ICANN|57
HYDERABAD
Agenda

- LGR Toolset – Marc Blanchet
- Best Practices for IDN LGR – Integration Panel
- Community Updates
  - Lao GP – Chittaphone Chansylilath
  - Chinese GP – Wang Wei
  - Japanese GP – Hiro Hotta
  - Korean GP – KIM Kyongsook
- Q/A
Label Generation Rulesets (LGRs) specify code point repertoire, variant rules and Whole Label Evaluation (WLE) rules, in addition to meta-data, to generated 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 allows for the following:
- Create a LGR
- Use a LGR to validate a label and determine its variants
- Manage LGRs, by comparing or combining them
- Review possible impact of a new or revised LGR on existing labels

For further details, visit the LGR Toolset webpage or www.icann.org/idn.
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 beta deployment
  - Visit [https://lgrtool.icann.org/](https://lgrtool.icann.org/)
  - Username: lgr and password: 37zEfM2LyN3DmSzjLaYoA

- Open source package(s) released with BSD license
  - Released at github: [lgr-core](https://github.com/icann/lgr-core), [lgr-django](https://github.com/icann/lgr-django), [munidata](https://github.com/icann/munidata)
  - Credits to developers: Audric Schiltknecht, Wil Tan, Julien Bernard, David Drouin
Create a French LGR

Add Ligatures: æ, œ (U+00E6, U+0153)

Validate with list of labels

Union/Diff of LGRs
Welcome to the LGR (Label Generation Ruleset) Editor

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

To begin using this application, you may use one of the following options:

- Import an existing XML file
- Start with a New blank XML file

Alternatively, you may select one of the built-in LGR's below as a starting point.

Built-in LGRs

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.
Create LGR

LGR Editor

<table>
<thead>
<tr>
<th>Name</th>
<th>French LGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validating repertoire</td>
<td></td>
</tr>
</tbody>
</table>

idna2008_6.3.0
msr-2-wle-rules-13apr15-en

Create
<table>
<thead>
<tr>
<th>Code point</th>
<th>Character Name</th>
<th>Comments</th>
<th>Action</th>
</tr>
</thead>
</table>

- Code points
- References
- Meta data
- Rules

- Expand range(s)
- Add code point(s)
Add code point(s)

**First code point**

- **a**

**Last code point**

- **z**

[Next]
Add Range - Validation

Add code point(s)

Code points:
- U+97 LATIN SMALL LETTER A
- U+98 LATIN SMALL LETTER B
- U+99 LATIN SMALL LETTER C
- U+100 LATIN SMALL LETTER D
- U+101 LATIN SMALL LETTER E
- U+102 LATIN SMALL LETTER F
- U+103 LATIN SMALL LETTER G
- U+104 LATIN SMALL LETTER H
- U+105 LATIN SMALL LETTER I
- U+106 LATIN SMALL LETTER J
- U+107 LATIN SMALL LETTER K
- U+108 LATIN SMALL LETTER L
- U+109 LATIN SMALL LETTER M
- U+110 LATIN SMALL LETTER N
### Range in LGR

#### LGR Editor / french-lgr

- **Import**
- **New**
- **Tools**

#### 26 code points added

<table>
<thead>
<tr>
<th>Code points</th>
<th>Character Name</th>
<th>Comments</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0061 (a) ... U+007A (z)</td>
<td>LATIN SMALL LETTER A ... LATIN SMALL LETTER Z</td>
<td><strong>See code point</strong></td>
<td><strong>Expand range</strong></td>
</tr>
</tbody>
</table>

**English (en)**

**Go**
## Expanded Range

<table>
<thead>
<tr>
<th>Code point</th>
<th>Character Name</th>
<th>Comments</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0061 (a)</td>
<td>LATIN SMALL LETTER A</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0062 (b)</td>
<td>LATIN SMALL LETTER B</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0063 (c)</td>
<td>LATIN SMALL LETTER C</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0064 (d)</td>
<td>LATIN SMALL LETTER D</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0065 (e)</td>
<td>LATIN SMALL LETTER E</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0066 (f)</td>
<td>LATIN SMALL LETTER F</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0067 (g)</td>
<td>LATIN SMALL LETTER G</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0068 (h)</td>
<td>LATIN SMALL LETTER H</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0069 (i)</td>
<td>LATIN SMALL LETTER I</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006A (j)</td>
<td>LATIN SMALL LETTER J</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006B (k)</td>
<td>LATIN SMALL LETTER K</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006C (l)</td>
<td>LATIN SMALL LETTER L</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006D (m)</td>
<td>LATIN SMALL LETTER M</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006E (n)</td>
<td>LATIN SMALL LETTER N</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006F (o)</td>
<td>LATIN SMALL LETTER O</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0070 (p)</td>
<td>LATIN SMALL LETTER P</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0071 (q)</td>
<td>LATIN SMALL LETTER Q</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0072 (r)</td>
<td>LATIN SMALL LETTER R</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0073 (s)</td>
<td>LATIN SMALL LETTER S</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0074 (t)</td>
<td>LATIN SMALL LETTER T</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0075 (u)</td>
<td>LATIN SMALL LETTER U</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0076 (v)</td>
<td>LATIN SMALL LETTER V</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0077 (w)</td>
<td>LATIN SMALL LETTER W</td>
<td></td>
<td>See code point</td>
</tr>
</tbody>
</table>
Add Code Point (œ)

<table>
<thead>
<tr>
<th>Code point</th>
<th>Code point range</th>
<th>Import from file</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0061 (a)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0062 (b)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0063 (c)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0064 (d)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0065 (e)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0066 (f)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0067 (g)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0068 (h)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+0069 (i)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+006A (j)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+006B (k)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+006C (l)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
<tr>
<td>U+006D (m)</td>
<td>0 Variant(s)</td>
<td></td>
</tr>
</tbody>
</table>

