

# Proposal for the Generation Panel for the Sinhala Script Label Generation Ruleset for the Root Zone

## 1 General information

Sinhala belongs to the Indo-European language family with its roots deeply associated with Indo-Aryan sub family to which the languages such as Persian and Hindi belong. Although it is not very clear whether people in Sri Lanka spoke a dialect of Prakrit at the time of arrival of Buddhism in the island, there is enough evidence that Sinhala evolved from mixing of Sanskrit, Magadi (the language which was spoken in Magada Province of India where Lord Buddha was born) and local language which was spoken by people of Sri Lanka prior to the arrival of Vijaya, the founder of Sinhala Kingdom. It is also surmised that Sinhala had evolved from an ancient variant of Apabramsa (middle Indic) which is known as 'Elu'. When tracing history of Elu, it was preceded by Hela or Pali Sihala.

Sinhala though has close relationships with Indo Aryan languages which are spoken primarily in the north, north eastern and central India, was very much influenced by Tamil which belongs to the Dravidian family of languages. Though Sinhala is related closely to Indic languages, it also has its own unique characteristics: Sinhala has symbols for two vowels which are not found in any other Indic languages in India: 'æ' (ඇ) and 'æ:' (ඈ).

Sinhala script had evolved from Southern Brahmi script from which almost all the Southern Indic Scripts such as Telugu and Oriya had evolved. Later Sinhala was influenced by Pallava Grantha writing of Southern India. Since 1250 AD, the Sinhala script had remained the same with few changes. Although some scholars are of the view that the Brahmi Script arrived with the Buddhism, Mahavamsa (Great Chronicle) speaks of written language even right after the arrival of Vijaya. Archeologists had found pottery fragments in Anuradhapura Sri Lanka with older Brahmi script inscriptions, which had been carbon dated to 5th century BC. The earliest artifacts with Brahmi Script found in India had been dated to 6th Century BC in Tamil Nadu though most of Brahmi writing found in India had been attributed to emperor Ashoka in the 3rd century BC.

Sinhala letters are round-shaped and are written from left to right and they are the most circular-shaped script found in the Indic scripts. The evolution of the script to the present shapes may have taken place due to writing on Ola leaves. Unlike chiseling on a rock, writing on palm leaves has to be more round-shaped to avoid the stylus ripping the Palm leaf while writing on it. When drawing vertical or horizontal straight lines on Ola leaf, the leaves would have been ripped and this also may have influenced Sinhala not to have a period or full stop. Instead a stylistic stop which was known as 'Kundaliya' is used. Period and commas were later introduced into Sinhala script after the introduction of paper due to the influence of Western languages.

(Excerpt from the Introduction section of "Numerations in the Sinhala Language", Authored by Harsha Wijayawardhana and edited by Aruni Goonathilleke, Published by ICTA, in October 2009)

## 1.1 Script for which the panel is to be established

ISO 15924 – Code: Sinh

ISO 15924 – Number: 348

ISO 15924 - English Name: Sinhala

Native Name: සිංහල

Sinhala Unicode range: U+0D80 – U+0DFF

The Sinhala Generation Panel will consider the subset of code points within this range which have been shortlisted in latest available version of Maximal Starting Repertoire.

## 1.2 Principal languages using that script

Sinhala script is used to write Sinhala (sin) language, which is one of the official languages of Sri Lanka. In addition, it is used to write Pali (pli) and Sanskrit (san) languages in Sri Lanka.

## 1.3 Geographic territories or countries with significant user communities for the script

Sinhala script is used on the Island of Sri Lanka (predominantly in the south) and Sinhala Diaspora in Middle East (Saudi Arabia, Kuwait, Qatar, and UAE), Britain, USA, Australia and Canada.

## 1.4 Are there related scripts? If so, which ones?

The scripts covered by the Neo-Brahmi GP are related to Sinhala script. Based on an initial analysis, Sinhala GP has found script similarity with Malayalam, Kannada and Telugu scripts. In addition, Myanmar script is also related. Sinhala GP will investigate cross-script variants with these scripts further.

