
ICANN|57
HYDERABAD

IDN Root Zone LGR Workshop

IDN Program | ICANN 57 | 6 November 2016

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 Kyongsok
- Q/A

LGR Toolset

Marc Blanchet
Viagenie

Introduction to LGR Toolset

- 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.

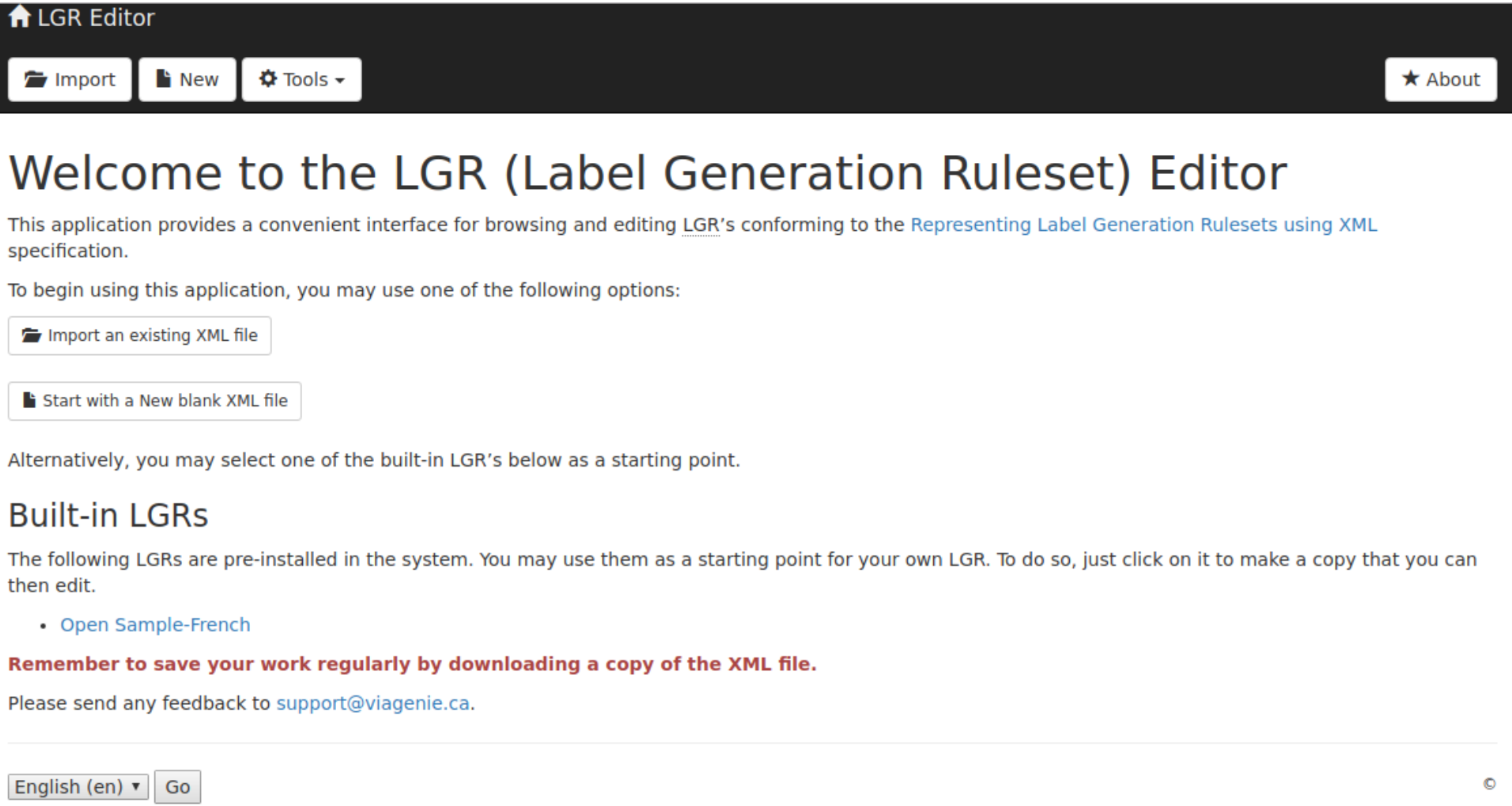
- Online beta deployment
  - Visit https://lgrtool.icann.org/
  - Username: lgr and password: 37zEfM2LyN3DmSzjLaYoA

- Open source package(s) released with BSD license
  - Released at github: lgr-core, lgr-django, munidata
  - Credits to developers: Audric Schiltknecht, Wil Tan, Julien Bernard, David Drouin

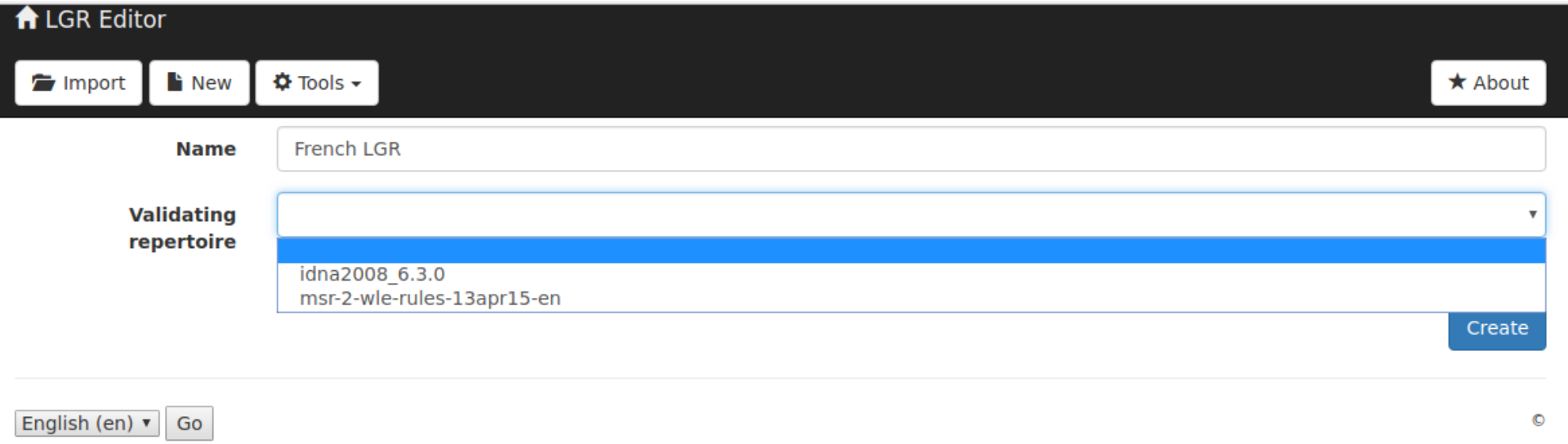
Walk-Through Example

- Create a French LGR
- Add Ligatures: æ, œ (U+00E6, U+0153)
- Validate with list of labels
- Union/Diff of LGRs

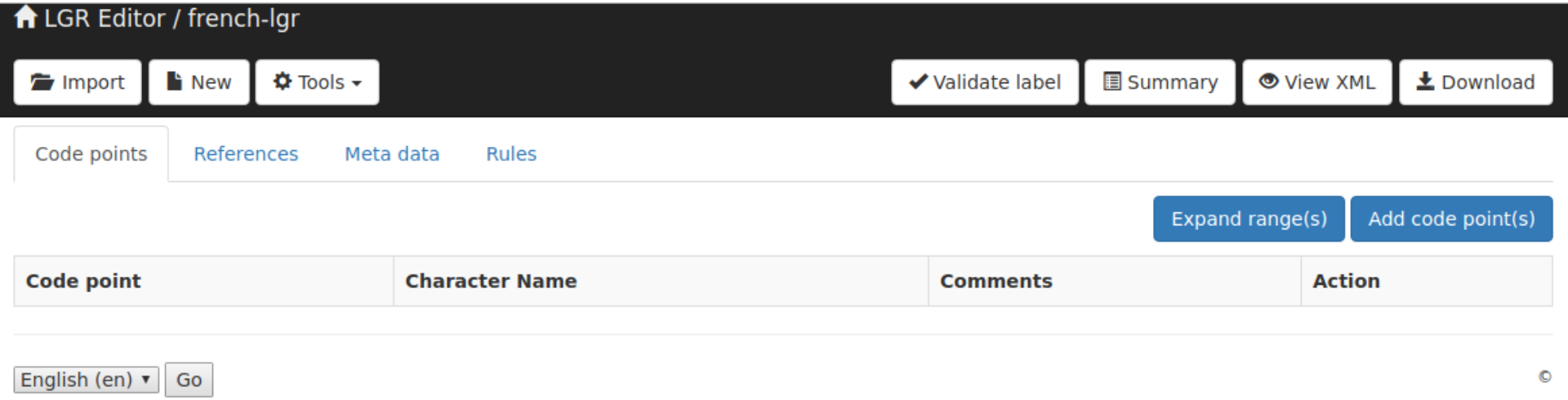
Home Page


LGR Editor
Import
New
Tools
About
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.
English (en)
Go
©

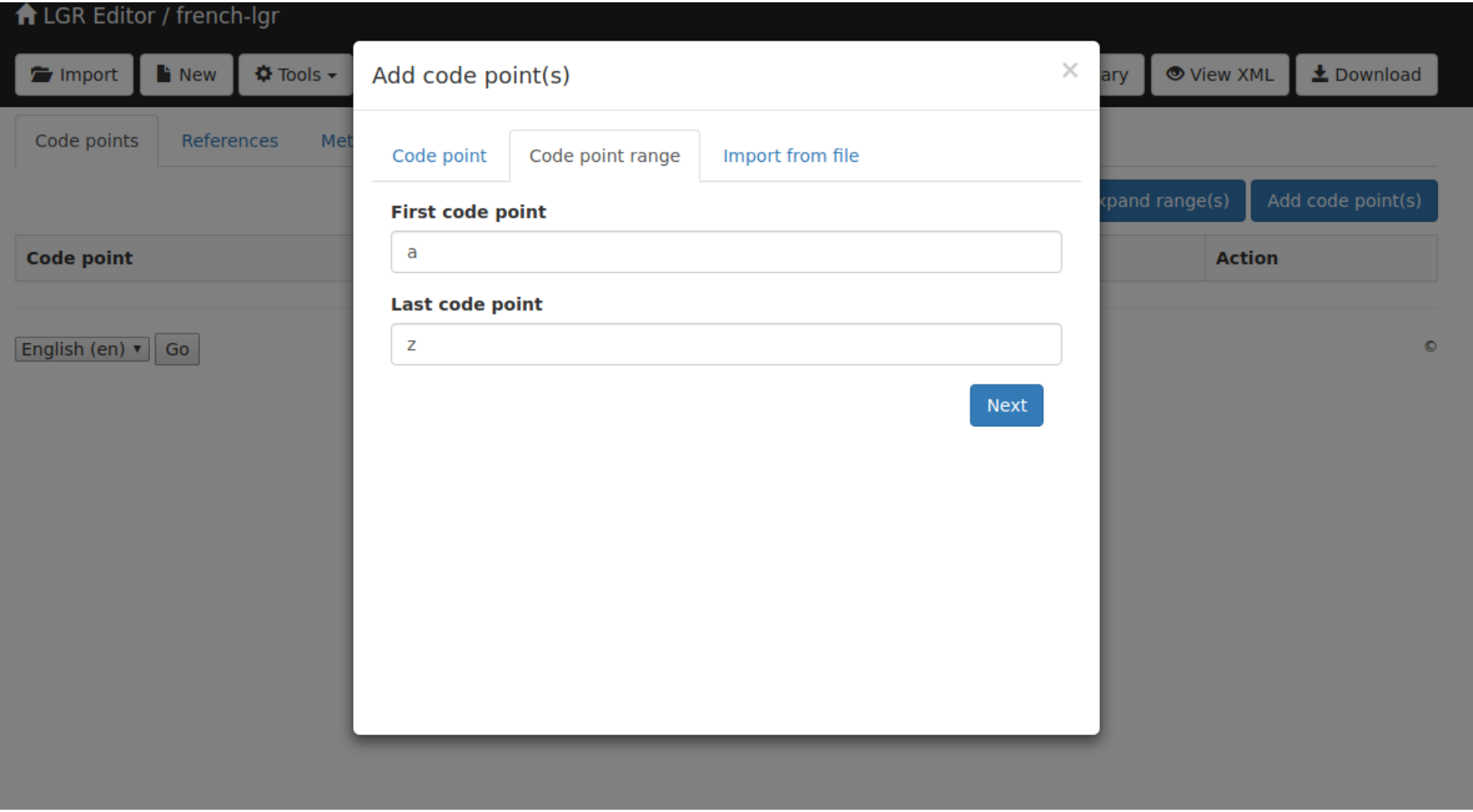
Create LGR


LGR Editor
Import
New
Tools
About
Name
French LGR
Validating repertoire
idna2008_6.3.0
msr-2-wle-rules-13apr15-en
Create
English (en)
Go

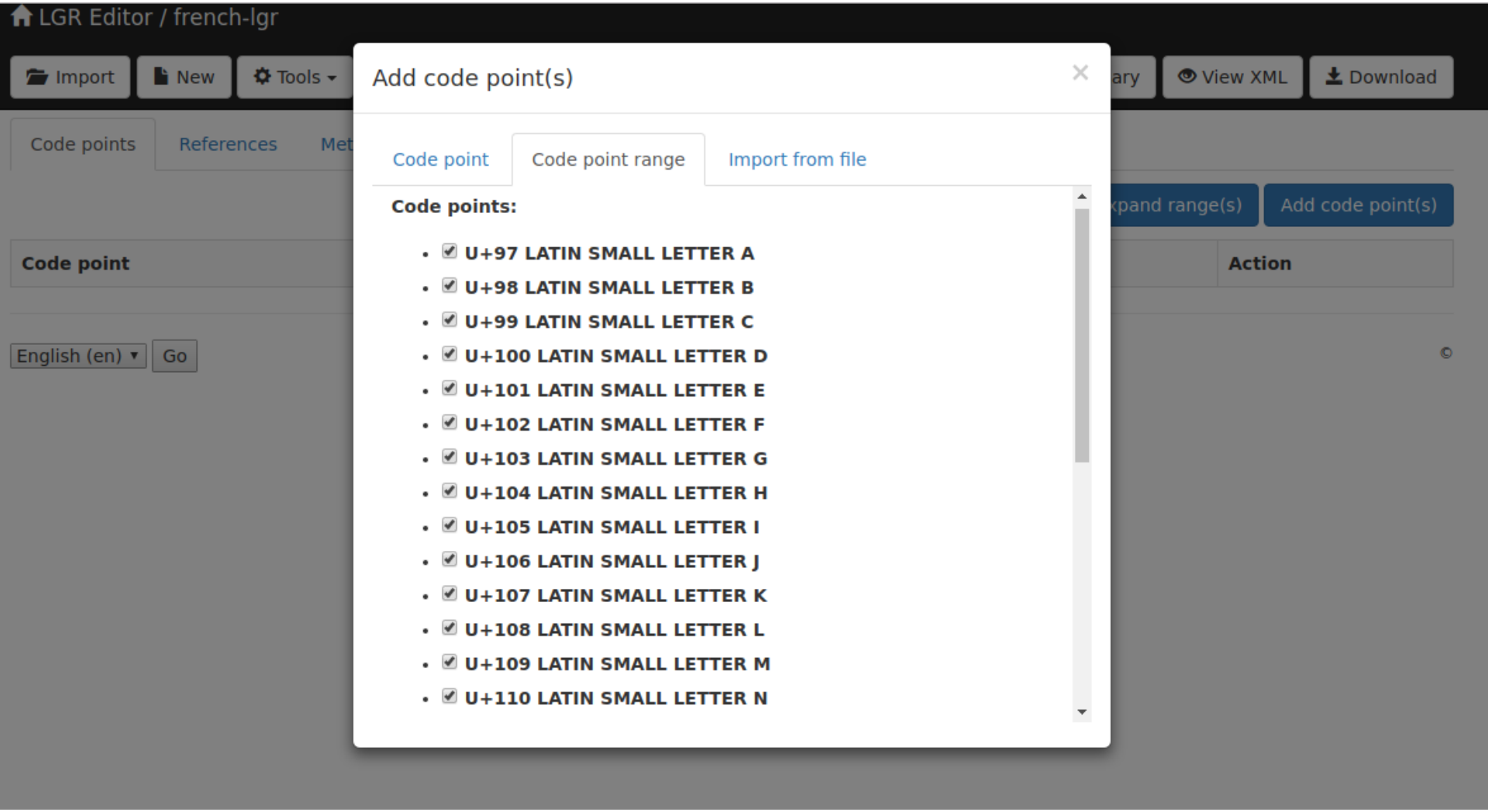
Empty LGR


LGR Editor / french-lgr
Import
New
Tools
Validate label
Summary
View XML
Download
Code points
References
Meta data
Rules
Expand range(s)
Add code point(s)
Code point
Character Name
Comments
Action
English (en)
Go
©