U+0061: LATIN SMALL LETTER H
U+0062: LATIN SMALL LETTER I
U+0063: LATIN SMALL LETTER J
U+0064: LATIN SMALL LETTER K
U+0065: LATIN SMALL LETTER L
U+0066: LATIN SMALL LETTER M

Add code point(s)

- Code point

œ

Override repertoire
Add Code Point
Add Code Point Sequence (o e)
Add Variant Òe to (o e)
Add Variant (o e) to œ
Label List

bonjour
œuf
œuf
bœuf
bœuf
æquo
aequo
cætera
cæterae
As the computing may be very long, we will warn by e-mail once the result can be downloaded. Please provide a valid e-mail address:

Provide your e-mail address
Computing annotations on labels provided in file french-lgr-label in LGR french-lgr

As the computing may be very long, once completed, an e-mail will be sent at the provided address: audric.schiltknecht@viagenie.cat
Welcome to the LGR (Label Generation Ruleset) Editor

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

Your LGRs
Previously, you edited the following LGR file(s). Click on its title to resume your editing session.
⚠️ Note that importing large LGR files may take significant time to load on your browser.

- View French-lgr

You may also use one of the following options:

- Import an existing XML file
- Start with a New blank XML file

Alternatively, you may select one of the built-in LGR’s below as a starting point.

Built-in LGRs
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.

Your saved results
The following files contains your tools computation results.
⚠️ Note that these files could be cleaned up regularly.

- Download annotation_french-lgr_20161026_160215.txt.gz

Please send any feedback to support@viagenie.ca.
Annotated Result (æ invalid)

bonjour: allocate
œuf: allocate
œuf: allocate
boeuf: allocate
bœuf: allocate
æquo: invalid
aequo: allocate
cæterae: invalid
cæterae: allocate
Create Another LGR

- (just for the purpose of showing the union and diff tools)
- Create another LGR, add a-z (do not add “oe”)
Add Code Point Sequence (a e)
Add Code Point (∅)
Add Variant æ to (a e)
**Add Variant (a e) to æ**

Code point appeared in Unicode version: 1.1.0.0

### Variants

<table>
<thead>
<tr>
<th>Code point</th>
<th>Type</th>
<th>Comments</th>
<th>When</th>
<th>Not When</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0061 (a) U+0065 (e) LATIN SMALL LETTER A LATIN SMALL LETTER E</td>
<td>block</td>
<td></td>
<td></td>
<td></td>
<td>Delete variant Edit references References</td>
</tr>
</tbody>
</table>

### Tags

Space-separated tags

**When**

**Not-When**

**Comment**

Save variants, tags, context rules and comment

### References

No references associated with code point.

Edit

Delete code point

English (en) → Go
Annotated Result (œ invalid)

bonjour: allocate
oeuf: allocate
œuf: invalid
boeuf: allocate
bœuf: invalid
æquo: allocate
aequo: allocate
cætera: allocate
cæterae: allocate
Compare

LGR Editor / french-lgr-updated

First LGR: french-lgr
First LGR to use in comparison

Second LGR: french-lgr-updated
Second LGR to use in comparison

Action to perform on LGRs:
- Intersection
- Union
- Intersection
- Diff

Compare button

English (en) dropdown
Result of Diff

** Compare Metadata **

Compare Version
Some version value for both LGR: '1'.
Some version comment value for both LGR: 'None'.

Compare Description
First LGR has no description
Second LGR has no description

Some scapes value for both LGR: '').
Some languages value for both LGR: '[]'.
Some date value for both LGR: 'None'.
Some validity start value for both LGR: 'None'.
Some validity end value for both LGR: 'None'.
Some unicode version value for both LGR: '6.3.0'.'

Some references value for both LGR: '[]'.'

** Compare repertoire **

Repertoire values differ:
Values only in first LGR: U+006F U+0065 U+0153.
Values only in second LGR: U+0001 U+0005 U+0065.
Common values: U+006F U+0074 U+0061 U+0068 U+0072 U+0074 U+0076 U+0068 U+006A U+006E U+0079 U+0065 U+0066 U+0062 U+0064 U+0066 U+0069

** Compare common code points in repertoire **

Compare code point 40109
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40107
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40101
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40103
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40103
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.

Compare code point 40110
Some comment value for both LGR: 'None'.
Some tags value for both LGR: '[]'.
Some variants value for both LGR: '[]'.
Union of LGRs

First LGR: french-lgr
First LGR to use in comparison

Second LGR: french-lgr-updated
Second LGR to use in comparison

Action to perform on LGRs: Union
Choose the action to perform on selected LGRs

Compare
### Result of LGR Union

<table>
<thead>
<tr>
<th>Code Point</th>
<th>Character Name</th>
<th>Comments</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0061 (a)</td>
<td>LATIN SMALL LETTER A</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0061 (a) U+0065 (e)</td>
<td>LATIN SMALL LETTER A LATIN SMALL LETTER E</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0062 (b)</td>
<td>LATIN SMALL LETTER B</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0063 (c)</td>
<td>LATIN SMALL LETTER C</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0064 (d)</td>
<td>LATIN SMALL LETTER D</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0065 (e)</td>
<td>LATIN SMALL LETTER E</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0066 (f)</td>
<td>LATIN SMALL LETTER F</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0067 (g)</td>
<td>LATIN SMALL LETTER G</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0068 (h)</td>
<td>LATIN SMALL LETTER H</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0069 (i)</td>
<td>LATIN SMALL LETTER I</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006A (j)</td>
<td>LATIN SMALL LETTER J</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006B (k)</td>
<td>LATIN SMALL LETTER K</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006C (l)</td>
<td>LATIN SMALL LETTER L</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006D (m)</td>
<td>LATIN SMALL LETTER M</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006E (n)</td>
<td>LATIN SMALL LETTER N</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006F (o)</td>
<td>LATIN SMALL LETTER O</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+006F (o) U+0065 (e)</td>
<td>LATIN SMALL LETTER O LATIN SMALL LETTER E</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0070 (p)</td>
<td>LATIN SMALL LETTER P</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0071 (q)</td>
<td>LATIN SMALL LETTER Q</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0072 (r)</td>
<td>LATIN SMALL LETTER R</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0073 (s)</td>
<td>LATIN SMALL LETTER S</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0074 (t)</td>
<td>LATIN SMALL LETTER T</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0075 (u)</td>
<td>LATIN SMALL LETTER U</td>
<td></td>
<td>See code point</td>
</tr>
<tr>
<td>U+0076 (v)</td>
<td>LATIN SMALL LETTER V</td>
<td></td>
<td>See code point</td>
</tr>
</tbody>
</table>
Annotated Result of the Union

bonjour: allocate
œuf: allocate
œuf: allocate
boeuf: allocate
bœuf: allocate
æquo: allocate
aequo: allocate
cætera: allocate
caetera: allocate
Best Practices for IDN LGR

