

Event: Reference LGRs for Indonesian Scripts (Balinese, Javanese, Pegon)

Date(s): 2-3 December 2023

Meeting Agenda

2 Dec 2023

- Opening Speeches
- LGR Overview Presentation
- Balinese, Javanese, Pegon LGR Presentation
- Visit Museum Wiswakarma
- Balinese keyboard presentation

3 Dec 2023

- Balinese Wikimedia Presentation
- Balinese script lonta workshop
- LGR Tool and XML overview training
- Discuss the detail and next steps for the Balinese, Javanese, Pegon LGRs

Attendees

- Balinese Community
 - I Nyoman Arya Wibawa, Dean FIB
 - Prof. Dra Ni Luh Watiniasih, Dead FMIPA
 - Dewa Ayu Carma Citrawati (Citra), Wikimedia
 - Cokorda Rai Adi Pramatha
- Javanese Community
 - Arif Budiarto
- Pegon Community
 - Diaz Nawaksara
- PANDI
 - John Sihar Simanjuntak, CEO
 - Ghea
 - Mira
- ICANN Org
 - Pitinan Kooarmornpatana
 - Sarmad Hussain

Summary Notes

Draft reference LGRs for Balinese Script ,Javanese script, Pegon (Javanese language using Arabic script were discussed.

In general updates required are:

1. In the Background section, the information on the current usage should be added. E.g. number of people using the script, number of newspapers, publication using the script.
2. Uniformly add the Unicode Code Point value and the Unicode Code Point Name along with the Glyph.
3. Add pronunciation of the word in examples.

Balinese script

The Balinese script is taught in school. Generally students learn the script for 12 years during their primary school and high school. After school, people can read Balinese script in general, but may already have forgotten how to write.

In 2013, only 50% of the total 4.2 million Balinese population knew the Balinese script. With the effort from the government to revive the script it is likely to have more usage in the future. Currently there are signs on the official buildings and big companies and restaurants. Hardware keyboard has been developed and used in the school, software keyboard is being developed. With these input mechanisms becoming available, the content in Balinese script is likely to be more.

The community planned to use Balinese script ref. LGR development process as a model for the remaining scripts.

Based on the discussions, followings are modifications needed for Balinese script LGR:

1. Add rules to only allow a maximum of two consecutive consonants.
2. Cross-script variant checks with Sinhala, Myanmar, Javanese should be added in the proposal document.
3. The rules of 1B02, 1B03 can follow consonants.
4. Code point 1B25 and 1B1F looks the same, however, only 1B25 is encoded in the keyboard, therefore, the repertoire should only include 1B25.
5. It was discussed that in case of Zero Width character needed within the word, the Zero Width Non-Joiner (U+200C) should be used instead of the Zero Width Space (U+200B) However Based on the conservatism principle, the first version of Balinese LGR will not include any Zero Width character.

Additional dataset from the modern usage should be provided. It will be useful to have two lists

1. The word list from dictionary
2. The word list from of names

Javanese script

Based on the discussions, followings are modifications needed for Javanese script LGR:

1. The information that Microsoft Locale supports Javanese script should be included in the background section.
2. The code point A9B4 needs further clarification whether it is in repertoire or not, if it is in, variant definition may be required with A9B5
3. Rules for combining marks e.g. section 6.3, 6.4, 6.5 needs to be reviewed. Currently it states what can follow the code point, but missing the information of what code point can the letter follow.

Besides the LGR discussion, the topic related to the category of the Javanese script as categorized by the Unicode consortium was discussed.

It was shared that Sarmad attended the Unicode Consortium Technical Conference in November 2023 and requested the Unicode to publish the criteria for updating the category of script. The Unicode Consortium will look into the request. In the meantime, the community is encouraged to compile the supporting data on the current usage of the script and create a regular Unicode proposal.

Pegon script

Pegon based on the Arabic script with additional code points required for Javanese, Sudanese, Madura languages. The community will first focus on the LGR of Javanese language using Arabic script. The language tag will be 'jv-Arab'

Many of the diacritics code points required for Pegon are included in the Unicode version 14.0.0. And the LGR tool should be updated to support.

In addition to the LGR discussion, the IDN ccTLD Fast Track was discussed. It was shared that as the requirement for the root zone, the requested label will need to be a valid label based on the Arabic script RZ-LGR, that is the Arabic script diacritic can not be used. It was shared that other supporting documentation as per IDN ccTLD Fast Track is required, e.g. the support that Pegon is used as an administrative language or script by the government.

Action Items

#	Task	Owner
1	Update the Balinese script, Javanese script, Pegon script LGR document as discussed.	PANDI and Script Experts
2	Provide additional wordlist to cover current usage e.g. words from the news articles including words and names.	PANDI and Script Experts
3	Add the requirement to support newer version of the Unicode version in the LGR Tool	ICANN org