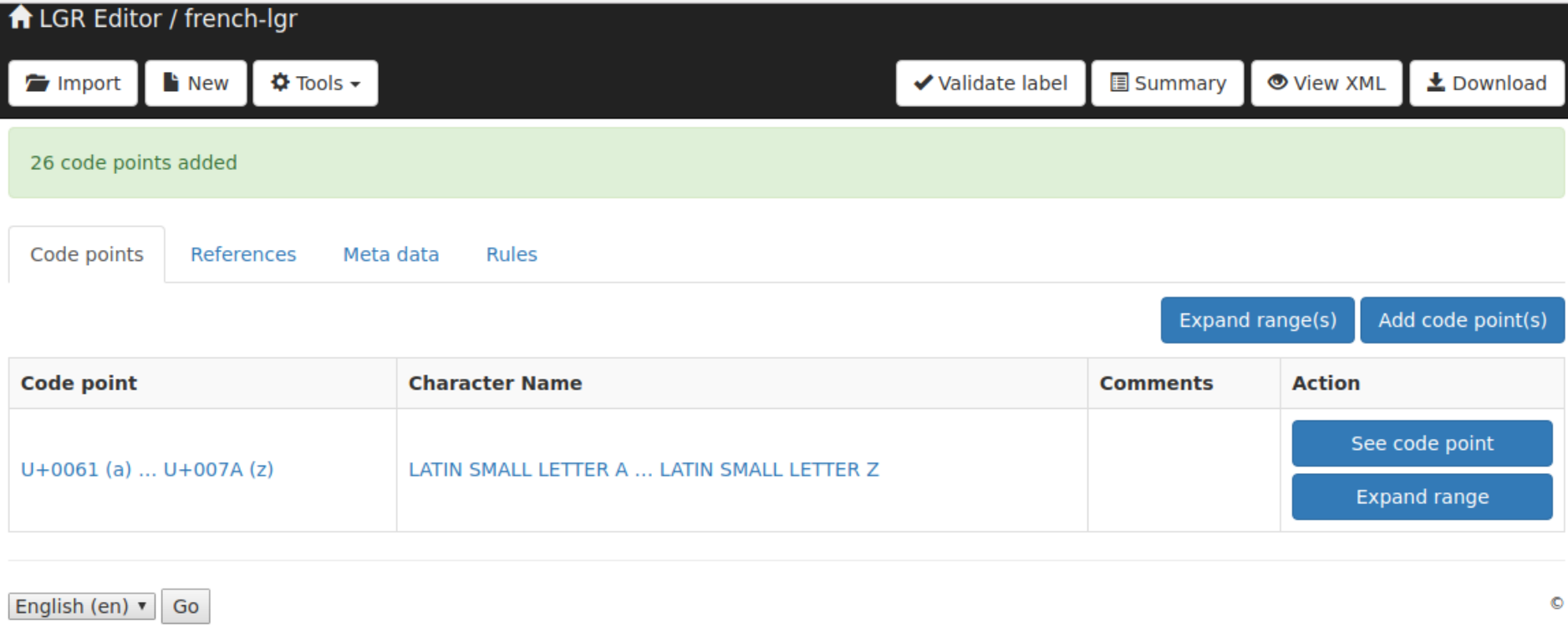
Add Range


LGR Editor / french-lgr
Import
New
Tools
View XML
Download
Code points
References
Add code point(s)
Code point
Code point range
Import from file
First code point
a
Last code point
z
Next
Add code point(s)
Code point
Action
English (en)
Go

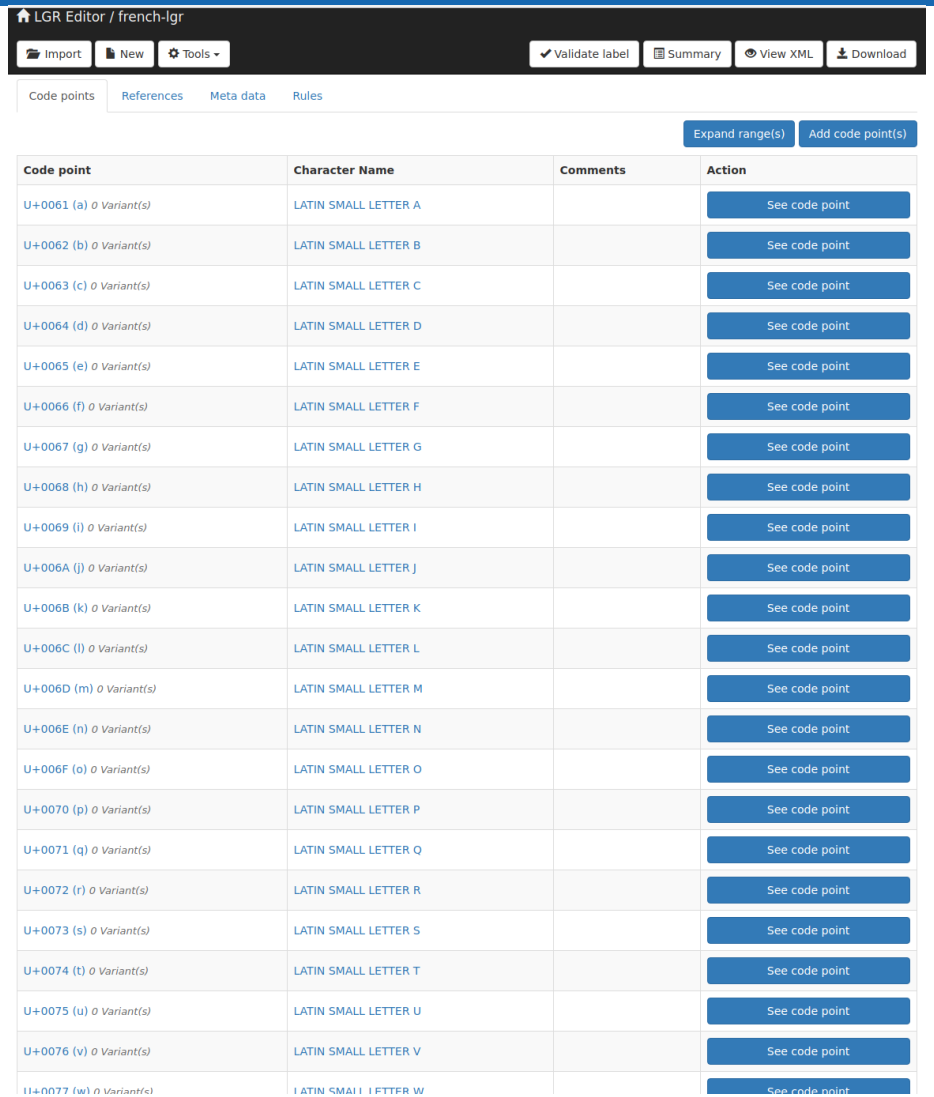
Add Range - Validation



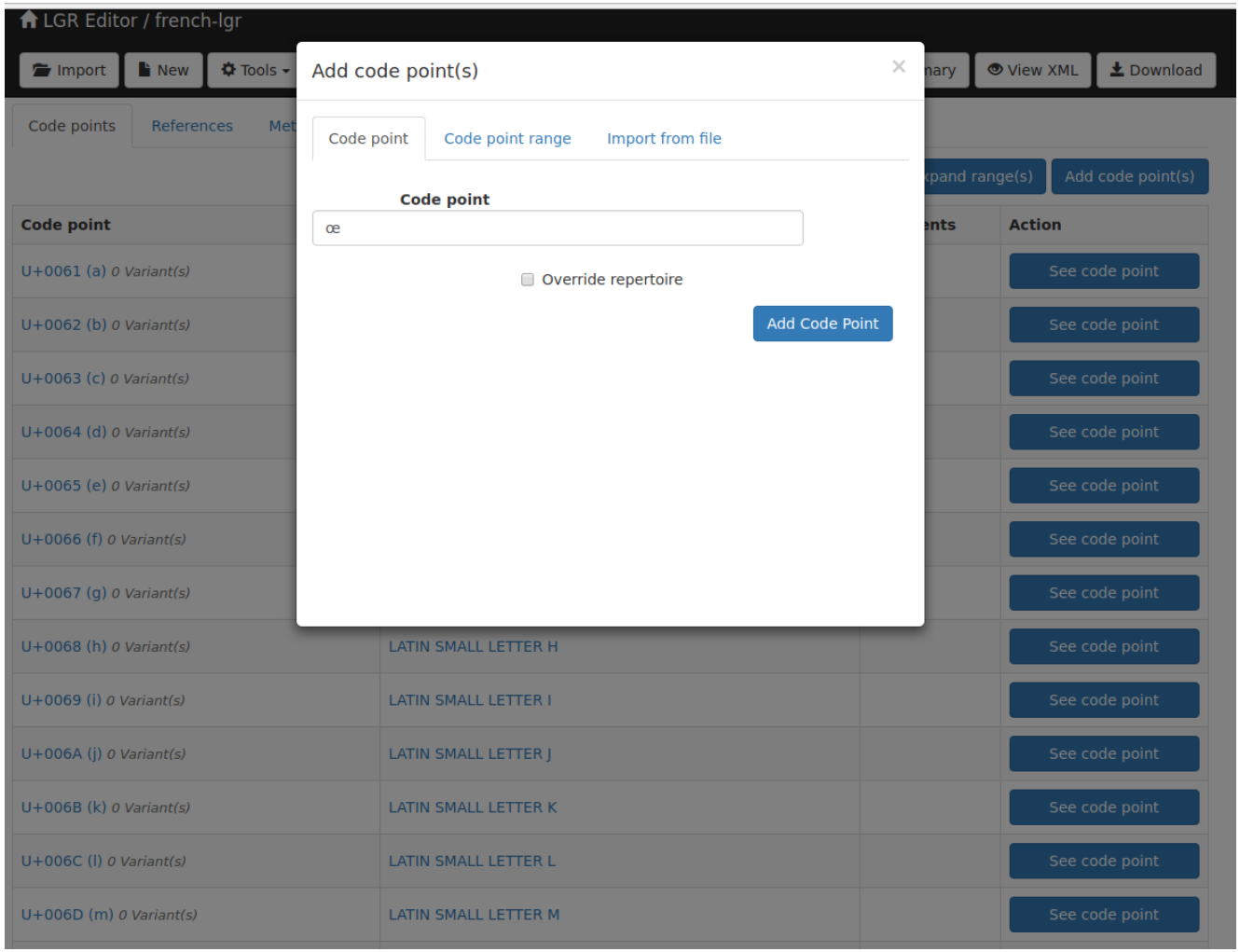
Range in LGR


LGR Editor / french-lgr
Import
New
Tools
Validate label
Summary
View XML
Download
26 code points added
Code points
References
Meta data
Rules
Expand range(s)
Add code point(s)
Code point
Character Name
Comments
Action
U+0061 (a) ... U+007A (z)
LATIN SMALL LETTER A ... LATIN SMALL LETTER Z
See code point
Expand range
English (en)
Go
©

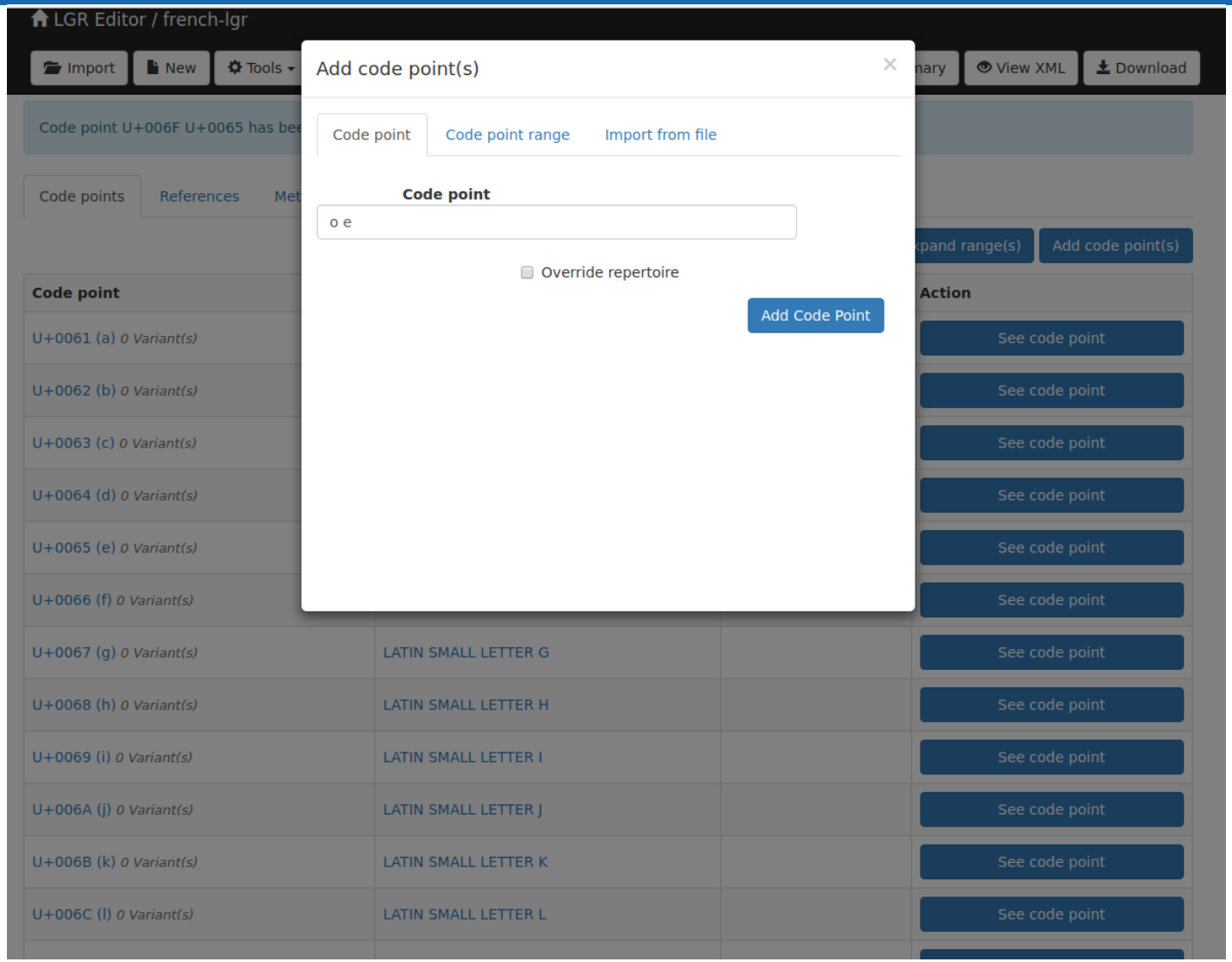
Expanded Range



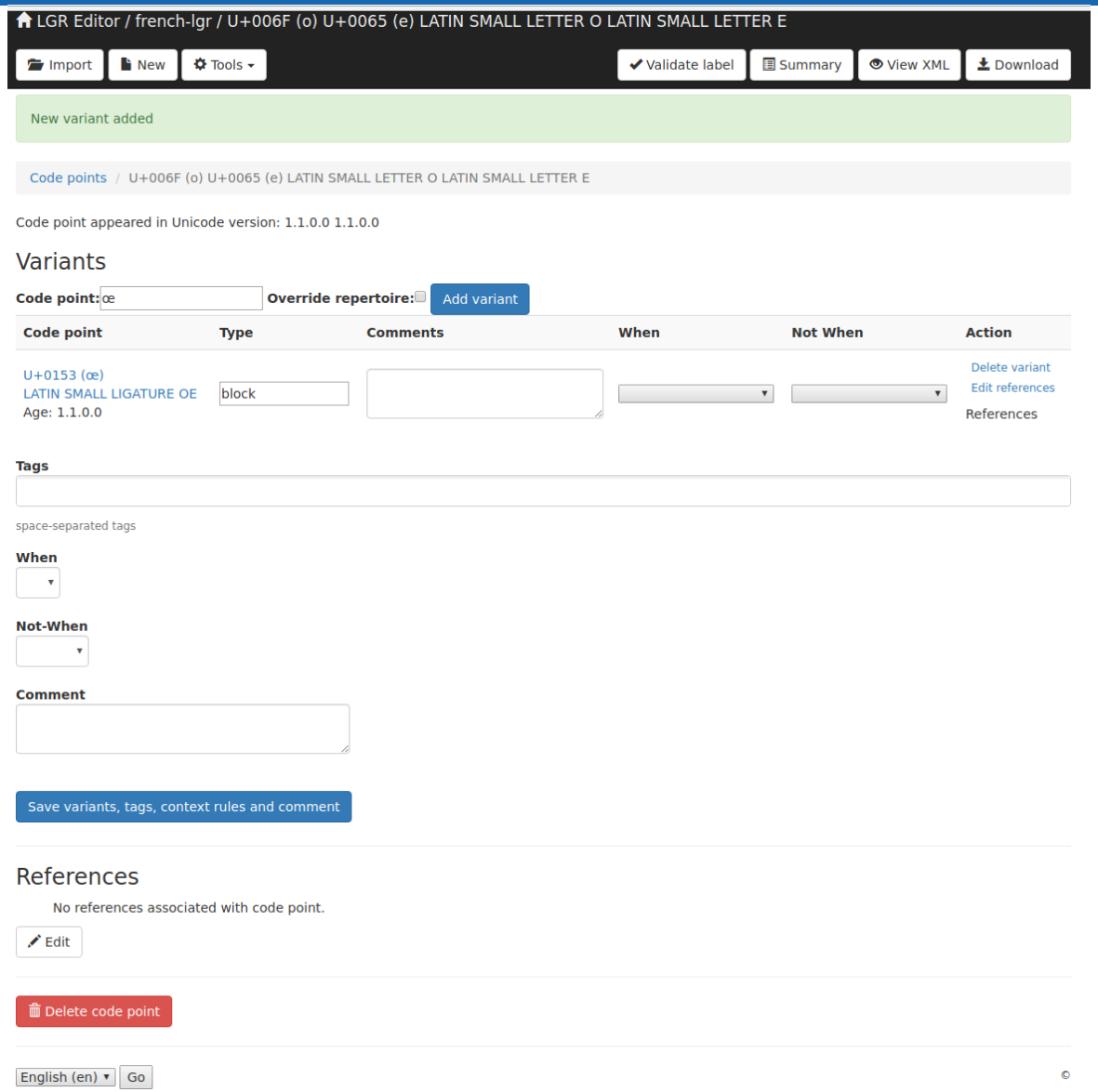
Add Code Point (œ)



