

Title: Discussion of Liaison Report from ISO/TC46/SC4/WG1 (L2/02-143)

Source: Joan M. Aliprand, Research Libraries Group (US)

Status: Unconfirmed minutes of discussion at L2/UTC Joint Meeting on May 3, 2002

Action: For information

Date: 2002-05-05

Guide to Documents

ISO standard	Content of standard	Unmapped characters are:	L2 copy of ISO standard	Current status of mapping
5426-2	Miscellaneous Latin	Manuscript contractions	L2/98-185	L2/02-150
6630	Bibliographic control codes	Control codes	L2/98-189	No report
6861	Glagolitic	Letters	L2/98-190	L2/02-151
8957	Hebrew	Cantillation marks (<i>teamim</i>)	L2/98-192	L2/02-152
10754	Non-Slavic Cyrillic	Letters	L2/98-195	L2/02-153

Discussion

General Points

- Treatment as compatibility characters is not an option.
- Existence in an ISO standard is not sufficient demonstration of use
 - ISO/TC46/SC4/WG1 must have information on sources used for the standards in its archives

ISO 5426-2: Latin characters used in minor European languages and obsolete typography

- 31 characters without identifiable mappings
- All manuscript contractions
 - Tiny subset of extensive number of contractions (hundreds, perhaps thousands)
 - At least one of the unmapped characters is a superscript combining letter comparable to the medieval superscript letter diacritics (U+0363..U+036F).
- Only known bibliographic use by the British Library
 - BL's "Specials" character set was a source for ISO 5426-2 (60% overlap)
 - BL maps contractions to code points in Private Use Area (CHASE Project)
- Not used in leading databases of rare book cataloging

Outcome of discussion:

1. UTC/L2 requests evidence of widespread use of these contractions as characters in implementations.

2. UTC/L2 is against encoding characters to support the emulation of handwriting in plain text. Some of the contractions are clearly abbreviations for common letter combinations, used when writing Latin by hand.
3. If encoded, should be outside the BMP (but should superscript letter diacritic be separated?).

ISO 6630: Bibliographic control characters

- 4 sanctioned and definitely in use: PLU, PLD, NSB, NSE
 - PLU, PLD used to create super- and subscripts; cf. Murray Sargent's methodology for super- and subscripts in math
- Other 11: use unknown, not sanctioned.

Outcome of discussion:

1. Use of C1 controls constitutes a higher level protocol. A particular higher level protocol may associate ISO 6630 functionality with C1 controls (U+0080..U+009F).. Interpretation of the C1 controls to produce the desired results must be done by the application. Only bibliographic applications can be expected to interpret the C1 controls in line with ISO 6630 functionality. Non-bibliographic applications may interpret the C1 controls in accordance with ISO/IEC 6429 or not interpret them at all.
2. The mapping of the characters of ISO 6630 to UCS equivalents is accommodated through a higher level protocol. Therefore, ISO 6630 is eligible for transfer to WG3.

Note: Use of controls considered to be a higher level protocol is addressed in Section 2.8 of *The Unicode Standard, Version 3.0*, p. 29-30.

ISO 6861: Glagolitic

This standard was not included in the proposal on compatibility, but it was examined to complete the set of TC46's character set standards.

- 90 characters
- All discrete characters of this standard are in WG2 proposal N1931
- 5 additional case pairs in ISO 6861
 - 4 pairs annotated “variant”
 - 1 pair annotated “equivalent”

Outcome of discussion:

1. UTC/L2 considers it premature to consider mapping of this character set until WG2 proposal N1931 is progressed.
2. UTC/L2 notes that there are several possible ways to relate the 5 pairs of characters without mappings to the contents of N1931:
 - (a) unify by mapping both variants to the same character
 - (b) make a proposal to register the variants as Standardized Variants
 - (c) propose addition of the variants as truly unique characters if the evidence exists.

Note:

The UTC and L2 were unaware that the issue of the variants had been covered in WG2 N1931:

Several of the letters have variant glyph forms. These are not given separate character codes here. In ISO 6861:1996 five variant forms are explicitly encoded, and were in N1659. Discussion with the convener of ISO/SC4/WG2 indicated that in bibliographical contexts, probably *most* current use transliterates Glagolitic into Cyrillic in any case, and it is suggested that the variant characters in ISO 6861:1996 be unified with the basic letters at the present time. If in future evidence proves that the distinction must be made, there is room in the table to add them.

“ISO/SC4/WG2” presumably was intended to be “ISO/TC46/SC4/WG1”.

ISO 8957: Hebrew

Table 1 (Letters, vowel points, etc.): In active use (original source was RLG’s Hebrew character set). Fully mapped

Table 2 (Cantillation marks): Not fully mapped. Not used by the library community.

- 17 characters without identifiable mappings
 - If combining diacritical marks are used for mapping, 8 characters without identifiable mappings
- 17 unmapped characters of Table 2 are quite possibly real and unique.
- Assistance from experts on *teamim* (cantillation marks) is needed to determine appropriate names for the characters and to document use.

Outcome of discussion:

1. The UTC has contacts with a number of Hebrew experts and is willing to investigate the unmapped characters of Table 2.
2. Only Table 1 is used by the library community, and the characters in this table are all mapped to UCS equivalents. ISO/TC46/SC4/WG1 might wish to consider immediate transfer of ISO 8957 to WG3 on the grounds that demonstrable library needs have been met.

10754: Non-Slavic Cyrillic

Categories of unmapped characters

- (a) Letters with stroke/ring – 3 case pairs
 - In 1998, L2 suggested use of combining diacritical marks (U+0335, U+030A)
 - But inconsistent with precomposed approach to Cyrillic
- (b) Typographical ligatures – 2 case pairs
 - In 1998, L2 suggested encoding as component letters
- (c) Possible variants – 6 case pairs
- (d) Cyrillic clones of Latin letters – 4 case pairs
 - A problematic issue
- (e) YA IE – 1 case pair

Outcome of discussion:

UTC/L2 requests evidence of use of this character set in library applications.

Note: So far, there has been no response to a query sent to Russian experts on character sets.