

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation Internationale de Normalisation
 Международная организация по стандартизации

Doc Type: Working Group Document**Title: Final proposal for encoding the Warang Citi script in the SMP of the UCS****Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)****Author: Michael Everson****Status: Liaison Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N1958 (1999-01-29), N3411 (2008-04-08), N3668 (2009-08-05)****Date: 2012-04-19**

1. Introduction. The Warang Citi script is used to write the Ho language. Ho is a North Munda language, which family, together with the Mon-Khmer languages, makes up Austro-Asiatic. Warang Citi was devised by charismatic community leader Lako Bodra as part of a comprehensive cultural program, and was offered as an improvement over scripts used by Christian missionary linguists. Ho people live in the Indian states of Orissa and Jharkand. In Jharkand, they are found in Ranchi, Chaibasa and Jamshedpur, and in villages like Pardsa, Jaldhar, Tekasi, Tilupada, Baduri, Purtydigua, Roladih, Tupu Dana, Jetia, Dumbisai, Harira, Gitilpi, Karlajuri, Sarliburu, Narsana, Gidibas Kokcho, Lupungutu, Pandaveer, Jhinkapani, Kondwa. According to the SIL *Ethnologue*, there are 1,026,000 speakers of Ho. There are at present two publications in the script: a magazine *Ho Sanagam* ('meeting' from Hindi *saṅgam*), which comes out yearly and *Kolhan Sakam*, which comes out biweekly.

Today, the Ho community can be characterized as still a primarily oral community, with an emergent literary tradition. Many Ho do not write their language in any form. Among those who do, there are different camps. One camp that promotes the use of Devanagari for Ho; another camp knows how to use Warang Citi, and expresses a strong sentiment, even a kind of reverence towards it. And there are some people who seem to fall into both camps. There are also people who use Latin letters to write Ho on an ad-hoc basis and for things like e-mail or SMS. The use of Warang Citi seems to be stronger in the eastern part of the diaspora, where books that use it have been printed. (See samples below.) The expert we consulted, Mr Dobro Buriuli, teaches the script and over 200 students have passed his courses. In the opinion of Anna Pucilowski, who did fieldwork in February 2012, "the people that are interested in Warang Chiti are all ages I would say but they're mostly educated. Literacy is pretty low outside the towns and cities."

2. Structure. Warang Citi is a casing script, written from left to right. Unlike many other Indic scripts, it has features of an alphabet: vowels are written as full letters, with no vowel-modifiers. However, consonants may have an inherent vowel; it typically is pronounced [a] or [ɔ], and less often [ɛ], and never occurs in final position. Because the inherent vowels do not appear explicitly, there can be ambiguity in the reading of certain words. The script has no regular system of conjuncts nor an explicit virama. In the presentation here, the inherent vowel is transcribed *Aa* and the explicit vowel 𑄂𑄃 is transcribed *Āā*.

The letter 𑄂𑄃 *hiyo* is used after a vowel to indicate lengthening or aspiration: 𑄂𑄃𑄂𑄃 *ĀH āh* (*Ā ā*). Long *i* and *u* are represented by the individual letters 𑄂𑄃 *ī ī* and 𑄂𑄃 *ū ū*; these do not appear to be equivalent to 𑄂𑄃𑄂𑄃 *ih ih* or 𑄂𑄃𑄂𑄃 *uh uh*. The letters 𑄂𑄃 *ī ī* and 𑄂𑄃 *ū ū* may also, in certain cases, be pronounced as short vowels or diphthongs ([ai] for long *i* or [au] for long *u*). The details of how (or if) length is handled orthographically in Warang Citi is unclear at this point.

As in the Brahmi-derived scripts, Warang Citi includes the OM sign ॐ , which occasionally appears in words, such as ॐॐॐ *śoms*, representing Hindi *śvāms* ‘breathing’. The script does not include a diacritic mark for anusvara as in Devanagari, but rather has a separate grapheme, ॐ NGAA.

