To: Unicode Technical Committee and ISO/IEC JTC1/SC2 and WG2

From: Deborah Anderson, SEI, UC Berkeley

Date: 20 August 2016

Title: Summary of Meeting on Khitan Scripts, 20 August 2016 (Yinchuan, China) - Ad Hoc Report #1

Attendance: Arakawa Shintaro, Nie Hongyin, Sun Bojun, Jing Yongshi, Jia Changye, Tang Jun, Pierre Marsone, Deborah Anderson, Andrew West, Michael Everson, Viacheslav Zaytsev, Alla Sizova, Chen Zhuang, Lin Ying-Chin, You Chengyu, and Alan Downes.

Note: This is the first of two meetings on the Khitan scripts. Please also see the Ad Hoc Report #2 (WG2 N4737=L2/16-244), which documents later changes to the block size, font, and location of the REPEAT CHARACTER ("iteration character").

The following points were agreed by consensus by the meeting participants:

For Khitan Small Script (based on document WG2 N4725R)

◎ Character names will be in the format KHITAN SMALL SCRIPT CHARACTER-XXXXX and, for the radicals, KHITAN SMALL SCRIPT RADICAL-01, etc. The range of the block for Khitan Small Script will extend from 18B00..18CDF, and the radicals will be located in a separate block of two columns at 18CE0-18CFF. Khitan Ideographs (Khitan Large Script) will begin at 18D00.

◎ The encoding model will be as described in Section 4 of WG2 N4725R, but with the names 16FE2 KHITAN SMALL SCRIPT SINGLE CLUSTER INITIAL (formerly STACKER-2) and 16FE3 KHITAN SMALL SCRIPT DOUBLE CLUSTER INITIAL (formerly STACKER-1).

◎ The REPEAT MARK will be kept in main Khitan Small Script block.

◎ Twenty radicals will be encoded, comprising (a) the 19 radicals found in fifth column (with the header "N3918") of Table 8 of WG2 N4725R, but the glyph shape of J-0461 (on page 108) should be modified to have a dot on it, and (b) one additional radical, as shown in the row beneath J-0461 (271-278 on page 108). It was also noted that the radical 195-219 (page 108) was an error in the original source, and is not included because it is a mistake.

- A short document with a table, based on section 6.2 Discussion of Individual KSS Characters (pp. 75-78), will be written by West and Zaytsev, for review by Chinese experts.
- The characters listed in section 6.6 Allographs, Variants and Erroneous Forms (p. 110) will be separately encoded. The encoding of these characters follows principles used for Tangut and other scripts like Anatolian Hieroglyphs and Linear A, which have been encoded on the basis of lexical or catalogue inventories.
- The composite characters 377 and 14, which are discussed in section 6.7, will be separately encoded. Character 349 was also agreed to be separately encoded (pp. 111-112).
- The dotted and undotted forms, listed in Section 6.8 (pp. 113-115), will be separately encoded.
- The UTC can decide the linebreak properties.
- Jing has supplied his Khitan Small Script fonts to serve as the basis of the code chart and proposal.
- Experts are encouraged to check the repertoire, and verify the order is correct. Feedback should be sent to West by the end of August. Relevant questions for reviewers would be: are any characters missing or do any characters need to be added?
- West and Zaytsev will write a new proposal on encoding of Khitan Small Script, in conjunction with Sun, Jing, Wu and Jiruhe, and it will be forwarded to WG2 and the Unicode Technical Committee.

For Khitan Large Script

- West and Zaytsev will make available a background document on Khitan Large Script in the near future (before the end of 2016).
- Preference will be given to high-quality rubbings of the original texts as the basis of future work.
- West, Zaytsev, Sun, and other experts from China will cooperate on the Khitan Large Script.

For Tangut Variants

- There was agreement in principle to use variation sequences which will be registered in the Ideographic Variation Database.
- West will work with Jing and Jia to produce a list of variation sequences.
- West will gather evidence for 7 character forms currently unified in Unicode, but for which there may be a difference in how they are written. Based on the evidence, West

will propose what to do (i.e., disunify the forms or use variation sequences with the variation captured in a font).

Group photo: