Proposal for
Generation Panel for Neo-Brāhmī Scripts
Label Generation Ruleset for the Root Zone
Contents

1. Preamble ........................................................................................................................................3
2. General Information ....................................................................................................................4
3. Principal Neo-Brāhmī Languages ................................................................................................6
   3.1 Scripts of India: .......................................................................................................................6
   3.2 Central Asian Scripts ...............................................................................................................7
   3.3 Southeast Asian Scripts ..........................................................................................................7
4. Target Scripts for the Proposed Generation Panel .......................................................................10
   4.1 Principal Languages using the Script .....................................................................................10
5. Proposed Initial Composition of the Panel ..................................................................................12
   5.1 Panel Chair and Members (with Expertise) ..........................................................................12
   5.2 Panel Diversity ......................................................................................................................18
6. Work Plan ....................................................................................................................................19
   6.1 Suggested Timeline with Significant Milestones ..................................................................19
   6.2 Proposed Schedule of Meeting and Teleconferences ..........................................................22
   6.3 Sources of Funding for Travel and Logistics .......................................................................22
   6.4 Need for ICANN Provided Advisors ....................................................................................22
7. REFERENCES ................................................................................................................................23
1. **Preamble**

The present document is divided into 5 parts.

Part 1 furnishes the flow of this document.

Part 2 provides a historical overview of Brāhmī and attempts to describe very rapidly the evolution of the script through the ages.

Part 3 restricts itself to a synchronic picture and provides the present day scenario of the principal languages based on Neo-Brāhmī scripts. This part starts off with a description of the 10 major Scripts of India and its near neighbours\(^1\). Each of these scripts is dealt with succinctly. This is followed by a short picture of scripts and major languages in South-East Asia which have evolved partly or wholly from Brāhmī. These are the Central Asian Scripts and the South-East Asian scripts proper.

Part 4 is a logical corollary of the preceding two parts and proposes the target scripts for the proposed generation panel. Some of these scripts are shared with other countries and which have made a separate proposal to ICANN.\(^2\)

The proposal could have restricted itself to the target scripts but Parts 2 and 3 have been drafted with a didactic aim in view: inform the members of the panel as well as other users about the range and diversity of languages and scripts derived from the Brāhmī script and although quite compendious in nature provide a useful background to Neo-Brāhmī languages.

Part 5 provides the composition of the Panel and its diversity and the last part treats of the time line and the work-plan.

An exhaustive list of references (once again with a didactic point of view in mind) rounds up this document.

---

\(^1\) Sinhala although used in Sri Lanka is included here since its evolution is close to that of the modern Indian scripts.

\(^2\) Singapore and Bangladesh, for example, have made a separate proposal to ICANN for Tamil and Bengali respectively.
2. General Information

The Brāhmī script has been the progenitor of all scripts used to write Modern Indo-Aryan languages, Dravidian and to a lesser extent scripts of the Tibeto-Burman and Munda families. It was also adopted by a large number of cultures in Southeast Asia to transcribe their languages: Burmese, Thai, Lao, Khmer (in South-East Asia), and others in Central Asia (including Tibetan), some of which are no longer in use but attest to the spread of Brāhmī. Like Arabic and Latin scripts, it thus constitutes a script that was the mother of many others. The Neo-Brāhmī group is so named to cover all such scripts used today and which are based on Brāhmī.

The origin of Brāhmī is a debatable question. Some scholars treat it as a based on the Semitic writing system: Phoenician or Aramaic. Others view it as an indigenous Indian invention, often associated with the Indus Valley script.

Brāhmī is written from left to right (though several specimens running from right to left have been found) and has an angular shape. As it evolved this angular feature was gradually replaced by rounded shapes in cultures where palm leaves where used as a medium of written communication. The main feature of Brāhmī is the written syllable or akṣara. The akṣara system is based on the concept of admitting a full Consonant or Vowel as a node. Vowels can admit Vowel Modifiers such as nasals or vowel lengtheners. Consonants (which written simply imply a following vowel such as schwa or /a/) are at times modified by a combining mark functioning as "vowel-killer" (termed Halanta), truncating the following vowel. They also admit a set of modifiers which serve as Vowel Signs; and both vocalized consonants and simple vowel signs also accept further combining marks, which signify Nasalization of vowels. The adjuncts to the Vowel or Consonant nodes are appended in a strict rule-order. This feature has been remarkably stable over the evolution of Brāhmī and has been followed by all the later Indic and Southeast scripts derived from the script.

Brāhmī evolved around the 3rd Century BC. Six centuries later i.e. around the 3rd Century AD, Brāhmī script in India had already divided itself into two main styles commonly termed as Northern and Southern. It is around the 12th Century AD, under the influence and growth of vernacularisation (that went in parallel with Bhakti Movement in literature and philosophy) that the need was felt to transcribe texts in the vernacular, giving impetus to the growth of native scripts.

The Northern Branch of Brāhmī gave rise to what is termed as the Indo-Aryan family. Bengali and Odia (also spelt as "Oriya" in older documents) constituted one branch, whereas Devanāgarī (which later split into Gujarati and Devanāgarī script), constituted the other. These

---

3 The information provided is based on the detailed bibliography on Writing Systems provided in this document and especially on Daniels and Bright. 1996. The World’s Writing Systems. Oxford, Oxford University Press.

