	Hindi
Raiomond Doctor	English, Hindi, Marathi, Gujarati
N. DeivaSundaram	Tamil
Shantaram S. WardeWalawalikar	Konkani
Bal Krishna Bal	Nepali
Ganesh Murmu	Santali
Balaram Prasain	Nepali
Rajib Chakraborty	Bangla (Bengali)
Gurpreet Singh Lehal	Panjabi
SarojaBhate	Sanskrit
Shambhu Kumar Singh	Maithili
SwarnaPrabhaChainary	Bodo
Ghanashyam Nepal	Nepali

KalyanVasudeo Kale	Marathi
Shashi Pathania	Dogri
Santhosh Thottingal	Malayalam, Sourashtra, Tamil
Uma Maheshwar G	Telugu
Girish Chandra Mishra	Odia
K. C. Tikayat ray	Odia
Kuldeep Patnaik	Odia
Debajit Sharma	Assamese
DrDevasisaJethy	Oriya
Basanta Kumar Panda	Oriya
Arvind Bhandari	Gujarati
Harish Chowdhary	Hindi
Chitrita Chatterjee	Multiple languages represented by members of IAMAI
U.B. Pavanaja	Kannada
Hempal Shrestha	Nepali, Newari
Suraj Adhikari	Nepali

Gangadhar Panday	Telugu
Vinay Murarka	Hindi
Mukesh Saini	Hindi
Jay Paudyal	Hindi
PawanChitrakar	Nepali
NirajanParajuli	Nepali
Uttam Shrestha Rana	Nepali
Dev DassManandhar	Nepali,Newari
BhimDhoj Shrestha	Nepali, Newari
Rajiv Kumar	Hindi
Shubham Saran	Hindi
Anivar A. Aravind	Malayalam
Shanmugam R	Tamil
Prasad PK	Malayalam
SinnathambiShanmugarajah	Tamil

Table 7: Contributors and NBG Panel

9   References

[MSR] Integration Panel, "Maximal Starting Repertoire — MSR-3Overview and Rationale", 15 January 2018https://www.icann.org/sites/default/files/packages/lgr/msr/msr-3-wle-rules-15jan18-en.html

[NBGP] Neo-Brahmi Generation Panel

[Code Charts] The Unicode Standard 10.0 Character Code Chartshttp://www.unicode.org/charts/PDF/U0B00.pdf (Accessed on 12th Jan. 2018)

[1]   The Unicode Standard 1.0.0
        Any code point originally encoded in Unicode 1.0.0

[2]   The Unicode Standard 1.0.1
        Any code point originally encoded in Unicode 1.0.1

[4]   The Unicode Standard 4.0
        Any code point originally encoded in Unicode 4.0

[5]   The Unicode Standard 5.1
        Any code point originally encoded in Unicode 5.1

[101] Omniglot, "Oriya"https://www.omniglot.com/writing/oriya.htm

[102] Odia (Oriya) alphabet - Wikipedia (https://en.wikipedia.org/wiki/Odia_alphabet)

[103]Odia language - Wikipedia (https://en.wikipedia.org/wiki/Odia_language)

[104] Oriya (Unicode block) - Wikipedia (https://en.wikipedia.org/wiki/Oriya_(Unicode_block)

[105] Odisha State Govt. Primary School Grade 1 e-book "HasaKhela": by Odisha Primary Education Programme Authorityhttp://opepa.odisha.gov.in/website/Download/e-Text-Book/CLass%20I/Hasa%20Khela%20Part%20II/Haso%20Khelo-II-Page-112.pdf

[Add other websites and sources cited here]

Figure 4: Odisha State Govt. Primary School Grade 1 e-book (Page 112)

10 Appendix A: Cross-script Confusable Code Points

Oriya script has a set of possible cross-script confusable code points with the Gujarati, Bengali, Telugu, and Kannada. The Table 8 and Table 9 lists them.

10.1 Oriya and Gujarati

The following characters are visually confusable. The NBGP discussed and concluded that they are similar code points but should not be considered as variant code points.

Oriya	Gujarati
ଃ (0B03)	ઃ (0A83)
ପ (0B2A)	ઘ (0A98)
ଥ (0B25)	થ (0AA5)

Table 8: Confusable code points between Oriya script and Gujarati script

10.2 Oriya and Bengali

The following characters are visually confusable. The NBGP discussed and concluded that they are similar code points and should not be considered as variant code points.

Bengali	Oriya
ও U+0993	ଓ U+0B13

Table 9 Confusable code points between Oriya script and Bengali script

The following characters were discussed and the NBGP concluded that they are not variant code points nor confusable code points.