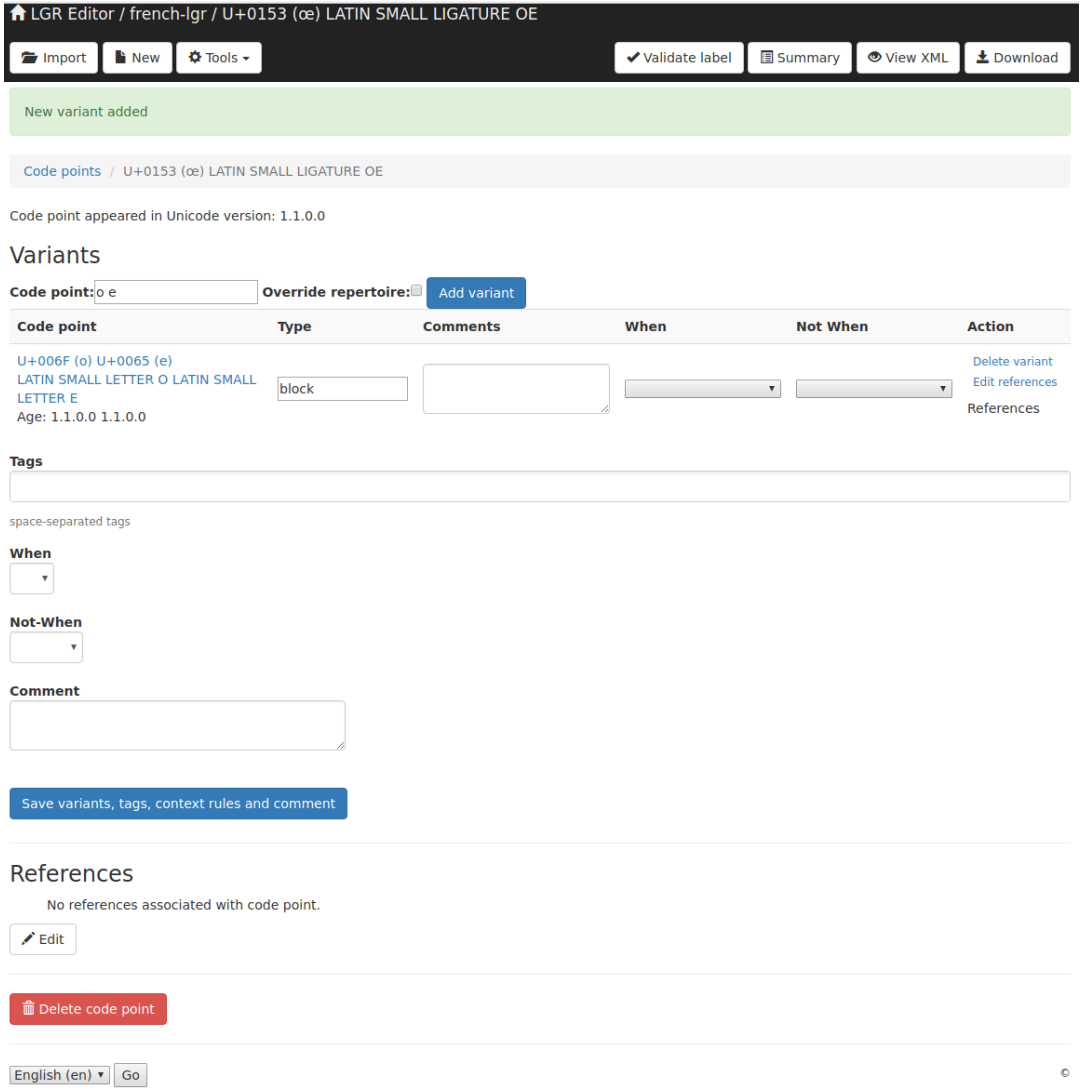
Add Code Point Sequence (o e)



Add Variant œ to (o e)



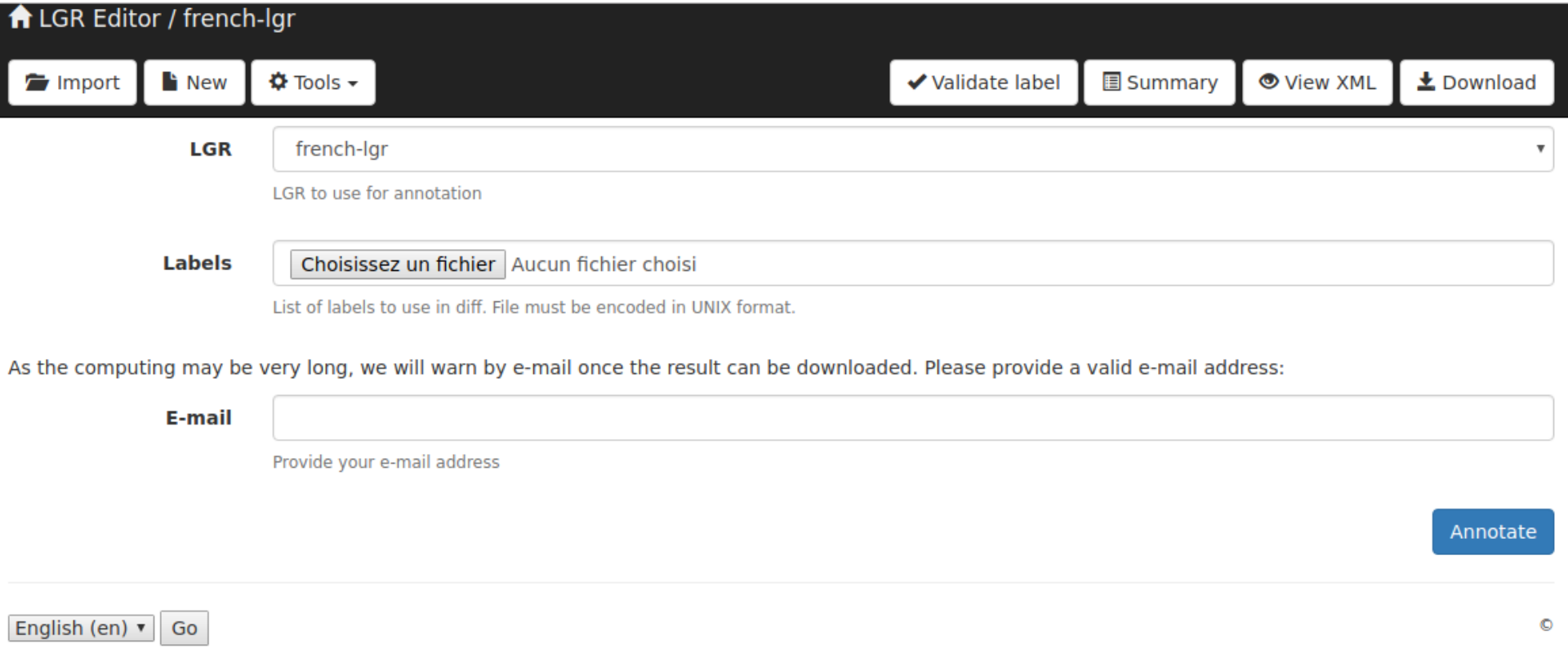
Add Variant (o e) to œ



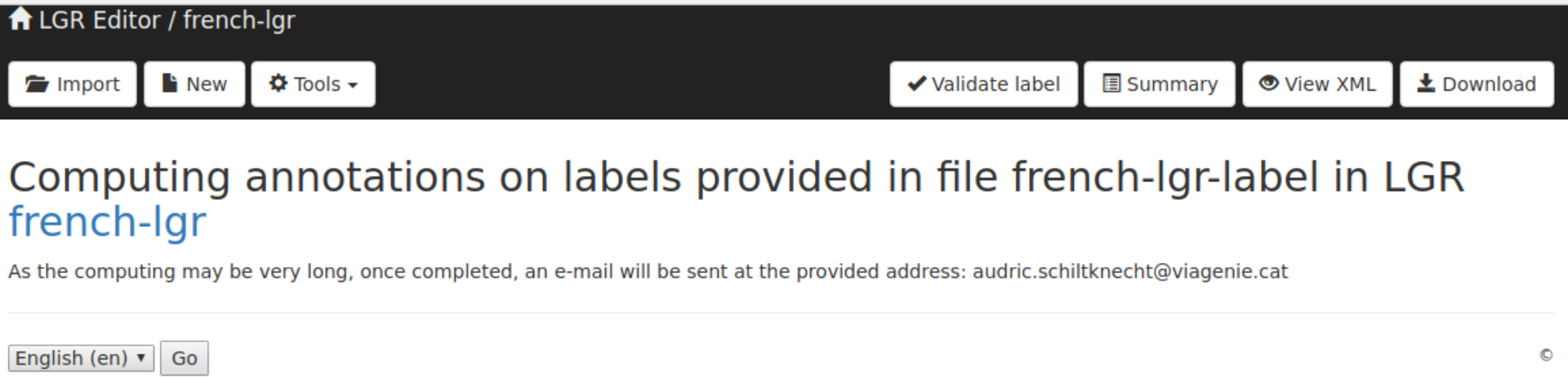
Label List

bonjour
œuf
œuf
bœuf
bœuf
æquo
aequo
cætera
caetera

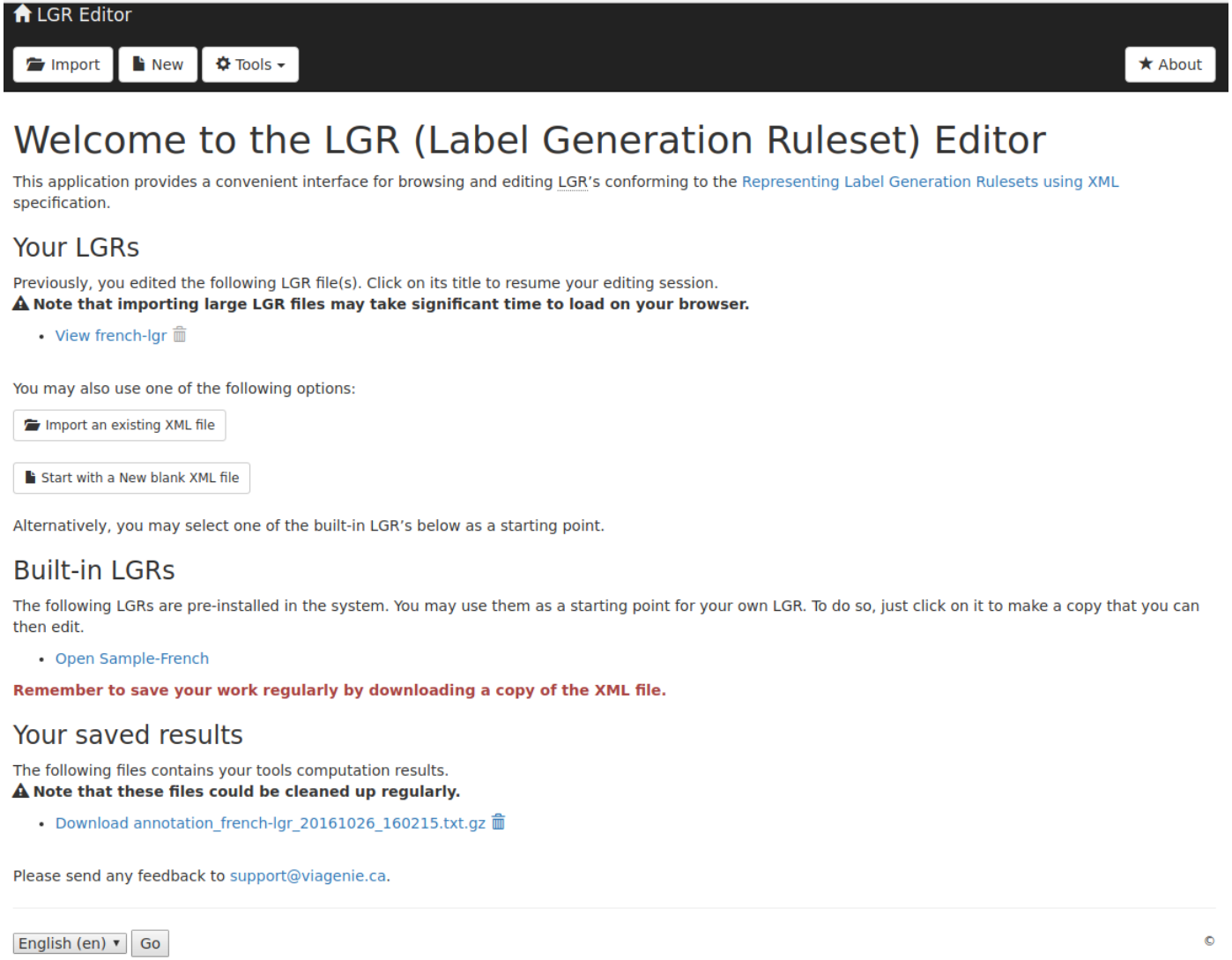
Annotation


LGR Editor / french-lgr
Import
New
Tools
Validate label
Summary
View XML
Download
LGR
french-lgr
LGR to use for annotation
Labels
Choisissez un fichier
Aucun fichier choisi
List of labels to use in diff. File must be encoded in UNIX format.
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:
E-mail
Provide your e-mail address
Annotate
English (en)
Go
©

Annotation (cont.)



Annotated Appearing in Home Page



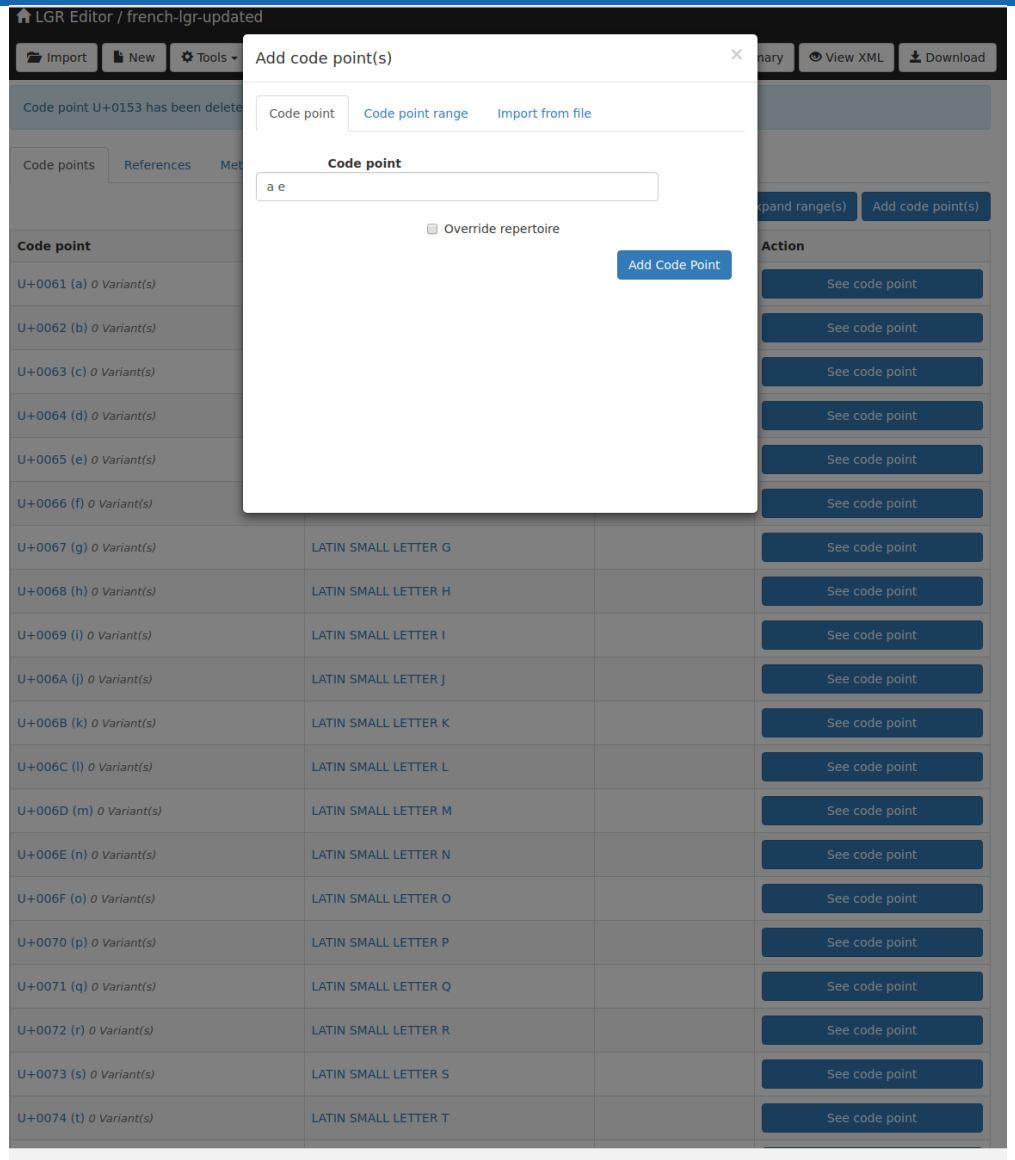
Annotated Result (æ invalid)

bonjour: allocate
oeuf: allocate
œuf: allocate
boeuf: allocate
bœuf: allocate
æquo: invalid
aequo: allocate
cætera: invalid
caetera: allocate

