Introduction to LGR Toolset

- Label Generation Rulesets (LGRs) specify metadata, code point repertoire, variant rules and Whole Label Evaluation (WLE) rules to generate labels.

- **RFC 7940** describes how LGR can be specified using XML, a machine readable format.

- LGR can be used to generate domain name labels for use in the internet’s root zone and other levels.

- LGR Toolset can be used to:
  - create an LGR
  - view LGR as an HTML webpage
  - merge multiple LGRs into a single LGR
  - validate single label or multiple labels using an LGR
  - determine cross-script variants of labels using a merged LGR
  - manage LGRs by comparing or combining them
  - review possible impact of a new or a revised LGR on existing labels
  - harmonize multiple LGRs
Availability of LGR Toolset

- LGR Toolset is available with the following disclaimer:

  THIS SOFTWARE IS PROVIDED BY ICANN AND CONTRIBUTORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL ICANN OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Online deployment
  - Visit https://lgrtool.icann.org/

- Open source package(s) release with BSD license
  - Released at github: lgr-core, lgr-django, munidata, picu

- For queries or feedback
  - Email to IDNProgram@icann.org

- For further details, visit the LGR Toolset webpage or www.icann.org/idn
Agenda

1. Importing an existing LGR in a toolset (pp. 9-12)
2. Importing multiple LGRs in a toolset (pp. 13-16)
3. Creating an LGR via toolset (pp. 17-43)
4. Validating LGR (pp. 44-48)
5. Viewing LGR as XML and HTML (pp. 49-60)
6. Validating a label using an LGR (pp. 61-66)
Agenda (cont.)

7. Reviewing impact on existing labels by revising an existing LGR (pp. 67-72)

8. Reviewing impact on existing labels by introducing a new LGR (pp. 73-78)

9. Validating multiple labels using LGR(s) (pp. 79-85)

10. Viewing cross-script variants of labels (pp. 86-91)

11. Comparing LGRs (pp. 92-96)

12. Harmonizing multiple LGRs (pp. 97-100)
Agenda Details

- **Import or Load LGR**
  - Import LGR
  - Select LGR with validating repertoire

- **Import Multiple LGRs**
  - Import LGR
  - Select multiple LGRs with validating repertoire
  - Specify name of the merged LGR

- **Create LGR**
  - Create LGR
  - Enter LGR Details
  - Add Code Points
  - Expand Ranges
  - Add References
  - Define Meta Data
  - Define Classes, Rules & Actions
  - Define Code Point Properties
  - Download LGR

- **Validate LGR**
Agenda Details (cont.)

- View LGR as XML and HTML
- Validate a label
  - Validate a label using a single LGR
  - Validate a label using a merged LGR
- Review Impact on Existing Labels by Revising an Existing LGR
  - Select “Diff labels of two LGRs”
  - Enter Details
  - Email Notification
  - Download Results
- Review Impact on Existing Labels by Introducing a New LGR
  - Select “Get collisions in a list of labels”
  - Enter Details
  - Email Notification
  - Download Results
- Validate Multiple Labels using a single or a merged LGR
  - Select “Generate disp. annotations”
  - Enter Details
  - Email Notification
  - Download Results
Agenda Details (cont.)

- **View Cross-script Variants of Labels**
  - Select “Cross-script variants”
  - Enter Details
  - Email Notification
  - Download Results

- **Compare LGRs**
  - Compare LGRs
  - Select Union, Intersection or Difference
  - With merged LGRs, “Diff” chosen by default

- **Harmonize multiple LGRs**
  - Select “Harmonization”
  - Enter Details
  - Email Notification
  - Download Results
Import or Load LGR
Import or Load LGR

To start by using an existing LGR file in XML format, click on the “Import” button.
Select LGR with Validating Repertoire

1. To import or load an existing LGR in XML format, click on “Choose Files”

2. Choose the “Validating repertoire” from the given options

3. Click on the “Import” button
The screen will look like this after successful import of existing LGR file in XML format.
Import Multiple LGRs
Import or Load LGRs

Welcome to the LGR (Label Generation Ruleset) Tool

This application provides a convenient interface for browsing and editing LGRs conforming to the Representing Label Generation Rulesets using XML specification.

Previously loaded LGR file(s)
No LGR has been previously loaded.

Create a new LGR file or import an existing one

- Import an existing XML file

⚠️ Note that importing large LGR files may take significant time to load on your browser.

- Start with a New blank XML file

Start from a built-in LGR
The following LGRs are pre-installed in the system. You may use them as a starting point for your own LGR. To do so, just click on it to make a copy that you can then edit.

- Open Sample-French

Remember to save your work regularly by downloading a copy of the XML file.

Please send any feedback to support@viagenie.ca.
Select LGRs with Validating Repertoire

1. To import or load multiple LGRs in XML format, click on “Choose Files” and select multiple files.

2. Choose the “Validating repertoire” from the two given options.

3. Enter a name for the set of LGRs.

4. Click on the “Import” button.
Imported LGRs As a Merged LGR

The screen will look like this after successful import of multiple LGR files in XML format.
Create an LGR
To start by creating an LGR file in XML format, click on the “New” button.
Enter LGR Details

1. Write name for the LGR being created

2. Select Validating repertoire from the given options for the “New” LGR. MSR for RZ-LGR and IDNA version for second level LGRs

3. Click on the “Create” button
Click on “Add code point(s)” to add code points to the newly created LGR.
To add code points, there are four different ways:

1. Add code points one by one
2. Add code points by giving a range
3. Add code points by specifying a script
4. Add code points from a file
Add Code Points

1. Write the code point to be added. The code point value or the actual character can be entered.

2. Click on the “Add Code Point” button.
Add Code Points

1. Add the first code point of the range
2. Add the last code point of the range
3. Click on “Next” button

Either code point value or the character can be added.
Add Code Points

1. Select a script from the given options
2. Select a Validating Repertoire from the given options
3. “Manual import” is optional – allows checking each code point in the script before adding to the LGR
4. Click on “Next” button
Add Code Points

1. Select the file containing code points
2. Select the file type from the given options
3. “Manual import” is optional – allows checking each code point in the file before adding to the LGR
4. Click on the “Next” button
Expand Ranges

To expand all the code points in all the ranges of the LGR, click on “Expand range(s)”.

