ISO/IEC JTC1/SC2/WG2 **N4561**

2014-02-25

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc type: **Working group document**
Title: **Nüshu ad hoc report**
Source: **Nüshu ad hoc committee**
Author: **Tero Aalto, CSC – IT Center for Science**
Status: **Ad hoc Report**
Action: **For consideration by JTC1/SC2/WG2**
Date: **2014-02-25**

An ad hoc committee on Nüshu met in San Jose, USA on February 25, 2014. The following experts were present:

> Tero Aalto (Finland)
> Deborah Anderson (USA)
> Chen Zhuang (China)
> Peter Constable (USA)
> Michael Everson (Ireland)
> Mike Ksar (WG2 convener)
> Roozbeh Pournander (USA)
> Michel Suignard (USA)
> Suzuki Toshiya (Japan)
> V. S. Umamaheshwaran (Canada)
> Wei Lin Mei (TCA)
> Ken Whistler (USA)

The meeting was chaired by Tero Aalto.

The discussion in the ad hoc group was based on these current documents:
- N4472 – Code table and names list for encoding Nüshu
- N4513 – Comments on Nushu script in N4484/PDAM1
  - N4533 – Addition comments on N4483/PDAM1 - Nushu
  - N4556 – Reply to N4513 and N4533 Comments on Nushu encoding
    - N4562 – Feedback To N4556

Prior documents related to the proposal:
- N3287 – Proposal for Encoding Nüshu in the SMP of the UCS
  - N3322 – Comments on Nüshu Proposal by the Chinese NB
    - N3337 – Reply to WG2 N3322 – UC Berkeley's Comments on Nüshu

- N3340 – Proposal for Encoding Nüshu in the SMP of the UCS
- N3426 – Revised Proposal for encoding Nushu in the SMP of the UCS
- N3598 – Secondly Revised Proposal for Encoding Nüshu
    - N3705 – Comments and Questions on N3598
        - N3719 – A Quick Response to WG2 N3705
- N3635 – Ad hoc report on Nüshu
- N4341 – Updated Proposal for Encoding Nyushu in the SMP of the UCS
    - N4366 – Input on encoding Nushu
    - N4376 – Nüshu ad hoc report
    - N4442 – Comments on Nushu
    - N4451 – Mapping of Nushu characters (from N4376 and N4341) and Additional Comments
- N4376 – Nüshu ad hoc report
- N4461 – Nüshu ad hoc report

In the documents N4513 and N4533, Japan had expressed concerns about the Nushu block as it is in the present proposal. The main issues concerned the method for selecting the representative glyph, the sufficiency of the data the statistics are based on, the stability of the character names, and sorting criteria. The discussion had continued in documents N4556 and N4562 from China and Japan, respectively.

Tero Aalto opened the discussion by going through some less controversial aspects of the proposal. The US ballot comments for PDAM 1 pointed out winding errors in a few characters that should be fixed in the font. Michael Everson suspected that one of them might not be an actual error.

The most controversial question about encoding the script turned out to be the naming of the characters. Suzuki Toshiya's main concern was that because character names in the standard are immutable, they should not contain any "uncertain" elements. In the existing proposals, each character name includes the phonetic value of the most frequent reading of the character, including the tone marks. However, since the sample data the statistics are based on is limited, all of the numbers may not be reliable, making the naming scheme potentially unstable. Some of the names had already been changed between the proposals from years 2007 to 2009 due to the statistics having been adjusted.

Everson pointed out that the statistics were based on the chosen source, the book "Nushu Duben", and that even if further knowledge would be gained at a later stage, compliance to the chosen source could still be referred to. Nushu Duben is the source the Chinese Nüshu experts attending the WG2 meeting had recommended and indicated as the one that would be used in teaching in schools. Suzuki, however, said the Japanese expert Endō Orie, author of N3705, had been unable to find evidence that Nushu Duben was authorized material for use in schools. Chen Zhuang replied that this meant that some teachers used the content of Nushu Duben without it being an actual textbook in Nushu schools. Suzuki suggested that "Nushu Yongzi Bijiao" would be a better source. Suzuki also quoted Endō (N3705) saying that Zhang Tie Bao had concluded the coin pictured in Chinese submissions as a forged object and that an

authorized book should not refer to it as evidence. Chen reported Zhao saying that Zhang is wrong about this issue.

Everson and Chen supported the inclusion of phonetic values in the character names because they would be useful to end-users. Ken Whistler, however, questioned the relevance of standardizing phonetic values of characters that can have up to ten different readings. It would be more sensible to provide the pronunciation information outside the names, in which case all possible pronunciations could be provided. That would be in accordance to the principle that the standard be a stable point onto which other information is anchored.

Whistler and Michel Suignard advised avoiding potential unstability when it comes to character naming. The names should be based on code points instead. This solution was acceptable to Japan and China. Everson, however, said that this naming scheme would produce characters with no actual name. The present names list also includes additional information, including stroke counts and cross references to Chinese characters, in cases where the characters are not "real Nüshu" but skewed versions of Chinese characters.

The other big question was the character repertoire. Suzuki pointed out that there were a number of characters not included in the proposed set that were on He Yanxin's glyph list, and vice versa. Chen explained the differences were caused by unification. Providing specific unification rules would be very difficult, but the general principles can be described. There is an appendix document concerning character comparison, and it is currently available in Chinese and Japanese. China was requested to provide an English version of the appendix in order to facilitate other national bodies' work.

The ad hoc group unanimously agreed that Nüshu should remain on the ballot because the user community had already been waiting for their script to be encoded for such a long time and removing the script and starting over would mean a delay of at least one more year. It was considered unlikely that there would also be no funding for additional work. Therefore the group decided to work on the repertoire in order to define a stable character set supported by all sources. The rest of the characters presently in need of further verification could be added later, especially because Nüshu is not a script dependent on ordering. Suignard requested that, at this stage, characters not be moved on the chart but instead empty slots be left where a character was removed and any additions be done to the end of the list.

The ad hoc group arrived at the following conclusions:
- Any errors in the font should be verified and corrected.
- China will provide an English version of the character comparison document.
- The interested national bodies (China, Japan, USA, Ireland) should work together to define the stable character set.
- The characters will be named based on code points.
- Nüshu will be included in Amendment 2.