Document type:	National Body Contribution
Title:	Proposal to start the development of CJK Regional Supplementary Ideographs and terminate the development of CJK Unified Ideographs
Status:	Circulated for information
Source	Japan
Actions	For consideration by SC2, WG2 and IRG
Date:	June 19, 2018

Note: Changes in this revision, which are to reflect the comments received during IRG #50 Beijing meeting in May, are indicated by "|" (vertical bar).

The Japan National Body proposes to start the development of "CJK Regional Supplementary Ideographs" which are independently assigned per each IRG member body in ISO/IEC 10646. In addition to the introduction of CJK Regional Supplementary Ideographs, we also propose to basically terminate the development of CJK Unified Ideographs. Since the unification review process for CJK Regional Supplementary Ideographs and CJK characters that have already been standardized is not conducted, it is expected that the development of CJK ideographs will take a much shorter period of time. We would like to start the development of CJK Regional Supplementary Ideographs from the IRG Working Set 2017. Therefore, we request SC2 and WG2 to reserve a coding space for CJK Regional Supplementary Ideographs.

1. Background and Rationale

As described in IRG N 2216 at the IRG #48 Seoul meeting, the Japan NB has been very concerned about the proliferation of variant characters and rarely used ancient/academic characters in CJK Unified Ideographs extensions, then has proposed to revise the CJK unification rules in order to minimize the number of glyph variants and to establish a separate framework to encode rarely used characters. However, the discussion has not progressed very far up until today. One of the reasons must be that it takes a big effort, it is very difficult to reach a consensus on the revision of the existing unification rules.

The following table shows the percentage of single source characters in the entire repertoire of

each CJK Unified Ideograph Extension. As shown in this table, after CJK Extension C, almost all characters in the CJK Extensions are single source characters, for example, 99.53 % of CJK Extension E characters are single source characters. This fact indicates the following:

1) Almost all demanded CJK characters will be not for global or universal purposes but for regional ones, especially for the use of regional, historical and/or archives of old documents.

2) There is very little possibility that a proposed character can be unified with other characters proposed by other IRG member bodies, or the current unification mechanism has not been functioning effectively.

Therefore we concluded that CJK ideographs to be proposed in the future could be assigned separately per each IRG member body, and the unification review between proposed repertoires could be avoided.

Block name	# of chars	# of single source chars	% of single source chars
URO	20,983	1,983	9.45%
CJK Extension A	6,582	567	8.61%
CJK Extension B	42,711	19,004	44.49%
CJK Extension C	4,149	3,783	91.18%
CJK Extension D	222	218	98.20%
CJK Extension E	5,762	5,735	99.53%
CJK Extension F	7,473	7,301	97.70%

Note: Data in this table is calculated from CJKSrc.txt.

By the way, there is the possibility in the future to increase the multiple source characters by the Horizontal Extension, so that the data in "% of single source chars" column in the table will be decreased. However, it does not affect the discussion if the unification review in the initial working set is required. Note that this proposal does not inhibit the Horizontal Extension.

2. Some details of CJK Regional Supplementary Ideographs

Basically, each repertoire of CJK Regional Supplementary Ideographs is developed on each IRG member body's responsibility, because they will be implemented for a specific purpose of each IRG member body, not for any general purposes. However, in order to avoid standardization of

any needless characters, we propose the following guidelines.

- Each repertoire of CJK Regional Supplementary Ideographs should be proposed with their source information and evidence, same as the existing development process of CJK Unified Ideographs described in the IRG Principles and Procedures.
- It is strongly recommended that proposed glyphs should be reviewed with CJK Unified Ideographs that have already been standardized.
- As shown in the table, there is a slight possibility that new characters can be theoretically unified with other characters in the same working set. As for these characters, basically they could be separately encoded in each CJK Regional Supplementary Ideographs, because the purposes of usage are different. However, for few cases such as the usage as human name, it may be considered to share it.
- A character might be proposed in a few CJK Regional Supplementary Ideographs with same origin/source information. For this character, it is not recommended to encode separately.
- It is encouraged to adopt the IVS/IVD mechanism for the distinction between unifiable glyphs needed regionally.

As CJK Regional Supplementary Ideographs are defined for the use of specific purpose, some limitations should be considered not to prevent the implementation that does not support it, just like the ones for the R-zone characters specified in previous version of ISO/IEC 10646.

3. Development process of CJK Regional Supplementary Ideographs

Steps to standardize CJK Regional Supplementary Ideographs are outlined in the following itemized list.

1) Once a repertoire of CJK Regional Supplementary Ideographs is submitted, IRG will verify its source information and evidence for each proposal.
2) Then, for a certain period of time, the technical content will be reviewed within IRG. Any comments or questions will be handled by the IRG member body that proposed the repertoire, but no unification review is conducted between the proposed repertoire and CJK Regional Supplementary Ideographs as well as CJK characters standardized already.
3) After the review is completed, the repertoire is moved to WG2 for it to be included in the text of the committee draft.

By rough estimation, we spend one thru two years to complete the developing process of new CJK Extension in IRG. Although new process additionally takes a review over a certain period,

it does not take the unification review and the discussion at IRG meeting. As for the unification review with CJK Unified ideographs already encoded, it is recommended to conduct it before submitting each repertoire of CJK Regional Supplementary Ideographs. So that it is expected that the development of CJK ideographs will speed up considerably.

(End of Document)