The script has four so-called “ligatures” ॐ YA ($\text{ॐ} + \text{ॐ}_M + \text{ॐ}$), ॐ YO ($\text{ॐ} + \text{ॐ}_A + \text{ॐ}$), ॐ II ($\text{ॐ}_I + \text{ॐ}_I$), and ॐ UU ($\text{ॐ}_U + \text{ॐ}_U$). None of these “ligatures” are decomposable, and so are encoded separately. The digraph hb occurs occasionally and is used for [b] or [w].

3. Digits and numbers. Warang Citi now makes use of a zero, though this may not have been the case previously; no zero is given in Pinnow 1972 or Zide 1996, but more recent materials use zero, as for instance in the date $\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}$ ‘15-10-2003’ (see Figure 6). Note that the higher digits may have fallen into disuse. In the date, for example, 118EA WARANG CITI NUMBER TEN is not used, but rather 118E1 WARANG CITI DIGIT ONE and 118E0 WARANG CITI DIGIT ZERO. No symbols for 100 or 1000 are known, though Pinnow postulated that they may exist.

4. Sorting. Collation order is as in the code chart, with casing pairs interfiling as in Latin. An exception here is U+118FF WARANG CITI OM, which does not file like $\text{ॐ}\text{ॐ}$ *OM om*, but rather comes at the beginning of the alphabet, although it is placed last in the code chart so that the alphabet itself can have a constant offset. The order is as given in Figure 5, which is considered authoritative. In the code chart all the capital letters are given, then all the small letters, as in ASCII, Armenian, Cyrillic, Deseret, Glagolitic, and Greek.

ॐ om < ॐ M <<< ॐ m < ॐ Ā <<< ॐ ā < ॐ I <<< ॐ i < ॐ U <<< ॐ u < ॐ P <<< ॐ p < ॐ Y <<< ॐ y < ॐ Ī <<< ॐ ī < ॐ Ū <<< ॐ ū < ॐ E <<< ॐ e < ॐ O <<< ॐ o < ॐ Ñ <<< ॐ ñ < ॐ G <<< ॐ g < ॐ K <<< ॐ k < ॐ Ñ̄ <<< ॐ ñ̄ < ॐ J <<< ॐ j < ॐ C <<< ॐ c < ॐ N̄ <<< ॐ n̄ < ॐ D̄ <<< ॐ d̄ < ॐ T̄ <<< ॐ t̄ < ॐ N̄ <<< ॐ n̄ < ॐ D <<< ॐ d < ॐ O <<< ॐ o < ॐ M <<< ॐ m < ॐ B <<< ॐ b < ॐ P <<< ॐ p < ॐ H <<< ॐ h < ॐ L <<< ॐ l < ॐ R <<< ॐ r < ॐ T <<< ॐ t < ॐ S <<< ॐ s < ॐ V <<< ॐ v

5. Character names. The name of the script is $\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}$ *Hbârṇ citi* ‘undying writing’, where $\text{ॐ}\text{ॐ}$ hb is a digraph representing /w/. The name *Varang Kshiti* found in some literature is a sanskritized form. Letter names are traditional. Pinnow 1972 gives the following transcriptions. After those are given the recommended spellings for English and French. The name VIYO for the last letter ॐ [v] has been constructed; it is an original ligature of ॐ A and ॐ HIYO.