Bengali	Oriya	Resolution
ঘ U+0998	ଷ U+0B38	Distinguishable

10.3 Table 10: Other resolutions between Oriya script and Bengali scriptOriya and Telugu

The following characters were discussed and the NBGP concluded that they are not variant code points nor confusable code points

Oriya	Telugu	Resolution
ଠ (0B20)	ర (0C30)	Distinguishable
ଠ (0B20)	ఠ (0C20)	Distinguishable

Table 11: Other resolutions between Oriya script and Telugu script

10.4 Oriya and Kannada

The following characters were discussed and the NBGP concluded that they are not variant code points nor confusable code points

Oriya	Kannada	Resolution
ଠ (0B20)	ರ (0CB0)	Distinguishable
ଠ (0B20)	ಠ (0CA0)	Distinguishable

Table 12: Other resolutions between Oriya script and Kannada script

11 Appendix B: Oriya Dialects

There are different ways of speaking and meaning of words in local Oriya Language. However the script remains the same.[2]

11.1.1.1 Standard Odia

Kataki Odia or *The Odia of Mughalbandi region*

considered as Standard Odia due to literary traditions. It is spoken mainly in the eastern half of the state of Odisha, with little variation, in districts like

Khurdha, Puri, Cuttack, Jajpur, Jagatsinghpur, Kendrapada, Dhenkanal, Angul and Nayagarh district.

11.1.1.2 Major forms, or dialects

Midnapori Odia:

Spoken in the undivided Midnapore District of West Bengal.

Singhbhumi Odia:

Spoken in East Singhbhum, West Singhbhum and Saraikela-Kharsawan district of Jharkhand

Baleswari Odia:

Spoken in Baleswar, Bhadrak and Mayurbhanj district of Odisha.

Ganjami Odia:

Spoken in Ganjam and Gajapati districts of Odisha and Srikakulam district of Andhra Pradesh.

Sambalpuri Odia:

Spoken in
Bargarh, Bolangir, Boudh, Debagarh, Jharsuguda, Kalahandi, Nuapada, Sambalpur and Subarnapur districts of Odisha and by some people in Raigarh, Mahasamund, Raipur districts of Chhattisgarh state.

Desiya Odia:

Spoken in
Koraput, Rayagada, Nowrangpur and Malkangiri Districts of Odisha and in the hilly regions of Vishakhapatnam, Vizianagaram District of Andhra Pradesh.

Bhatri:

---

[2] Extracted from Wikipedia (https://en.wikipedia.org/wiki/Odia_language#Major_forms_or_dialects)

Spoken in South-western Odisha and eastern-south Chhattisgarh.

Halbi:

Spoken in undivided Bastar district of Chhattisgarh. Halbi is a mixture of Odia and Marathi with influence of Chatishgarhi tribal languages.

Phulbani Odia:

Spoken in Phulbani, Phulbani Town, Khajuripada block of Kandhamal, and in nearby areas bordering Boudh district .This language gained momentum during the amalgamation of Kandhamal (Phulbani), and Boudh, region into a single district Phulabani,

11.1.1.3 Minor non literary and tribal forms or dialects

Sundargadi Odia : Variation of Odia

Spoken in Sundargarh district of Odisha and in adjoining pockets of Jharkhand and Chhattisgarh.

Kalahandia Odia : Variation of Odia

spoken in undivided Kalahandi District and neighboring districts of Chhattisgarh.

Kurmi: Spoken in Northern Odisha and South west Bengal.

Sounti: Spoken in Northern Odisha and South west Bengal.

Bathudi: Spoken in Northern Odisha and South west Bengal.

Kondhan: A tribal dialect spoken in Western Odisha..

Laria: Spoken in bordering areas of Chatishgarh and Western Odisha.

Aghria: Spoken mostly by the ingenious people of Aghria caste in Western Odisha.

Bhulia: Tribal form spoken in Western Odisha.

Sadri: A mixture of Odia and Hindi language with major regional tribal influence.

Bodo Parja / Jharia: Tribal dialect of Odia spoken mostly in Koraput district of Southern Odisha .

Matia: Tribal dialect of Odia spoken in Southern Odisha.

Bhuyan: Tribal dialect of Odia spoken in Southern Odisha.

Reli: Spoken in Southern Odisha and bordering areas of Andhra Pradesh.

Kupia: Spoken by Valmiki caste people in the Indian state of Telangana and Andhra Pradesh, mostly in Hyderabad, Mahabubnagar, Srikakulam, Vizianagaram, East Godavari and Visakhapatnam districts.