To expand all the code points in this range, click on “Expand range”.

[Diagram of LGR Tool/new-lgr interface with highlighted actions for expanding ranges]
Add References

1. Add “Reference id”
2. Add comments
3. Add detailed reference
4. Click on the “Add” button

Click on “References” tab to add references to the LGR
Add References

“Existing references” tab shows the added references
### Define Meta Data

Click on “Meta data” tab to add meta information about the LGR.

<table>
<thead>
<tr>
<th>Code points</th>
<th>References</th>
<th>Meta data</th>
<th>Tags</th>
<th>Rules</th>
</tr>
</thead>
</table>

| **Version** | 1 |
| **Version comment** |  |
| **Date** | 2018-08-08 |
| **Language** | und-Khmer |
| **Scope** | example |
| **Scope type** | domain |
| **Validity start** |  |
| **Validity end** |  |
| **Unicode version** | 10.0.0 |
| **Description** |  |
Define Meta Data

Second half of the “Meta data” tab

- **Add any text for describing the LGR**
- **Select “Description type” from the given options**
- **Select “Validating repertoire” from the given options**
- **After filling every detail, click on the “Save” button**

### LGR Tool / new-lgr

<table>
<thead>
<tr>
<th>Scope type</th>
<th>domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity start</td>
<td></td>
</tr>
<tr>
<td>Validity end</td>
<td></td>
</tr>
<tr>
<td>Unicode version</td>
<td>10.0.0</td>
</tr>
</tbody>
</table>
| Description      | <h1>Label Generation Rules for Khmer</h1>
| Description type | text/html |
| Validating repertoire | msr-3-wie-rules-28mar18-en |

---

[Image of the LGR Tool interface]

---

[ICANN Logo]
Click on “Tags” tab to view all tag names and associated code points.
Define Classes, Rules & Actions

Click on “Rules” tab to add certain rules to the LGR.

Click on “New class” button to add classes to the LGR.

Click on “New rule” button to add rules to the LGR.

Click on “New action” button to add actions to the LGR.
**Define Classes**

1. Add classes in the relevant box
2. Click on the “Save” button
Define Rules

1. Add the rule in the relevant box
2. Click on the “Save” button
Define Actions

1. Add action in the relevant box
2. Click on the “Save” button
Apply Batch Action

1. Select multiple code points to apply the batch action
2. Select the type of batch action
Apply Batch Action – Add WLE

1. Add when-rule / not-when-rule from the list of WLE rules
2. Click on “Next” button
Apply Batch Action – Add Tags

1. Type new tags, separated by a space, or select the existing ones

2. Click on “Next” button
Define Code Point Properties

Click on the "See code point" button to add the code point details/properties
Define Code Point Properties - 1

1. Add variant for the code point
2. Click on the “Add variant button”
3. Add details for the added variant – type, comments, when-rule and not-when rule
4. Add tags for the code point
5. Add when-rule/not-when rule from the list of rules it provides
6. Add any description for the code point
Define Code Point Properties - 2

7. Click on “Save variants, tags, context rules and comment” button

Note: If the added code point is not in the repertoire, the system will automatically add the out-of-repertoire mapping.

8. Click on the “Edit” button to add references to the code point

To delete the code point & its details, click on the “Delete code point” button
Define Code Point Properties - 3

The screen looks like this after successfully defining code point properties.

If the variant definitions are not symmetric, this button will show up. Click “Populate variants” to automatically populate variant mappings.
Click on “Output” and select “Download” from the given options to download the created LGR. Always download before closing the browser as the LGRs are not saved on the server.
Validate LGR
Validate LGR

Click on “Validate LGR” button to validate LGR and get a summary of the entire LGR. It is important to note that this function should be used for checking an LGR before use, e.g. when it is created or imported.
Validate LGR

Summary of checks performed, including symmetry and transitivity
Second half of the “Validate LGR” output

