X.2 EXAMPLE OF PERSON NAME VALUE REPRESENTATION IN THE CHINESE LANGUAGE

Person names in the Chinese language may be written in Pinyin (phonetic characters), Hanzi (ideographic characters), or English Name (alphabetic characters). The three component groups should be written in the order of phonetic, alphabetic, and ideographic (English name).

Specific Character Set

(0008,0005) \ ISO 2022 IR 58

Character String:

Zhang^XiaoDong=张小东=

Encoded String:

Zhang^XiaoDong=ESC 02/04 02/09 04/01 张小东 ESC 02/08 04/02=

Character Encoded representation (GB2312):

Update PS3.5 Chinese encoding examples from CP1234
Notes: 1. The underlined byte codes correspond to double byte characters, the bold byte codes to escape sequences.

2. The multi-byte character set (ISO-IR 58) and single-byte character set (ISO 646) can be used intermixed without any explicit escape sequence after the initial escape sequence, up to the next delimiter (^ or =) or the end of the value field. Once ISO 646 has been designated to G0 and ISO-IR 58 to G1, each character set has a different code area, thus can be used intermixed. The decoder will check the most significant bit of a character to know whether it is a two byte character in the G1 area (high bit one) or a one byte character in the G0 area (high bit zero). There does not need to be an explicit escape to invoke ISO 646 into G0 at the end of the string prior to a delimiter (^ or =) or the end of the value field. However, there does need to be a new invocation of ISO-IR 58 in each name component in which it is used.

X.3 EXAMPLE OF LONG TEXT VALUE REPRESENTATION IN THE CHINESE LANGUAGE WITH EXPLICIT ESCAPE SEQUENCES BETWEEN GB 2312 G0 AND GB 2312 G1

Chinese (ISO 2022 IR 58) and ASCII (ISO 646) character sets can be used intermingled without explicit escape sequences between them. The Chinese character set ISO IR 58 is invoked to the G1 area, and the ASCII character set is invoked to the G0 area. The following is an example of a Long Text value representation which includes ASCII and Chinese character set. Every line must start in ASCII, end in ASCII is presumed to start in the default character set and requires an explicit invocation of GB 2312 into G1, but does not require re-invocation of the default ASCII character set into G0.

Specific Character Set:

(0008,0005) \ISO 2022 IR 58

Character String (with CR LF after each line):

1) 第一行文字。

2) 第二行文字。

3) 第三行文字。

Encoded String:

1) ESC 02/04 02/09 04/01 第一行文字。 ESC-02/08-04/02

2) ESC 02/04 02/09 04/01 第二行文字。 ESC-02/08-04/02

3) ESC 02/04 02/09 04/01 第三行文字。 ESC-02/08-04/02

Character encoded representation (GB2312):

0x31 0x2e 0x1B 0x24 0x29 0x41 0xB5 0xDA 0xD2 0xBB 0xD0 0x0D 0x0A 0xC4 0xD7 0xD6 0xA1 0xA3 0x1B-0x28-0x42 0x0D 0x0A

0x32 0x2e 0x1B 0x24 0x29 0x41 0xB5 0xDA 0xB6 0xFE 0xD0 0x0D 0x0A 0xC4 0xD7 0xD6 0xA1 0xA3 0x1B-0x28-0x42 0x0D 0x0A

0x33 0x2e 0x1B 0x24 0x29 0x41 0xB5 0xDA 0xC8 0xFD 0xD0 0x0D 0x0A 0xC4 0xD7 0xD6 0xA1 0xA3 0x1B-0x28-0x42 0x0D 0x0A 0x20
Note: The underlined byte codes correspond to double byte characters, the bold byte codes to escape sequences.

### Table X-1

**CHARACTER SETS AND ESCAPE SEQUENCES USED IN THE EXAMPLES OF PERSON NAME**

<table>
<thead>
<tr>
<th>Character Set Description</th>
<th>Component Group</th>
<th>Value of (0008,0005) Defined Term</th>
<th>ISO registration number</th>
<th>Standard for Code Extension</th>
<th>ESC Sequence</th>
<th>Code Element</th>
<th>Character Set: Purpose of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>First: Phonetic</td>
<td>Value 1: none</td>
<td>ISO-IR 6</td>
<td></td>
<td>G0</td>
<td>ISO 646:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alphabetic (English name)</td>
<td>Value 1: none</td>
<td>ISO-IR 6</td>
<td>ISO 2022</td>
<td>G0</td>
<td>ISO 646: For delimiters</td>
<td></td>
</tr>
</tbody>
</table>