

ISO/IEC JTC 1/SC 2/WG 2/N4111

Doc Type:	Working Group Document
Title:	Sources for Orphaned CJK Ideographs
Source:	U.S. National Body
Status:	National Body Contribution
Action:	For consideration by JTC1/SC/WG2
Date:	June 6, 2011

(This document supersedes US comments in N4058 = INCITS/L2/11-199.)

Clause 23.1 of ISO/IEC 10646:2011 states, “A CJK Ideograph is always referenced by at least one source reference... The source reference information establishes the character identity for CJK Ideographs.” Accordingly, the US notes that every CJK ideograph must have at least one unique source reference that establishes its identify.

Currently, this is a problem for ideographs that have lost their source references, because multiple ideographs share the same source reference ID. For example, the following two entries in CJKC_SR.txt (from ISO/IEC 10646:2011) have non-unique source references:

0F92C;090CE;163.6;;;K0-0000;;

0F9B8;096B8;171.9;;;K0-0000;;

Thus, the requirements of Clause 23.1 are not being met.

To remedy this, the US proposes that U-Source references be created to track the distinct identity of orphaned ideographs that have lost their original source references. Effectively, the de facto source of these characters has become ISO/IEC 10646. In order to avoid confusion with U-Source references for characters with real sources independent of ISO/IEC 10646, a distinct label “UCI” will be used.

In accordance with this proposal, the US will request that the Unicode Consortium revise UTR #45 “U-Source Ideographs” at the earliest possible opportunity to accommodate source references using the “UCI” label for this purpose.