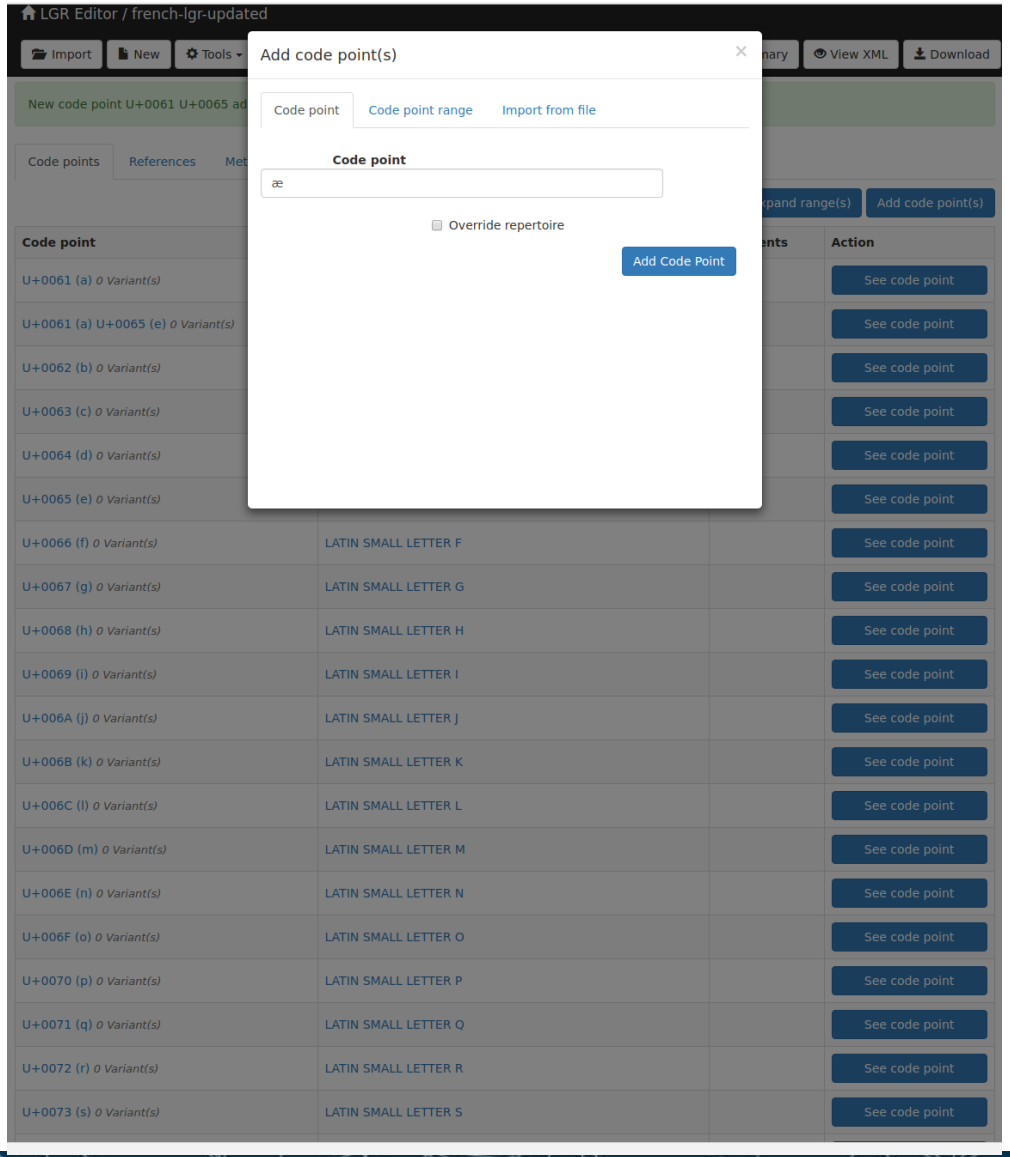
Create Another LGR

- (just for the purpose of showing the union and diff tools)
- Create another LGR, add a-z (do not add “oe”)

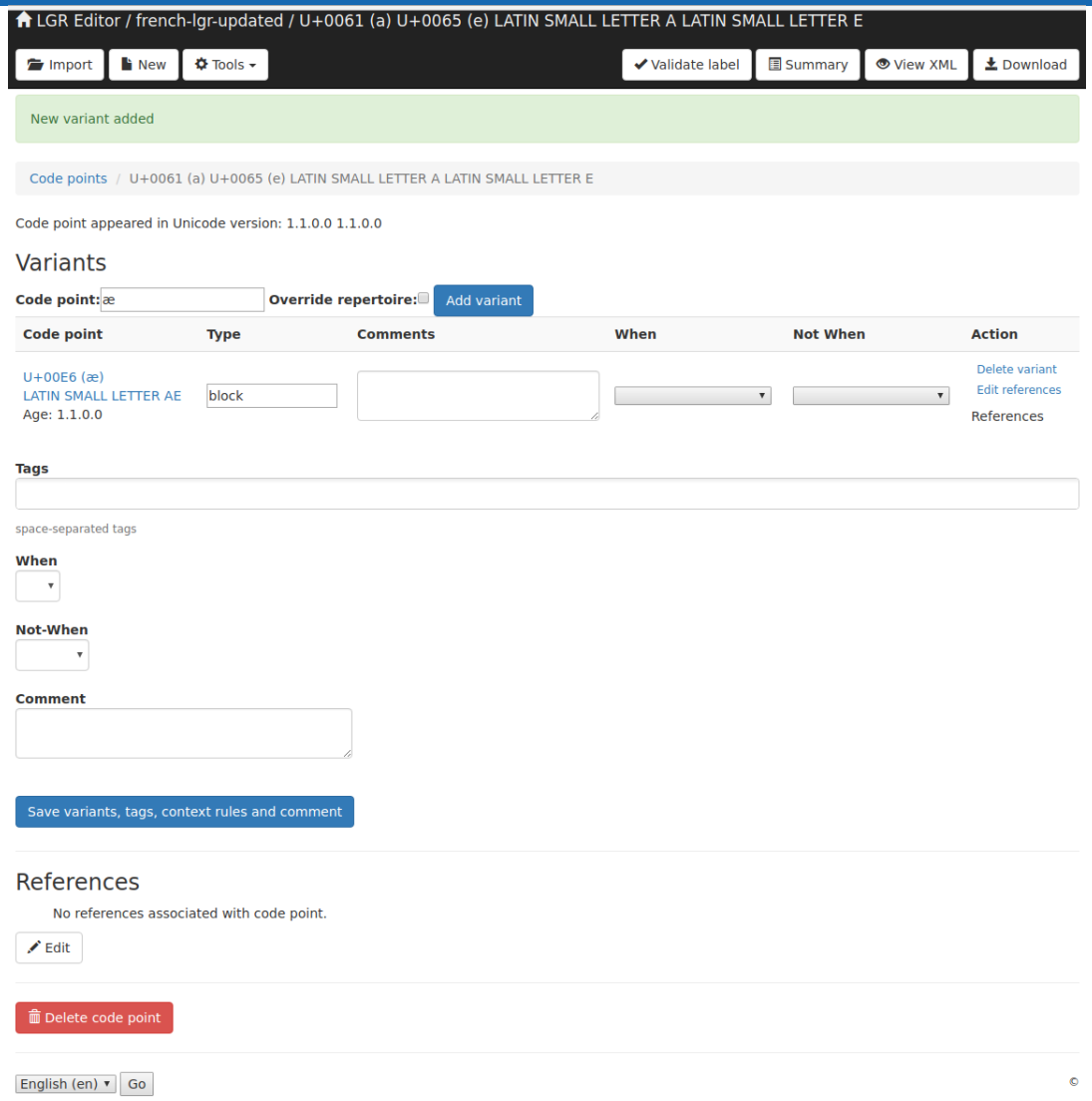
Add Code Point Sequence (a e)



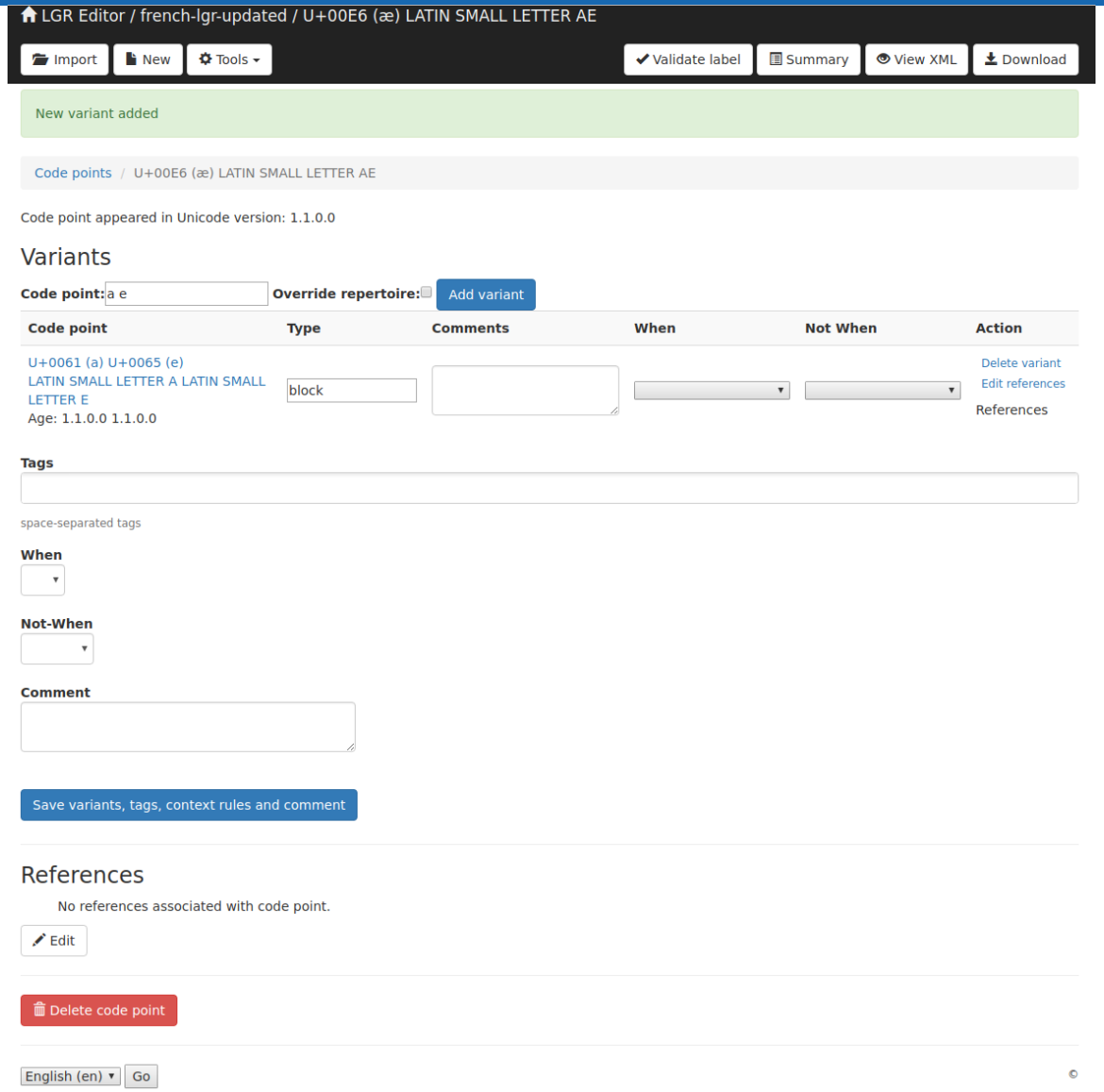
Add Code Point (æ)



Add Variant æ to (a e)



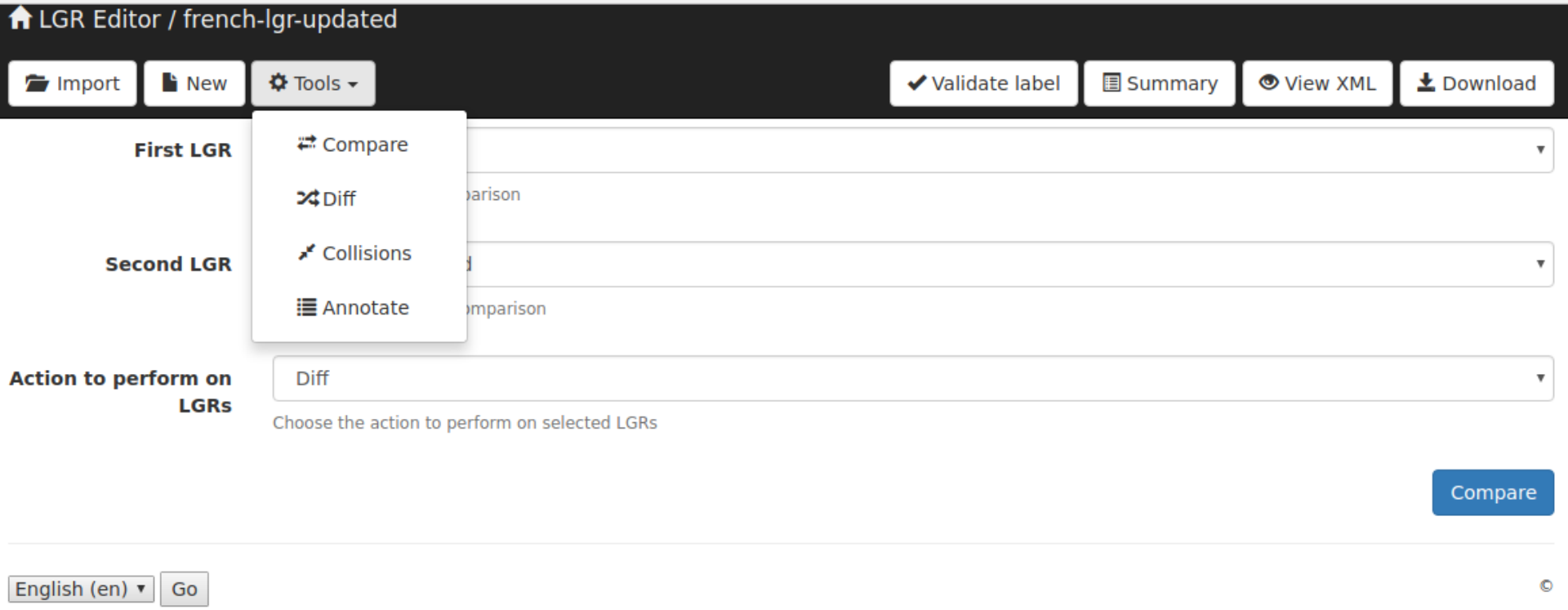
Add Variant (a e) to æ



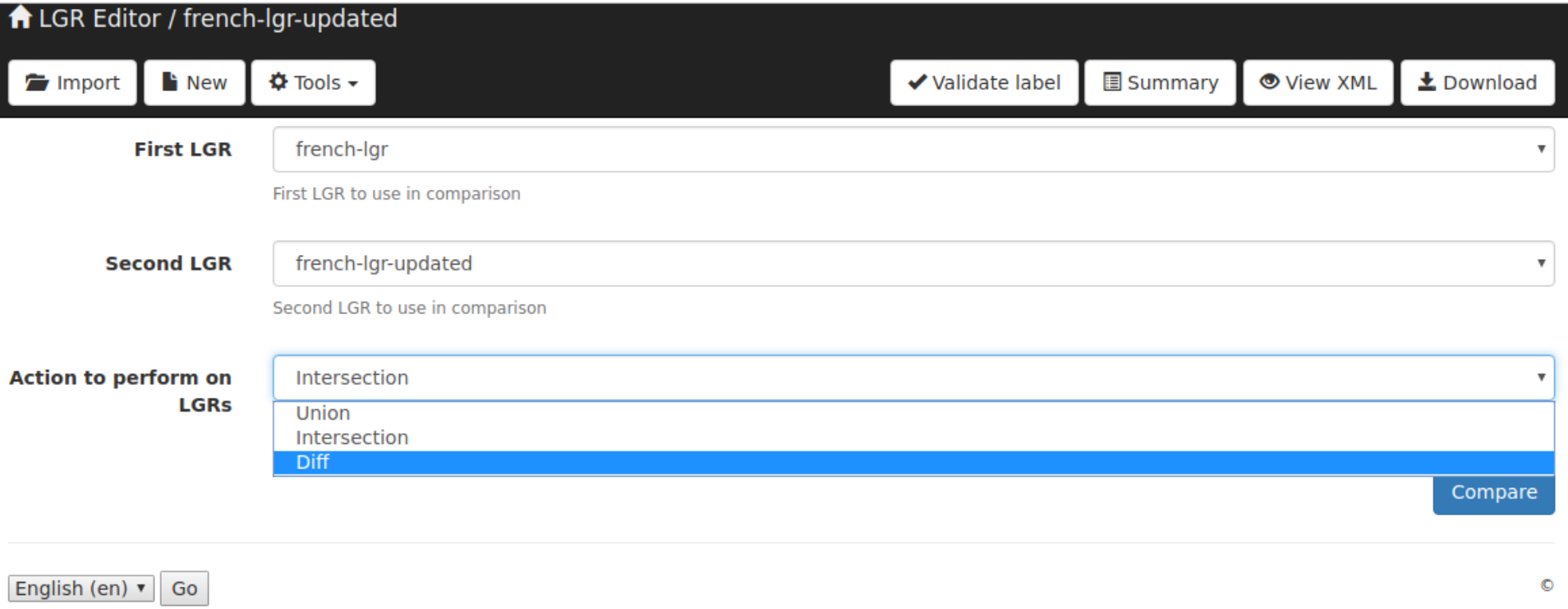
Annotated Result (œ invalid)

bonjour: allocate
oeuf: allocate
œuf: invalid
boeuf: allocate
bœuf: invalid
æquo: allocate
aequo: allocate
cætera: allocate
caetera: allocate