4 The name of Oriya was changed to Odia in March 2011, by passing the Orissa (Alteration of Name) Bill, 2010 and the Constitution (113th Amendment) Bill, 2010 in the Parliament. Wherever old documents have used Oriya, the name has been retained as such.
developed in succeeding centuries into sub-varieties which can be broadly divided into Western and Eastern varieties in the north and into Deccan and Peninsular varieties in the south. Other branches of this family include the scripts of the western Himalayan regions, of which the modern Gurmukhi or Punjabi script is the only important modern survivor. The Śāradā script of Kashmir also belongs to this group.

The Southern Branch is characterized by the Dravidian family. Although the evolution itself started around the 3rd Century C.E., by the 12th Century the Southern branch of Brāhmī had already split into smaller script families: Two large families evolved which one gave rise to Telugu and Kannada and the other to Tamil and Malayalam.

The Sinhala script was imported from North India but was influenced at various stages of its development by south Indian scripts.

The distinct forms of modern scripts of India derived from Brāhmī took a stable form between the tenth and fifteenth centuries.

The diagram below (excerpted and adapted from Daniels and Bright, *The World’s Writing Systems*. Oxford, Oxford University Press, 1996, p. 380) shows the major stages in evolution of this family of scripts, illustrated with different forms of the retroflex nasal: \textit{nna} from Brāhmī to the neo-Brāhmī scripts.
3. Principal Neo-Brāhmī Languages

In what follows a short overview of the scripts derived from Brāhmī is given. This section is divided into two parts: Scripts of India and those of South-East Asia. All of these share the major features of Brāhmī described above.

3.1 Scripts of India:

*Devanāgarī*: The script called Nāgarī or Devanāgarī is written from left to right. Historically it derives from the Brāhmī alphabet of the Ashokan inscriptions. Devanāgarī is currently used for 11 out of 22 official languages of India (Boro/Bodo, Dogri, Hindi, Kashmiri, Konkani, Maithili, Marathi, Nepali, Sanskrit, Santhali and Sindhi) and around 45 other languages especially the related Indo-Aryan languages: Bagheli, Bhilli, Bhojpuri, Himachali dialects, Magahi, Newari and Rajasthani and its dialects: Marwari, Mewati, Shekhwatı, Bagri, Dhundhari, Harauti and Wagri. Closely associated with Sanskrit and Pnakrit, it is an alternative script for Kashmiri and Sindhi. It is growing in use by speakers of tribal languages of Arunachal Pradesh, Bihar and Andaman & Nicobar Islands. The script is also used in Fiji to represent Fiji Hindi. Hindi is also used in Mauritius, Malaysia, England, Canada, South Africa, Indonesia as well as emigrant communities around the world.

*Gujarati*: Used for writing Gujarati and Kacchi, it is a variant of Devanāgarī, the main difference being the absence of the shirorekha or the line above the character and also 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. Gujarati is extensively spoken in large parts of Africa, Madagascar, UK and the USA as well as by emigrant communities around the world.

*Gurmukhi* which evolved separately in the Northern family is used to write the Punjabi language in the Indian state of Punjab and elsewhere in India. Gurmukhi which literally means 'from the mouth of the Guru' stabilised around the 16th century when it was used to transcribe the holy Granth Sahib.

*Bengali*: Often termed as Bangla by linguists and grammarians, is historically related and similar in design to the Devanāgarī script and with one or two exceptions has the same consonant and vowel set. Like Devanāgarī, Bengali Script is used to transcribe quite a few languages of which the most prominent are Assamese and Manipuri. The former differs from Bengali in a few consonant characters. The same is the case with Manipuri which today is also written in Meetei Mayek.

*Oriya* [Odia] can be traced back to the Ashokan inscriptions: 3rd century B.C.E. The earliest inscription in Oriya/Odia is in Kalinga script, from which the modern Oriya/Odia script has

---

5 In what follows we have restricted ourselves to living languages with an EGIDS scale of 1-4 and have not taken into account languages based on Brāhmī which have died out such as Sharada, Tirhuta, Kaiti, Mahajani, Modi and some which are used for ritual purposes such as Grantha, Avestan (in Gujarati script); nor does this list take into account new emerging scripts which have yet to be “mandated” such as TaniLipi. A rather exhaustive list of such scripts is to be found on Anshuman Pandey's website linguistics.berkeley.edu/~pandey. A detailed bibliography provides references to materials which can be used for reference. This is in conformity with the guide-lines laid down by MSR 1

6 Unicode prefers the term South Asian Scripts cf. http://www.unicode.org/charts/

7 cf. footnote 4 supra on the names Oriya and Odia
Because of the prevalence of a large number of tribal languages belonging to the Munda and Dravidian families in the state of Odisha (Orissa), the Oriya/Odia script is used in writing these languages.

Sinhala used for writing Sinhala language and at times also Pali, is derived from Brāhmī as early as the third-second century B.C.E. Although it belongs historically to the Northern family, it has been considerably influenced by the early Grantha script of South India.

