TO: ISO TC 46 SC 4

FROM: Sally McCallum

DATE: May 20, 2003

RE: Revision of ISO 2709

The following is a contribution from Mr Sten Hedberg:

ISO 2709 was conceived when one byte was enough to represent a character, and ISO2022 was used to shift between byte tables. With the arrival of Unicode, and especially when Unicode is served in a variety of formats, it is impossible to say generally what character encoding is used in a record, so that each record must have a marker for that. This marker must be placed in a part of the record that is universally available, so that the possibility to use a variety of character schemes is restricted to that part of the record that comes after the directory and the Base address of data marker.

The following details the changes that might be made to the text of ISO 2709 to address this matter. New or changed text is underlined; text to be deleted is ~~crossed out.~~

3.6A octet : a byte of eight bits, that may represent one character or be part of a representation of a character

4.1Record label
The record label shown in figure 2 is fixed in length to 24 octets, each representing one character according to ISO 6937/2 and defined as follows.

4.1.3 Implementation codes (character positions 6 to ~~9~~8)

4.1.3A Codes for the character implementation scheme (position 9)
ISO 6937/2 or other techniques using at a maximum one octet per character
a   Unicode in UTF8 presentation
b   Unicode in UTF16BE presentation
c   Unicode in UTF16 presentation

4.2.1 Directory entry
c) starting ~~character~~ position
All octets in the directory represent Arabic digits according to ISO6937/2. The length of the tag is three ~~characters~~ octets. No part of the entry shall exceed nine ~~characters~~ octets in length. All entries in a directory shall have the same structure.

4.2.2. Tag
Three ~~characters~~ octets, which specify, according to definition in an implementation International Standard the name of any associated datafield.

4.2.3   Length of field
This length is either:
a) the total number of ~~characters~~ octets including indicator(s) and field separator) in the datafield indicated by the preceding tag; or

4.2.4 Starting ~~character~~ position
A decimal number giving the position of the first ~~character~~ <u>octet</u> of the datafield identified by the preceding tag, relative to the base address of data (i.e. the starting ~~character~~ position of the first datafield following the directory is 0 (zero).)

525/2709-rev2.doc