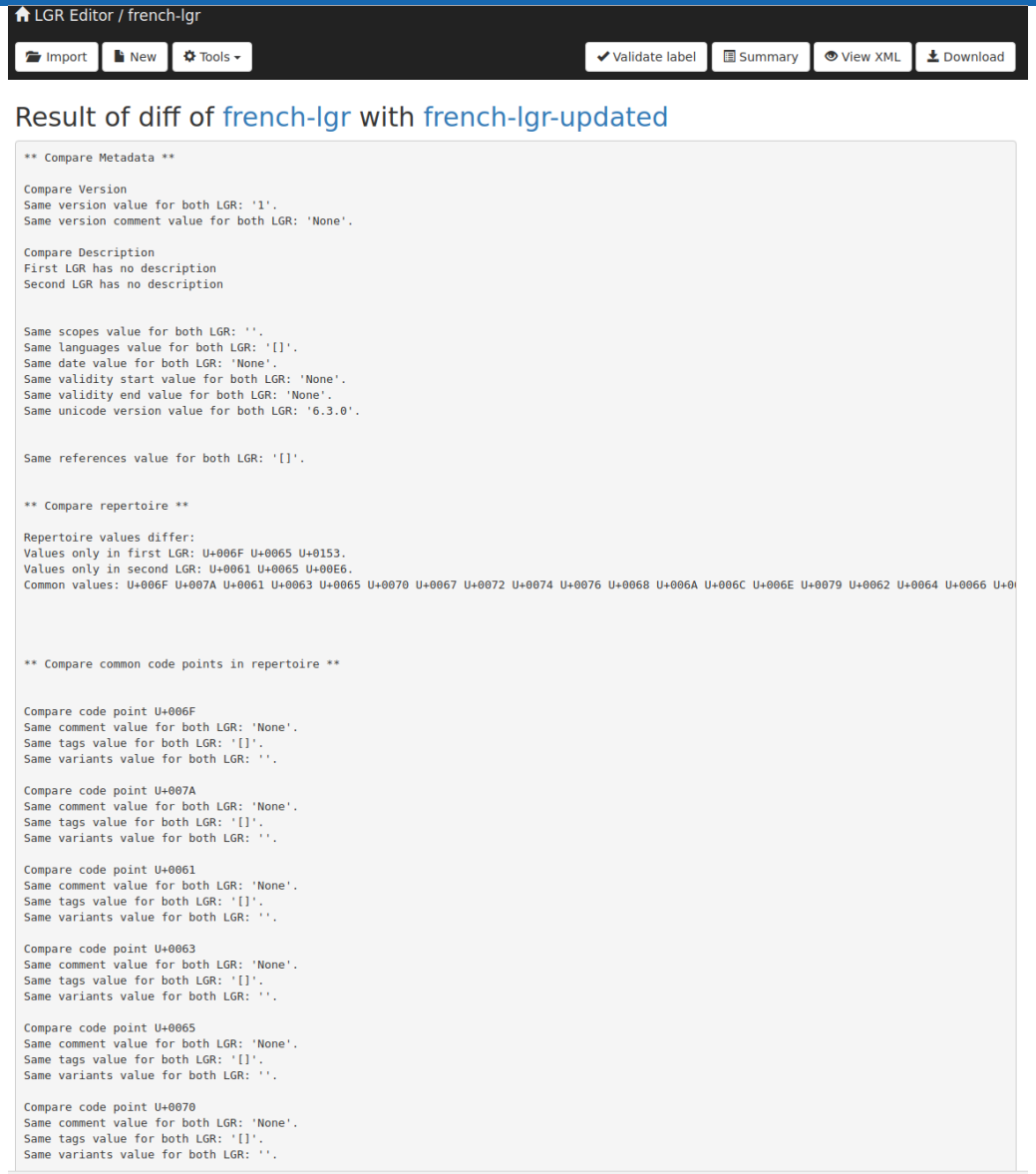
LGR Processing Tools



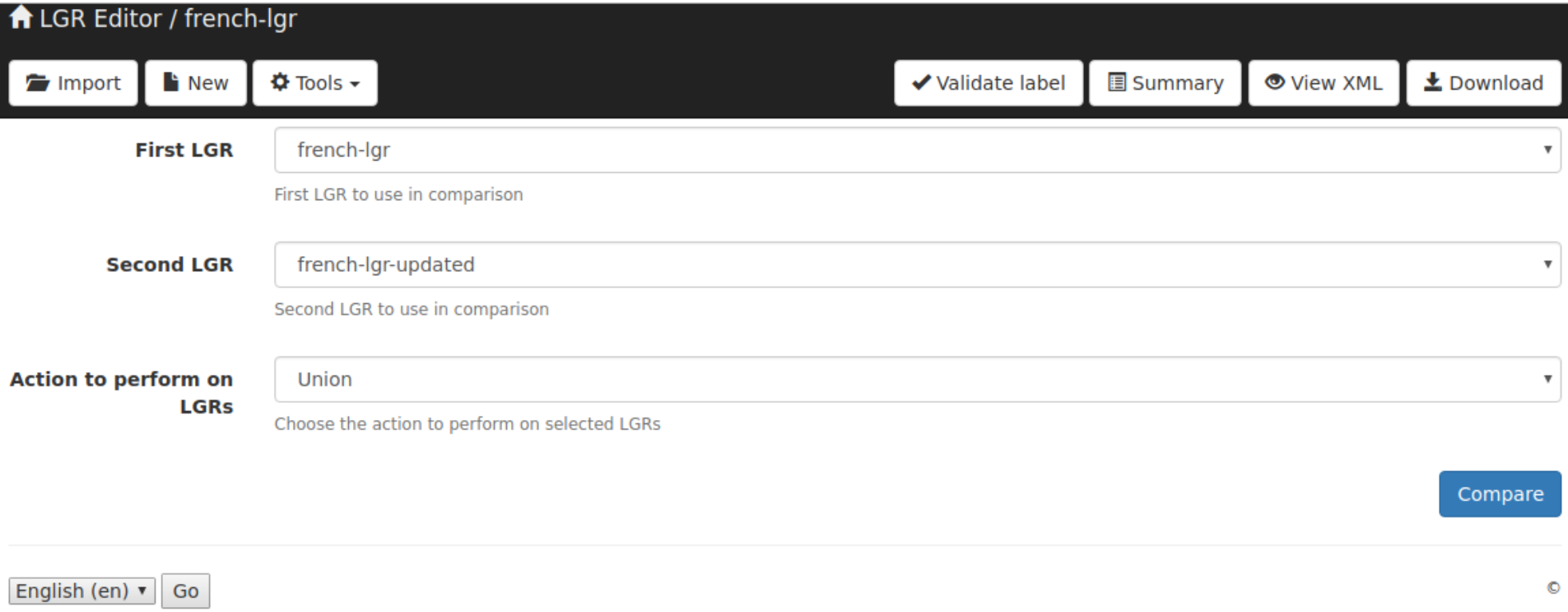
Compare


LGR Editor / french-lgr-updated
Import
New
Tools
Validate label
Summary
View XML
Download
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
English (en)
Go

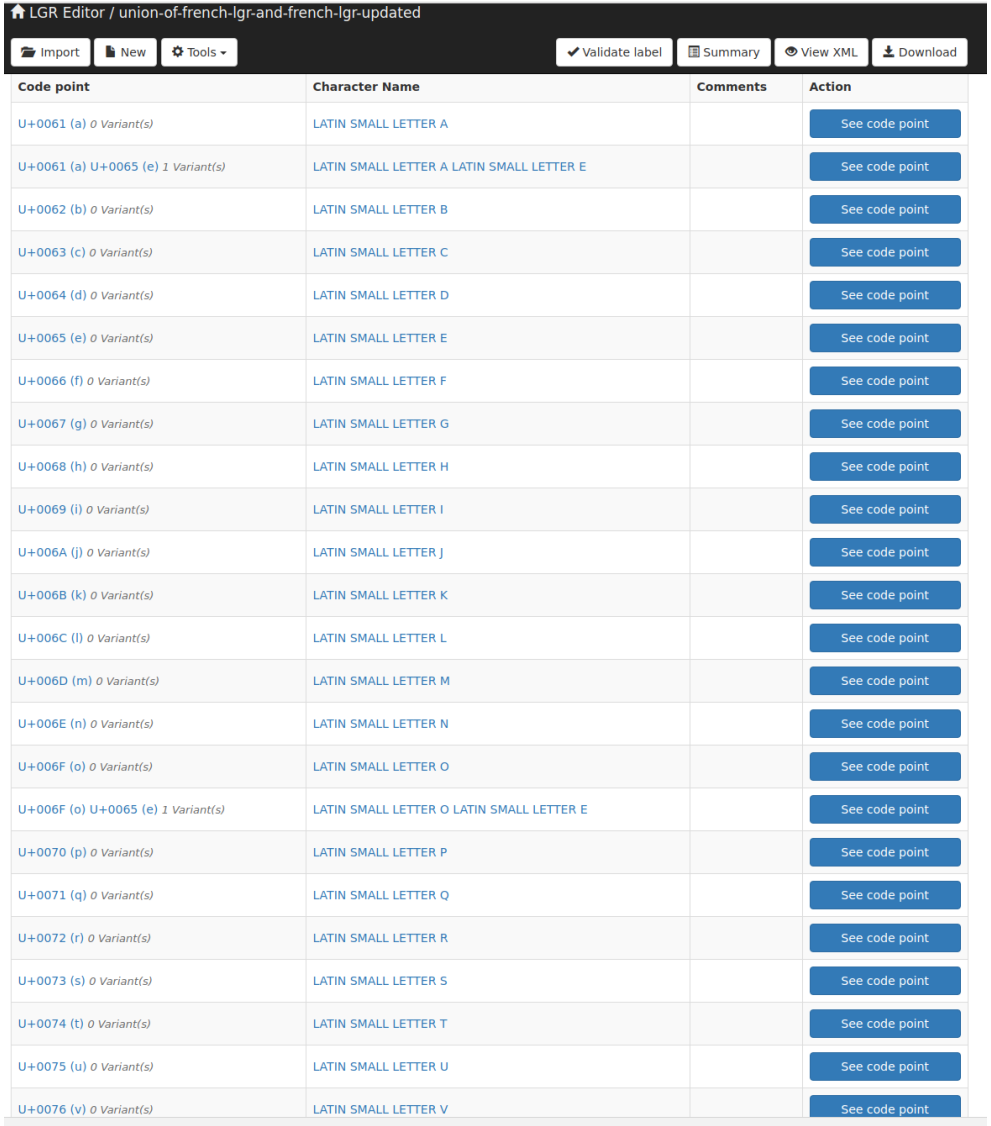
Result of Diff



Union of LGRs



Result of LGR Union



Code point	Character Name	Comments	Action
U+0061 (a) *0 Variant(s)*	LATIN SMALL LETTER A		See code point
U+0061 (a) U+0065 (e) *1 Variant(s)*	LATIN SMALL LETTER A LATIN SMALL LETTER E		See code point
U+0062 (b) *0 Variant(s)*	LATIN SMALL LETTER B		See code point
U+0063 (c) *0 Variant(s)*	LATIN SMALL LETTER C		See code point
U+0064 (d) *0 Variant(s)*	LATIN SMALL LETTER D		See code point
U+0065 (e) *0 Variant(s)*	LATIN SMALL LETTER E		See code point
U+0066 (f) *0 Variant(s)*	LATIN SMALL LETTER F		See code point
U+0067 (g) *0 Variant(s)*	LATIN SMALL LETTER G		See code point
U+0068 (h) *0 Variant(s)*	LATIN SMALL LETTER H		See code point
U+0069 (i) *0 Variant(s)*	LATIN SMALL LETTER I		See code point
U+006A (j) *0 Variant(s)*	LATIN SMALL LETTER J		See code point
U+006B (k) *0 Variant(s)*	LATIN SMALL LETTER K		See code point
U+006C (l) *0 Variant(s)*	LATIN SMALL LETTER L		See code point
U+006D (m) *0 Variant(s)*	LATIN SMALL LETTER M		See code point
U+006E (n) *0 Variant(s)*	LATIN SMALL LETTER N		See code point
U+006F (o) *0 Variant(s)*	LATIN SMALL LETTER O		See code point
U+006F (o) U+0065 (e) *1 Variant(s)*	LATIN SMALL LETTER O LATIN SMALL LETTER E		See code point
U+0070 (p) *0 Variant(s)*	LATIN SMALL LETTER P		See code point
U+0071 (q) *0 Variant(s)*	LATIN SMALL LETTER Q		See code point
U+0072 (r) *0 Variant(s)*	LATIN SMALL LETTER R		See code point
U+0073 (s) *0 Variant(s)*	LATIN SMALL LETTER S		See code point
U+0074 (t) *0 Variant(s)*	LATIN SMALL LETTER T		See code point
U+0075 (u) *0 Variant(s)*	LATIN SMALL LETTER U		See code point
U+0076 (v) *0 Variant(s)*	LATIN SMALL LETTER V		See code point

Annotated Result of the Union

bonjour: allocate
oeuf: 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

An Example of Related Scripts: Abugida

- 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

An Example: Realism in Coding Lao LGR

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: ◌ົ 0EBC
   1. SEMI-CONSONANT LO - final [r] in cons cluster: ຫ [h] vs. ຫ ົ [hr]
   2. Previously, had appeared after (under) various consonants
   3. Lao spelling reforms (1960s): only after HO SUNG (ຫ) – (ຫ ົ)
   4. Modern Lao has borrowed words from other languages

      ຟ ົ (0E9F **0EBC** 0EB5) “f**r**ee” , also “**pr**o.g**r**am” and “e.lec.t**r**o.nic”

   ⊙ LGR rule for context of SEMI-CONSONANT LO must follow usage

Specification – Starting Point

- 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
- Examples:
  - Arabic: https://www.icann.org/en/system/files/files/arabic-lgr-proposal-18nov15-en.pdf
  - Khmer (draft): https://www.icann.org/en/system/files/files/proposal-khmer-lgr-15apr16-en.pdf

Code Points – Specification

- 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
    - ฤๅ (U+0E24 U+0E45)
    - ฦๅ (U+0E26 U+0E45)
    - Adding singletons ฤ and ฦ, but not ◌ๅ — ensures that ◌ๅ can only be used after ฤ and ฦ

Documentation – Consistency

- 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

Variant Set (cont.)

- 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

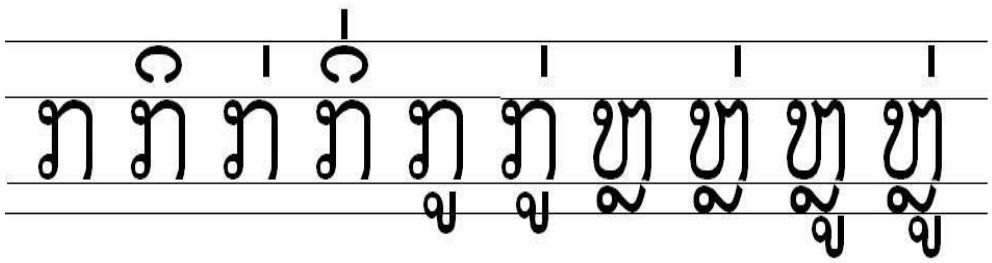
Agenda

- 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

		X5					
		X4					
X0	X1	X	X6	X7	X8	X9	X10
		X2					
		X3					



ເຫຼືອມ ເຮືອນ ໂຮງຮຽນ

Overview of Lao Generation Panel

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

Overview of Lao Generation Panel Cont.