Summarized LGR
View LGR as XML and HTML
Click on “Output” and select “View XML” from the given options to get an XML view of the LGR.
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
<lgr xmlns="urn:ietf:params:xml:ns:ldr-1.0">
  <meta>
    <version comment="Thai Script Root Zone LGR Version 6.9">2</version>
    <date>2017-05-25</date>
    <unicode-version>6.3.0</unicode-version>
    <language>th</language>
    <scope type="domain" /></scope>
  </meta>
  <![CDATA[
  <h1>Label Generation Rules for the Thai Script</h1>
  <p>This file contains Label Generation Rules (LGR) for the Thai script as would be appropriate for the Root zone. For more details on this LGR see "Proposal for a Thai Script Root Zone LGR [Proposal]"</p>
  <h2>Overview</h2>
  <p>In addition to the 68 code points according to Section 5 “Repertoire” in [Proposal], three sequences have been defined. The sequence U+0E4D U+0E32 was defined to replace the disallowed U+0E33 (THAI CHARACTER SARA AM) and to facilitate implementation of WLE rule <b>follows-consonant-tone</b> as a context rule. The other two sequences were defined to restrict U+0E45 (THAI CHARACTER LAKKHYANGYO) from appearing in any context other than these sequences. Accordingly, while U+0E45 is not listed by itself it brings the total of distinct code points to 69.</p>
  <h2>Variants</h2>
  <p>According to Section 6 "Variants", in "[Proposal]", this LGR defines no variants.</p>
  <h2>Character Classes</h2>
  <p>The Thai Script is an abugida in which consonant-vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel, tone mark or diacritic notation are secondary. It is written with the combining marks stacked above or below the base consonant, like diacritics in European languages. However, although the concepts are quite similar, the implementations are significantly different.</p>
  <p>There are 44 characters that are classified as consonants, code points from this subset have been given the tag "cons".</p>
  <p>The 18 vowel symbols pronounced after a consonant are non-sequential: they can be located before (lv), after (fv), above (av) or below (bv) the consonant, or in a combination of these positions, code points from this subset have been given the tag "fv1", "fv2", "fv3", "av", "bv", "lv". There are three code point sequences defined that include vowels. (Code point sequences do not carry tag values; instead, for code point sequences the subset values are indentified in comments).</p>
  <p>There are 5 phonemic tones: mid, low, falling, high, and rising. These 5 tones are represented by 4 tone marks plus the absence of a mark. Code points from this subset have been given the tag "tone".</p>
  <p>There are 3 diacritic symbols that have been included here and given the tag "ad". They differ in their frequency and purpose of usage. See also the discussion in section 5.4 in [Proposal].</p>
  ]]>
</lgr>
```
View LGR As HTML

Click on “Output” and select “HTML Output” from the given options to view the HTML output of the LGR.
HTML Output - Meta Data - 1

**Merged-Lgr-1**

<table>
<thead>
<tr>
<th>Meta data in LGR</th>
</tr>
</thead>
</table>

| LGR Version       | 2          |
| Date              | 2017-10-21 |
| Language(s)       | und-Ethi   |
|                   | und-Laoo   |
|                   | und-Thai   |
| Scope(s)          | domain: .  |
| Unicode Version   | 6.3.0      |

This document is mechanically formatted from the XML file for the LGR. It provides additional summary data and explanatory text. The XML file remains the sole normative specification of the LGR.

**Table of Contents**

1. Description
2. Repertoire
3. Variant Sets
4. Classes, Rules and Actions
   4.1 Character Classes
   4.2 Whole label evaluation and context rules
   4.3 Actions
5. Table of References

**Description**
Label Generation Rules for the Thai Script

Overview

This file contains Label Generation Rules (LGR) for the Thai script as would be appropriate for the Root zone. For more details on this LGR see "Proposal for a Thai Script Root Zone LGR [Proposal]."

Repertoire

In addition to the 68 code points according to Section 5 “Repertoire” in [Proposal], three sequences have been defined. The sequence U+0E4D U+0E32 was defined to replace the disallowed U+0E33 (THAI CHARACTER SARA AM) and to facilitate implementation of WLE rule follows-consonant-tone as a context rule. The other two sequences were defined to restrict U+0E45 (THAI CHARACTER LAKKHANGYAO) from appearing in any context other than these sequences. Accordingly, while U+0E45 is not listed by itself it brings the total of distinct code points to 69.

Variants

According to Section 6 "Variants", in "[Proposal]", this LGR defines no variants.

Character Classes

The Thai Script is an abugida in which consonant–vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel, tone mark or diacritic notation are secondary. It is written with the combining marks stacked above or below the base consonant, like diacritics in European languages. However, although the concepts are quite similar, the implementations are significantly different.

There are 44 characters that are classified as consonants, code points from this subset have been given the tag "cons".

The 18 vowel symbols pronounced after a consonant are non-sequential: they can be located before (lv), after (fv), above (av) or below (bv) the consonant, or in a combination of these positions, code points from this subset have been given the tag "fv1", "fv2", "fv3", "av", "bv", "lv". There are three code point sequences defined that include vowels. (Code point sequences do not carry tag values; instead, for code point sequences the subset values are identified in comments).

There are 5 phonemic tones: mid, low, falling, high, and rising. These 5 tones are represented by 4 tone marks plus the absence of a mark. Code
**Repertoire**

**Summary**

<table>
<thead>