Integration Panel
Agenda

- An example of related scripts: Abugida
- An example: realism in coding Lao LGR
- LGR specification issues
  - Starting point
  - Code points
  - Documentation consistency
- Variant set
- Rules simplification
Abugida scripts include:

- Ethiopic; Neo-Brahmi (S. Asia); and Thai, Lao, Khmer etc. (SE Asia)

Historically related, and structural similarities remain:

- each syllable: has leading consonant, & satellite vowels: → ← ↑ ↓
- any cons. with no following vowel: derived from basic C, by Halant
- (unlike Arabic) all vowels obligatorily marked

Unicode encoding models are different:

- Ethiopic – code for each syllable (almost like Korean)
- Neo-Brahmi – vowel CPs: always follow cons, as combining marks
- SE Asian – vowel CPs: some independent, like consonants; some occur before cons in syl; use of combining marks for signs and tones

These differences result in different LGR designs
1. Strictly, tones marked only on consonants at head of syllable:
   — But LGR rules based on near context of CPs in label: syllabs invisible
2. In Lao, there are also linguistic constraints on syllable structures:
   — But relaxed for labels: initialisms are arbitrary strings of consonants
3. Evolving treatment of a letter: ０ 0EBC
   1. SEMI-CONSONANT LO - final [r] in cons cluster: 部副 [h] vs. 副 [hr]
   2. Previously, had appeared after (under) various consonants
   3. Lao spelling reforms (1960s): only after HO SUNG (部副) – (部副)
   4. Modern Lao has borrowed words from other languages
      𝙧 (0E9F 0EBC 0EB5) “free”, also “pro.gra.m” and “e.lec.tra.nic”
   ○ LGR rule for context of SEMI-CONSONANT LO must follow usage
Use LGR proposal template for a consistent layout and appearance of the main document

Template:
- [https://community.icann.org/download/attachments/43989034/LGR-Proposal-Template.docx](https://community.icann.org/download/attachments/43989034/LGR-Proposal-Template.docx)

Examples:
Be conservative:
  • Problematic or doubtful code points should not be included
Use tag to create subset for context or rules
Code points used only in specific sequences (example: combining sequences) should only be included as sequences, not separately
  • Example: Thai
    • ฦๅ (U+0E24 U+0E45)
    • ฦๅ (U+0E26 U+0E45)
    • Adding singletons ฦ and ฦ, but not ฿ — ensures that ฿ can only be used after ฦ and ฦ
Provide one or more references for each code point

Use same reference numbers in Documentation and XML
  - For repertoire (mandatory)
  - For variants (if applicable)
  - For WLE rules (if applicable)

Notation for tag, context, and rules should be identical

WLE rules should be enumerated in same order and have the same content

Discrepancy makes review difficult; danger of wrong interpretation
Variant Set

- Provide rationale
  - Document source reference (such as existing IDN table)
  - Provide rationale for any deviation from existing practice

- Consistency
  - Specify all mappings: symmetric and transitive
  - If reflexive mapping is used, apply it to all repertoire elements

- For root zone:
  - Must use a type value on all variant mappings
  - Must not use context on variants
Allocatable variants: Limit as much as possible

- [Procedure] A.3.3: “From the Conservatism Principle, it follows that the number of allocatable variants should be minimized

- Read carefully [RFC 7490], especially sections:
  - 5.3 Variants
  - 7.2 Actions with Variants Type Triggers
  - Appendix B: How to Translate Tables on RFC 3743 into the XML Format
Rules Simplification

- LGR is not a spelling specification
  - Typically allows acronyms and initialisms
  - Brand names may use innovative spelling
  - Rules define syntax of script as a whole, not languages

- Rules should aim at improving security
  - One goal is to avoid ambiguous rendering
  - Restrict use of combining sequences to meaningful context

- Keep rules simple
  - Some over/underproduction of labels OK
  - Generally prefer rules based on local (immediate) context
  - Generally prefer sequences over single-code point rules
Rules Simplification – Test Labels

- Complete LGR proposal includes **Test Labels** provided by the GP
  - Valid labels covering a good cross section of code points
  - Invalid labels, classified by what rule or context they break
  - If appropriate: test labels for variant generation
- IP will source **real world data** (word lists)
  - Provide a useful check on over/under generation of labels
  - Allow tracking of changes as proposal matures
  - Verify effects of simplification
- IP will compare proposed LGR’s effect on delegated TLDs
Thank You – Questions?
Update by the Lao GP

Chittaphone Chansylilath
Lao GP Coordinator
Agenda

- Introduction to Lao script
- Overview of Lao Generation Panel
- Challenges in developing the LGR
- Current progress
- Timeline
Introduction to Lao script

- The Lao script is used to write the official language of Laos
- Syllables are written around the main consonant. Vowels occur above, below, before, after or around this consonant
- Writing without spaces between words and syllables
- Lao is written from left to right in horizontal lines
- Lao script and Thai script have many characters which have some similarities (see Appendix A), but Lao has fewer letters which are in a more rounded form, making them visually different from Thai
- The Lao dialect is differentiated into five main areas in Laos - Vientiane, Luang Prabang, Xieng Khuang, Khammuan and Champassak provinces
# Overview of Lao Generation Panel