Kannada and Telugu are closely related scripts used to write two Dravidian languages: Kannada in the state of Karnataka, and Telugu in Telangana and Andhra. Over the centuries, Brāhmī evolved with marked characteristics in the south. Around the tenth century, these crystallised into the Old Kannada script, used where both Kannada and Telugu are now spoken. By around 1500, this script divided into Kannada and Telugu. As a result, there are very few differences between these two scripts.

Malayalam: Subject to reforms, modern Malayalam has introduced alphabetic writing into the script, although the main structure of Malayalam still adheres to the akṣara.

Tamil (also spelt as "Tamizh" in this document) is derived from the southern branch of Ashokan Brāhmī. More than any other script derived from Brāhmī, it is highly alphabetical in nature and admits no ligatures with the exception of two consonant conjuncts. Apart from being the official language of Tamil Nadu, Tamil is also an official and national language of Sri Lanka and one of the official languages of Singapore. It is also one of the languages of medium of education in Malaysia along with English, Malay and Mandarin. It is spoken in the states of Kerala, Karnataka, Andhra Pradesh and Andaman and Nicobar Islands as one of the secondary languages. Tamil is also spoken by significant minorities in Malaysia, England, Mauritius, Canada, South Africa, Fiji, Indonesia, as well as emigrant communities around the world.

3.2 Central Asian Scripts

Due to societal and liturgical reasons, Brāhmī influenced the writing systems of the Himalayan region and Tibetan and Dzongkha are among the languages of the Himalayan region whose scripts are based on Brāhmī.

3.3 Southeast Asian Scripts

As in the case of Central Asian Scripts, this part provides a very rapid and succinct overview of these scripts since they do not fall within the purview of the Neo-Brāhmī panel per se and in what follows a short picture of the major scripts and languages has been provided.

The spread of Brāhmī in Southeast Asia originates from the Southern family, especially the influence of the Pallava script named after the dynasty of the Pallavas. In the initial phase, this was restricted to writing Indian languages in these scripts. Subsequently this gave rise to adaptation of the scripts to the requirements of the local languages (use of tone marker in Thai for example). Thus by the late eighth century; the scripts began to acquire separate identities.

---

8 Cf. footnote 5 supra
9 Cf. footnote 5 supra
Brāhmī-based scripts are used to write languages of several different families in South-East Asia: Mon (Burma), Khmer (of Cambodia); Thai, Lao, Shan; Karen (Burma) and Burmese Javanese and other languages of Indonesia and the Philippines. Some of the major South-East languages using these scripts belong to the Austronesian (Jawi, Cham), Tai (Thai) and Tibeto-Burman (Karen, Mon, Shan, Burmese) families.

The Thai alphabet is derived from the Old Khmer script which itself is derived from Pallava (பல்லவ).\(^{10}\)

The Lao alphabet, Akson Lao (Lao: ສາທາລະ ralluak), is the main script used to write the Lao language and other minority languages in Laos. It is of Indic origin and is traditionally written from left to right.\(^ {11}\)

The Shan alphabet based on Southern Brāhmī is the native language of Shan people and spoken mostly in Shan State, Burma. It is also used in pockets of Kachin State in Burma, in northern Thailand, and decreasingly in Assam.

The Burmese script is an abugida based on the Brāhmī family and is used for writing Burmese. It is an adaptation of the Old Mon or the Pyu script. In recent decades, other alphabets using the Mon script, including Shan and Mon itself, have been restructured according to the standard of the now-dominant Burmese alphabet. Besides the Burmese language, the Burmese alphabet is also used for the liturgical languages of Pali and Sanskrit.\(^ {12}\)

The figure below shows the evolution of the akṣara "ta" in major Southeast scripts influenced by Brāhmī. (Adapted from The World’s Writing System, edited by Peter T. Daniels and William Bright, New York - Oxford, Oxford University Press, 1996, p. 448.)

\(^{10}\)http://en.wikipedia.org/wiki/Thai_alphabet

\(^{11}\)Modified from the article on Lao: http://en.wikipedia.org/wiki/Lao_alphabet

\(^{12}\)Modified from the article on Burmese alphabet with references to Mon
http://en.wikipedia.org/wiki/Burmese_alphabet
4. Target Scripts for the Proposed Generation Panel
As per prior experience, the proposed Generation Panel anticipates that the work on LGR will be divided in two major parts.

i. Generic principles applicable to all the scripts (and thereby languages)
ii. Script specific detailed rules

The Neo-Brāhmī Generation Panel has large number of scripts (and thereby languages) under its ambit. At the proposal submission stage, the Generation Panel has representation from majority of scripts (languages). The Generation Panel plans to initiate work on all the LGRs, while remaining open to taking experts on-board. It is estimated that the Generic principles can be finalized with the current expertise available with the group. If experts for some script LGR remain unavailable till the completion of other LGRs, a set of guiding principles will be created for the subsequent LGR(s).

4.1 Principal Languages using the Script
The table given below provides a list of scripts and the languages which the Neo-Brāhmī GP plans to target.

<table>
<thead>
<tr>
<th>Script</th>
<th>ISO 15924 Code</th>
<th>Language</th>
<th>ISO 639-3 Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bengali</td>
<td>Beng</td>
<td>Assamese – অসমীয়া</td>
<td>asm</td>
</tr>
<tr>
<td>Bengali</td>
<td>Beng</td>
<td>Bengali – বাংলা</td>
<td>ben</td>
</tr>
<tr>
<td>Bengali</td>
<td>Beng</td>
<td>Manipuri – মণিগুর্জি</td>
<td>mni</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Bodo – বড়ো</td>
<td>brx</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Dogri – डोगरी</td>
<td>dgo</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Hindi – हिन्दी</td>
<td>hin</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Kashmiri – कॉशुर, Kāšur, Koshur</td>
<td>kas</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Konkani – कॉंकणी,</td>
<td>knn</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Maithili – मैथिली, मমিলী</td>
<td>mai</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Marathi – मराठी</td>
<td>mar</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Nepali – नेपाली</td>
<td>npi</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Sanskrit – संस्कृतम्, संस्कृतावाक्</td>
<td>san</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Santali/Santhali – संथाली</td>
<td>sat</td>
</tr>
<tr>
<td>Devanāgarī</td>
<td>Deva</td>
<td>Sindhi – सिंधी</td>
<td>snd</td>
</tr>
<tr>
<td>Gujarati</td>
<td>Gujr</td>
<td>Gujarati – ગુજરાતી</td>
<td>guj</td>
</tr>
<tr>
<td>Gurmukhi</td>
<td>Guru</td>
<td>Punjabi – ਪੰਜਾਬੀ</td>
<td>pan</td>
</tr>
<tr>
<td>Kannada</td>
<td>Knda</td>
<td>Kannada – ಕನ್ನಡ</td>
<td>kan</td>
</tr>
<tr>
<td>Malayalam</td>
<td>Mlym</td>
<td>Malayalam – മലയാളം</td>
<td>mal</td>
</tr>
<tr>
<td>Language</td>
<td>Code</td>
<td>Native Name</td>
<td>Short Name</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Oriya(Odia)</td>
<td>Orya</td>
<td>Odia – ଓଡ଼ିଆ</td>
<td>ory</td>
</tr>
<tr>
<td>Tamil</td>
<td>Tamil</td>
<td>Tamil – தமிழ், Tamizh</td>
<td>tam</td>
</tr>
<tr>
<td>Telugu</td>
<td>Telu</td>
<td>Telugu – తెలుగు</td>
<td>tel</td>
</tr>
</tbody>
</table>
5. Proposed Initial Composition of the Panel

The initial contribution of the Neo-Brāhmī group was towards the development of the Devanāgarī VIP report (q.v.) which catered to the problems and issues of implementing Devanāgarī script driven language gTLD's. Presentations and discussions at meetings and workshop conducted by ICANN as well as the Indian Ministry of Communications and Information Technology have proved useful and have made (we hope) a fruitful contribution to ICANN.

With this as a background, the panel is widening its ambit in two major areas.

In the first place, more scripts/languages are being added to the fold. This will ensure that scripts derived from classical Brāhmī and which are widely used in India as well as in South-East Asia and some other countries where the scripts/languages spread owing to societal reasons will eventually get recognition at the international level and will permit users of these scripts to have ccTLD's (in some cases) and gTLD's in their native scripts.

The second area on which focus is being laid comprises technical issues which will define the problems and provide solutions for secure and easy use of Neo-Brāhmī based IDN's for this large community.

These are in common with the other panels and include the following:
- Neo-Brāhmī based Scripts Label Generation Ruleset (LGR) for the Root Zone
- Second level LGRs for the Neo-Brāhmī based Scripts.
- Evaluating the MSR report and identifying valid Neo-Brāhmī Script characters.
- Technical challenges around registration of Neo-Brāhmī based Scripts IDNs and variants

5.1 Panel Chair and Members (with Expertise)

Below is the composition of the Neo-Brahmi Scripts Generation Panel (NBSGP).

