IP Feedback on the Neo-Brahmi Governing Principles document

Date: 01.07.2016

# General Point on Presentation

To facilitate review and the ability to provide comments in context, the Integration Panel usually prefers to receive draft documents as Word document or in some other editable format, if practical. In some cases, where documents rely on particular fonts, a PDF version may ensure that what the reader sees is what the GP intended. However, note that the PDF/A format in particular appears to prevent annotations, even using PDF editors, and should perhaps be avoided for draft documents.

In the document as submitted, the table of contents did not match the actual structure and contents of the document. Our comments here are structured according to the sections of the document as it was delivered.

## PRELIMINARIES .

### "Although Script is favoured over language".

This passage is imprecise. A better way to put it would be: "While each LGR is intended to be valid for an entire script, the Generation Panel will consider the code point repertoire and other requirements based on individual languages."

This should be followed up by a section 6 (to be added) on "Methodology". This (and further additional sections which are recommended for addition) will be listed in their numerical places below.

### “For clarity's sake, the chart below sums up the script-language association among the scripts…”

The document, in Preliminaries, says "the GP will initially go language-wise". Yet the writers do not explain how they have reached their list of qualifying languages.

By contrast, Scriptsource [201]

http://scriptsource.org/cms/scripts/page.php?item\_id=script\_detail&key=Deva

lists 184 languages that use Devanagari, which indicates that the 11 languages given here are a result of a certain selection process, the principles of which should perhaps be covered in the suggested “Methodology” section.

Likewise Scriptsource

http://scriptsource.org/cms/scripts/page.php?item\_id=script\_detail&key=Beng

lists 39 for Bengali script, as against the 3 listed here.

(These lists also overlap, showing that some languages are written with more than 1 script.)

Similar issues are likely to affect the 7 other scripts which are accepted.

The IP expects better accounting of languages - even those that are ruled out. Please provide a justification (positive or negative) for the selected subset of languages the GP wants to support / feels must be supported. Once that task is done, the fact that a supported language needs a code point for everyday writing is all the justification that said code point will require.

## GUIDING PRINCIPLES

## Environment Limitations

## i. The Unicode Chart:

## ii. IDNA Protocol:

## iii. Maximal Starting Repertoire:

In general, the Guiding Principles here all seem correct and compatible with the ICANN Generation and Integration Panels’ Principles. In particular, the "Environment Limitations", with a key role for the Maximal Starting Repertoire (MSR), are likely to be useful in setting conceptual bounds for new members of their GP not yet oriented to our framework.

The governing principles should arguably refer explicitly to the most recent version of the MSR, so as to rule out the marginal possibility that a newer version may introduce changes that might affect them.

There follow some suggested additions.

## 2. Punctuation Marks:

## 3. Symbols and Abbreviations:

## 4. Rare and Obsolete Characters:

## 5. Stress Markers of Classical Sanskrit and Vedic:

The content of these sections is generally good, even if the Table of Contents seems to be lacking some entries, and its numbering is odd.

More importantly, the document is incomplete, lacking any consideration of Variants and / or Whole Label Evaluation rules (WLE), The following more general sections should be added:

**6. Methodology**

<TBD – contents to be created>

This section should describe in more detail the methodology used by the Neo Brahmi Generation Panels in constructing a repertoire by following the Inclusion Principle laid out in [101].

The Procedure that governs the work of the Generation Panel (GP) lays down that only code points that are actively selected by the GP will be included in the LGR. For each script's LGR, the GP will investigate the languages that use this script and define what code points are used by these languages --- while adhering to the restrictions laid out earlier. The set of code points to be included in the LGR will then be the superset of the required code point repertoires for the languages.

The first task, therefore, will be defining which language(s) to support for a given script. This choice matters for two reasons. One is in deciding the repertoire, as certain languages may require different code points. The other is in ensuring that any variant relations or WLE rules defined work equally well for all supported languages.

For this reason, the GP will need an explicit criterion for determining the set of supported languages. Presumably, the languages’ degree of prominence will be justified by reference to an indicator of language vitality such as EGIDS (Expanded Graded Intergenerational Disruption Scale, defined by Joshua Fishman, and operationalized in the SIL Ethnologue [202]), to be linked with the languages’ actual use of a given script or scripts. The Generation Panel will make sure to not bias their selection of supported languages against smaller or minority languages - as long as their written use in everyday life generally matches the conditions described in the MSR [203].

