

Comments on ISO 10646 fixed subset 288 Multilingual Latin Set

L2/13-171

Denis Moyogo Jacquerye <moyogo@gmail.com>

2013-07-30

1. Comments

Standardized subsets such as Microsoft's WGL-4 or those defined in ISO 10646 such as Multilingual European Subset (MES-1, MES-2 and previously MES-3) are commonly used to provide support to language groups when developing fonts. The Multilingual Latin Set (MLS) was recently added as a fixed subset to ISO 10646 to provide a collection of characters which can be input by means of the current revision of ISO/IEC 9995-3 "Information Technology – Keyboard layouts for text and office systems – Part 3: Complementary layouts of the alphanumeric zone of the alphanumeric section". MLS does not cover all the languages that use the Latin script. The current description in ISO 10646 does not accurately describe the purpose or scope of that subset and could lead to believe it supports all language using the Latin script. A complete collection would be difficult to define in a fixed subset as many minor languages are difficult to document. A non fixed subset would be more adequate and a clarification of the scope of these subsets would make it clear what they support and don't support.

Microsoft's WGL-4 and the ISO 10646 subsets are useful to or commonly requested to font developers to target language support based on user or client requirements. However these Latin subsets are mainly focused on European languages and only the Multilingual Latin Subset covers a bit more but not consistently nor globally (for example it covers some Native American languages but lacks support for those using U+0245 LATIN CAPITAL LETTER TURNED V and U+028C LATIN SMALL LETTER TURNED V like Oneida or Tepehuan, and covers some African orthographies but lacks support for those many using U+0190 LATIN CAPITAL LETTER OPEN E or U+025B LATIN SMALL LETTER OPEN E). A larger non fixed subset would offer a better coverage but would require extensive

work. Initiatives like the character list of African Localization Network, ScriptSource.org or the Character Exemplars of the CLDR offer information of possible extensions. Considering the scarcity of information, people defining font requirements are rarely capable to do the extensive research required to support even language spoken by millions. A clarification of the scope of the ISO 10646 subsets and note of other possible collections would help font developers and their clients to identify.

In ISO 10646, the MLS is currently described as :

The fixed collection 288 MULTILINGUAL LATIN SUBSET is an international Latin subset.

2. Suggestion

Clearly state that MLS only supports the characters which can be input with the current version of ISO/IEC 9995-3 and does not support all languages using the Latin script.