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Name</th>
<th>Initials</th>
<th>Organization</th>
<th>Country /Region</th>
<th>Language Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anupam Agrawal</td>
<td>Mr.</td>
<td>Tata Consultancy Service</td>
<td>India</td>
<td>Hindi, Bengali</td>
</tr>
<tr>
<td>2</td>
<td>Akshat S. Joshi</td>
<td>Mr.</td>
<td>C-DAC</td>
<td>India</td>
<td>Hindi, Marathi</td>
</tr>
<tr>
<td>3</td>
<td>Abhijit Dutta</td>
<td>Mr.</td>
<td>Wikimedia</td>
<td>India</td>
<td>Bengali, Hindi</td>
</tr>
<tr>
<td>4</td>
<td>Mahesh D. Kulkarni</td>
<td>Mr.</td>
<td>C-DAC</td>
<td>India</td>
<td>Marathi, Hindi</td>
</tr>
<tr>
<td>5</td>
<td>Neha Gupta</td>
<td>Ms.</td>
<td>C-DAC</td>
<td>India</td>
<td>Hindi</td>
</tr>
<tr>
<td>6</td>
<td>Nishit Jain</td>
<td>Mr.</td>
<td>C-DAC</td>
<td>India</td>
<td>Hindi</td>
</tr>
<tr>
<td>7</td>
<td>Prabhakar Pandey</td>
<td>Mr.</td>
<td>C-DAC</td>
<td>India</td>
<td>Hindi</td>
</tr>
<tr>
<td>8</td>
<td>Raimond Doctor</td>
<td>Dr.</td>
<td>C-DAC</td>
<td>India</td>
<td>English, Hindi, Marathi, Gujarati</td>
</tr>
<tr>
<td>9</td>
<td>Udaya Narayana</td>
<td>Dr.</td>
<td>Holder of <strong>Professorial</strong></td>
<td>India</td>
<td>Bengali, Maithili,</td>
</tr>
<tr>
<td>No.</td>
<td>Name</td>
<td>Role</td>
<td>Relevant Experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1   | Anupam Agrawal                | Community Representative    | • Part of various working groups of Govt. of India  
• Member, ISOC Kolkata Chapter  
• ICANN fellow at ICANN 40  
• Active participation in ICANN forums.  
• Native Bengali language speaker |
| 2   | Akshat S. Joshi               | IDNA/Unicode Expert, Policy Expert | • Working on Indian language computing since last 6 years  
• Part of India’s National initiative on IDNs  
• Speaker at various National & International conferences on Indian IDN implementation  
• Part of ICANNs Devanagari VIP team  
• Part of ICANNs Integrated Issues Report project |
<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Position</th>
<th>Experience and Achievements</th>
</tr>
</thead>
</table>
| 3 | Abhijit Dutta      | Community Representative                      | • Part of RootLGR Procedure drafting team  
• Representative in IBM for South Asian Languages (including Urdu and Sinhala)  
• Part of Government of India (GoI) committees forming e-Governance standards |
| 4 | Mahesh D. Kulkarni | Expert in Linguistics and Unicode              | • Specializing in hardware & electronics as well as in software and team management.  
• Has to his credit large number of publications and patents in the areas of computing.  
• Expertise in the area of standardization specially dealing with Indian language  
• Member of committees and organizations both in India and at international level  
• Awardee of various honours at the national level.  
• Country manager. W3C India  
• Has successfully transformed C-DAC GIST into a multinational company. |
| 5 | Neha Gupta         | IDNA/Unicode Expert, Policy Expert             | • Working on Indian language computing since last 7 years  
• Part of India’s National initiative on IDNs  
• Speaker at various National & International conferences on Indian IDN implementation  
• Part of ICANNs Devanagari VIP team  
• Part of ICANNs Integrated Issues Report project  
• Part of RootLGR Procedure drafting team |
| 6 | Nishit Jain        | IDNA/Unicode Expert, Policy Expert             | • Working on Indian language computing since last 3 years  
• Part of India’s National initiative on IDNs  
• Speaker at various National & International conferences on Indian IDN implementation  
• Part of ICANN Whois IRD Expert WG |
| 7 | Prabhakar Pandey   | Expert in Linguistics                          | • Hindi Officer at C-DAC Pune  
• Worked in Wordnet and other NLP projects in IIT Bombay as a Research Associate |
| 8 | Raimond Doctor     | Expert in Linguistics and Unicode              | • Specialising in Linguistics and trained in Germany and France, Raimond Doctor has lectured extensively in Europe and was Directeur d'Etudes at the Ecole des Hautes Etudes en Sciences Sociales, Paris and also Maître de Conférences at the prestigious Collège de France.  
• Has six books and over a hundred papers dealing with various aspects of language and computational linguistics to his credit.  
• Has been an advisor to UNESCO on Information Technology and also to the prestigious BULAC in Paris for adopting ISO for their library content as well as their meta-data.  
• Involved with the development of ISCII, also |
closely associated with Unicode and ISO both for neo-Brahmi scripts and Urdu.
- Also associated with the Government of India’s IDN project since its inception and has worked extensively on the framework of the policy for Indian Languages.
- His main area of expertise is Natural Language processing with a strong bent towards Morphology and Syntax and also String Theory and Chaos Maths in their application to Linguistic theories.

