

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

Doc Type: Working Group Document

Title: IRG and IVD

Authors: Peter Constable, Richard Cook, John Jenkins, Ken Lunde, Ken Whistler

References: WG2 N4084 (= L2/11-205)

Status: Expert Contribution

Action: For consideration by IRG

Date: 2011-06-08

Background

Encoding proposals often undergo a binary and possibly lossy evaluation process. The process is binary when the result is either (A) the encoding of a new abstract character, or (B) rejection of the candidate as unsuitable for encoding (possibly reflecting unification with an already encoded abstract character). The process is lossy when the decision discards potentially useful information. For example: shape differences (subtle or not) may be considered unifiable for general purposes (any one of the forms in the abstract character class may represent the class, and the other forms may be discarded); but shape differences may also be useful for special purposes (for context-specific font support, for defining the encoding model).

For Han (CJK), the limitations of such a binary process are especially acute. A significant amount of metadata is submitted to the IRG (*Ideographic Rapporteur Group*) in support of encoding proposals. If the IRG decides that a submitted form is unifiable with an encoded character, or that two submitted forms are unifiable with each other, some supporting metadata may be lost at the end of the encoding process. Such information would then be unavailable to support future encoding decisions. The decision to encode or unify in any given case presumes an encoding model based on reference data. Decisions made for each case have the potential to introduce confusion. Recording those decisions in an accessible form helps define a more consistent and robust encoding model.

The IVD (*Ideographic Variation Database*) provides a standardized mechanism for registering essential metadata about collections of variant glyphs for CJK characters. Use of the IVD can help streamline the encoding process by tracking information about variant glyphs which the IRG has determined should be unified with existing or prospective unified CJK ideographs. It can thereby simplify multi-column Code Chart production, reduce implementation costs, and eliminate end-user confusion about variation in glyphs. (See [UTS #37](#) for documentation about the IVD.)

Recommendation

We request that IRG investigate the use of IVD registrations as a means of tracking and standardizing CJK glyph variants and their metadata. Such IVD registrations will be based on candidate collections prepared by IRG participants, with special focus on cases involving shape differences important to the definition of the Han character encoding model.