<tr>
<th>Number of elements in repertoire</th>
<th>434</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ranges in repertoire</td>
<td>0</td>
</tr>
<tr>
<td>Number of code point sequences</td>
<td>4</td>
</tr>
</tbody>
</table>

**Repertoire by Code Point**

The following table lists the repertoire by code point (or code point sequence). The data in the Script and Name column are extracted from the Unicode character database. Where the comment in the original LGR is equal to the character name, it has been suppressed.

For any code point or sequence for which a variant is defined, the link to the associated variant set, or if mapped to itself, the variant type of that mapping is provided in the Variants column.

<table>
<thead>
<tr>
<th>#</th>
<th>Code Point</th>
<th>Glyph</th>
<th>Script</th>
<th>Name</th>
<th>Tags</th>
<th>Required Context</th>
<th>Variants</th>
<th>Comment</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>U+0E01</td>
<td>น</td>
<td>Thai</td>
<td>THAI CHARACTER KO KAI</td>
<td>Thai,und-Thai-cons</td>
<td></td>
<td></td>
<td></td>
<td>[5], [100], [101]</td>
</tr>
<tr>
<td>2</td>
<td>U+0E02</td>
<td>ง</td>
<td>Thai</td>
<td>THAI CHARACTER KHO KHAI</td>
<td>Thai,und-Thai-cons</td>
<td></td>
<td></td>
<td></td>
<td>[5], [100], [101]</td>
</tr>
<tr>
<td>3</td>
<td>U+0E03</td>
<td>ช</td>
<td>Thai</td>
<td>THAI CHARACTER KHO KHUAT</td>
<td>Thai,und-Thai-cons</td>
<td></td>
<td></td>
<td></td>
<td>[5], [100], [101]</td>
</tr>
<tr>
<td>4</td>
<td>U+0E04</td>
<td>ฑ</td>
<td>Thai</td>
<td>THAI CHARACTER KHO KHWAI</td>
<td>Thai,und-Thai-cons</td>
<td></td>
<td></td>
<td></td>
<td>[5], [100], [101]</td>
</tr>
</tbody>
</table>

---

**Repertoire section in LGR**
HTML Output - Variant Sets

Variant Sets

Summary

<table>
<thead>
<tr>
<th>Number of variant sets</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Largest variant set</td>
<td>4</td>
</tr>
<tr>
<td>Ordinary Variants by Type</td>
<td>blocked (98)</td>
</tr>
</tbody>
</table>

The following tables list all variant sets defined in this LGR, except for singleton sets. Each table lists all variant mapping pairs of the set; one per row. Mappings are assumed to be symmetric: each row documents both forward (→) and reverse (←) mapping directions. In each table, the mappings are sorted by Source value in ascending code point order; shading is used to group mappings from the same source code point or sequence.

Where the type of both forward and reverse mappings are the same, a single value is given in the Type(s) column, otherwise the types for forward and reverse mappings, as well as comments and references are listed above one another.

A mapping where source and target are the same is reflexive. Variant sets consisting of only a single reflexive mapping are not shown as a set. Instead, the variant type of the mapping is listed in the Variants column of the Repertoire by Code Point table. Reflexive mappings that are part of a larger set are indicated with a “=”. In any LGR with variant specifications that are well behaved, all members within each variant set are defined as variants of each other; the mappings in each set are symmetric and transitive; and all variant sets are disjoint.

Common Legend

Source: Source of the mapping pair.
Target: Destination of the mapping pair.
Glyph: The shape displayed for source or target depends on the fonts available to your browser.
→ - forward: Indicates that variant Type, References and Comment apply to the mapping from source to target.
← - reverse: Indicates that variant Type, References and Comment apply to the reverse mapping from target to source.
↔ - both: Indicates that variant Type, References and Comment apply to both forward and reverse mapping.
= - reflexive: Indicates that variant Type, References and Comment are for a reflexive mapping where source equals target.
□ - not in LGR: Indicates that variant is not in LGR.
Type: The type of the variant mapping. There are some predefined variant types such as "allocatable" and "blocked", while others are defined specifically for each LGR.
References: One or more reference IDs (optional). A "/" separates references for reverse / forward mappings, if different.
Comment: A descriptive comment (optional). A "/" separates comments for reverse / forward mappings, if different.

Variant Set 1 — 3 Members - 3 Mappings

<table>
<thead>
<tr>
<th>#</th>
<th>Source</th>
<th>Glyph</th>
<th>Target</th>
<th>Glyph</th>
<th>Type(s)</th>
<th>References</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>U+1200</td>
<td>U</td>
<td>U+1210</td>
<td>r</td>
<td>← blocked</td>
<td>[4]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>U+1200</td>
<td>U</td>
<td>U+1280</td>
<td>r</td>
<td>← blocked</td>
<td>[4]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>U+1210</td>
<td>r</td>
<td>U+1280</td>
<td>r</td>
<td>← blocked</td>
<td>[4]</td>
<td></td>
</tr>
</tbody>
</table>
### Classes, Rules and Actions

#### Character Classes

The following table lists all top-level classes with their definition and the regular expression defining their members.

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Count</th>
<th>Members</th>
<th>References</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>und-Lao-Cf</td>
<td>Tag= und-Lao-Cf</td>
<td>14</td>
<td>{U+0E81 U+0E87 U+0E8A U+0E8D U+0E94 U+0E97 U+0E99 U+0E9A U+0E9F U+0EA1 U+0EA3 U+0EA5 U+0EA7 U+0EAA}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>und-Lao-consonant</td>
<td>Tag= und-Lao-consonant</td>
<td>27</td>
<td>{U+0E81 U+0E82 U+0E84 U+0E87 U+0E88 U+0E8A U+0E8D U+0E94 U+0E95 U+0E96 U+0E97 U+0E99 U+0E9A U+0E9B U+0E9C ...}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>und-Lao-semi-consonant</td>
<td>Tag= und-Lao-semi-consonant</td>
<td>1</td>
<td>{U+0EBC}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>und-Lao-tone-mark</td>
<td>Tag= und-Lao-tone-mark</td>
<td>4</td>