<table>
<thead>
<tr>
<th>No.</th>
<th>Name and Surname</th>
<th>Organization</th>
<th>Role</th>
<th>Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mr. Phonpasit Phissamay</td>
<td>Director General of E-Government Center</td>
<td>Chair</td>
<td>Lao localization projects since 2003 and Integration of Lao in E-government</td>
</tr>
<tr>
<td>2.</td>
<td>Mr. Khamphanh Souvannakha</td>
<td>Deputy Director of National Internet Center</td>
<td>Co-Chair on DNS</td>
<td>Supervision of .la domain name registration</td>
</tr>
<tr>
<td>3.</td>
<td>Mr. Valaxay Dalaloy</td>
<td>Cabinet Office</td>
<td>Policy Member</td>
<td>ICT policy and localization since 2003</td>
</tr>
<tr>
<td>4.</td>
<td>Mr. Bualy Paphaphanh</td>
<td>National University of Laos</td>
<td>Linguistic Member</td>
<td>Linguistic expert and advisor to Lao localization</td>
</tr>
<tr>
<td>5.</td>
<td>Mr. Sengfa Holanouphab</td>
<td>National University of Laos</td>
<td>Linguistic Member</td>
<td>Linguistic expert</td>
</tr>
<tr>
<td>6.</td>
<td>Mr. Bounmy Kongmany</td>
<td>National University of Laos</td>
<td>Linguistic Member</td>
<td>Linguistic expert</td>
</tr>
<tr>
<td>7.</td>
<td>Mrs. Chittaphone Chansylilath</td>
<td>E-Government Center</td>
<td>Technical member</td>
<td>Lao localization specialist, Font, Keyboard, OCR, TTS Projects.</td>
</tr>
</tbody>
</table>
## Overview of Lao Generation Panel Cont.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name and Surname</th>
<th>Organization</th>
<th>Role</th>
<th>Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.</td>
<td>Mr. Thonglor Douansouvanh</td>
<td>Vientiane times newspaper</td>
<td>Community member</td>
<td>Media</td>
</tr>
<tr>
<td>9.</td>
<td>Mr. Phouthong Sisavath</td>
<td>National Internet Center</td>
<td>Technical member</td>
<td>DNS operation</td>
</tr>
<tr>
<td>10.</td>
<td>Ms. Phavanhna Douangboupha</td>
<td>National Internet Center</td>
<td>Technical member</td>
<td>Coordinator for international cooperation</td>
</tr>
<tr>
<td>11.</td>
<td>Mr. Khamphay Inthara</td>
<td>E-Government Center</td>
<td>Technical member</td>
<td>Lao localization specialist, Lao Font, Lao Keyboard project</td>
</tr>
<tr>
<td>12.</td>
<td>Mr. Saysomvang Souvannavong</td>
<td>National Internet Center</td>
<td>Technical member</td>
<td>DNS operation</td>
</tr>
<tr>
<td>13.</td>
<td>Mr. Phousana Silivong</td>
<td>E-Government Center</td>
<td>Technical member</td>
<td>Lao localization specialist, Lao Font, Lao Keyboard project</td>
</tr>
</tbody>
</table>
Challenges in Developing the LGR

- No national standard for writing Lao context
  - There is no rule or standard of using semi-consonant ة 0EBC and the sign ة (0ECC)

- Complexity of syllable or writing structure, especially for foreign words like
  - 唰革新 (Vietnamese’s name), three consonants to form a consonant cluster
  - วกะ (Lao word), two after vowel come together to form diphthong
Timeline – 2016

Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec

- Develop principles
- Determine the code points
- Determine (any) variants
- Determine label rules
- Write proposal and create XML
- Submit
Thank You
Update by the Chinese GP

Wang Wei and Kenny Huang
Chinese GP Co-Chairs
## Repertoire

<table>
<thead>
<tr>
<th></th>
<th>CDNC</th>
<th>19559</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normalized Hanzi for Common Use (NHCU)</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>IICORE</td>
<td>145</td>
</tr>
<tr>
<td>3</td>
<td>dotAsia</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19738</td>
</tr>
</tbody>
</table>
### Repertoire

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>CDNC</td>
<td>19561 = 19559 + 2</td>
</tr>
<tr>
<td>1</td>
<td>dotAsia</td>
<td>124 = 62 + 62</td>
</tr>
<tr>
<td>2</td>
<td>NHCU</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>IICORE \cap JGP, IICORE \cap KGP</td>
<td>43</td>
</tr>
</tbody>
</table>

**MSR**

- CDNC 19561
  - NHCU 18
  - dotAsia 124

- IICORE \cap JGP
  - GE-202A
  - H-987D
  - T3-4B4C

- IICORE \cap KGP
  - 58B5
  - 32.12
<table>
<thead>
<tr>
<th></th>
<th>CDNC 2015</th>
<th>19561=19559+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>dotAsia</td>
<td>124=62+62</td>
</tr>
<tr>
<td>2</td>
<td>NHCU</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>IICORE∩JGP IICORE∩KGP</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>19746</th>
<th></th>
</tr>
</thead>
</table>

**Notes:**
- The table lists domain names with their corresponding codes and亚洲 (Asia) suffix.
- The values 19561, 124, 18, and 43 are used to illustrate the domain names and their codes.
<table>
<thead>
<tr>
<th></th>
<th>CDNC 2015</th>
<th>19561 = 19559 + 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>dotAsia</td>
<td>124 = 62 + 62</td>
</tr>
<tr>
<td>2</td>
<td>NHCU</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>IICORE ∩ JGP IICORE ∩ KGP</td>
<td>43</td>
</tr>
</tbody>
</table>

