

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

**Doc Type:** ISO/IEC JTC1/SC2/WG2/IRG  
**Title:** UK Activity Report for IRG #59  
**Source:** Andrew West  
**Status:** Member Body Contribution  
**Action:** For consideration by IRG  
**Date:** 2022-10-14

## 1. Review of WS2021

UK experts provided comments and feedback on WS2021 using the online review tool. We have the following suggestions to improve the quality of future IRG submissions, and to make reviewing easier and more efficient:

- A.** Characters found in quotations from classical or pre-20th-century texts in a modern typeset edition should if possible also provide an image of an original edition of the text in order to be sure that the character form given in the modern edition is not an error form.
- B.** Characters for which the evidence is a pre-modern woodblock printed text should if possible provide at least two evidence images from different sources (may comprise the same or different content) in order to demonstrate that the character is not a one-off error in a single edition.
- C.** Complete page evidence showing the submitted character should be provided if possible, and clipping of the image to show only a small incomplete section of the relevant text should be avoided (clipping of a dictionary entry is acceptable if the complete entry is shown). If text relating to the submitted character extends over more than one page, then images of all relevant pages should be provided.
- D.** Complete references for source evidence should be supplied (author, title, publisher, year, page number, etc.) for each evidence image.
- E.** The red square used to highlight the submitted character should avoid touching any part of the submitted character.

## 2. TCA submission for WS2021

It is simply impossible to review the TCA submissions that only provide a glyph, but no evidence of actual usage, as there is no way of verifying whether the glyph is valid or an error form. Presumably personal name characters were originally provided in hand-written form by

the person using the character, and were converted to a computer font glyph at some later date by someone completely unrelated to the original informant. If the original character was carelessly written then the person creating the computer font glyph could easily misread the hand-written character and create an erroneous font glyph. This error-prone process could explain many of the unusual-looking characters in the TCA submission, but we have no way of knowing, which is why actual evidence of the character in written (preferably printed) texts is so important. We strongly agree with Ken Lunde's comments that such evidence is insufficient, and we would support the removal from WS2021 of all TCA character which lack sufficient evidence.

### **3. Daoist-usage Characters**

We welcome the contribution by TCA on “Suggestions for Taoism Sacral Character Encoding” (IRGN2579), but note that IRG Recommendation M58.9 states that “Daoist characters in IRG WS 2021 will be processed based on current IRG principles and procedures”. Based on current IRG principles and procedures, we believe that the evidence provided for Daoist-usage characters submitted by the UK, showing their usage in printed editions of key Daoist texts, is more than sufficient for encoding. Therefore we oppose only encoding a subset of 44 of the 152 submitted characters, and maintain that all submitted characters (other than those that have been withdrawn) are eligible for encoding, and should be accepted for encoding in the current IRG working set.

### **4. Modern Self-created Characters**

We are not convinced that it is necessary to update the IRG Principles and Procedures to explicitly discuss modern self-created characters, and we could not support the imposition of a rule that formally prohibits the encoding of modern self-created characters. We consider that IRG PnP 2.2.1 (d) 2 parts a) [Original Source] and b) [Multiple Sources] already provides sufficient guidance for dealing with modern self-created characters.