ॐ a?	NGAA	NGĀ’	ॐ ε?	E	’E’	ॐ ε̄	ENN	’ENN	ॐ u?	PU	PU’
ॐ a	A	’A	ॐ o?	O	’O’	ॐ ō	ODD	’ODD	ॐ iy	HIYO	HIYO
ॐ ü?	WI	Ū’I	ॐ ang	ANG	’ANG	ॐ ε?	TTE	TTE’	ॐ o	HOLO	HOLO
ॐ u?	YU	YU’U	ॐ a?	GA	GA’	ॐ uṅ	NUNG	NUNG	ॐ o	HORR	HORR
’ya?	YA	YA’	ॐ o?	KO	KO’	ॐ a?	DA	DA’	ॐ o	HAR	HAR
’yo?	YO	YO’	ॐ ēñ	ENY	’EÑ	ॐ a?	AT	’AT	śu:	SSUU	ÇŪ
ॐ i:	II	’Ī	ॐ ü’j	YUJ	YŪ’J	ॐ a?	AM	’AM	śī:	SII	ŚĪ
ॐ u:	UU	’Ū	ॐ üc	UC	’ŪC	ॐ u	BU	BU	ॐ iy	VIYO	VIYO

6. Punctuation. Pinnow states explicitly that European punctuation is used as in English. The following characters have been observed: . , : ... ‘ ’ “ ” - — ? and !. In some fonts the dots tend to be drawn rather high—but this is still FULL STOP, not MIDDLE DOT. The apostrophe ’ (RIGHT SINGLE QUOTATION MARK) is used very frequently in words ending in -’y -’ṛ, as in $\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}\text{ॐ}$ *rājā’ṛ*.

7. Conjunct forms. A small number of conjunct forms are known, most of these represent doubled consonants. Pinnow calls them “ligatures” and says:

Sonstige Ligaturen kommen in den älteren Büchern gar nicht vor; in den späteren werden gelegentlich Ligaturen gebraucht, besonders zur Umschrift von Hindi-Wörtern. Die Ligaturen bestehen einfach darin, daß des jeweils zweite der ohne durch Vokal getrennten Konsonantenzeichen über das erste gesetzt wird, z.B. $b(a)r^n$, eigentlich br^n ... Da die Ligaturen nur selten vorkommen, sind Druck und Schrift des Alphabets relativ einfach.

‘Other ligatures do not appear at all in the older books; in later ones ligatures were occasionally used, especially for the transliteration of Hindi words. The ligatures consist of the simple positioning of the second of two consonants which are not separated by a vowel on top of the first, e.g. $b(a)r^n$, actually br^n ... Since the ligatures occur only rarely, printing and writing the alphabet are relatively simple.’

The upper letter in the conjunct form is supposed to be drawn upside down. In the example below, where 𑌖 rra contrasts with 𑌗 rra , the latter is evidently a typographical error. (If this is the case then perhaps 𑌘 cca , 𑌙 nda , 𑌚 nta , and 𑌛 tta are errors too.) When capitalized, only the lower letter takes a capital shape, and is reduced in size. (No examples of a capital letter with an inverted conjunct has been seen in the source materials) In the table of contents of the book *Bha butu woṅga buru* (1984), the poem “*Bha Āṭeḍā kana Diṣuma Bano*” is written $\text{𑌗 𑌖 𑌗 𑌖 𑌗 𑌖 𑌗 𑌖 𑌗 𑌖 𑌗 𑌖}$, but on p. 49 it is written with a conjunct in the first word, as $\text{𑌗 𑌖 𑌗 𑌖 𑌗 𑌖 𑌗 𑌖 𑌗 𑌖 𑌗 𑌖}$ “*BHA ĀṬEḌĀ KANA DIṢUMA BANO*”. *Bha butu woṅga buru* shows the following conjunct forms:

𑌖 Āṃ	$\text{𑌖 𑌖 𑌖 𑌖 Āṃde /ande/}$	𑌛 Tsa	𑌛 𑌛 Tsina /tisiṅ/
𑌘 cca	$\text{𑌖 𑌖 𑌖 𑌖 Ūccuḍa alongside 𑌖 𑌖 𑌖 𑌖 Ūccuḍa /ucuḍ/}$	𑌗 Bha	$\text{𑌗 Bha and 𑌘 bha alongside 𑌗 Bha~Baha /ba:/}$
𑌙 ṭṭa	$\text{𑌖 𑌖 𑌖 𑌖 Ṭṭuṭ; 𑌖 𑌖 𑌖 𑌖 ṭṭala /aṭal/}$	𑌗 ppa	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 juppuāḥa (juppuāḥa) /jupuwa?/}$
𑌚 ṭra	$\text{𑌖 𑌖 𑌖 𑌖 ṭṭaraṣaṭra /udrast/}$	𑌛 rra	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 purra /pur/; 𑌖 𑌖 𑌖 𑌖 𑌖 porra /por/; 𑌖 𑌖 𑌖 𑌖 𑌖 Torra, 𑌖 𑌖 𑌖 𑌖 𑌖 torra /tor/; 𑌖 𑌖 𑌖 𑌖 𑌖 haserra /heser/}$
𑌛 nda	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 Sippinduri /sipinduri/ alongside 𑌖 𑌖 𑌖 𑌖 𑌖 Sindurii /sinduri/}$	𑌗 rra	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 parra /par/; 𑌖 𑌖 𑌖 𑌖 𑌖 hirra /hir/}$
𑌛 nta	$\text{𑌖 𑌖 𑌖 𑌖 Ṃnanta /anante/}$	𑌛 lla	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 sulla /sul/}$
𑌛 tta	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 Paruṣottama /puru sutam/}$	𑌛 lla	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 olla alongside 𑌖 𑌖 𑌖 𑌖 𑌖 olla /ol/}$
𑌛 tra	$\text{𑌖 𑌖 𑌖 𑌖 𑌖 triyu /tiriyu/}$		

The choice of a conjunct form does not appear to be predictable. The recommended mechanism for representing these conjuncts is to make use of U+200D ZERO-WIDTH JOINER, so that users can specify the discretionary ligatures they prefer to have. There is no difference in meaning between 𑌖 𑌖 𑌖 𑌖 Ūccuḍa and 𑌖 𑌖 𑌖 𑌖 Ūccuḍa , but because the variation should be preserved in interchange, ZERO-WIDTH JOINER is preferred to OpenType font features.

8. Linebreaking. Letters and digits behave as in Latin and other alphabetic scripts. U+118FF 𑌗 WARANG CITI OM behaves like a letter.

9. Unicode Character Properties.

- 118A0;WARANG CITI CAPITAL LETTER NGAA;Lu;0;L; ; ; ; ;N; ; ; ;118C0;
- 118A1;WARANG CITI CAPITAL LETTER A;Lu;0;L; ; ; ; ;N; ; ; ;118C1;
- 118A2;WARANG CITI CAPITAL LETTER WI;Lu;0;L; ; ; ; ;N; ; ; ;118C2;
- 118A3;WARANG CITI CAPITAL LETTER YU;Lu;0;L; ; ; ; ;N; ; ; ;118C3;
- 118A4;WARANG CITI CAPITAL LETTER YA;Lu;0;L; ; ; ; ;N; ; ; ;118C4;
- 118A5;WARANG CITI CAPITAL LETTER YO;Lu;0;L; ; ; ; ;N; ; ; ;118C5;
- 118A6;WARANG CITI CAPITAL LETTER II;Lu;0;L; ; ; ; ;N; ; ; ;118C6;
- 118A7;WARANG CITI CAPITAL LETTER UU;Lu;0;L; ; ; ; ;N; ; ; ;118C7;
- 118A8;WARANG CITI CAPITAL LETTER E;Lu;0;L; ; ; ; ;N; ; ; ;118C8;
- 118A9;WARANG CITI CAPITAL LETTER O;Lu;0;L; ; ; ; ;N; ; ; ;118C9;
- 118AA;WARANG CITI CAPITAL LETTER ANG;Lu;0;L; ; ; ; ;N; ; ; ;118CA;
- 118AB;WARANG CITI CAPITAL LETTER GA;Lu;0;L; ; ; ; ;N; ; ; ;118CB;
- 118AC;WARANG CITI CAPITAL LETTER KO;Lu;0;L; ; ; ; ;N; ; ; ;118CC;
- 118AD;WARANG CITI CAPITAL LETTER ENY;Lu;0;L; ; ; ; ;N; ; ; ;118CD;
- 118AE;WARANG CITI CAPITAL LETTER YUJ;Lu;0;L; ; ; ; ;N; ; ; ;118CE;
- 118AF;WARANG CITI CAPITAL LETTER UC;Lu;0;L; ; ; ; ;N; ; ; ;118CF;
- 118B0;WARANG CITI CAPITAL LETTER ENN;Lu;0;L; ; ; ; ;N; ; ; ;118D0;
- 118B1;WARANG CITI CAPITAL LETTER ODD;Lu;0;L; ; ; ; ;N; ; ; ;118D1;
- 118B2;WARANG CITI CAPITAL LETTER TTE;Lu;0;L; ; ; ; ;N; ; ; ;118D2;
- 118B3;WARANG CITI CAPITAL LETTER NUNG;Lu;0;L; ; ; ; ;N; ; ; ;118D3;
- 118B4;WARANG CITI CAPITAL LETTER DA;Lu;0;L; ; ; ; ;N; ; ; ;118D4;
- 118B5;WARANG CITI CAPITAL LETTER AT;Lu;0;L; ; ; ; ;N; ; ; ;118D5;

