ISO/IEC JTC 1/SC 2/WG 3
7-bit and 8-bit codes and their extension
SECRETARIAT : ELOT

DOC TYPE :	Comment
TITLE :	Rationale against the use of the C1 space for standard graphic character usage for data interchange
SOURCE :	Canadian national body
PROJECT:	JTC 1.02.02
STATUS :	Proposal
ACTION ID :	ACT
DUE DATE :	-
DISTRIBUTION :	P, O and L Members of ISO/IEC JTC 1/SC 2 WG Conveners, Secretariats WG 3 MembersISO/IEC JTC 1 Secretariat ISO/IEC ITTF
MEDIUM :	P, DOC file
NO OF PAGES :	5

Canada wants to document the multiple reasons why populating the C1 control character space is not the panacea to solving data interchange in the 8-bit character set world. The following are the reasons why this method can not work in the real world among the multiple 8-bit platforms:

a. ISO-8-extended coded character sets are used primarily in the PC platforms -- starting with DOS, later in Windows and in OS/2 systems. There are already several PC DOS and OS/2 code pages that are FULL utilizing the full C1 space and even the C0 space and DEL position (the Empty House symbol). MS Windows code pages are the only ones which still have some room in the C1 area since they did not fill all of the C1 space when they were (relatively recently) defined.

b. The far-east PC codes such as S-JIS, use this space as valid first bytes of a two byte code. The Chinese equivalent uses all of the C1 and G1 space for valid first bytes of two bytes.

c. The C1 space has the two key Shift controls SS2 and SS3 - in order to be able to access the G2 and G3 sets -- used in TCP/IP and Internet Mime protocols especially in the far east. Teletex service did re-define the SS2 and SS3 to C0 area -- and if one follows the ITU-T example, it will force a change to C0.

d. While 6429 permits use of ESC yy form for representing C1 controls, this is not used in pure 8-bit modes in protocols such as those used in popular UNIX system terminals such as DEC VT 100 VT 220 X-Windows protocols etc. Same is true of the ISO and ITU standards on Office Document Interchange ODA (though it is not widely implemented).

e. The available C1 space is grossly inadequate to meet the real needs of the full LATIN set -- 6937 and corresponding Teletex recommendation from ITU to resort to combining sequences (Accent + Letter encoding for all the Latin accented letters) were defined in the past as a result.

f. As to the IBM EBCDIC to ISO-8 compatibility, IBM has gone through a lot of careful disciplined approach to maintain the equivalence in space between the controls and graphics of ISO-8 (4873) and EBCDIC SBCS codes. There are places in EBCDIC-based software, where code position values less than x3F are discarded as invalid Graphic Characters.

g. If we are going to TINKER with the 8-bit coding structures of ISO 2022/4873 etc. (for example, adding graphics only to C1 space), the requirements that are to be met must be spelled out clearly. Canada strongly believes that it will be inadequate from day no. 1 -- because there are several FULL codepages that are candidates to go in, for example, IBM CP 850 (and several others).

A real case in point:

The first requirement that Canada made -- need for integral support of French OEs and Y DIAERESIS for standard character data interchange can not be made with character sets that go beyond 191 characters in the 8-bit world (for multiple reasons stated above).

The solution for all existing environments is to have a 191-character standard character set that will make possible upgrading current character sets without changing architecture. The ISO/IEC 8859-15 caters for this requirement perfectly well while being architecturally compatible with legacy systems (and for most environments which won't require the extra characters, it is also data-compatible; furthermore it allows roundtrip integrity respect for all others which require them and which would already face the ubiquitously spreaded MS-Windows character set that already provides the required characters on the PC platform).

ISO/IEC 8859-15 also corrects historical mistakes in Latin 1 for Finnish, as well as for French, while bringing a replacement character for the EURO SIGN that will eventually allow this sign's codepoint to be shared with other standard Latin and non-Latin codes which will be convenient even for bad

situations where tagging of character sets is lots or made very difficult. The full superset of these character sets go much beyong the bossibilities of an 8-bit character set (see g above).

CONCLUSION

Any requirements calling for more than 190 graphic characters in a single-octet encoding (including combining marks or non-spacing marks etc.plus SPACE, should be recommended strongly by SC 2 and its working groupsto be satisfied and met only by the UCS.