<td>{U+0EC8 U+0EC9 U+0ECA U+0ECB}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>und-Lao-vowel-above</td>
<td>Tag= und-Lao-vowel-above</td>
<td>7</td>
<td>{U+0EB1 U+0EB4 U+0EB5 U+0EB6 U+0EB7 U+0EBB U+0ECD}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>und-Lao-vowel-below</td>
<td>Tag= und-Lao-vowel-below</td>
<td>2</td>
<td>{U+0EB8 U+0EB9}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>und-Thai-above-vowel</td>
<td>Tag= und-Thai-above-vowel</td>
<td>5</td>
<td>{U+0E31 U+0E34 U+0E35 U+0E36 U+0E37}</td>
<td>Any above vowel</td>
<td></td>
</tr>
<tr>
<td>und-Thai-below-vowel</td>
<td>Tag= und-Thai-below-vowel</td>
<td>2</td>
<td>{U+0E38 U+0E39}</td>
<td>Any below vowel</td>
<td></td>
</tr>
<tr>
<td>und-Thai-c-av-bv</td>
<td>([:und-Thai-consonant:][:und-Thai-above-vowel:][:und-Thai-below-vowel:])</td>
<td>51</td>
<td>{U+0E01 U+0E02 U+0E03 U+0E04 U+0E05 U+0E06 U+0E07 U+0E08 U+0E09 U+0E0A U+0E0B U+0E0C U+0E0D U+0E0E U+0E0F ...}</td>
<td>Any consonant, vowel-above or vowel-below</td>
<td></td>
</tr>
<tr>
<td>und-Thai-cons</td>
<td>Tag= und-Thai-cons</td>
<td>44</td>
<td>{U+0E01 U+0E02 U+0E03 U+0E04 U+0E05 U+0E06 U+0E07 U+0E08 U+0E09 U+0E0A U+0E0B U+0E0C U+0E0D U+0E0E U+0E0F ...}</td>
<td>Any Consonant</td>
<td></td>
</tr>
<tr>
<td>und-Thai-tct</td>
<td>([:und-Thai-consonant:][:und-Thai-tone:])</td>
<td>48</td>
<td>{U+0E01 U+0E02 U+0E03 U+0E04 U+0E05 U+0E06 U+0E07 U+0E08 U+0E09 U+0E0A U+0E0B U+0E0C U+0E0D U+0E0E U+0E0F ...}</td>
<td>Any consonant or tone</td>
<td></td>
</tr>
<tr>
<td>und-Thai-tta</td>
<td>([:und-Thai-consonant:][:und-Thai-tone:][:und-Thai-sara-aa:])</td>
<td>49</td>
<td>{U+0E01 U+0E02 U+0E03 U+0E04 U+0E05 U+0E06 U+0E07 U+0E08 U+0E09 U+0E0A U+0E0B U+0E0C U+0E0D U+0E0E U+0E0F ...}</td>
<td>Any consonant, tone or sara-aa</td>
<td></td>
</tr>
<tr>
<td>und-Thai-sara-aa</td>
<td>Tag= und-Thai-sara-aa</td>
<td>1</td>
<td>{U+0E32}</td>
<td>SARA AA</td>
<td></td>
</tr>
<tr>
<td>und-Thai-tone</td>
<td>Tag= und-Thai-tone</td>
<td>4</td>
<td>{U+0E48 U+0E49 U+0E4A U+0E4B}</td>
<td>Any tone mark</td>
<td></td>
</tr>
</tbody>
</table>

Character classes in LGR
Whole label evaluation and context rules in LGR

The following table lists all the top-level, or named rules defined in the LGR and indicates whether they are used as trigger in an action or as context (when or not-when) for a code point. (Any use of context rules for variants is not indicated).

<table>
<thead>
<tr>
<th>Name</th>
<th>Regular Expression</th>
<th>Used as Trigger</th>
<th>Used as Context</th>
<th>Anchor</th>
<th>References</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common-leading-combining-mark</td>
<td>(start) [:class property:gc=Mn:]u[:class property:gc=Mc:]</td>
<td>True</td>
<td>False</td>
<td>False</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>und-Laoo-follows-consonant</td>
<td>([:und-Laoo-consonant:])←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 1; semi-consonant must follow a consonant</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-precedes-consonant</td>
<td>←([:und-Laoo-consonant:])</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 2; vowel-before precedes a main consonant cluster</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-follows-main-consonant</td>
<td>([:und-Laoo-consonant:][[:und-Laoo-semi-consonant:]])←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 3; vowel-above, and vowel-below follow a main consonant C</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-follows-C-tonemark-vabove</td>
<td>([:und-Laoo-consonant:][[:und-Laoo-semi-consonant:]][[:und-Laoo-vowel-above:]][[:und-Laoo-tone-mark:]][[:und-Laoo-vowel-above:]]):←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 4; vowel-after follows a main consonant, tone-mark or vowel-above</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-consonant-cluster</td>
<td>([[:und-Laoo-consonant:]][1,2] [:und-Laoo-semi-consonant:][0,1])←</td>
<td>False</td>
<td>False</td>
<td>False</td>
<td>Defining consonant cluster for Rule No. 5</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-follows-Vbefore-consonant-cluster</td>
<td>(U+0EO0([:und-Laoo-consonant-cluster:]))←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 5; The sequence (0EB2 0EB0) follows a vowel before, and a consonant cluster</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-follows-C-vabove-vbelow</td>
<td>([[:und-Laoo-consonant:]][[:und-Laoo-semi-consonant:]][[:und-Laoo-vowel-above:]][[:und-Laoo-vowel-below:]]):←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 6; A tone-mark follows a main consonant, vowel-above or vowel-below</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-follows-Cf</td>
<td>([[:und-Laoo-Cf:]])←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 7; The sign 0ECC can only occur after final consonants</td>
<td></td>
</tr>
<tr>
<td>und-Laoo-repetition-mark-limit</td>
<td>←((U+0EC6)(0,2)(end))</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE Rule No. 8; The sign 0EC6 can only occur 0 to 3 times at the end of the label</td>
<td></td>
</tr>
<tr>
<td>und-Thai-precedes-consonant</td>
<td>←([[:und-Thai-consonant:]])</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE 7.2: check if current cp is preceding a consonant</td>
<td></td>
</tr>
<tr>
<td>und-Thai-follows-consonant</td>
<td>([[:und-Thai-consonant:]])←</td>
<td>False</td>
<td>True</td>
<td>True</td>
<td>WLE 7.3: check if current cp is following a consonant</td>
<td></td>
</tr>
</tbody>
</table>
Actions

The following table lists the actions that are used to assign dispositions to labels and variant labels, based on the specified conditions. The order of actions defines their precedence: the first action triggered by a label is the one defining its disposition.

<table>
<thead>
<tr>
<th>#</th>
<th>Condition</th>
<th>Rule / Variant Set</th>
<th>Disposition</th>