118B6;WARANG CITI CAPITAL LETTER AM;Lu;0;L;;;N;;;118D6;
118B7;WARANG CITI CAPITAL LETTER BU;Lu;0;L;;;N;;;118D7;
118B8;WARANG CITI CAPITAL LETTER PU;Lu;0;L;;;N;;;118D8;
118B9;WARANG CITI CAPITAL LETTER HIYO;Lu;0;L;;;N;;;118D9;
118BA;WARANG CITI CAPITAL LETTER HOLO;Lu;0;L;;;N;;;118DA;
118BB;WARANG CITI CAPITAL LETTER HORR;Lu;0;L;;;N;;;118DB;
118BC;WARANG CITI CAPITAL LETTER HAR;Lu;0;L;;;N;;;118DC;
118BD;WARANG CITI CAPITAL LETTER SSUU;Lu;0;L;;;N;;;118DD;
118BE;WARANG CITI CAPITAL LETTER SII;Lu;0;L;;;N;;;118DE;
118BF;WARANG CITI CAPITAL LETTER VIYO;Lu;0;L;;;N;;;118DF;
118C0;WARANG CITI SMALL LETTER NGAA;Ll;0;L;;;N;;;118A0;;118A0
118C1;WARANG CITI SMALL LETTER A;Ll;0;L;;;N;;;118A1;;118A1
118C2;WARANG CITI SMALL LETTER WI;Ll;0;L;;;N;;;118A2;;118A2
118C3;WARANG CITI SMALL LETTER YU;Ll;0;L;;;N;;;118A3;;118A3
118C4;WARANG CITI SMALL LETTER YA;Ll;0;L;;;N;;;118A4;;118A4
118C5;WARANG CITI SMALL LETTER YO;Ll;0;L;;;N;;;118A5;;118A5
118C6;WARANG CITI SMALL LETTER II;Ll;0;L;;;N;;;118A6;;118A6
118C7;WARANG CITI SMALL LETTER UU;Ll;0;L;;;N;;;118A7;;118A7
118C8;WARANG CITI SMALL LETTER E;Ll;0;L;;;N;;;118A8;;118A8
118C9;WARANG CITI SMALL LETTER O;Ll;0;L;;;N;;;118A9;;118A9
118CA;WARANG CITI SMALL LETTER ANG;Ll;0;L;;;N;;;118AA;;118AA
118CB;WARANG CITI SMALL LETTER GA;Ll;0;L;;;N;;;118AB;;118AB
118CC;WARANG CITI SMALL LETTER KO;Ll;0;L;;;N;;;118AC;;118AC
118CD;WARANG CITI SMALL LETTER ENY;Ll;0;L;;;N;;;118AD;;118AD
118CE;WARANG CITI SMALL LETTER YUJ;Ll;0;L;;;N;;;118AE;;118AE
118CF;WARANG CITI SMALL LETTER UC;Ll;0;L;;;N;;;118AF;;118AF
118D0;WARANG CITI SMALL LETTER ENN;Ll;0;L;;;N;;;118B0;;118B0
118D1;WARANG CITI SMALL LETTER ODD;Ll;0;L;;;N;;;118B1;;118B1
118D2;WARANG CITI SMALL LETTER TTE;Ll;0;L;;;N;;;118B2;;118B2
118D3;WARANG CITI SMALL LETTER NUNG;Ll;0;L;;;N;;;118B3;;118B3
118D4;WARANG CITI SMALL LETTER DA;Ll;0;L;;;N;;;118B4;;118B4
118D5;WARANG CITI SMALL LETTER AT;Ll;0;L;;;N;;;118B5;;118B5
118D6;WARANG CITI SMALL LETTER AM;Ll;0;L;;;N;;;118B6;;118B6
118D7;WARANG CITI SMALL LETTER BU;Ll;0;L;;;N;;;118B7;;118B7
118D8;WARANG CITI SMALL LETTER PU;Ll;0;L;;;N;;;118B8;;118B8
118D9;WARANG CITI SMALL LETTER HIYO;Ll;0;L;;;N;;;118B9;;118B9