No.	Name and Surname	Organization	Role	Expertise
8.	Mr. Thonglor Douansouvanh	Vientiane times newspaper	Community member	Media
9.	Mr. Phouthong Sisavath	National Internet Center	Technical member	DNS operation
10.	Ms. Phavanhna Douangboupha	National Internet Center	Technical member	Coordinator for international cooperation
11.	Mr.Khamphay Inthara	E-Government Center	Technical member	Lao localization specialist, Lao Font, Lao Keyboard project
12.	Mr. Saysomvang Souvannavong	National Internet Center	Technical member	DNS operation
13.	Mr.Phousana Silivong	E-Government Center	Technical member	Lao localization specialist, Lao Font, Lao Keyboard project

Challenges in Developing the LGR

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

- Complexity of syllable or writing structure, especially for foreign words like
  - ຫ្ງວ្ງ (Vietnamese's name), three consonants to form a consonant cluster
  - ◌ກ◌ໍ◌ະ (Lao word), two after vowel come together to form diphthong

Current Progress

1 **Develop Principles**

2 **Determine Code Point**

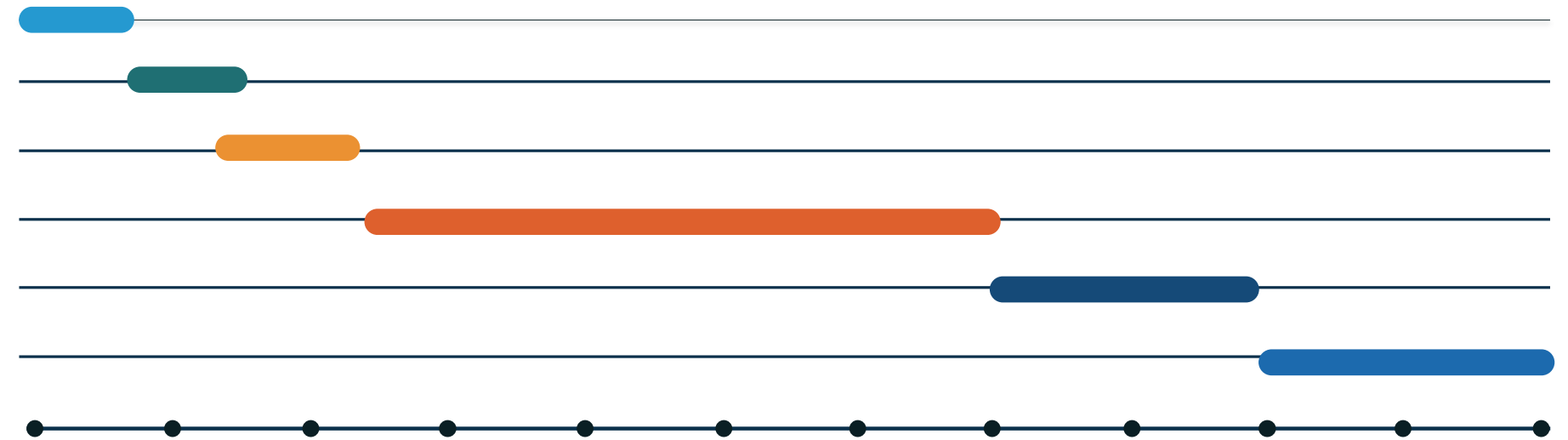
3 **Determine (any) Variants**

4 **Determine Label Rule**

5 **Write Proposal and Create XML**

6 **Submit**

Timeline – 2016



- Develop principles
- Determine the code points
- Determine (any) variants
- Determine label rules
- Write proposal and create XML
- Submit

Thank You

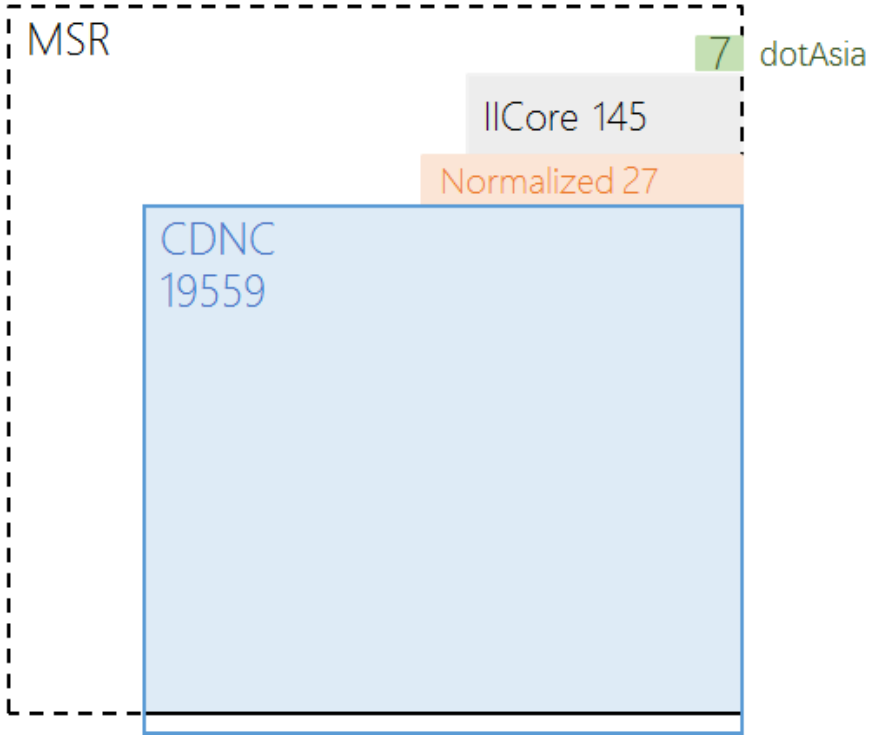
Update by the Chinese GP

Wang Wei and Kenny Huang
Chinese GP Co-Chairs

CGP LGR Proposal 1 （201606）

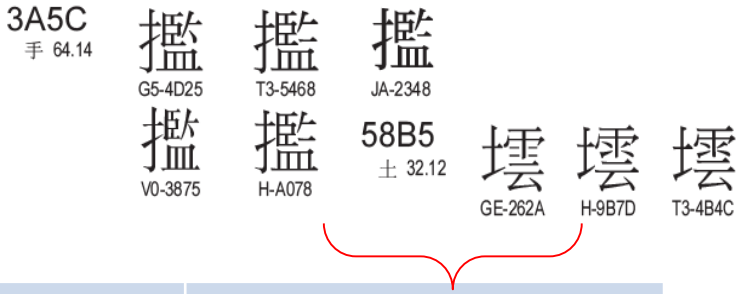
⊙ Repertoire

0	CDNC	19559
1	Normalized Hanzi for Common Use （NHCU）	27
2	IICORE	145
3	dotAsia	7
		19738

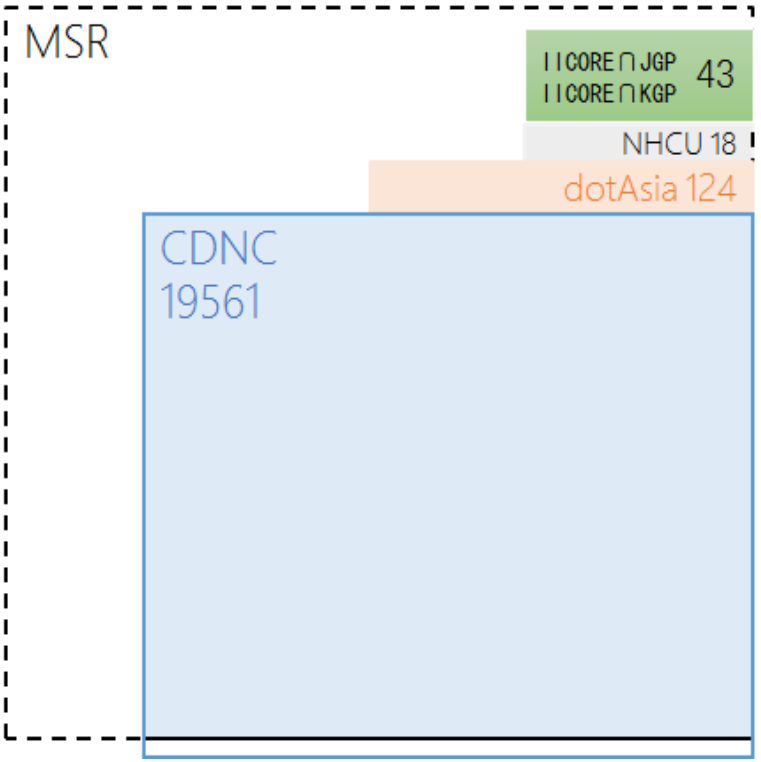


CGP LGR Proposal 2 （201610）

⊙ Repertoire



0	CDNC	19561=19559+2
1	dotAsia	124=62+62
2	NHCU	18
3	IICORE∩JGP IICORE∩KGP	43
		19746



0	CDNC 2015	19561=19559+2
1	dotAsia	124=62+62
2	NHCU	18
3	IICORE∩JGP IICORE∩KGP	43
		19746

65FF	旰	HK2015	.ASIA
4C81	鰷		.ASIA
5605	嘅	HK2015	.ASIA
6335	挻	HK2013	.ASIA
656D	敿	HK2013	.ASIA
681E	栞	HK2013	.ASIA
7460	瑠	HK2015	.ASIA
74C8	璨	HK2015	.ASIA
9771	靱	HK2015	.ASIA
34E4	刮		.ASIA
3577	咱		.ASIA
35A1	唊		.ASIA
35AD	唻		.ASIA
35BF	啷		.ASIA
35CE	喿		.ASIA
35F3	嗒		.ASIA
35FE	嚼		.ASIA
39F8	揨		.ASIA
39FE	揔		.ASIA
3A18	揹		.ASIA
3A52	撙		.ASIA
3A67	攝		.ASIA
3B39	胼		.ASIA
3DE7	煡		.ASIA
3DEB	殼		.ASIA
3E74	狻		.ASIA
3ED0	珒		.ASIA
4065	瞹		.ASIA
406A	瞕		.ASIA
40BB	碌		.ASIA
40DF	礒		.ASIA

44EA	葛		.ASIA
4606	蠵		.ASIA
47F4	踕		.ASIA
4AB8	顕		.ASIA
4C7D	鰄		.ASIA
4C85	鯥		.ASIA
4EEE	仮	HK2015	.ASIA
51B4	冴	HK2015	.ASIA
5689	嚫	HK2015	.ASIA
57DE	埕	HK2015	.ASIA
60E3	惣	HK2013	.ASIA
62A6	拆	HK2015	.ASIA
637F	捿	HK2015	.ASIA
6667	晧	HK2013	.ASIA
701E	瀞	HK2015	.ASIA
7534	由	HK2015	.ASIA
757A	畺	HK2015	.ASIA
7AC3	竃	HK2015	.ASIA
8420	萠	HK2015	.ASIA
9244	鉄	HK2015	.ASIA
932C	錬	HK2015	.ASIA
98C7	颸	HK2015	.ASIA
98E1	飡	HK2015	.ASIA
99C5	駅	HK2013	.ASIA
39DB	挙		.ASIA
3BA3	槩		.ASIA
43D3	朊		.ASIA
4443	朦		.ASIA
4882	牡		.ASIA
4C9D	鮑		.ASIA
4C9E	鮭		.ASIA

0	CDNC 2015	19561=19559+2
1	dotAsia	124=62+62
2	NHCU	18
3	IICORE∩JGP IICORE∩KGP	43
		19746

Supplementary Ideographic Plane

2070E		.ASIA
20731		.ASIA
20779		.ASIA
20C53		.ASIA
20C78		.ASIA
20C96		.ASIA
20CCF		.ASIA
20CD5		.ASIA
20D15		.ASIA
20D7C		.ASIA
20D7F		.ASIA
20E0E		.ASIA
20E0F		.ASIA
20E77		.ASIA
20E9D		.ASIA
20EA2		.ASIA
20ED7		.ASIA
20EF9		.ASIA
20EFA		.ASIA
20F2D		.ASIA
20F2E		.ASIA
20F4C		.ASIA
20FB4		.ASIA
20FBC		.ASIA
20FEA		.ASIA
2105C		.ASIA
2106F		.ASIA
21075		.ASIA
21076		.ASIA
2107B		.ASIA
210C1		.ASIA
210C9		.ASIA
211D9		.ASIA
220C7		.ASIA
227B5		.ASIA
22AD5		.ASIA
22B43		.ASIA
22BCA		.ASIA
22C51		.ASIA
22C55		.ASIA
22CC2		.ASIA
22D08		.ASIA
22D4C		.ASIA
22D67		.ASIA
22EB3		.ASIA
23CB7		.ASIA
244D3		.ASIA
24DB8		.ASIA
24DEA		.ASIA
2512B		.ASIA
26258		.ASIA
267CC		.ASIA
269F2		.ASIA
269FA		.ASIA
27A3E		.ASIA
2815D		.ASIA
28207		.ASIA
282E2		.ASIA
28CCA		.ASIA
28CCD		.ASIA
28CD2		.ASIA
29D98		.ASIA

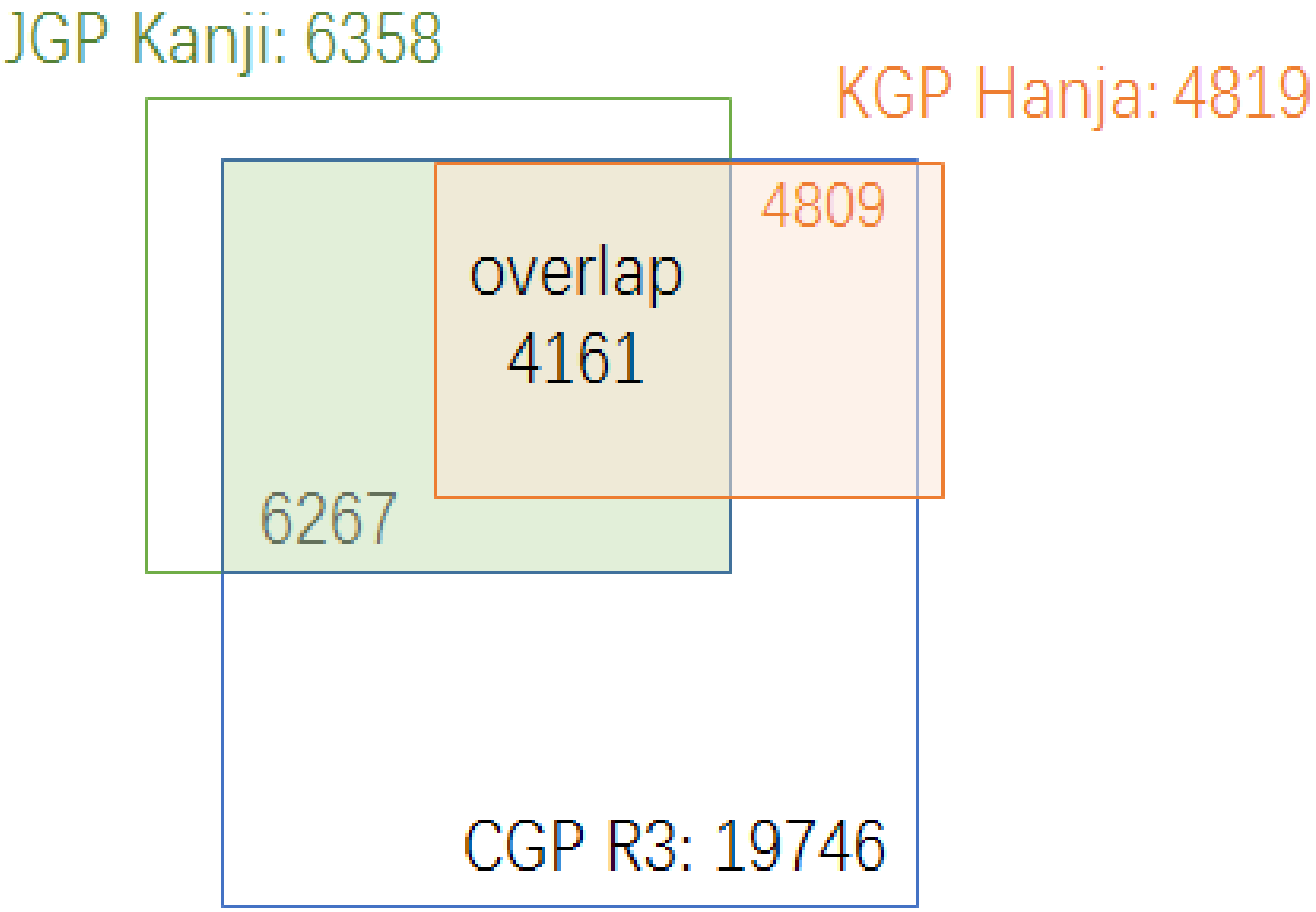
0	CDNC 2015	19561=19559+2
1	dotAsia	124=62+62
2	NHCU	18
3	IICORE∩JGP IICORE∩KGP	43
		19746

48BC	邽	N
732F	猏	N
9EB9	麭	N
5227	刧	V
524F	刼	V
6060	恮	V
74A2	璔	V
750E	甄	V
754A	畊	V
7ADA	竧	V
8262	艢	V
88B5	衹	V
894D	褋	V
8B0C	謌	V
8F19	輙	V
945A	鑽	V
984B	顋	V
9DC0	鷁	V

0	CDNC 2015	19561=19559+2
1	dotAsia	124=62+62
2	NHCU	18
3	IICORE∩JGP IICORE∩KGP	43
		19746

967A	険	JGP	
7E4A	繊	JGP	
9421	鐵	JGP	
9D8F	鶏	JGP	
4FAD	侭	JGP	
6442	摂	JGP	
685F	桟	JGP	
7E4B	繋	JGP	
81D3	臓	JGP	
8217	舗	JGP	
9039	達	JGP	
9271	鉱	JGP	
9EBA	麺	JGP	
3960	惰		KGP
51E6	処	JGP	

61F4	懴	JGP	
685C	桜	JGP	
6D9C	涜	JGP	
6E8C	溌	JGP	
731F	猟	JGP	
784F	研		KGP
7C14	簔	JGP	
7D9A	続	JGP	
8133	脳	JGP	
86CD	蛍	JGP	
8E99	躙	JGP	
91A4	醤	JGP	
91C8	釈	JGP	
96B2	隲	JGP	
982C	頬	JGP	
98EE	飲	JGP	KGP
9A12	騒	JGP	
9A13	験	JGP	
9C2E	鰛	JGP	
9D0E	鴎	JGP	
9D2C	鴬	JGP	
56A2	嚢	JGP	
663B	昜		KGP
7A36	穐		KGP
7B86	篭	JGP	
839F	茖	JGP	
83B5	莵	JGP	
9A28	騨	JGP	


