

CJK Integration Procedure (v0.5)

Precondition

- Each of CJK Generation Panels (GPs) generates LGR for each language TLD before integration.
 - CJK GPs pick up ideographic variants for CJK from domain name usage perspective.
 - CJK GPs don't elaborate ideographic variants for CJK from linguistic perspective.
- CJK GPs agree on the mechanism (steps) to integrate and extract each language (script) LGR.

Step 1: Each CJK GP generates its own LGR (hereinafter, LGR-alpha)

- LGR-alpha generation process is left to each CJK GP.
- LGR-alpha format must follow XML schema for LGR.
<<https://datatracker.ietf.org/doc/draft-ietf-lager-specification/>>
- Each code point must have individual <char> element (i.e. don't use <range> element).
- Each <char> element in LGR-alpha must have reflexive mapping as <var> element (i.e. each code point must have explicit variant type/subtype).
- WLE of Each LGR-alpha must not conflict (Conflict of WLE must be coordinated and solved at this step).

Step 2: CJK GPs collectively generate a merged table of each LGR-alpha (hereinafter, LGR-M)

1. Extract every <char> element tagged "sc:Hani" from each LGR-alpha.
 2. For each extracted <char> element, check the existence of another <char> element with the same code point ("cp" value), and if exists, merge them into one element. At this time, "type" attribute of <var> element must be removed.
 3. After the check was finished, record every merged <char> elements to LGR-M.
- Repertoire of LGR-M is the union of all sc:Hani in each CJK LGR-alpha.

- Variants of each <char> elements in LGR-M is the union of all variants defined for the code point in each CJK LGR-alpha.
- LGR-M does not have following information.
 - Language tag.
 - Variant type/subtype attribute of <var> elements.
 - WLE (Whole Label Evaluation rules).

Step 3: Each CJK GP extract its original repertoire with integrated variants from LGR-M.

1. For each <char> element in its LGR-alpha (hereinafter, $\{\text{char}::\text{LGR-alpha}\}$), extract <char> element of the same code point (“cp” value) from LGR-M (hereinafter, $\{\text{char}::\text{LGR-M}\}$).
 - 1.1. For each <var> element in $\{\text{char}::\text{LGR-M}\}$ (hereinafter, $\{\text{var}::\text{LGR-M}\}$), compare with <var> elements in $\{\text{char}::\text{LGR-alpha}\}$ (hereinafter, $\{\text{var}::\text{LGR-alpha}\}$).
 - 1.1.1. If $\{\text{var}::\text{LGR-M}\}$ has the same code point (“cp” value) with one of $\{\text{var}::\text{LGR-alpha}\}$, then copy “type” attribute of corresponding $\{\text{var}::\text{LGR-alpha}\}$ to $\{\text{var}::\text{LGR-M}\}$.
 - 1.1.2. Otherwise, if $\{\text{var}::\text{LGR-M}\}$ has the same code point (“cp” value) with any of <char> elements in LGR-alpha, set “type” attribute with value “review” to $\{\text{var}::\text{LGR-M}\}$.
 - 1.1.3. Otherwise, set “type” attribute with value “blocked” to $\{\text{var}::\text{LGR-M}\}$, and record the code point of $\{\text{var}::\text{LGR-M}\}$ to Out-of-Repertoire list (hereinafter, $\{\text{OoR-list}\}$).
 - 1.2. Record $\{\text{char}::\text{LGR-M}\}$ to Integrated-Repertoire list (hereinafter, $\{\text{IR-list}\}$).
 - 1.3. If $\{\text{IR-list}\}$ includes “review” type, **responsible GP** execute either (a) return to Step 1 and redefine appropriate type/subtype to corresponding <var> elements, or (b) replace all “review” by “blocked”.
- **Whether “review” type is included in $\{\text{IR-list}\}$ is differ by each GP. Therefore, each GP should decide “review” type treatment independently.**

Step 4: Each CJK GP add “Out of Repertoire” code points for symmetry.

1. For each code point in $\{\text{OoR-list}\}$ (hereinafter, $\{\text{cp}::\text{OoR-list}\}$), extract <char>

- 72 element of the same code point (“cp” value) from LGR-M (hereinafter,
 73 $\{\text{char}::\text{LGR-M}\}$).
- 74 1.1. For each $\langle\text{var}\rangle$ element in $\{\text{char}::\text{LGR-M}\}$ (hereinafter, $\{\text{var}::\text{LGR-M}\}$),
 75 compare $\{\text{cp}::\text{OoR-list}\}$ and code point (“cp” value) of $\{\text{var}::\text{LGR-M}\}$.
- 76 1.2. If the two code points are the same, add “type” attribute with value “out-of-
 77 repertoire-var” to $\{\text{var}::\text{LGR-M}\}$.
- 78 1.3. Otherwise, add “type” attribute with value “blocked” to $\{\text{var}::\text{LGR-M}\}$.
- 79 1.4. Record $\{\text{char}::\text{LGR-M}\}$ to $\{\text{IR-list}\}$.

80

81 **Step 5: Each CJK GP merge WLE in LGR-alpha into one.**

- 82
- 83 1. Each GP extract $\langle\text{rules}\rangle$ element from each LGR-alpha and merge them into one
 84 WLE (generate integrated $\langle\text{rules}\rangle$ element, hereinafter, $\{\text{rules}::\text{LGR-M}\}$).
- 85 2. Each GP add following rule to $\{\text{rules}::\text{LGR-M}\}$ for handling “out-of-repertoire-
 86 var” variant type.
- 87 $\langle\text{action disp=“invalid” any-variant=“out-of-repertoire-var”}/\rangle$

88

89 **Step 6: Each CJK GP generates integrated LGR (hereinafter, LGR-beta).**

- 90
- 91 1. Each GP extract preambles from its LGR-alpha (hereinafter, $\{\text{preamble}::\text{LGR-}$
 92 $\text{alpha}\}$).
- 93 2. Each GP extract all $\langle\text{char}\rangle$ elements with “tag” value other than “sc:Hani” and
 94 record to IR-list.
- 95 3. Each GP merge $\{\text{preamble}::\text{LGR-alpha}\}$, IR-list and $\{\text{rules}::\text{LGR-M}\}$ into LGR-
 96 beta.
- 97
- 98 ● In other words, this step replaces body of $\langle\text{data}\rangle$ element and $\langle\text{rules}\rangle$ element of
 99 LGR-alpha to IR-list and $\{\text{rules}::\text{LGR-M}\}$ respectively.