<table>
<thead>
<tr>
<th></th>
<th>Udaya Narayana Singh</th>
<th>Chair and Co-Investigator of several NLP Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specializes in Linguistics, Translation Studies, Creative Writing, Culture Studies, Lexicography &amp; ELT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Professor, Rabindra Bhavana (since June 2009) &amp; Chair-person, Centre for Endangered Languages, Visva-Bharati</td>
<td></td>
</tr>
<tr>
<td></td>
<td>He has been the first Pro-Vice-Chancellor of Visva-Bharati between Aug 20, 2010 and June 15, 2012.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>He was earlier the Director of Central Institute of Indian Languages (CIIL), Ministry of HRD, Govt. of India (GOI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>As a poet, playwright, essayist and editor in Maithili and Bangla and as a Linguist, he has 49 books to his credit, and has over 180 papers published in National and International journals.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lectured and taught at Universities and Institutions both in India and all over the world.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was the Chief Editor of Indian Linguistics, 1988-1990, Editor (with P.P.Giridhar as Co-Editor) of Translation Today, and a Visiting Professor at the IIAS-Shimla (1989), besides being the Series Editor of eight books under Language and Development (LAD) under Sage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Has been the General Editor of Longman-CIIL Series of eleven Bilingual Dictionaries. Co-Editor of International Journal of Inter-Cultural Relations; 2008-onwards, plus a Member of the Editorial Board of the Brill’s series titled Studies in South and Southwest Asian Languages.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Was the Founder-Director, Study India Program (SIP), University of Hyderabad (1995-2000) &amp; Chief Coordinator, Centre for Distance Education there</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Founder-Head, Centre for Applied Linguistics &amp; Translation Studies (CALTS), University of Hyderabad (1989-1994)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chief Editor, Mithila Darshan (Bi-monthly), the largest circulating literary magazine in Maithili</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Name</td>
<td>Role</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------</td>
<td>-------------------------------</td>
</tr>
</tbody>
</table>
| 10  | N. DeivaSundaram      | Expert in Linguistics         | • Retired Professor, University of Madras  
• MD, NDS Lingsoft Solutions Pvt Ltd.  
• Holds a doctorate in Linguistics and Masters in Tamil language                                                                                       |
| 11  | Shantaram S. Warde Walawalikar | Expert in Linguistics         | • Expert on Konkani language  
• Part of various Gov initiatives on Indian languages as Konkani expert  
• Member, Official Language Terminology, Govt. of Goa                                                                                               |
| 12  | Bal Krishna Bal       | Expert in Linguistics         | • Specialization and Expertise: Software Localization and Natural Language Processing, Unicode, CLDR, Standardisation etc.  
• Currently, Assistant Professor and Team Lead, Information and Language Processing Research Lab, Department of Computer Science and Engineering, Kathmandu University  
• Chief Technical Officer – Language Technology Kendra, Lalitpur, PatanDhoka, Nepal.  
• Project Manager and Team Leader, PAN Localization Project, Nepal Country Component (2004-2009)                                          |
| 13  | Ganesh Murmu          | Expert in Linguistics         | • Works at the Department of Tribal and Regional Languages, University of Ranchi in Jharkhand, India.  
• Expert in the tribal languages of Jharkhand and in particular a tireless advocate and activist for his native Santali language  
• Has worked on the Documentation of Koro Aka Project, on the poorly known Koro Aka language of Arunachal Pradesh, north eastern India, and on three Enduring Voices trips to Arunachal Pradesh.  
• In addition to Koro Aka, Ganesh served as primary or secondary eliciting linguist in the recordings Living Tongues has made of numerous languages of the region, including Hruso Aka, Miji, Apatani, Hill Miri and Nishi.  
• He also serves as the Local Project Coordinator and liaison for Living Tongues Institute for Endangered Languages in Jharkhand State, India. |
| 14  | Balaram Prasain       | Expert in Linguistics         | • Member, Language Technology Kendra  
• Lecturer, Central Department of Linguistics, Tribhuvan University, Kathmandu, Nepal                                                                                                                   |
| 15  | Rajib Chakraborty     | Community Representative      | • Working as a Linguist at the Society for Natural Language Technology Research, Kolkata (a Society under the aegis of Ministry of Information Technology, Govt. of West Bengal).  
• Presently working on Digitization of Bangla literary texts for online preservation, designing                                                                 |
<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Expert in</th>
<th>Achievements</th>
</tr>
</thead>
</table>
| 16  | Gurpreet Singh Lehal        | Expert in Linguistics | • Training a UNICODE-compliant Bangla Spell Checker system for Windows and Linux  
• Editing an International Journal on Linguistics & Language Technology (Bhasha Bijnan O Prayukti)  
• Chief Coordinator of Indo-UK joint project “Enhancing Communication and Co-operation across South Asia: An ICT Solution to Script Barriers” with University of Manchester, UK during 2014-2015  
• Indo-US collaborative project, “Transliteration of Punjabi words in the English-Punjabi Dictionary from the AP2171 Romanization into the modified Perso-Arabic script”, with University of Maryland, Maryland, USA (2012-2013)  
• Chief Coordinator of Indo-UK joint project “Web based Transliteration and Translation System between Urdu and Hindi Languages” with University of Manchester, UK during 2009-2010. Project sponsored under ISIF grants, Australia  
• Chief Coordinator of Indo-UK joint project “Shahmukhi to Gurmukhi Transliteration Solution for Networking” with University of Manchester, UK during 2006-2008. Project sponsored under Pan-Asia grants, Singapore  
• Working in Consortium mode with 15 premier institutes in India,  
• Chief coordinator of project “Punjabi Text to Speech Help for people with cognitive disabilities” in collaboration with IIT Delhi and GNEC Ludhiana during 2011-2013  
| 17  | Saroja Bhatे               | Expert in Linguistics | • M.A. Ph.D. in Sanskrit Grammar  
• Hon. Secretary, Bhandarkar Oriental Research Institute Pune, 2005-08 & 2008-10 |
| 18  | Shambhu Kumar Singh        | Expert in Linguistics | • Working as a Consultant (Academic-Maithili) in National Translation Mission (NTM) of Central Institute of Indian Languages, since July 01, 2008 to till date.  
• Worked as an Associate Editor for Pearson Longman’s Basic English-English-Maithili dictionary under NTM.  
• Associated with ‘Certification of Translators Programme’ of NTM. |
<p>| 19  | Swarna Prabha              | Expert in Linguistics | • Working as Language Expert for Bodo association with LANGUAGE TECHNOLOGY |</p>
<table>
<thead>
<tr>
<th>Member</th>
<th>Expertise</th>
<th>Contributions</th>
</tr>
</thead>
</table>
| Chainary | DEVELOPMENT PROJECT (LTDP) | - Member, LIS Project for Bodo  
- Authored one Bodo grammar for Under Graduate & Post Graduate classes in Bodo, one school grammar for classes from VI-VIII, three volumes of article collection book on language in Bodo.  
- Presented Tibeto-Burman language papers in different symposiums and seminars in India and abroad. |
| 20 Ghanashyam Nepal | Expert in Linguistics | - Assistant Professor of Nepali, University of North Bengal from April 1983 to October 1990  
- Associate Professor of Nepali, Department of Indian Languages, Banaras Hindu University from November 1990 to December 1999  
- Professor of Nepali, Department of Nepali, University of North Bengal, Darjeeling from January 2000 till date. |
| 21 Kalyan Vasudeo Kale | Expert in Linguistics | - M.A. Sanskrit Linguistics  
- M.A. Marathi Ardhamaghdi  
- Ph. D. Marathi Literature  
- Reader, Department of Marathi, University of Pune 1981-97 |
| 22 Shashi Pathania | Expert in Linguistics | - Has two books and 95 research papers and publications to her credit  
- Has done Localization of software tools in Dogri.  
- Worked on various research projects.  
- Currently working on Shallow Parser Tool for Dogri language |

