

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  
**Source:** Andrew West  
**Status:** Member Body Contribution  
**Action:** For consideration by IRG  
**Date:** 2016-05-15

The UK has been working with experts from UK, USA and China to help encode Han characters that are currently unencoded, and are required by users. This resulted in the UK's first ever submission to IRG, IRGN2107, which comprises 1,640 characters for inclusion in IRG Working Set 2015. We have been working on encoding characters in the following categories.

### **Oracle Bone Transcription Characters**

IRGN2107 includes 164 characters used for the transcription of Oracle bone script characters in modern sources. These are required by scholars working on Oracle bone script, and the characters sourced to Takashima's *A Concordance to Fascicle Three of the Inscriptions from the Yin Ruins* were requested to us for encoding by Dr. Nathan Hill of SOAS, University of London.

### **Bamboo Text Transcription Characters**

IRGN2107 includes 182 characters used for the transcription of Warring States period texts written on bamboo slips or silk. 164 of these characters are used in the important scholarly book *Buried Ideas* (2015) by Prof. Sarah Allan of Dartmouth University. The Chinese translation of this book is scheduled to be published by the Commercial Press (Shangwu yinshuguan), and having all the transcription characters used in the book encoded in Unicode will obviously facilitate the translation and publication process, and will be essential for scholars who wish to quote from or critique Prof. Allan's book.

Prof. Allan is also editor of the academic journal *Early China* published by Cambridge University Press, and we have been working with Prof. Allan and the technical staff at Cambridge University Press to identify and prepare for encoding transcription characters for Warring States period texts that are used in articles written by a number of different scholars. In order to facilitate the editorial and publishing processes it is important these characters are encoded in Unicode, especially as many of the characters are used in multiple different articles.

Since submitting IRGN2107 we have identified 232 additional unencoded transcription characters for Warring States texts that are used in articles and books edited by Prof. Allan and her colleagues. These are candidates for a future UK submission to IRG.

### **Simplified Chinese Characters**

We have been working with experts in China to identify unencoded simplified Chinese characters attested in dictionaries and other print sources. IRGN2107 includes 1,146 simplified Chinese characters, mostly attested in important PRC dictionaries such as 漢語大字典 *Hànyǔ Dàzìdiǎn*, 中华字海 *Zhōnghuá Zìhǎi* and 现代汉语词典 *Xiàndài Hànyǔ Cídiǎn*. The encoding of these simplified characters is of great importance to users in China, who want to be able to represent texts consistently with simplified characters.

### **Characters used in Chinese Dialects**

We have been working with experts in China to identify unencoded characters used for writing various Chinese dialects, including Cantonese, Hakka, and Fukienese. Since submitting IRGN2107 we have found printed evidence for several hundred additional dialect characters (both simplified and traditional form characters), and these are candidates for a future UK submission to IRG.

### **Other Characters**

Also included in IRGN2107 are a number of characters in modern and historical use that have been requested to us for encoding by a variety of individual users. The UK submission of IRGN2107 has attracted the attention of users in China, and since submitting IRGN2107 we have been contacted by a number of different users and experts who want our help in getting various unencoded characters encoded. Those characters for which evidence of use in printed source can be found are candidates for a future UK submission to IRG.