# 2 Proposed initial composition of panel

The Sinhala Generation Panel is composed of individuals actively involved in research and development of Sinhala language computing, its standardization and the domain name system, including the internationalized domain names. The panel also has experts in Sinhala linguistics, font development and community members. The panelist come from government, academia, private sector, ccTLD registry and the community.

## 2.1 List prospective chair and members

Name	Role	Expertise	Organization	Relevant Experience
Ruvan Weerasinghe	Co-chair	Unicode /Policy Technical	University of Colombo School of Computing (UCSC)	Involved in Standardization of SLS1134, Sinhala Unicode Implementation, and NLP research in Sinhala
Harsha Wijayawardhana	Co-chair	Unicode /Standardization Technical	Theekshana	Involved in Standardization of Sinhala SLS1134, Standardization of Sinhala Numerals in Unicode, Sinhala font creation and Sinhala implementation work

Rajeewa Abegunaratna	Member	Community Member	Lanka Education and Research Network	Network Operations, Community Engagements, Internet Development, and Event Coordination
Champika Wijayatunga	Member	DNS/Cyber Security	ICANN	Internet Identifier Systems, DNS Operations and Management, Cybersecurity, Internet Policy Development and ICANN Community Engagement
Pathum Egodawatta	Member	Community /Unicode	Textual	Type designer and font engineer, Researcher on Sinhala typography, Consultant on typography of South Asian scripts, and Contributor to free software typography and font development tools
Viraj Welgama	Member	Unicode /Technical	UCSC	Computational linguist, with over 10 years of experience in local language processing
Chamila Liyanage	Member	Linguist	UCSC	Linguist and involved in NLP Research
Prof. J.B. Disanayaka	Member	Linguist	Professor Emeritus, University of Colombo	Involved in encoding of Sinhala in Unicode and expert in Sinhala Language
Sagarika Wickramasekera	Member	Community	ISOC	Network specialist
Nimasha Dilshani	Member	Linguist	University of Moratuwa	Works as a language expert at SiTa, Sinhala Tamil translation project at university of Moratuwa. Worked as an interpreter at Indian high commission (2016 & 2017) Completed one month training of natural language processing at IIIT - Hyderabad, India (in 2017)
Rev. Mettavahiri	Member	Unicode /Community	Dhammavahiri	

			Foundation	
Chamara Disanayake	Member	DNS/Registry Operations	LK Domain Registry	Worked at LK Domain Registry and involved .lk equivalent Sinhala and Tamil labels (IDN)
Aruni Goonetilleke	Member	Policy	ICT Agency	Coordinated localization efforts of Sinhala and Tamil in Sri Lanka

### 3 Work plan

#### 3.1 Suggested Timeline, listing of significant milestones

Activity	Description	Start Date	Duration
Develop Principles	Principles to be used to determine valid code points, variants and labels	11/12/17	4 weeks
Determine Code Points	Select the code points from MSR which are needed for Root Zone LGR	17/01/18	2 weeks
Determine (any) Variants	From the codes points selected, determine if the end-user may confuse two code points	31/01/18	4 Weeks
Determine Label Rules	Determine if there are any label level constrains on the use of selected code points	28/02/18	4 Weeks
Write Proposal and Create XML	Write up the Root Zone LGR proposal, including references to each code point included, why variants needed and details of label rules developed + XML file	28/03/18	6 Weeks
Finalize	Get feedback from integration panel and the community, Finalize and Submit	9/05/18	6 Weeks
Submit	Get Public Comments, Finalize and Submit	20/06/18	2 Weeks

### 3.2 Proposed schedule of meeting and teleconferences

The panel aims to conduct face to face interactions, along with the email list provided by ICANN for sharing administrative information and discussions. The panel outputs will also be updated on its wiki page, provided by ICANN.

### 3.3 Anticipated need for logistic support, such as mailing list?

The panel anticipates additional support from ICANN for the following:

- Sinhala GP mailing list and wiki page
- Online meetings, as planned
- Public consultation organized in Sri Lanka
- Travel support to ICANN meeting for panel representative to interact with the community