19746

Supplementary Ideographic Plane
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>CDNC 2015</td>
<td>19561=19559+2</td>
</tr>
<tr>
<td>1</td>
<td>dotAsia</td>
<td>124=62+62</td>
</tr>
<tr>
<td>2</td>
<td>NHCU</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>IICORE∩JGP IICORE∩KGP</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19746</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>48BC</td>
<td>邳</td>
<td>N</td>
</tr>
<tr>
<td>732F</td>
<td>猃</td>
<td>N</td>
</tr>
<tr>
<td>9EB9</td>
<td>麴</td>
<td>N</td>
</tr>
<tr>
<td>5227</td>
<td>刀</td>
<td>V</td>
</tr>
<tr>
<td>524F</td>
<td>刚</td>
<td>V</td>
</tr>
<tr>
<td>6060</td>
<td>怅</td>
<td>V</td>
</tr>
<tr>
<td>74A2</td>
<td>璿</td>
<td>V</td>
</tr>
<tr>
<td>750E</td>
<td>罿</td>
<td>V</td>
</tr>
<tr>
<td>754A</td>
<td>畴</td>
<td>V</td>
</tr>
<tr>
<td>7ADA</td>
<td>疑</td>
<td>V</td>
</tr>
<tr>
<td>8262</td>
<td>舴</td>
<td>V</td>
</tr>
<tr>
<td>88B5</td>
<td>祈</td>
<td>V</td>
</tr>
<tr>
<td>894D</td>
<td>裳</td>
<td>V</td>
</tr>
<tr>
<td>8B0C</td>
<td>証</td>
<td>V</td>
</tr>
<tr>
<td>8F19</td>
<td>輓</td>
<td>V</td>
</tr>
<tr>
<td>945A</td>
<td>鑴</td>
<td>V</td>
</tr>
<tr>
<td>984B</td>
<td>頤</td>
<td>V</td>
</tr>
<tr>
<td>9DC0</td>
<td>鳀</td>
<td>V</td>
</tr>
<tr>
<td>0</td>
<td>CDNC 2015</td>
<td>19561=19559+2</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
<td>dotAsia</td>
<td>124=62+62</td>
</tr>
<tr>
<td>2</td>
<td>NHCU</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>II CORE (\cap) JGP II CORE (\cap) KGP</td>
<td>43</td>
</tr>
</tbody>
</table>

| 967A | 険 | JGP |
| 7E4A | 縄 | JGP |
| 9421 | 鐵 | JGP |
| 9D8F | 雞 | JGP |
| 4FAD | 佇 | JGP |
| 6442 | 摂 | JGP |
| 685F | 構 | JGP |
| 7E4B | 繋 | JGP |
| 81D3 | 臟 | JGP |
| 8217 | 舗 | JGP |
| 9039 | 達 | JGP |
| 9271 | 鉛 | JGP |
| 9EBA | 麥 | JGP |
| 3960 | 慄 | JGP |
| 51E6 | 劃 | JGP |
| 61F4 | 懺 | JGP |
| 685C | 桜 | JGP |
| 6D9C | 洮 | JGP |
| 6E8C | 洮 | JGP |
| 731F | 獨 | JGP |
| 784F | 砕 | JGP |
| 7C14 | 篁 | JGP |
| 7D9A | 統 | JGP |
| 8133 | 腦 | JGP |
| 86CD | 蚕 | JGP |
| 8E99 | 蹀 | JGP |
| 91A4 | 醡 | JGP |
| 91C8 | 釀 | JGP |
| 96B2 | 昴 | JGP |
| 982C | 頬 | JGP |
| 98EE | 飲 | JGP |
| 9A12 | 騾 | JGP |
| 9A13 | 驗 | JGP |
| 9C2E | 鰲 | JGP |
| 9D0E | 鴞 | JGP |
| 9D2C | 鴯 | JGP |
| 56A2 | 囊 | JGP |
| 663B | 昂 | JGP |
| 7A36 | 槂 | JGP |
| 7B86 | 瓠 | JGP |
| 839F | 茗 | JGP |
| 83B5 | 蕁 | JGP |
| 9A28 | 驢 | JGP |
Variant Mappings
  - Types and Sub-Types

<table>
<thead>
<tr>
<th>&quot;simp&quot;</th>
<th>Allocatable</th>
<th>preferred simplified variant char;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;r-simp&quot;</td>
<td>Allocatable</td>
<td>reflexive preferred simplified variant char;</td>
</tr>
<tr>
<td>&quot;trad&quot;</td>
<td>Allocatable</td>
<td>preferred traditional variant char</td>
</tr>
<tr>
<td>&quot;r-trad&quot;</td>
<td>Allocatable</td>
<td>reflexive preferred traditional variant char</td>
</tr>
<tr>
<td>&quot;both&quot;</td>
<td>Allocatable</td>
<td>preferred simplified and traditional variant chars are the same</td>
</tr>
<tr>
<td>&quot;r-both&quot;</td>
<td>Allocatable</td>
<td>reflexive preferred simp and trad variant chars are the same</td>
</tr>
<tr>
<td>&quot;blocked&quot;</td>
<td>Blocked</td>
<td>Non-allocatable variant char</td>
</tr>
</tbody>
</table>
Variant Mappings

<table>
<thead>
<tr>
<th>Variant</th>
<th>Allocatable Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;simp&quot;</td>
<td>Allocatable</td>
<td>preferred simplified variant char;</td>
</tr>
<tr>
<td>&quot;r-simp&quot;</td>
<td>Allocatable</td>
<td>reflexive preferred simplified variant char;</td>
</tr>
<tr>
<td>&quot;trad&quot;</td>
<td>Allocatable</td>
<td>preferred traditional variant char</td>
</tr>
<tr>
<td>&quot;r-trad&quot;</td>
<td>Allocatable</td>
<td>reflexive preferred traditional variant char</td>
</tr>
<tr>
<td>&quot;both&quot;</td>
<td>Allocatable</td>
<td>preferred simplified and traditional variant chars are the same</td>
</tr>
<tr>
<td>&quot;r-both&quot;</td>
<td>Allocatable</td>
<td>reflexive preferred simp and trad variant chars are the same</td>
</tr>
<tr>
<td>&quot;r-neither&quot;</td>
<td>Blocked</td>
<td>Non-allocatable reflexive/original char</td>
</tr>
<tr>
<td>&quot;blocked&quot;</td>
<td>Blocked</td>
<td>Non-allocatable variant char</td>
</tr>
</tbody>
</table>
88, China Country Code, preferred simplified variant
886, Taiwan Country Code, preferred traditional variant

All variants, including reflexive one.

