From:	Ken Lunde <lunde@adobe.com>
Sent:	Wednesday, 17 October 2018 2:19 AM
To:	LU, Qin [COMP]
Cc:	irgeditors; 陈永聪; chen zhuang; CMEX-Selena (selena@cmex.org.tw); beartsn@gmail.com
Subject:	Response to IRG N2340

Dr. Lu,

Please post to the IRG #51 document register this somewhat long email as my personal response to IRG N2340.

I'd like to thank those who prepared this document, because I also agree that references to "IVD" have been overused/abused in review comments for the two current IRG Working Sets.

About Section 1 (pp 1 and 2), I *completely* agree with the three numbered bullet points, and especially with the final paragraph. About the first numbered bullet point, registering IVSes is not a requirement at all, but rather a choice. That choice is solely for the submitter to make, not the revisers. The reviewers can only suggest unification based on the unification rules that apply (they are different for the two current IRG Working Sets).

The same can be said for horizontal extensions. Some unifications are best handled as a horizontal extension, specifically when the ideograph to which it is unified does not yet have a source reference for the region of the submitter. But again, this is a choice that needs to be made by the submitter, because it requires effort.

The importance of a unifiable ideograph, or rather a collection of unifiable ideographs, needs to rise to a level at which the effort that it takes to register an IVD collection is worthwhile and beneficial.

About Section 2 (pp 2 and 3), the following rewording of the first sentence is more accurate:

"The IVD is a solution for representing unifiable variants of an ideograph in "plain text" through the use of a sequence, and is not to be used for variants in general, particularly non-unifiable variants."

Of the examples that are provided in the table that spans pp 2 and 3, the ideographs for IRG Working Set 2017 serial numbers 02636 and 02964 are obvious cases of non-unifiable variants due to the significant structural differences. I would classify them as simplified variants.

The ideographs for serial numbers 02732, 03052, and 03140 could fall under one of the two new unification rules that are reflected in Section 2 of UTS #37 (as a side note, I do not yet see these unification rules reflected in the current IRG P&P, though maybe I simply couldn't find them), but only if they are cognate:

https://secure-web.cisco.com/10XlcbQvjNYU2fS4xj7jTcAVn_L8BLrJIBhSwlInc46jxwR0CMKkVy6mGLDb9h9TNEwAHCjRV8Z2rurX4Kp
MdTbkARK7oX4v7IWzwX5G4zH9KmgmvMsc6wsFC369Zswh7DZRi5yUEqEqaI9nXvGruSNWCA-
QG5uD4ZMGpcFaxGYf7XCB3Xmo5GU55hRWGCaXr1QIINm1SsDLWUn87vzZjpRkms6WvLqTpni80Ni-
IsXx5nl6FEjW3iXteZn0FP05beTW0RCdE83ePplprQjJ6TkJabQI14itYXlChMim-s7AkFJku6yW0xGiSuG8NJDU-
_CQ_tWd0R6X8lfcFw9ysvw/https%3A%2F%2Fwww.unicode.org%2Freports%2Ftr37%2F#w1aab9b1

The burden of proof, in terms of whether a submitted ideograph is cognate with an existing (encoded) ideograph, should be primarily on the reviewer, though the discussion necessarily involves the submitter, especially if additional evidence is needed.

About Section 3 (pp 3 and 4), to provide an answer for the first question in the first numbered bullet, "cognate" should mean "unifiable (by shape) and cognate." In other words, it means a "variant," so the answer would be "yes,"

though I would consider "unifiable variant" to be a more accurate way of stating this. This means that the answer to the second question in the same numbered bullet would be a definitive "no."

Thank you for submitting IRG N2341, which I found to be very helpful:

http://secure-web.cisco.com/1ebd80K5-wqw6n-1JhNL_aEhG0U_3BlHkUuiWWE6J8kFRO0u26l6c__CmWojfNW0OAD_I6HEYtG8R-l0xUq810nwDbetM5unMqmfrniQuO1cFiSHVPtqOzybURtY6CG627XGgaNsFHv10M39HOAKoo1rAGf3RN78TuRM0ITTzQr1LxvrxBrVWHmM_SS6aW8U1rZWlBpqrNHAYQjKu4soc9wNew2W3K_S4s4ZRB6CnUjEK98qdWmE7tM0adT9akH32wp3igbyxRcp8HrUWCPpCQzsgQBgYO9tRDxchenJ2jz_OpAgcNnYcGsylUnXB3ocg/http%3A%2F%2Fappsrv.cse.cuhk.edu.hk%2F%7Eirg%2Firg%2Firg51%2FIRGN2341IntroMOEVariantsDictionary.pdf

This particular dictionary could easily serve as the basis for a new registered IVD collection, in terms of handling its ideographs that can be considered unifiable variants. Registering such an IVD collection would require a lot of work, but the end result would be useful in that all of the ideographs, regardless of the variant status, would be represented in "plain text."

Although the scope and intention of this dictionary is different from Japan's "Moji Jōhō Kiban Project" (文字情報基盤整備事業), I see a lot of parallels, in terms of how variants were handled, either as non-unifiable (encoded as CJK Unified Ideographs) or unifiable (registered as IVSes in the "Moji_Joho" IVD collection).

Regards...

-- Ken Lunde