5.2 Panel Diversity

The Generation Panel (GP) for the Neo-Brāhmī Scripts LGR gathers experts from a variety of backgrounds including language experts/epigraphers/linguists; those who have vast expertise in the scripts of their domain. As is the case with languages using Neo-Brāhmī scripts, some of the experts are polyglots and have expertise in more than one script.

In addition to linguistic expertise, the Generation panel comprises experts from other domains: community experts, policy designers, experts with a deep knowledge of the Akṣara formalism and also the technical community directly working with the DNS (e.g. registries) and security. Members from academia round up the list.
6. Work Plan

6.1 Suggested Timeline with Significant Milestones

The Generation Panel for Neo-Brāhmī intends to divide the work on LGR for the Root zone into four phases. These are logically ordered as under:

1. Finalization of Code Points
2. Finalization of Variants
3. Finalization of Whole Label Evaluation Rules (WLEs)
4. Finalization of LGR for Neo-Brāhmī scripts and submission to ICANN

Each of these phases will have three sub-phases.

Phase I: Defining and determining the general principles which in turn will act as a framework, on which further decisions will be based. This underlies all four phases

Phase II: Once the principles are determined, in this phase the data (character repertoire, variants, and labels) will be analyzed.

Phase III comprises outreach where the data analysed and arrived at will be released to the wider community ICANN as well as the user community of the given script.

At all levels feedback from community will be solicited.

Time-Line

Given the complexity of languages subsumed under the Neo-BrāhmīGP, the time-frame required for the above will take at least twelve months starting from January 2015. (This is with the assumption that the process of formation of the GP will be complete before January 2015)

A tentative work plan of Phase 1 is presented in the chart below:

<table>
<thead>
<tr>
<th>Broad timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of days</td>
</tr>
<tr>
<td>Start Date</td>
</tr>
<tr>
<td>End Date</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detailed timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
</tr>
<tr>
<td>Introduce Members</td>
</tr>
<tr>
<td>Discussion, Revision and Finalization of Proposal</td>
</tr>
<tr>
<td>Application to ICANN for formation of Neo-Brahmi GP for creation of Neo-Brahmi LGR for Root Zone</td>
</tr>
<tr>
<td>---</td>
</tr>
</tbody>
</table>

**Interaction within Neo-Brahmi GP**

<table>
<thead>
<tr>
<th>10</th>
<th>01-06-2015</th>
<th>10-06-2015</th>
</tr>
</thead>
</table>

**Meeting with Chair**

**Meeting for Devanagari, Gujarati, Punjabi script members**

**Meeting for Bengali script members**

**Meeting for Tamil, Telugu script members**

**Character Set**

|---|---|---|---|

- For inclusion
- For Exclusion
- For Deferral

**Interaction with IG for Feedback**

<table>
<thead>
<tr>
<th>10</th>
<th>26-06-2015</th>
<th>05-07-2015</th>
</tr>
</thead>
</table>

**Finalization of character sets**

<table>
<thead>
<tr>
<th>30</th>
<th>06-07-2015</th>
<th>04-08-2015</th>
</tr>
</thead>
</table>

- Included in Neo-Brahmi Scripts IDNs
- Excluded from Neo-Brahmi Scripts IDNs
- Interaction with IG for feedback
- Documenting Character Set on MSR

**Release for Public Comments:**

<table>
<thead>
<tr>
<th>30</th>
<th>05-08-2015</th>
<th>03-09-2015</th>
</tr>
</thead>
</table>