118DA;WARANG CITI SMALL LETTER HOLO;Ll;0;L;;;N;;;118BA;;118BA
118DB;WARANG CITI SMALL LETTER HORR;Ll;0;L;;;N;;;118BB;;118BB
118DC;WARANG CITI SMALL LETTER HAR;Ll;0;L;;;N;;;118BC;;118BC
118DD;WARANG CITI SMALL LETTER SSUU;Ll;0;L;;;N;;;118BD;;118BD
118DE;WARANG CITI SMALL LETTER SII;Ll;0;L;;;N;;;118BE;;118BE
118DF;WARANG CITI SMALL LETTER VIYO;Ll;0;L;;;N;;;118BF;;118BF
118E0;WARANG CITI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
118E1;WARANG CITI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
118E2;WARANG CITI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
118E3;WARANG CITI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
118E4;WARANG CITI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
118E5;WARANG CITI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
118E6;WARANG CITI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
118E7;WARANG CITI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
118E8;WARANG CITI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
118E9;WARANG CITI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
118EA;WARANG CITI NUMBER TEN;No;0;L;;;10;N;;;;;
118EB;WARANG CITI NUMBER TWENTY;No;0;L;;;20;N;;;;;
118EC;WARANG CITI NUMBER THIRTY;No;0;L;;;30;N;;;;;
118ED;WARANG CITI NUMBER FORTY;No;0;L;;;40;N;;;;;
118EE;WARANG CITI NUMBER FIFTY;No;0;L;;;50;N;;;;;
118EF;WARANG CITI NUMBER SIXTY;No;0;L;;;60;N;;;;;
118F0;WARANG CITI NUMBER SEVENTY;No;0;L;;;70;N;;;;;
118F1;WARANG CITI NUMBER EIGHTY;No;0;L;;;80;N;;;;;
118F2;WARANG CITI NUMBER NINETY;No;0;L;;;90;N;;;;;
118FF;WARANG CITI OM;Lo;0;L;;;N;;;;;

10. Bibliography.

Pinnow, Heinz-Jürgen. 1972. “Schrift und Sprache in den Werken Lako Bodras im Gebiet der Ho von Singbhum (Bihar)”, in *Anthropos* 67:822–857.
Zide, Norman. 1996. “Scripts for Munda languages”, in Peter T. Daniels and William Bright, eds. *The world’s writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
[s.n.]. 2003. ᱥᱟᱱ ᱦᱚᱱ (Şişu halam). [s.l.]: [s.n.]. (See Figure 25.)

11. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Warang Citi encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.