JGP Kanji: 6358
KGP Hanja: 4819
4809
overlap
4161
6267
CGP R3: 19746

CGP LGR Proposal 1

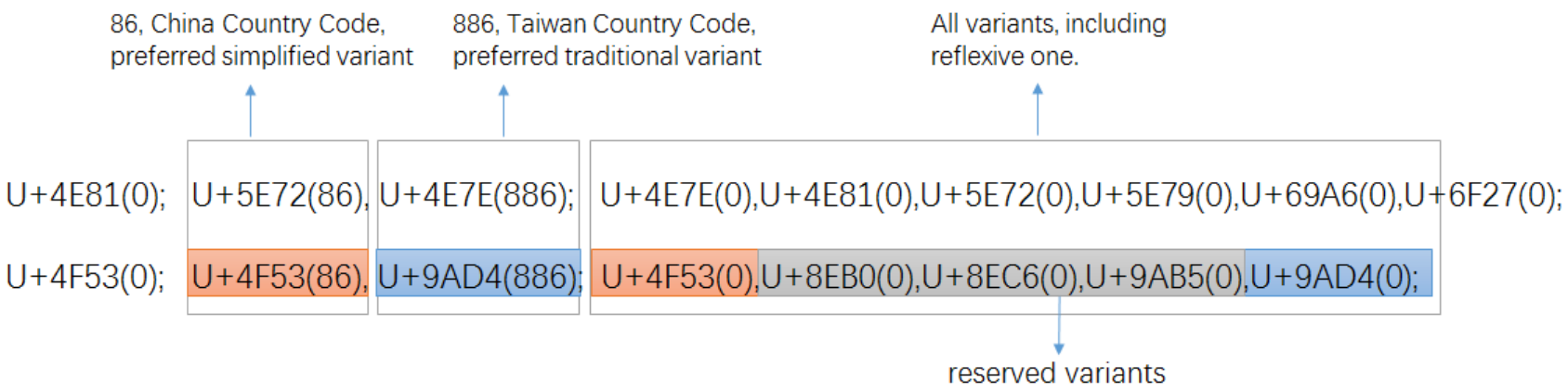
- Variant Mappings
  - Types and Sub-Types

“simp”	Allocatable	preferred simplified variant char;
“r-simp”	Allocatable	reflexive preferred simplified variant char;
“trad”	Allocatable	preferred traditional variant char
“r-trad”	Allocatable	reflexive preferred traditional variant char
“both”	Allocatable	preferred simplified and traditional variant chars are the same
“r-both”	Allocatable	reflexive preferred simp and trad variant chars are the same
“blocked”	Blocked	Non-allocatable variant char

CGP LGR Proposal 2

- Variant Mappings
  - Types and Sub-Types

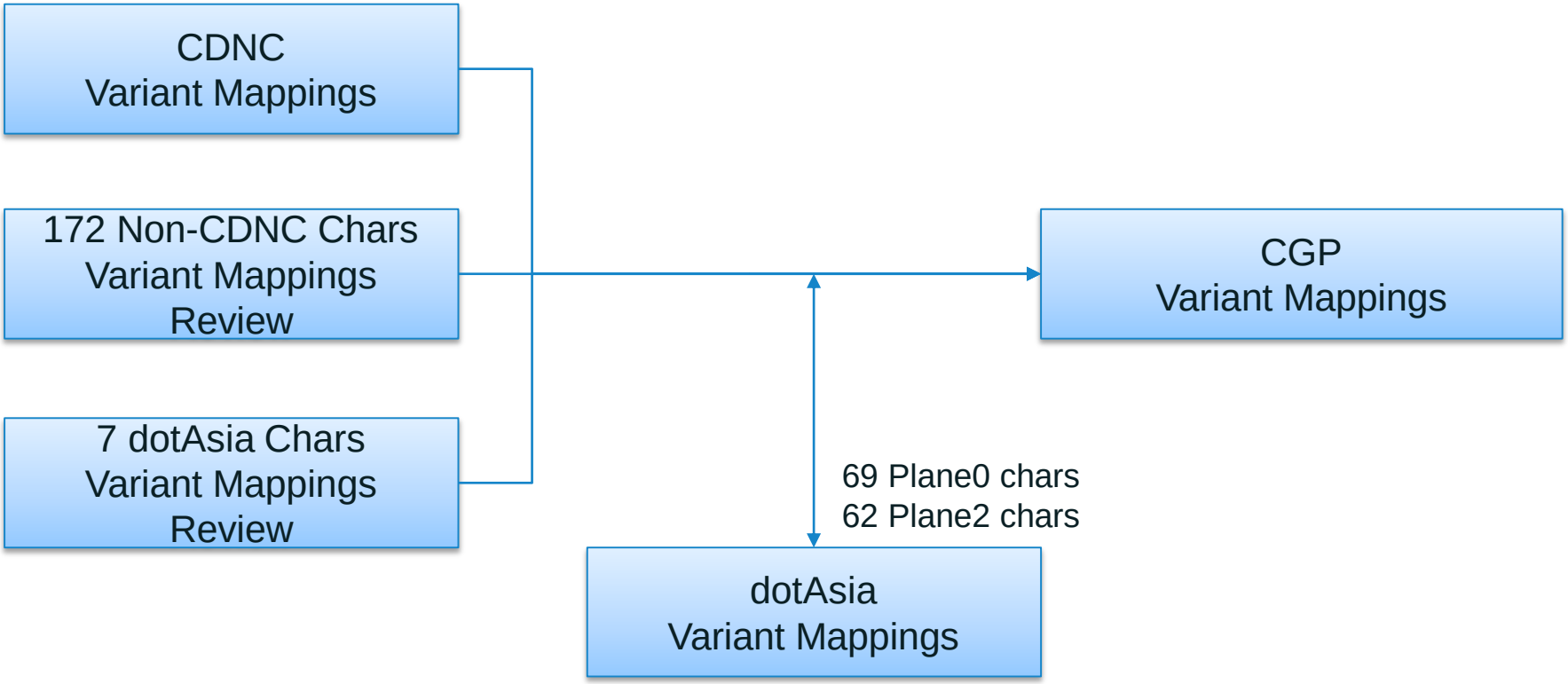
“simp”	Allocatable	preferred simplified variant char;
“r-simp”	Allocatable	reflexive preferred simplified variant char;
“trad”	Allocatable	preferred traditional variant char
“r-trad”	Allocatable	reflexive preferred traditional variant char
“both”	Allocatable	preferred simplified and traditional variant chars are the same
“r-both”	Allocatable	reflexive preferred simp and trad variant chars are the same
“r-neither”	Blocked	Non-allocatable reflexive/original char
“blocked”	Blocked	Non-allocatable variant char




```
<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="blocked" />
      <var cp="9AD4" type="traded" />
    </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>
```

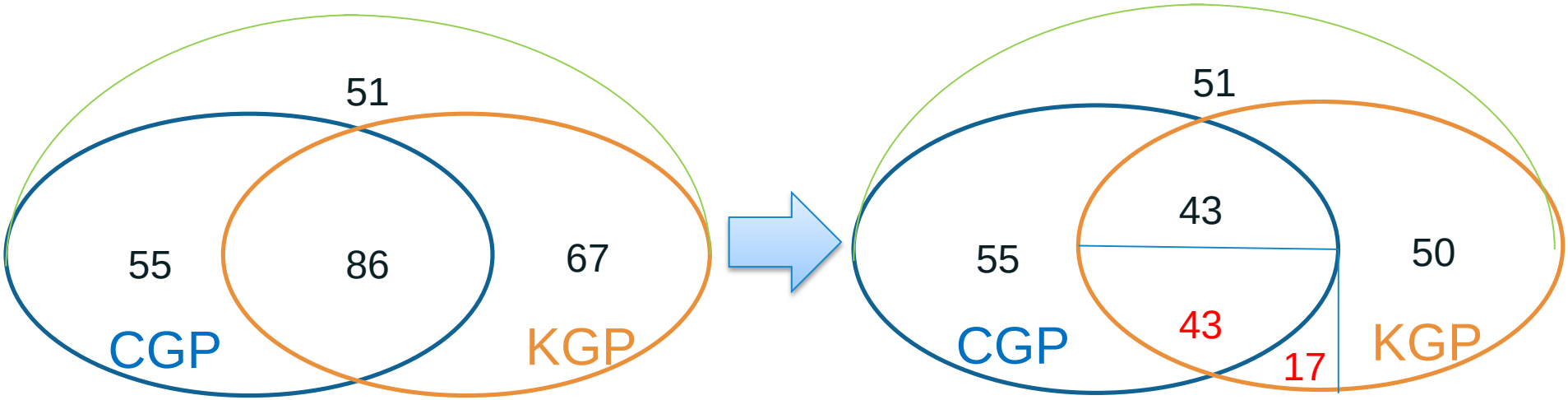
CGP LGR Proposal 2

- Variant Mappings
  - CGP interior coordination



CGP LGR Proposal 2

- Variant Mappings
  - 259 unacceptable variant groups coordination between C and K



CGP LGR Proposal 2

- 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)

台湾 (53F0 6E7E) >>

台湾 (53F0 6E7E)

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

CGP LGR Proposal 2

- Variant Mappings
  - Counter Example to IP's suggestion on eliminating variant mappings

Original	Simplified	Traditional
台(53F0)	台(53F0)	**台(53F0)** 檯(6AAF)臺(81FA)颱(98B1)
檯(6AAF)	台(53F0)	檯(6AAF)
薹(7C49)	台(53F0)	薹(7C49)
臺(81FA)	台(53F0)	臺(81FA)
颱(98B1)	台(53F0)	颱(98B1)

Original	Simplified	Traditional
鉄(9244)	铁(94C1)	鐵(9435)
銕(9295)	铁(94C1)	銕(9295)
鐵(9421)	铁(94C1)	鐵(9435)
鐵(9435)	铁(94C1)	鐵(9435)
铁(94C1)	铁(94C1)	鐵(9435)

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

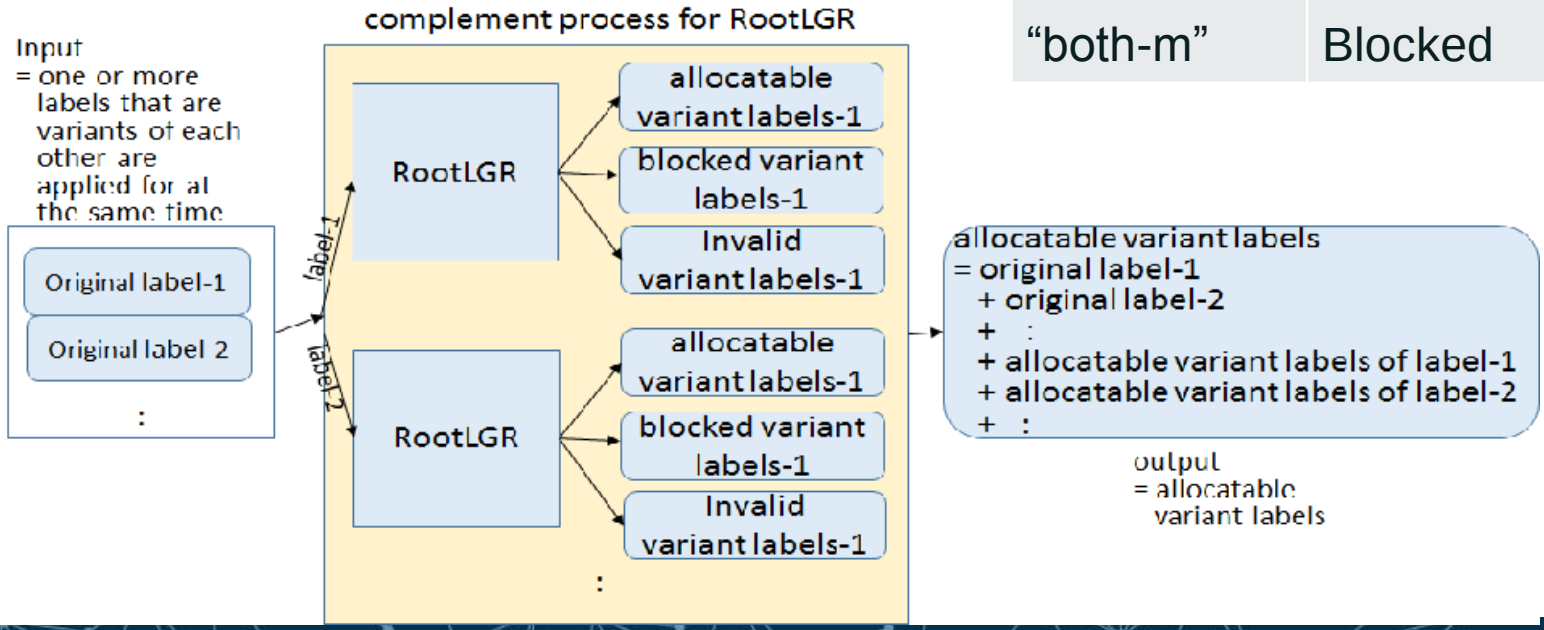
台(53F0)鐵(9435) >>台(53F0)铁(94C1), 檯(6AAF)鐵(9435)/臺(81FA)鐵(9435)/颱(98B1)鐵(9435)

CGP LGR Proposal 2

- Variant Mappings
  - Identify multiple mappings

Original	Simplified	Traditional
台(53F0)	台(53F0)	**台(53F0)** 檯(6AAF)**臺(81FA)颱(98B1)**

  - Execute multiple times



Sub-Type	Type
“r-simp-m”	Blocked
“simp-m”	Blocked
“r-trad-m”	Blocked
“trad-m”	Blocked
“r-both-m”	Blocked
“both-m”	Blocked

CGP LGR Proposal 2

- 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

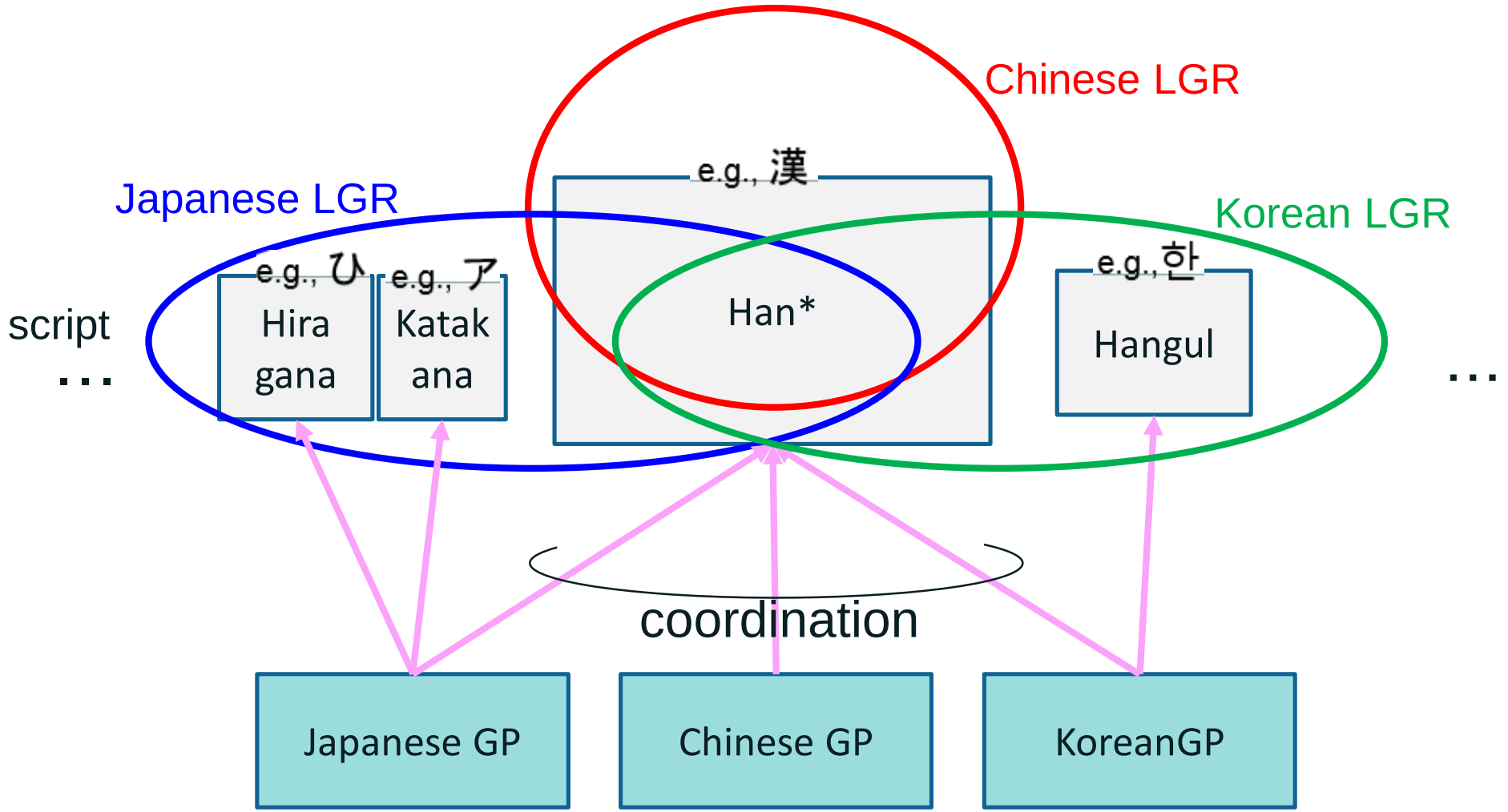
JGP meetings & related events

- 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
  - January16 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)
  - October18-22 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

