

Title: Summary of mappings of ISO/TC46 character set standards
Source: Joan M. Aliprand, Research Libraries Group (US)
Status: Expert Contribution
Action: For information
Date: 2002-05-05

Key to Documents

ISO No.	Content of standard	Unmapped characters are	L2 copy of ISO standard	Current status of mapping	WG2 proposals	WG2 decision on proposed characters
5426-2	Miscellaneous Latin	Manuscript contractions	L2/98-185	L2/02-150	N1746, N1747	N1884R2: Further study of those not accepted
6630	Bibliographic control codes	Control codes	L2/98-189	No report	Not proposed	n/a
6861	Glagolitic	Letters	L2/98-190	L2/02-151	N1931	No action
8957 Table 2	Hebrew	Cantillation marks (<i>teamim</i>)	L2/98-192	L2/02-152	N1749	N1884R2: Further study of all
10754	Non-Slavic Cyrillic	Letters	L2/98-195	L2/02-153	Included in N1744	N1884R2: 2 rejected, further study of those not accepted

ISO 5426-2, 6630, 6861, 8957 and 10754 are character set standards developed by ISO/TC46/SC4/WG1 that do not yet have complete mappings to ISO/IEC 10646 and Unicode. All other TC46 standards have complete mappings and have been transferred to ISO/IEC JTC1/SC2/WG3. Table 1 of ISO 8957 is completely mapped.

WG2 proposals N1744, N1746, N1747, N1749 include characters from TC46 standards that could not be mapped to ISO/IEC 10646 equivalents (according to mappings prepared by ISO/TC 46/SC 4/WG 1). WG2 considered the proposals at its meeting #35 in September 1998. N1884R2 is the report of the ad hoc that evaluated the proposals. N1931 is the latest version of a proposal for Glagolitic script. ISO 6861 was one of the sources used for this proposal.

Statistics

ISO No.	Content of standard	Total characters	Mapped N254	Mapped WG1 mtg	Additional May 2002	Total mapped	Unmapped L2/02-150+	Unmapped L2/02-143
5426-2	Miscellaneous Latin	70	32	34	5	WG1 = 32 L2 = 39	31	37
6630	Bibliographic control codes	15	1	Not considered	0	0	15	14
6861	Glagolitic	90	All (proposed)	Not considered	80 (to N1931)	80 (to N1931)	10 not in N1931	90
8957 Table 2	Hebrew cantillation marks	51	43	Decision deferred	n/a	WG1 = 43 L2 = 34	17	8
10754	Non-Slavic Cyrillic	93	49	Decision deferred	16	WG1 = 49 L2 = 89	2	44

Explanation of Headings in Table

Mapped N254: ISO/TC46/SC4/WG1 N254 (revised, 2000-05-09). Only the mapping for ISO 5428 (Greek) was updated from N227 (revised, 1998-05-29).

Mapped WG1 mtg.: Consolidation of mappings in N254 and recommendations from L2, approved by WG1 at its meeting in May 2000. Includes characters from TC46 standards accepted by WG2 in September 1998. Other WG1 actions also shown.

Additional May 2002: Mappings to characters added after May 2000. These mappings have not been reviewed by WG1.

Discrepancies

Unmapped Characters:

Unmapped characters in L2/02-143 are apparently based upon counts from ISO/TC46/SC4/WG1 N254 (revised, 2000-05-09).

Unmapped characters in L2/02-150 through 153 are mapping decisions made at the May 2000 WG1 meeting augmented by mappings to characters added since. (Or, in the case of ISO 6861, mappings to characters proposed in WG1 N1931.)

Individual Standards:

ISO 5426-2: There are 38 unmapped characters for this standard in N254. L2 total of mapped characters includes 9 characters accepted by WG2 in 1998, 5 additional mappings, and no contractions. WG2 total includes mappings for 7 contractions; it does not include the characters added in 1998, nor the 5 added since.

ISO 6630: N254 maps character 0x87 Close-up for sorting (CUS) to U+200B ZERO WIDTH SPACE. L2 did not consider mappings for ISO 6630 characters.

ISO 6861: There have been two versions of the proposal for Glagolitic. The variant characters were removed in the revision.

ISO 8957, Table 2: WG1 mapped some cantillation marks to combining diacritical marks based on appearance. L2 disagreed.

ISO 10754: L2 proposed mapping some Cyrillic characters using composed character sequences, and mapping the clones of Latin letters to the actual Latin letters (thus unifying them). WG1 disagreed. The L2 total for mapped characters also includes 16 additional mappings. The 44 unmapped characters reported in L2/02-143 include two characters that cannot be mapped because they were rejected by WG2.