<th>References</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>if label match</td>
<td>Common-leading-combining-mark</td>
<td>invalid</td>
<td>References</td>
<td>any variant label with a code point out of repertoire is invalid</td>
</tr>
<tr>
<td>2</td>
<td>if at least one variant is in {out-of-repertoire-var}</td>
<td>invalid</td>
<td>References</td>
<td>invalid</td>
<td>labels must not commence with a combining mark ⋅</td>
</tr>
<tr>
<td>3</td>
<td>if label match</td>
<td>Common-leading-combining-mark</td>
<td>invalid</td>
<td>References</td>
<td>any variant label with a code point out of repertoire is invalid ⋅</td>
</tr>
<tr>
<td>4</td>
<td>if at least one variant is in {out-of-repertoire-var}</td>
<td>invalid</td>
<td>References</td>
<td>invalid</td>
<td>any variant label with a code point out of repertoire is invalid ⋅</td>
</tr>
<tr>
<td>5</td>
<td>if label match</td>
<td>Common-leading-combining-mark</td>
<td>invalid</td>
<td>References</td>
<td>any variant label with a code point out of repertoire is invalid ⋅</td>
</tr>
<tr>
<td>6</td>
<td>if at least one variant is in {out-of-repertoire-var}</td>
<td>invalid</td>
<td>References</td>
<td>invalid</td>
<td>any variant label with a code point out of repertoire is invalid ⋅</td>
</tr>
</tbody>
</table>

Legend

{...} - variant type set: In the "Rule/Variant Set" column the notation {...} means a set of variant types.

Table of References

[0] The Unicode Standard 1.1, The Unicode Consortium, Mountain View, CA. 1993

Ethiopic Script Versions for the Eight Languages

[2] Corpus Analysis performed by crawling 598 html and 40 PDF files with Tigrigna Contents published online, August 2016
Cited as Auxiliary Evidence for Tigrigna Code Points

[3] Corpus Analysis performed by crawling 14,850 html Amharic Contents of size 1.8 GB published online, August 2016
Cited as Auxiliary Evidence for Amharic Code Points

Cited for Amharic-Driven Variants in Ethiopic Script

[5] The Unicode Standard 1.1


<table>
<thead>
<tr>
<th>Table of References</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0] The Unicode Standard 1.1, The Unicode Consortium, Mountain View, CA. 1993</td>
</tr>
<tr>
<td>[2] Corpus Analysis performed by crawling 598 html and 40 PDF files with Tigrigna Contents published online, August 2016</td>
</tr>
<tr>
<td>[3] Corpus Analysis performed by crawling 14,850 html Amharic Contents of size 1.8 GB published online, August 2016</td>
</tr>
<tr>
<td>[5] The Unicode Standard 1.1</td>
</tr>
<tr>
<td>[201] Lao grammar book published by the Ministry of Education in 1967, see Appendix B, Figure 1</td>
</tr>
<tr>
<td>[202] Lao grammar book published by the Ministry of Education in 1967, see Appendix B, Figure 2</td>
</tr>
<tr>
<td>[203] Lao grammar book published by the Ministry of Education in 1967, see Appendix B, Figure 3</td>
</tr>
<tr>
<td>[204] Lao grammar book published by the Ministry of Education in 2000, see Appendix B, Figure 4</td>
</tr>
<tr>
<td>[205] Lao grammar book published by the Ministry of Education in 2000, see Appendix B, Figure 5</td>
</tr>
<tr>
<td>[206] Lao grammar book published by the Ministry of Education in 2000, see Appendix B, Figure 6</td>
</tr>
<tr>
<td>[207] Lao grammar 1935, see Appendix B, Figure 7</td>
</tr>
</tbody>
</table>
Validate a Label
Validate a Label

1. Click on “Import” to load single or multiple existing LGR files. See “Import” for details.

2. Always view “Validate LGR” output as the tool checks the loaded LGR(s) during this process.

3. For validating a label, click on “Validate label” button.
Validate a Label with a Single LGR

1. Enter the label to be validated
2. Click on the “Validate” button
Validate a Label with a Single LGR

Result given by “Validate label” shows whether the label is valid or invalid and also lists its variants, their dispositions and the rules/actions against which the label or any of its variants is valid or invalid.
Validate a Label with a Merged LGR

1. Enter the label to be validated
2. Select a script from the list of the scripts of different LGRs forming the merged LGR
3. Click on the “Validate” button
Validate a Label with a Merged LGR

Result given by “Validate label” shows whether the label is valid or invalid and also lists its variants, their dispositions and the rules/actions of which the label or any of its variants is valid or invalid.
Review Impact on Existing Labels by Revising an Existing LGR
Review Impacts on Existing Labels Caused by Reviewed LGR

Click on “Diff labels of two LGRs” button to determine differences caused by modifying an LGR.
Enter Details

1. Select first LGR

2. Select second LGR

3. Select file containing labels

4. Enter the email address to receive the result notification

5. Check “Check collisions” if you want to check label collisions as well

6. Check “Output rules” if you want to check output rules for each label

7. Click on “Get diff” button
Email Notification

Hi,
The processing of diff from labels provided in the file 'test-labels.txt' between LGR 'proposed-arabic-lgr-18092016-en' and LGR 'proposed-arabic-lgr-18092017-en' has been successfully completed. You should now be able to download it from your home screen under the name: '20180823_032644_diff_proposed-arabic-lgr-18092016-en_proposed-arabic-lgr-18092017-en.txt.gz'. Please refresh the home page if you don't see the link.
Best regards

Email that gives information about the results
2. Click on the download link on the homepage to get the “Diff labels of two LGRs” results
Results

# Labels not in LGR 1 #

# Labels not in LGR 2 #

Label نادي

# LGR comparison #
Label نادي not in LGR proposed-arabic-lgr-18092017-en
Label نادي not in LGR proposed-arabic-lgr-18092017-en

## Comparison on label 'كلكت' [U+06A9 U+0644 U+0643 U+062A U+06C1]

### Test dispositions: ###
```
...
```
No changes in disposition.
```
```

### Test number of variants: ###
```
...
```
No changes in number of variants.
```