The Generation Panel will document, for each code point, at least one supported language that requires the code point for everyday writing (which thus represents a sufficient condition for selection). There is no need to document "all" languages’ use of the code point.

**7. Variants**

<TBD – contents to be created>

This section should describe the governing principles for assessing the need for both in-script and cross-script variants.

For cross-script variants, if a selection of code points in two scripts is of identical or very close to identical appearance, it would be possible to register a label in one script and make it look like a label in another script. Assigning cross-script variants would ensure mutual exclusion of such labels. This approach is preferred for cases where code points are indistinguishable in normal circumstances and may have additional, systematic or historical relations. Cases of more limited or accidental similarity would best be handled outside the LGR.

Sometimes, the forms which look alike are combining marks only. As these cannot be used without letters to combine with, it is not possible to create full labels using combining marks only. Therefore, no cross-script variants would be needed in this situation.

As for in-script variants, the Guiding Principles should discuss, in broad terms, whether variants are expected to exist and if so, under what circumstances. To this end, the discussion from the VIP study group might be a useful input to discussion (200).

Note that in the Khmer script, it was found that the appearance of certain code points was identical when they appeared in certain combined sequences, but not when standing alone. From the perspective of variants, attention must be paid to the possibility of such cases.

**8. Whole Label Evaluation (WLE) rules**

<TBD – contents to be created>

This section should describe the general guiding principles for devising Whole Label Evaluation rules.

The expectation from the Integration Panel is that such rules will be broadly similar to the rules devised for Khmer (in public comment) and Lao (in preparation).

Their purpose is to ensure that all elements in an akshara (fully specified syllable) occur in a standard order (and only in that standard order) so that no structurally malformed syllables are permitted within a label.

This is especially important in cases where rendering systems might fail to render a mal-formed syllable in a predictable way, leading to potential user confusion. (Note that it is never safe to assume that all registrants will be competent users of the scripts for which they submit applications.)

In addition, for some scripts, alternative orderings of the elements of an akshara exist that could lead to similar or identical presentation on some systems. To avoid ambiguities, WLE rules may be used to enforce a canonical ordering or, alternatively, blocked variants might be defined for the sequences in question. If there is no possibility of ambiguous ordering in a given script, this should be noted as a fact found by the Generation Panel.

Bear in mind that the language of a label is not known at the time a WLE rule is evaluated, and, in fact, labels may not necessarily be in any particular language. Therefore, WLE rules should only reflect restrictions inherent to any use of the script. Where multiple languages share a script, care must be taken to ensure that the WLE rules are not biased towards a few prominent languages.

**9. References**

Finally, it is always a good idea to have citations for documents referred to in the text and a references section listing them, as in the following examples:

1. The Unicode Consortium. The Unicode Standard, Version 6.3.0, (Mountain View, CA: The Unicode Consortium, 2013. ISBN 978-1-936213-08-5

100. The Unicode Consortium. The Unicode Standard, Version 8.0.0, (Mountain View, CA: The Unicode Consortium, 2015. ISBN 978-1-936213-10-8), Chapter 16: Southeast Asia, section 16.4: Khmer, pages 616-618. <http://www.unicode.org/versions/Unicode8.0.0/ch16.pdf>

101. Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013) <http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf>

102. Integration Panel “Requirements for LGR Proposals from Generation Panels” available online at <https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-20150424.pdf>

103. The Unicode Consortium, Unicode Character Database, available online as <http://www.unicode.org/Public/UCD/latest/>

200. Devanāgarī Vip Team Issues Report: <https://archive.icann.org/en/topics/new-gtlds/devanagari-vip-issues-report-03oct11-en.pdf>

201. Scriptsource: <http://scriptsource.org/cms/scripts/>

202. Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the World, Seventeenth edition.* Dallas, Texas: SIL International. Online version available as <http://www.ethnologue.com>.

203. Internet Corporation for Assigned Names and Numbers, "[Maximal Starting Repertoire - MSR-2-Overview and Rationale”, (Los Angeles, California:](https://www.icann.org/en/system/files/files/msr-2-overview-14apr15-en.pdf) April 2015)
https://www.icann.org/en/system/files/files/msr-2-overview-14apr15-en.pdf