U+4E81(0); U+5E72(86); U+4E7E(886); U+6F27(0);
U+4F53(0); U+4F53(86); U+9AD4(886); U+9AD4(0);

reserved variants

<char cp="4F53" tag="sc:Hani" >
  <var cp="4F53" type="r-simp" comment="identity" />
  <var cp="8EB0" type="blocked" />
  <var cp="8EC6" type="blocked" />
  <var cp="9AB5" type="traded" />
  <var cp="9AD4" type="neither" />
</char>

<char cp="4E81" tag="sc:Hani" >
  <var cp="4E7E" type="trad" />
  <var cp="4E81" type="r-neither" comment="identity" />
  <var cp="5E72" type="simp" />
  <var cp="5E79" type="blocked" />
  <var cp="69A6" type="blocked" />
  <var cp="6F27" type="blocked" />
</char>
Variant Mappings
  • CGP interior coordination

CGP Variant Mappings

CDNC
Variant Mappings

172 Non-CDNC Chars
Variant Mappings
Review

7 dotAsia Chars
Variant Mappings
Review

dotAsia
Variant Mappings

69 Plane0 chars
62 Plane2 chars
Variant Mappings
- 259 unacceptable variant groups coordination between C and K
Variant Mappings

- Limit the number of allocable labels

"under the conservatism principle, LGRs should strive to minimize allocatable variants ... can be fixed by not having multiple simp/trad mappings. It may be an acceptable trade-off to eliminate the multiple mappings, and let applicants who need a specific all-simplified or all-traditional variant label apply for just the specific label."

台 (53F0); 台 (53F0), 台 (53F0) 檯 (6AAF) 臺 (81FA) 颱 (98B1)
湾 (6E7E); 湾 (6E7E), 灣 (7063)

台灣 (53F0 6E7E) >>
台湾 (53F0 6E7E)

台灣 (53F0 7063), 檯灣 (6AAF 7063), 臺灣 (81FA 7063), 颱灣 (98B1 7063)
Variant Mappings

Counter Example to IP’s suggestion on eliminating variant mappings

<table>
<thead>
<tr>
<th>Original</th>
<th>Simplified</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>台(53F0)</td>
<td>台(53F0)</td>
<td>台(53F0)</td>
</tr>
<tr>
<td>檜(6AAF)</td>
<td>台(53F0)</td>
<td>檜(6AAF)</td>
</tr>
<tr>
<td>篱(7C49)</td>
<td>台(53F0)</td>
<td>篱(7C49)</td>
</tr>
<tr>
<td>臺(81FA)</td>
<td>台(53F0)</td>
<td>臺(81FA)</td>
</tr>
<tr>
<td>鱉(98B1)</td>
<td>台(53F0)</td>
<td>鱉(98B1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original</th>
<th>Simplified</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>鉄(9244)</td>
<td>鉄(94C1)</td>
<td>鐵(9435)</td>
</tr>
<tr>
<td>鍊(9295)</td>
<td>鉄(94C1)</td>
<td>鐵(9435)</td>
</tr>
<tr>
<td>鐵(9421)</td>
<td>鉄(94C1)</td>
<td>鐵(9435)</td>
</tr>
<tr>
<td>鐵(9435)</td>
<td>鉄(94C1)</td>
<td>鐵(9435)</td>
</tr>
<tr>
<td>鐵(94C1)</td>
<td>鉄(94C1)</td>
<td>鐵(9435)</td>
</tr>
</tbody>
</table>

台(53F0)鉄(9244) >> 台(53F0)鉄(94C1), 台(53F0)鐵(9435)/檜(6AAF)鐵(9435)/臺(81FA)鐵(9435)/鲱(98B1)鐵(9435)

台(53F0)鐵(9435) >> 台(53F0)鉄(94C1), 檜(6AAF)鐵(9435)/臺(81FA)鐵(9435)/鲱(98B1)鐵(9435)
Variant Mappings
- Identify multiple mappings
- Execute multiple times

<table>
<thead>
<tr>
<th>Original</th>
<th>Simplified</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>台(53F0)</td>
<td>台(53F0)</td>
<td>台(53F0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>檯(6AAF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>臺(81FA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>飈(98B1)</td>
</tr>
</tbody>
</table>

Sub-Type | Type  
---|---
“r-simp-m” | Blocked 
“simp-m” | Blocked 
“r-trad-m” | Blocked 
“trad-m” | Blocked 
“r-both-m” | Blocked 
“both-m” | Blocked
○ WLE Rules

<rules>
<!--Action elements - order defines precedence-->
<action disp="invalid" match="leading-combining-mark" comment="labels with leading combining marks are invalid"/>
<action disp="blocked" any-variant="blocked" comment="default action for blocked variant"/>
<action disp="allocatable" only-variants="simp r-simp both r-both" comment="simplified label"/>
<action disp="allocatable" only-variants="trad r-trad both r-both" comment="traditional label"/>
<action disp="allocatable" only-variants="r-simp r-trad r-both r-neither" comment="original label"/>
<action disp="blocked" only-variants="simp simp-m r-simp r-simp-m both both-m r-both r-both-m" comment="multiple simplified label"/>
<action disp="blocked" only-variants="trad trad-m r-trad r-trad-m both both-m r-both r-both-m" comment="multiple traditional label"/>
<action disp="blocked" any-variant="simp trad both r-simp r-trad r-both simp-m trad-m both-m r-simp-m r-trad-m r-both-m r-neither" comment="block any other mixed labels"/>
<action disp="allocatable" comment="catch-all"/>
</rules>
Next Step

- Reduce the number of unacceptable variant mappings to $K$
  - from 60 to ?