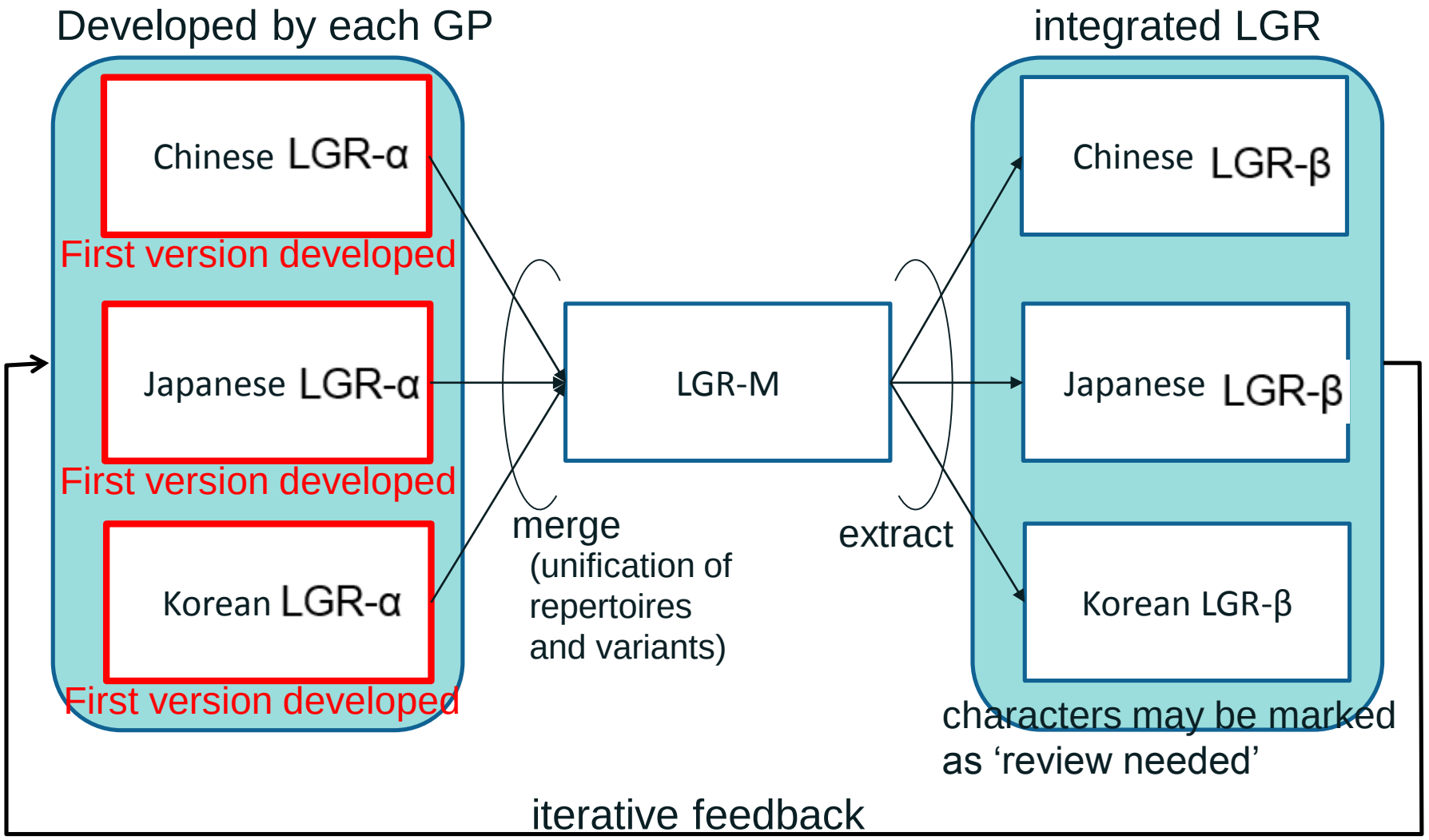
Relationship among CJK language LGRs



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

Framework of CJK LGR integration for Han characters

(revised by agreement in Buenos Aires)



Overview of Japanese LGR-α (J-LGR-α)

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

Script	# of characters
Han	6358
Hira	85
Kana	89
Total	**6532**

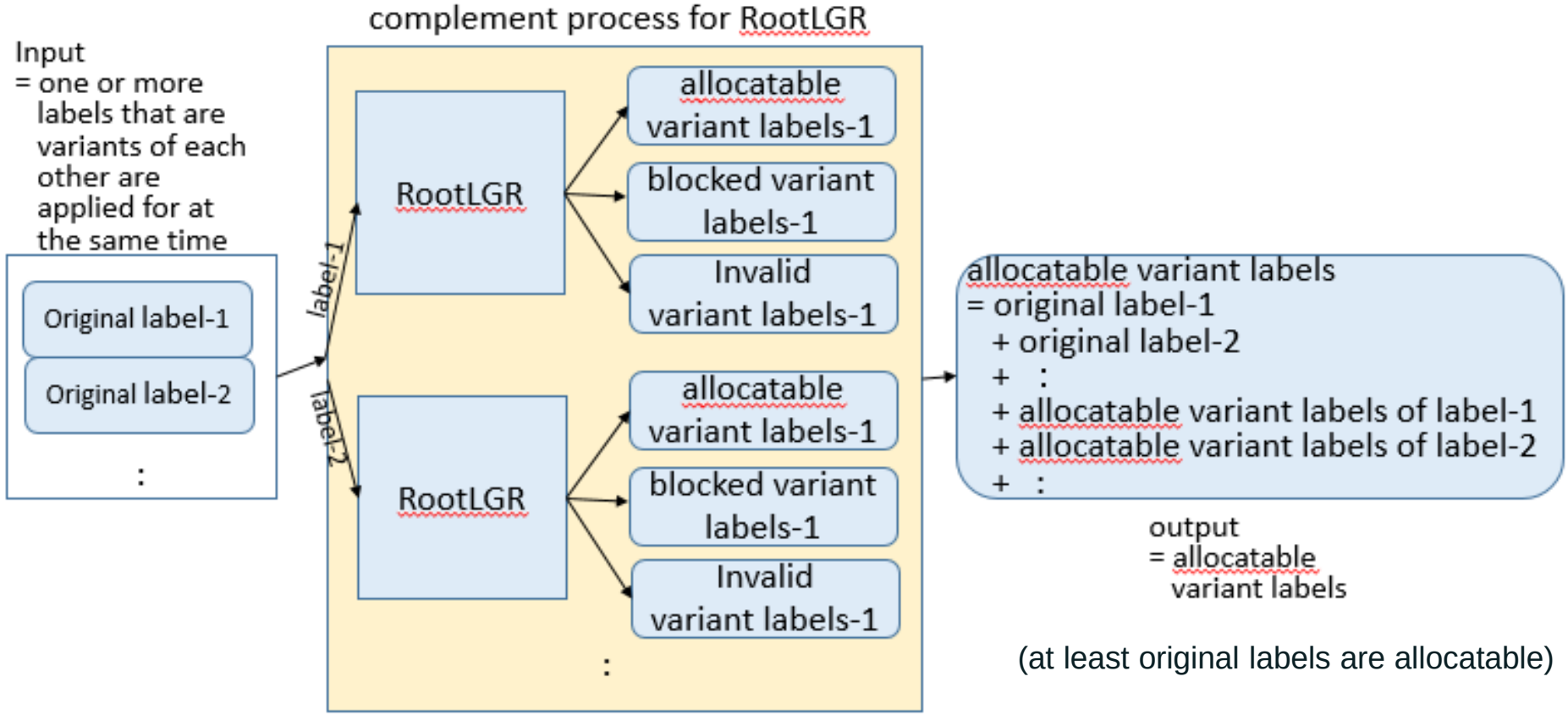
- Variants & their types
  - No variants
  - types of imported variants will be investigated and determined after LGR-α from CGP and KGP are proposed
- WLE
  - Rules (although not very many) are under discussion

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



Update by the Korean GP

KIM Kyongsok
Korean GP Chair

Agenda

- 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

K-LGR v0.5 (2016.09.28.)

- **A list of Hangul Syllables for K-LGR v0.5 (2016.09.28.)**
  - 11172 Hangul Syllbles (U+AC00 ~ U+D7A3) ← KS X ISO/IEC 10646

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

Source of Hanja Character Set	# chars
1) KS X 1001 (268 comptb. chars excluded)	4620
2) KPS 9566	4653
3) IICORE - K column marked	4743
4) IICORE - KP column marked (= KPS 9566)	4653
5) Qualifying Test of Korean Hanja Proficiency (한국 한자 능력 검정 시험)	4641
K-LGR v0.5 (2016.09.28.): Hanja List (Union of 1), 2), 3), 4), and 5))	4819

Review of K (Korean) and C (Chinese) Variant Groups (1)

- **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

Review of K (Korean) and C (Chinese) Variant Groups (2)

K position	# C variant groups
acceptable	46
unacceptable	258
total	304

- 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)]

Review of K (Korean) and C (Chinese) Variant Groups (3)

- **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

Review of K (Korean) and C (Chinese) Variant Groups (4)

- E.g., C variant group containing a simplitional char 91CC 里
  [K: vg (88CF 裏, 88E1 裡), indep (91CC 里)]
  [C: vg (88CF 裏, 88E1 裡, 91CC 里)]
  → C included 91CC 里 in C vg since it is a simplified char of traditional
  characters 88CF 裏 and 88E1 裡 in Chinese community.
  K position: 91CC 里 is completely different from (88CF 裏 = 88E1 裡) 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

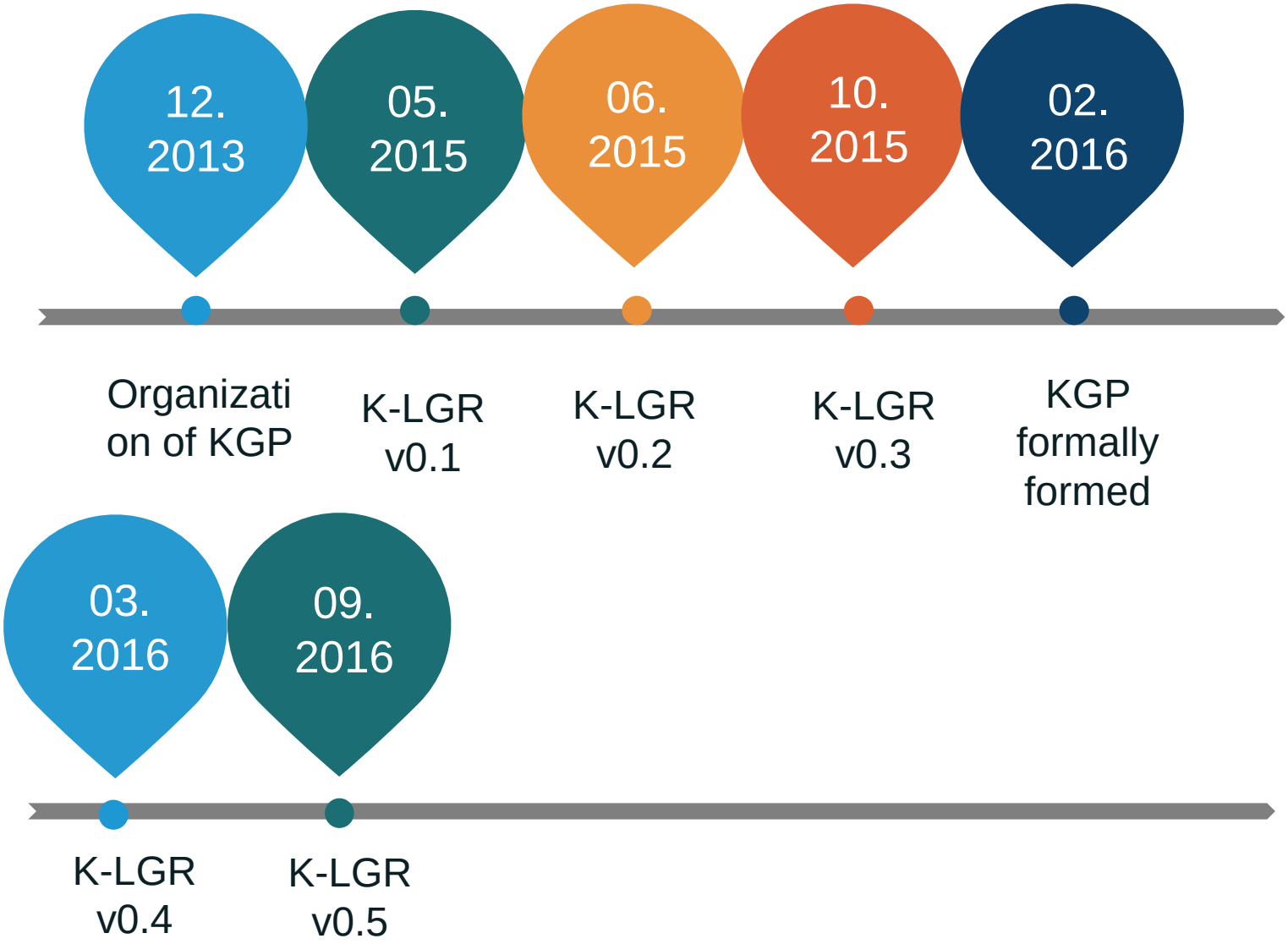
Oct: 21st and 22nd KGP meetings

Nov: Participating in ICANN57 @ Hyderabad, India

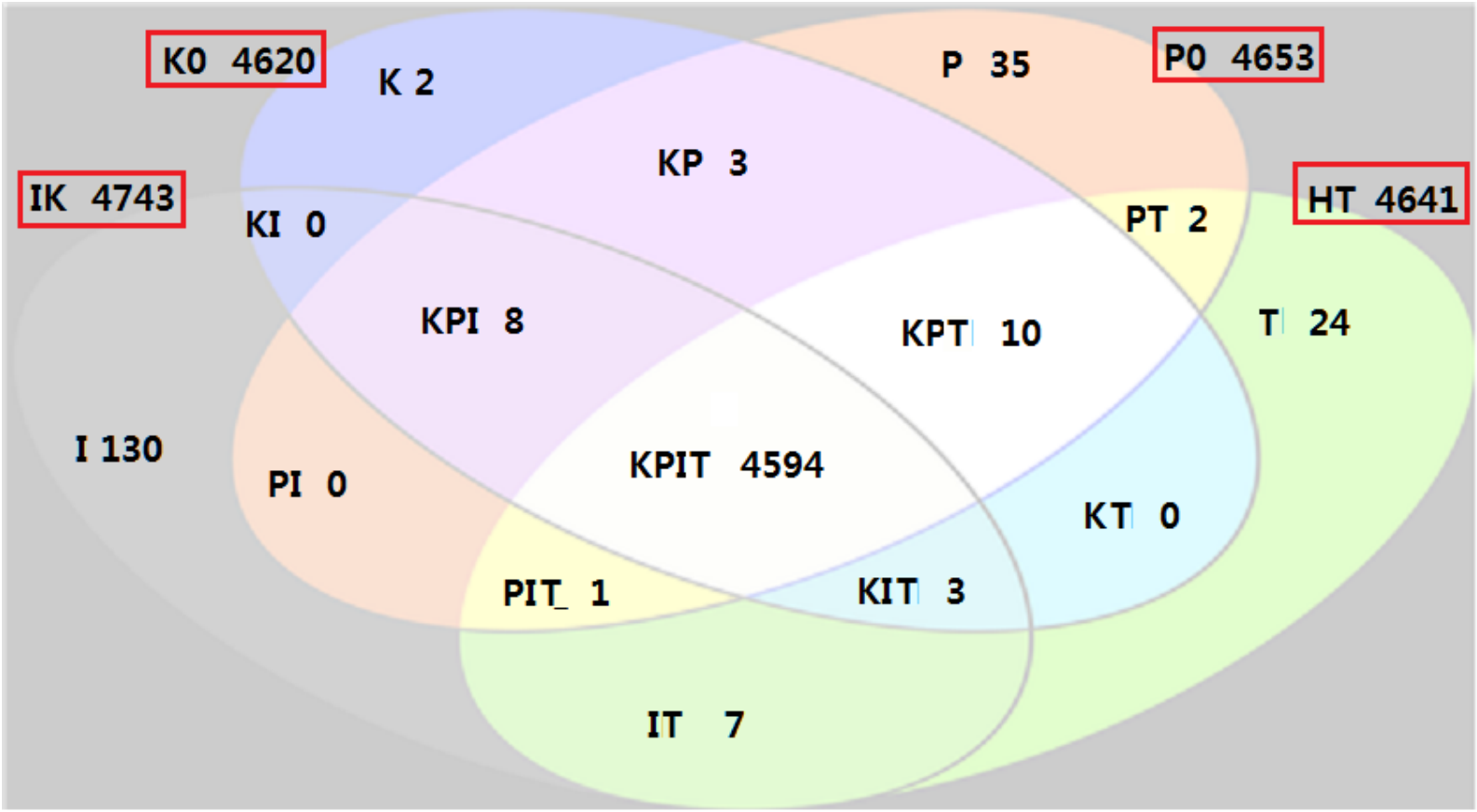
2017

Jan:

Timeline of KLGP Activities



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