## Comparison on label 'نادي' [U+0644 U+0642 U+0627 U+0621]

### Test dispositions: ###
```
...
```
No changes in disposition.
```
```

### Test number of variants: ###
```
...
```
No changes in number of variants.
```

Specifies changes in disposition of labels by the revised LGR

Specifies new variants of labels formed by the revised LGR
Review Impact on Existing Labels by Introducing a New LGR
Click on the “Get collisions in a list of labels” button to determine label collisions from an existing file if a new LGR is introduced – for example, two unique labels become variants of each other.
Enter Details

1. Select LGR
2. Select Labels file
3. Enter the email address to receive the result notification
4. Check “Full Dump” to get summary of each operation done on the labels
5. Check “Output rules” to get rules that the label has gone through
6. Click on “Get collisions”
Hi,

The processing of collisions from labels provided in the file 'test-labels.txt' in LGR 'proposed-arabic-lgr-18092017-en' has been successfully completed. You should now be able to download it from your home screen under the name: '20180823_031221_collisions_proposed-arabic-lgr-18092017-en.txt.gz'. Please refresh the home page if you don't see the link.

Best regards

Email that gives information about the results
Download Results

Click on the “LGR Tool” Home link to get to this page

Welcome to the LGR (Label Generation Ruleset) Tool

Previously loaded LGR file(s)
Previously, you edited the following LGR file(s). Click on its title to resume your editing session.

LGRs
- View proposed-arabic-lgr-18092016-en
- View proposed-arabic-lgr-18092017-en

Create a new LGR file or import an existing one

- Import an existing XML file

⚠️ Note that importing large LGR files may take significant time to load on your browser.

Start with a New blank XML file

Start from a built-in LGR

The following LGRs are pre-installed in the system. You may use them as a starting point for your own LGR. To do so, just click on it to make a copy that you can then edit.

- Open Sample-French

Remember to save your work regularly by downloading a copy of the XML file.

Please send any feedback to support@viagenie.ca.

Your saved results
The following files contains your tools computation results.

⚠️ Note that these files could be cleaned up regularly.

- Download 20180823_031221_collisions_proposed-arabic-lgr-18092017-en.txt.gz

Click on the download link to get the “Collisions” results
Results

List of labels which have become invalid by introducing a new LGR

Details of a label and its corresponding variant(s)
Validate Multiple Labels Using LGR(s)

Agenda Item #9
Select Generate Disposition Annotations

1. Click on “Import” to load existing LGR file(s). See “Import” for details

2. Click on “Generate disp. annotations” button to validate labels given in a text file
Enter Details for Single LGR

1. Select LGR

2. Select Labels file for validation. The labels file contains single label per line in UTF-8 format

3. Enter the email address to receive the result notification

4. Click on “Annotate” button
Enter Details for Merged LGR

1. Select LGR
   - Optional file of existing labels to check for collisions

2. Select a script from the list of the scripts of different LGRs forming the merged LGR

3. Select Labels file for validation

4. Enter the email address to receive the result notification

5. Click on “Annotate” button
Hi,
The processing of annotation from labels provided in the file 'lao-test-labels.txt' in LGR 'proposal-lao-lgr-31jan17-en' has been successfully completed. You should now be able to download it from your home screen under the name: '20180823_033613_annotation_proposal-lao-lgr-31jan17-en.txt.gz'. Please refresh the home page if you don't see the link.
Best regards

Email that gives information about the results
Welcome to the LGR (Label Generation Ruleset) Tool

This application provides a convenient interface for browsing and editing LGRs conforming to the Representing Label Generation Rulesets using XML specification.

Previously loaded LGR file(s)
Previously, you edited the following LGR file(s). Click on its title to resume your editing session.
LGRs
- View proposal-lao-igr-31jan17-en

Create a new LGR file or import an existing one
[Import an existing XML file]

Note that importing large LGR files may take significant time to load on your browser.

Start from a built-in LGR
The following LGRs are pre-installed in the system. You may use them as a starting point for your own LGR. To do so, just click on it to make a copy that you can then edit.
- Open Sample-French

Remember to save your work regularly by downloading a copy of the XML file.

Please send any feedback to support@viagenie.ca.

Download Results
Click on the download link to get the “Generate disp. annotations” results

Click on the “LGR Tool” home link to get to this page
Validation result of each label is written next to it. For invalid labels, the tool also specifies the reason/rule due to which the label is invalid.
View Cross-script Variants of Labels
Select Cross-script Variants

1. Click on “Import” to load single or multiple LGR file(s). See “Import” for details.

2. Click on “Cross-script variants” button to view cross-script variants of labels given in a text file.

<table>
<thead>
<tr>
<th>Code point</th>
<th>Description</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0E81</td>
<td>0 Variant(s)</td>
<td>LAO LETTER KHO</td>
</tr>
<tr>
<td>U+0E82</td>
<td>0 Variant(s)</td>
<td>LAO LETTER KHO TAM</td>
</tr>
<tr>
<td>U+0E84</td>
<td>0 Variant(s)</td>
<td>LAO LETTER NGO</td>