- Reach a consensus on how to handle the disagreed variant mapping
  - Do not allow character to be applied for?
  - Allow as separate characters?
  - Allow IDN variant?

- Limit the number of allocatable labels
  - Multiple LGR execution process
Thanks

Q&A
Update by the Japanese GP

Hiro Hotta
Japanese GP Chair
JapaneseGP (JGP) update

October 2016
Hiro Hotta <hotta@jprs.co.jp>

red vertical line shows the progress from March
JGP meetings & related events

• 2014
  – August 29 preparatory JGP meeting (1)
  – September 12 preparatory JGP meeting (2)
  – September 24 JGP meeting (1)
  – October 24 JGP meeting (2)
  – November 26 JGP meeting (3)
  – December 18 JGP meeting (4)

• 2015
  – January 16 JGP meeting (5)
  – February 4 JGP meeting (6)
  – February 6 submission of JGP proposal to ICANN
  – February 20 JGP meeting (7)
  – March 10 JGP establishment approved by ICANN
  – March 18 JGP meeting (8)
  – April 15 JGP meeting (9)
  – May 15-16 CJK coordination meeting in Seoul
  – May 20 JGP meeting (10)
  – June 17 JGP meeting (11)
  – June 21-25 CJK coordination meeting during ICANN
  – September 29 JGP meeting (12)
  – October 18-22 CJK coordination meeting during ICANN
JGP meetings & related events (cont’d)

- 2016
  - March 6-10  CJK coordination meeting during ICANN
  - March 20 -21 CJK coordination meeting in Beijing
  - June 27-30  CJK coordination meeting during ICANN
  - September 24 JGP meeting (13)
  - August 29-30 CJK coordination meeting in Taipei
Relationship among CJK language LGRs

Japanese LGR

Chinese LGR

Korean LGR

coordination

* “Han” is called “Kanji” in Japan, “Hanja” in Korea

Copyright © 2016 Japanese Generation Panel
Framework of CJK LGR integration for Han characters
(revised by agreement in Buenos Aires)

Developed by each GP

- Chinese LGR-α
  - First version developed

- Japanese LGR-α
  - First version developed

- Korean LGR-α
  - First version developed

Integrated LGR

- Chinese LGR-β

- Japanese LGR-β

- Korean LGR-β

Characters may be marked as ‘review needed’

Iterative feedback

Copyright © 2016 Japanese Generation Panel
Overview of Japanese LGR-α (J-LGR-α)

• **Repertoire**
  – Consists of characters from 3 scripts (Han, Hira and Kana – Jpan in ISO 15924)

<table>
<thead>
<tr>
<th>Script</th>
<th># of characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Han</td>
<td>6358</td>
</tr>
<tr>
<td>Hira</td>
<td>85</td>
</tr>
<tr>
<td>Kana</td>
<td>89</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>6532</strong></td>
</tr>
</tbody>
</table>

• **Variants & their types**
  – No variants
  – types of imported variants will be investigated and determined after LGR-α from CGP and KGP are proposed

• **WLE**
  – Rules (although not very many) are under discussion
Development at & after Marrakech

• Reduction of the number of allocatable labels
  – Variant labels will exist by importing CGP variant and KGP variant characters, although JGP defines no variants
  – Strings containing any combinations of characters are allowed as natural Japanese words
  – Therefore, many variant labels may exist for a label
    • waiting for Chinese and Korean GP’s definition of variant characters, which CGP and KGP are coordinating to converge

  – However, IP (Integration Panel) requests JGP to reduce the number of allocatable labels
Development at & after Marrakech (cont’d)

- Planning to propose ICANN to redesign TLD application/evaluation process

Complement process for RootLGR

Input = one or more labels that are variants of each other are applied for at the same time

Original label-1
Original label-2

RootLGR

Allocatable variant labels-1
Blocked variant labels-1
Invalid variant labels-1

Allocatable variant labels = original label-1 + original label-2 + : + allocatable variant labels of label-1 + allocatable variant labels of label-2 + :

Output = allocatable variant labels

(at least original labels are allocatable)
Update by the Korean GP

KIM Kyongsook
Korean GP Chair
Introduction

A list of Hangul Syllables, Hanja characters for K-LGR v0.5 (2016.09.28.)

Review of K (Korean) and C (Chinese) Variant Groups (Sets)

Timeline of KGP activities
Introduction

Characters included for “Kore" (Korean Label)
- Both Hangeul (Hangul) syllables and Hanja chars are included in K-LGR

K-LGR v0.5 (2016.09.28.)
- 11172 Hangeul syllables
- 4819 Hanja chars, 50 variant groups

The number of variant groups will probably change (increase) according to the discussion and conclusion between KGP and CGP
A list of Hangul Syllables for K-LGR v0.5 (2016.09.28.)
- 11172 Hangul Syllables (U+AC00 ~ U+D7A3) ← KS X ISO/IEC 10646

A list of Hanja characters for K-LGR v0.5 (2016.09.28.)

<table>
<thead>
<tr>
<th>Source of Hanja Character Set</th>
<th># chars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) KS X 1001 (268 comptb. chars excluded)</td>
<td>4620</td>
</tr>
<tr>
<td>2) KPS 9566</td>
<td>4653</td>
</tr>
<tr>
<td>3) IICORE - K column marked</td>
<td>4743</td>
</tr>
<tr>
<td>4) IICORE - KP column marked (= KPS 9566)</td>
<td>4653</td>
</tr>
<tr>
<td>5) Qualifying Test of Korean Hanja Proficiency (한국 한자 능력 검정 시험)</td>
<td>4641</td>
</tr>
<tr>
<td>K-LGR v0.5 (2016.09.28.): Hanja List (Union of 1), 2), 3), 4), and 5))</td>
<td>4819</td>
</tr>
</tbody>
</table>
3518 C vg’s in C-LGR (2016.07.20.) analyzed based on K chars in K-LGR v0.5 (2016.09.28.)