**Character sets for Neo-Brahmi Scripts LGR**

<table>
<thead>
<tr>
<th>5</th>
<th>04-09-2015</th>
<th>08-09-2015</th>
</tr>
</thead>
</table>

**Variants**

<table>
<thead>
<tr>
<th>Definition of General principles</th>
<th>20</th>
<th>09-09-2015</th>
<th>28-09-2015</th>
</tr>
</thead>
</table>

- Variant qualification
- Typology of Variants
<table>
<thead>
<tr>
<th>Task</th>
<th>Duration</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction with IG for Feedback</td>
<td>5</td>
<td>29-09-2015</td>
<td>03-10-2015</td>
</tr>
<tr>
<td>Analysis of Data</td>
<td>30</td>
<td>04-10-2015</td>
<td>02-11-2015</td>
</tr>
<tr>
<td>- Variants identification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Finalization of Variants with disposition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Documenting Variants on MSR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Release for Public Comment</td>
<td>30</td>
<td>03-11-2015</td>
<td>02-12-2015</td>
</tr>
<tr>
<td>Incorporation of Comments by Public and IG</td>
<td>5</td>
<td>03-12-2015</td>
<td>07-12-2015</td>
</tr>
<tr>
<td>Whole Label Rules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definition of General principles</td>
<td>20</td>
<td>08-12-2015</td>
<td>27-12-2015</td>
</tr>
<tr>
<td>- Finalization of guiding principles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Interaction with IG for Feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finalization of Rules</td>
<td>60</td>
<td>28-12-2015</td>
<td>25-02-2016</td>
</tr>
<tr>
<td>- Initial Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Second Review of Label Rules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Finalization of Whole Label Rules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Documenting Whole Label Rules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Release for Public Comment:</td>
<td>30</td>
<td>26-02-2016</td>
<td>26-03-2016</td>
</tr>
<tr>
<td>Incorporation of Comments by Public and IG</td>
<td>10</td>
<td>27-03-2016</td>
<td>05-04-2016</td>
</tr>
<tr>
<td>Finalizing LGR Document</td>
<td>30</td>
<td>06-04-2016</td>
<td>05-05-2016</td>
</tr>
<tr>
<td>Finalizing document</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finalizing LGR XML Structure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submission to ICANN</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.2 Proposed Schedule of Meeting and Teleconferences
Given the large number of experts and also their location in distant geographical regions, a large amount of discussions will be carried out by e-mail. As and when needed to ensure the principle of “Get in touch Keep in touch” we will try and also have an initial face-to-face meeting, followed up by teleconferences for the period of the work.

6.3 Sources of Funding for Travel and Logistics
Although the members of the Neo-Brāhmī panel will be volunteers and provide their time and expertise on a purely voluntary basis, issues of logistics such as travel and stay necessitated by members in a face-to-face meeting as well as support for conference calls, posting on the site will require support. ICANN may need to support these logistics of the group to a limited extent. The group will also try and identify alternate sources of funding.

6.4 Need for ICANN Provided Advisors
Apart from the linguistic and policy level expertise available with the GP, it is anticipated that in some cases there would be need for Advisors from ICANN for some specific areas like DNS and ICANN process imperatives. As and when needed, such help may be sought by the GP.
7. REFERENCES

Given the complexity of Neo-Brāhmī Scripts and the complex writing system we have provided a rather “extensive” set of references which cover material used/consulted in compiling this document as well as for those interested in further reading.

ON WRITING

BRĀHMĪ SCRIPT

NEO-BRĀHMĪ SCRIPTS AND LANGUAGES

BENGALISCRIPT

MANIPURI

ASSAMESE
Kakati, Banikanta; Goswami, Golock Chandra. 1962. Assamese, its formation and development: a scientific treatise on the history and philology of the Assamese language. Gauhati, Assam.

13 The references are provided for all the Official Languages of India using the respective scripts and in some instances important dialects of the language.

**DEVANĀGARĪ SCRIPT**

**General Overview**


**Languages**

*Bhojpuri*


*Boro/Bodo*


*Dogri*


*Hindi*


केंद्रीयहिंदी निदेशालय. 2014. ‘हिंदी लिखि तथा हिंदी वर्तनी का मानकीकरण नई दिल्ली’.

*Kashmiri in Devanāgarī script*


*Konkani*


*Maithili*


*Marathi*


*Nepali*


*Mewati and its dialects*

Marwari


Mewati


Shekhawati


Bagri


Dhundhari


Harauti


Wagri


*Sanskrit*


*Sindhi in Devanāgarī script*


**GUJARATI SCRIPT**

Cardona, George and Babu Suthar. “Gujarati” in Cardona and Dhanesh. 2003


Lambert H.M. *Gujarati Language Course*. Cambridge University press. 1971

**GURMUKHI SCRIPT**


**ODIA SCRIPT**


**DRAVIDIAN FAMILY**

**GENERAL OVERVIEW**


**TELUGU SCRIPT**


**KANNADA SCRIPT**


**MALAYALAM SCRIPT**


**TAMIL SCRIPT**


**Webography**

Indo-Germanisch classification of Languages:  
https://homepages.thm.de/~hg8429/sprachen.html#sprachfam2

ICANN. Devanāgarī VIP Issues Report