</tr>
<tr>
<td>U+0E87</td>
<td>0 Variant(s)</td>
<td>LAO LETTER CO</td>
</tr>
<tr>
<td>U+0E88</td>
<td>0 Variant(s)</td>
<td>LAO LETTER SO TAM</td>
</tr>
<tr>
<td>U+0E8A</td>
<td>0 Variant(s)</td>
<td>LAO LETTER YNO</td>
</tr>
<tr>
<td>U+0E94</td>
<td>0 Variant(s)</td>
<td>LAO LETTER DO</td>
</tr>
<tr>
<td>U+0E95</td>
<td>0 Variant(s)</td>
<td>LAO LETTER TO</td>
</tr>
</tbody>
</table>
Enter Details

1. Select single LGR or a merged LGR

2. Select a script from the list of the scripts of different LGRs forming the merged LGR

3. Select Labels file for validation

4. Enter the email to receive the result notification

5. Click on “Launch” button
Email Notification

Hi,
The processing of cross-script variants from labels provided in the file 'test-labels.txt' in LGR 'proposed-armenian-lgr-05nov15-en' has been successfully completed. You should now be able to download it from your home screen under the name: '20180823_073155_cross_script_variants_proposed-armenian-lgr-05nov15-en.txt.gz'. Please refresh the home page if you don't see the link.

Best regards

Email that gives information about the results
Download Results

Welcome to the LGR (Label Generation Ruleset) Tool
This application provides a convenient interface for browsing and editing LGRs conforming to the Representing Label Generation Rulesets using XML specification.

Previously loaded LGR file(s)
Previously, you edited the following LGR file(s). Click on its title to resume your editing session.
LGR sets
- View merged-lgr
- View embedded LGRs

Create a new LGR file or import an existing one
- Import an existing XML file
- Start with a New blank XML file

Start from a built-in LGR
The following LGRs are pre-installed in the system. You may use them as a starting point for your own LGR. To do so, just click on it to make a copy that you can then edit.
- Open Sample-French

Remember to save your work regularly by downloading a copy of the XML file.
Please send any feedback to support@viagenie.ca.

Your saved results
The following files contains your tools computation results.
⚠️ Note that these files could be cleaned up regularly.
- Download
  20180823_073155_cross_script_variants_proposed-armenian-lgr-05nov15-en.txt.gz

Click on the download link to get the “Cross-script variants” results
Results

Cross-script variants of the label

Disposition of cross-script variants

Specifies the LGR (from a set of multiple LGRs) that is used to generate these cross-script variants

Input label U+0561 U+0575 U+0580 (ωη) has cross-script variants:
- Cross-variant U+0448 U+0575 U+0580 (ωη), disposition invalid:
  + U+0448 (ω): Cyril
Input label U+0581 U+0561 U+0579 U+0586 (γωξ) has cross-script variants:
- Cross-variant U+0581 U+0448 U+0579 U+0586 (γωξ), disposition invalid:
  + U+0448 (ω): Cyril
- Cross-variant U+0067 U+0561 U+0579 U+0586 (γωξ), disposition invalid:
  + U+0067 (γ): Latin
Input label U+0566 U+0578 U+057D (ξνω) has cross-script variants:
- Cross-variant U+0566 U+0578 U+0075 (ξνω), disposition invalid:
  + U+0075 (ω): Latin
- Cross-variant U+0566 U+006E U+0578 (ξνω), disposition invalid:
  + U+006E (n): Latin
- Cross-variant U+0566 U+006E U+0075 (ξνω), disposition invalid:
  + U+0075 (ω): Latin
  + U+006E (n): Latin
- Cross-variant U+0071 U+0578 U+057D (ξνω), disposition invalid:
  + U+0071 (ξ): Latin
- Cross-variant U+0071 U+0578 U+0075 (ξνω), disposition invalid:
  + U+0071 (ξ): Latin
  + U+0075 (ω): Latin
- Cross-variant U+0071 U+006E U+057D (ξνω), disposition invalid:
  + U+0071 (ξ): Latin
  + U+006E (n): Latin
- Cross-variant U+0071 U+006E U+0075 (ξνω), disposition invalid:
  + U+0071 (ξ): Latin
  + U+0075 (ω): Latin
  + U+006E (n): Latin
Compare LGRs
Compare LGRs

Click on “Compare two LGRs” under the “Tools” tab to compare LGRs.
Union, Intersection or Difference

1. Select first LGR
2. Select second LGR
3. Select “Union”, “Intersection” or “Diff” to perform relevant function on the two LGRs
4. Click on the “Compare” button
With Merged LGRs, Only Difference

### LGR Tool / merged-lgr-2 - Initial Release

**First LGR**
- merged-lgr-1
  - First LGR to use in comparison

**Second LGR**
- merged-lgr-2
  - Second LGR to use in comparison

**Action to perform on LGRs**
- Diff
  - Choose the action to perform on selected LGRs

**Note:**
Comparison will be performed only between two LGR sets or two simple LGRs, union and intersection are not available for LGR sets.

1. **Select first LGR**
2. **Select second LGR**
   - “Diff” selected by default
3. **Click on the “Compare” button**
Result of Difference function

Differences of two LGRs

Important Note: These operations only provide provisional results which must be manually reviewed and finalized.
Harmonize Multiple LGRs
Select Harmonization

1. Click on “Import” to load single or multiple LGR file(s). See “Import” for details.

2. Click on “Harmonization” button to harmonize two LGR files.
Enter Details

1. Select first LGR
2. Select second LGR
3. Click on “Harmonize” button

Optional RootZone LGR to infer new variant sets
Results of harmonization for proposal-lgr-knda-20180602 and proposal-lgr-telugu-20180531

Resulting LGRs
- View harmonized version of proposal-lgr-knda-20180602
- View harmonized version of proposal-lgr-telugu-20180531

The Harmonized LGRs will also be listed on the homepage.
Engage with ICANN and IDN Program

Thank You and Questions
Reach us at: IDNProgram@icann.org
Website: icann.org/idn

- twitter.com/icann
- gplus.to/icann
- facebook.com/icannorg
- weibo.com/ICANNorg
- linkedin.com/company/icann
- flickr.com/photos/icann
- youtube.com/user/icannnews
- slideshare.net/icannpresentations