

Title: Report on ISO TC46/SC4 N517 Status of Character Set Work

Source: Michael Everson, ISO/IEC JTC1/SC2 liaison to ISO TC46/SC4

Date: 2003-05-20

Action: For information to ISO/IEC JTC1/SC2, ISO/IEC JTC1/SC2/WG2, and UTC

ISO TC46/SC4 N517 "Status of Character Set Work" is Sally McCallum's report to ISO TC46/SC4 (Technical Interoperability) regarding TC46/SC4 character sets. I attended this meeting in Rome as ISO/IEC JTC1/SC2 liaison to TC46/SC4; the text of the SC4 contribution is given below, with my own responses interspersed.

ISO 5426-2:1996 Extension of the Latin alphabet coded character set for bibliographic information interchange - Part 2: Latin characters used in minor European languages and obsolete typography.

This standard has a number of Latin abbreviation marks which have remained unmappable to ISO/IEC 10646. The only choice may be for the library community to abandon the hope of getting these into ISO/IEC 10646. We had hoped they would be let in as compatibility characters, but there has been stiff resistance within JTC1.

Response: There is, in fact, ongoing work to investigate this character set. In June I will be in Oxford meeting with a number of medievalists who have themselves expressed an interest in encoding some of these characters as characters, not as ligatures of other characters.

ISO 8957:1996 Hebrew alphabet coded character set for bibliographic information interchange

This standard has a small repertoire of cantillation marks which have yet to be added to ISO/IEC 10646. JTC1 is waiting for Israel to make a case for these, as they did when the TC46 standard was under development. If no case can be made, TC46 may have little choice but to give up support for them.

Response: It remains the case that experts on cantillation marks have not made a definitive mapping of existing ISO/IEC 10646. In the absence of such input little progress can be made.

ISO 10574:1996 Extension of the Cyrillic alphabet coded character set for non-Slavic languages for bibliographic information interchange.

There are 10 pairs of Latin-like Cyrillic characters which have been controversial in JTC1. We hope these will be added at the December 2003 meeting of JTC1/SC2/WG2, which will allow for ISO 10754 to JTC1. The other characters can be unified with ISO/IEC 10646.

Response: There were originally even more Cyrillic characters in this character set which have been independently verified since and which have been added to

ISO/IEC 10646. Work continues to independently verify the remaining characters; the clones of Qq and Ww are verified, but some members of JTC1 (and UTC) remain reluctant to disunify them from their Latin antecedents. It remains to be seen whether a proposal will be mature enough for processing by the end of the year.

Characters from the other TC46 character set standards have been fully mapped to ISO/IEC 10646, thus responsibility for the 8-bit sets has been transferred to JTC1/SC2/WG3.

Response: This omits mention of ISO 6861:1996 *Glagolitic alphabet coded character set for bibliographic information interchange*, which also has not been transferred to JTC1/SC2. Experts in JTC1/SC2 have been working on encoding Glagolitic, however, and have taken the contents of this standard into account.