IRGN 1372

Title: On better usage of IDS for IRG development process

Author: Taichi Kawabata, Tatsuo Kobayashi

Status: Expert Contribution

Action: For consideration and discussion

Background:

Currently, IRG requests every new character submission to be included in UCS Unified Ideograph part to attach accurate IDS data.

However, as a matter of fact, to develop all IDS data for huge number of characters is very severe load for each national body and this task makes the IRG developing process slower.

Considering this situation, it is very important to clearly acknowledge the Pros and Cons of IDS, and to find the better and optimized usage with least developing effort.

This paper discusses a new practical usage of IDS data for better development of Unified Ideographs with less workload.

Goal:

- Less workload to developing IDS
- Less workload to find possible unifiable pairs
- Better quality of whole Unified Ideograph part
- Recognizable effect of using IDS

Proposal:

- The attachment of IDS data will be only requested, not mandatory.
- Frequency of “Eye-ball check” for Characters with IDS data attached will be reduced.
- The number of IDS elements will be limited to 3 elements; also usage of overlap element will be prohibited.
- To ensure the quality of IDS data, the IDS data will be developed by two independent developers.

The methodology of IDS checking and the form of report

- Submitted characters with IDS data will be cross-checked against already standardized characters and other submitted characters with mechanical way (IDS checking program).

The characters which are not hit with this process will be marked with "better quality (default B)" and excluded from further "preliminary checking process".
NB: It does not means that the "better quality" is safe from possible unification. These characters should be checked with "eye-ball" at least once during the development process of IRG.

- Preliminary "eye-ball" check should be concentrated to following characters:
  - The characters, which is hit by the IDS checking program, with counter possibly unifiable characters.
  - The characters with more than 4 IDS elements, against other submitted characters and already standardized characters with more than 4 IDS elements and without IDS data.
  - The characters, to which it is impossible to add the IDS data, against other submitted characters and already standardized characters without IDS data or with more than 4 IDS elements.

Submitted Characters / Already Standardized Characters	Less than 3 IDS elements	More than 4 IDS elements Impossible to add IDS Irregular IDS
Less than 3 IDS elements	Mechanically check each other with IDS checking Program. Hit characters should be classified as "F". Passed characters should be classified as "default B", not "A".	No IDS check.
More than 4 IDS elements Impossible to add IDS Irregular IDS	No IDS check.	Subject to the "eye-ball check". The methodology of "eye-ball check" should be decided with other discussion.

NB:
Roughly saying, 80% of already standardized character have less than 3 IDS elements.

60% of submitted characters for Ext-D have less than 3 IDS elements.

It means that the number of the already encoded characters subject to the “heavy eye-ball check” is expected to be reduced to 20% , and the workload of “heavy eye-ball check combinations” for encoded characters against candidates is expected to be reduced to 20% * 40% = 8%.