- K-LGR v0.5 (2016.09.28.): 4819 Hanja chars and 50 variant groups
- C-LGR (2016.07.20.): 19738 Hanzi chars and 3518 variant groups
  (a variant group (set) is composed of two or more variant chars)

Analysis of 3518 C (Chinese) variant groups (sets)

- K extracted 304 C variant groups where there are two or more K characters
  → K need to review those 304 C vg’s
    - K character is a character belonging to K-LGR v0.5 (2016.09.28.)
  - No or just one K char in the remaining 3214 (= 3518 – 304) C vg’s
    → K need NOT review those 3214 C vg’s

Korea classified 304 C variant groups into acceptable/unacceptable categories. Summarized in the following table
<table>
<thead>
<tr>
<th>K position</th>
<th># C variant groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>acceptable</td>
<td>46</td>
</tr>
<tr>
<td>unacceptable</td>
<td>258</td>
</tr>
<tr>
<td>total</td>
<td>304</td>
</tr>
</tbody>
</table>

- KGP and CGP are discussing to decrease the number 258 to 50 or so.

Possible scenarios:

**scenario 1)** K accepts C position: C vg is maintained
- [K: indep (C1), indep (C2)]; [C: vg (C1, C2)] → [K & C: vg (C1, C2)]

**scenario 2)** C accepts K position: C vg will be modified. (usually) one char is removed from C vg and that char becomes an independent char.
- [K: indep (C4), indep (C5)]; [C: vg (C4, C5, C6)]; Note. C6 is NOT a K char
  → [K: indep (C4), indep (C5)]; [C: indep (C4), vg (C5, C6)]
A special class of variant groups in C-LGR

About 56 "Simplitional chars": [= SIMPLIfied + tradiTIONAL]

- Currently, the char is a simplified char in China
- However, the char has been used for a long time in Korea, China, etc. before PRC announced simp. chars in 1964 → a traditional char
- An example of Simplitional char: 机

1) In China:
   - 机: Currently, Simplified char, "machine"

   Simplified from Traditional char 機 (machine).

2) In Korea: the two chars are distinct
   - 机: desk (reading "gwe")
   - 機: machine (reading "gi")

It is very hard for K to accept (most) vg’s containing one of those 56 "Simplitional chars“

- The number 56 might be decreased by 10 or so
- E.g., C variant group containing a simplitional char 91CC 里 [K: vg (88CF 裏, 88E1 裏), indep (91CC 里)] [C: vg (88CF 裏, 88E1 裏, 91CC 里)]
  → C included 91CC 里 in C vg since it is a simplified char of traditional characters 88CF 裏 and 88E1 裏 in Chinese community.
  K position: 91CC 里 is completely different from (88CF 裏 = 88E1 裏) in meaning, usage, etc. in Korean community.
- CJK coordination meeting in Taipei, Taiwan in Sep. 2016:
  → KGP and CGP tried to reduce the number of unacceptable vg’s.
  → much progress although not finalized yet
KGP’s Activities History (1)

2013
Dec: organization of Korean LGP

2014
Mar: Participated in CJK joint meeting @ ICANN49 Singapore
Jun: Participated in ICANN50 @ London; KGP status update
Jun: 1st KGP meeting
Aug: 2nd KGP meeting
Oct: Participated in ICANN51 LA; KGP status update

2015
Jan: 3rd KGP meeting; KGP re-composed
Feb: Participated ICANN52 @ Singapore; KGP status update
Apr: 4th and 5th KGP meetings; KGP reorganized
May: 6th and 7th KGP meetings (K-LGR-1 v0.1); CJK Joint meeting in Seoul
Jun: 8th KGP meeting (K-LGR, v0.2); participated in ICANN53 @ Buenos Aires
Jul: 9th KGP meeting and workshop; participated in APrIGF Macau
Aug: 10th KGP meeting (K-LGR, v0.3)
Sep: 11th KGP meeting
Oct: Call for formal Generation Panel to ICANN and participated in ICANN54 @ Dublin
KGP’s Activities History (2)

2015
Nov: 12th KGP meeting

2016
Jan: 13th KGP meeting
Feb: The Korean Community “formally” Forms Generation Panel for Developing the Root Zone Label Generation Rules (LGR), 2016-02-01.
Mar: Participate ICANN55 @ Marrakesh, Morocco and present KGP status update
Mar: 14th KGP meeting (K-LGR v0.4)
Mar: Participate CJK coordination meeting @ Beijing
Apr: 15th KGP meeting
May: 16th KGP meeting
Jun: 17th KGP meeting
Jun: Participated in ICANN56 @ Helsinki
Jul: 18th KGP meeting
Aug: 19th KGP meeting
Sep: 20th KGP meeting; (K-LGR v0.5); CJK coordination meeting @ Taipei
## KGP’s Activities History (3)

### 2016
- **Oct:** 21\(^{st}\) and 22\(^{nd}\) KGP meetings
- **Nov:** Participating in ICANN57 @ Hyderabad, India

### 2017
- **Jan:**
Timeline of KLGP Activities

12. 2013
Organisation of KGP

05. 2015
K-LGR v0.1

06. 2015
K-LGR v0.2

10. 2015
K-LGR v0.3

02. 2016
KGP formally formed

03. 2016
K-LGR v0.4

09. 2016
K-LGR v0.5
Appendix. Hanja in K0, P0, IICORE/K, HT (Hanja Test, QTKHP)

Venn Diagram of 4 sets showing number of Hanja chars: (K-LGR v0.3, 2015.08.13.)

K0 (KS X 1001), P0 (KPS 9566), IK (IICORE: K), HT (Hanja Test) klgp168_2b_v03
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