DICOM Correction Item

Correction Number: CP-155
Submission Abstract: Add support for ISO-IR 149 Korean character sets

<table>
<thead>
<tr>
<th>Type of Change Proposal:</th>
<th>Name of Document:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>PS 3.3-1999: Information Object Definitions, PS 3.5-1999: Data Structures and Encoding</td>
</tr>
</tbody>
</table>

Rationale for change:
Korea uses its own characters, Hangul, and Hangul needs to be used in DICOM. Hangul can be implemented easily in DICOM by character encoding methods that PS 3.5 has defined. A Defined Term for Character set for Hangul needs to be added in Table C.12-4 of PS 3.3

Sections of document affected/ Suggest Wording of Change:

Part 3

1. C.12.1.1.2 Specific Character Set
Add the following entry to Table C.12-4.

<table>
<thead>
<tr>
<th>Character Set Description</th>
<th>Defined Term</th>
<th>Standard for Code Extension</th>
<th>ESC Sequence</th>
<th>ISO registration number</th>
<th>Number of characters</th>
<th>Code element</th>
<th>Character Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>ISO 2022 IR 149</td>
<td>ISO 2022</td>
<td>ESC 02/04 02/09 04/03</td>
<td>ISO-IR 149</td>
<td>94²</td>
<td>G1</td>
<td>KS X 1001: Hangul and Hanja</td>
</tr>
</tbody>
</table>

Part 5

1. Section 2
Add the following Hangul multi-byte character set to the end of section 2 “Normative references”:

   KS X 1001-1997 Code for Information Interchange (Hangul and Hanja)

2. Section 6.1.2.4 Code Extension Techniques
Modify the Note to read:

   2. Support for Japanese kanji (ideographic), hiragana (phonetic), and katakana (phonetic) characters, and Korean characters (Hangul) is defined in PS3.3. Definition of Chinese Korean, and other multi-byte character sets awaits consideration by the appropriate standards organizations.

3. Change the current Annex I to Annex J and add the followings as Annex I, “Character sets and person name value representation in the Korean language”: 
Annex I
(Informative)
Character sets and person name value representation In the Korean Language

I.1 CHARACTER SETS FOR THE KOREAN LANGUAGE IN DICOM

KS X 1001 (registered as ISO-IR 149) is used as a Korean character set in DICOM. This character set is the one most broadly used for the representation of Korean characters. It can be encoded by ISO 2022 code extension techniques, and is registered in ISO 2375.

Escape Sequence (for reference) (see PS 3.3)

<table>
<thead>
<tr>
<th>Escape Sequence</th>
<th>Character Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO-IR 149</td>
<td></td>
</tr>
<tr>
<td>G0 set</td>
<td>ESC 02/04 02/08 04/03</td>
</tr>
<tr>
<td>G1 set</td>
<td>ESC 02/04 02/09 04/03</td>
</tr>
</tbody>
</table>

Notes:
1. ISO-IR 149 is only used as a G1 set in DICOM.
2. The Korean character set (ISO IR 149) is invoked to the G1 area. This is different from the Japanese multi-byte character sets (ISO 2022 IR 87 and ISO 2022 IR 159) which use the G0 code area. Japan’s choice of G0 is due to the adoption of an encoding method based on “ISO-2022-JP”. ISO-2022-JP, the most familiar encoding method in Japan, and uses only the G0 code area. In Korea, most operating systems adopt an encoding method that invokes the Hangul character set (KS X 1001) in the G1 code area. So, the difference between code areas of Korean and Japanese character originates in convention, not a technical problem. Invocation of multi-byte character sets to the G1 area does not change the current DICOM normative requirements.

I.2 EXAMPLE OF PERSON NAME VALUE REPRESENTATION IN THE KOREAN LANGUAGE

Person names in the Korean language may be written in Hangul (phonetic characters), Hanja (ideographic characters), or English (single-byte characters). The three component groups should be written in the order of single-byte, ideographic, and phonetic (see Table 6.2-1).

(0008,0005) \ISO 2022 IR 149

Hong^Gildong=洪^古洞=洪^ 김동

Hong^Gildong= ESC 02/04 02/09 04/03 洪^ ESC 02/04 02/09 04/03 古洞= ESC 02/04 02/09 04/03  김동

Character String:
Encoded representation:
04/08 06/15 06/14 06/07 05/14 04/07 06/09 06/12 06/04 06/15 06/14 06/07 03/13
Result of representation by an ASCII-based machine which displays 01/11 as `033`:

```
Hong^Gildong=\033$)C\373\363^\033$)C\321\316\324\327=\033$)C\310\253^\033$)C\261\346\265\277
```

Notes: 1. The multi-byte character set (ISO-IR 149) and single-byte character set (ISO 646) can be used intermixed without any explicit escape sequence after the initial escape sequence. Once ISO 646 has been designated to the GL area and ISO-IR 149 to the GR area, each character set has 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 GR area (high bit one) or a one byte character in the GL area (high bit zero).

2. In the above example of person name representation, explicit escape sequences precede each Hangul and Hanja string. These escape sequences are to meet the requirements of the code extension technique that specifies a switch to the default character repertoire before delimiters. In the previous example, it is assumed that the default character repertoire (ISO-646) is invoked to G0 code area and no character set to G1 area after delimiters ("^" and "=" signs). See 6.1.2.5.3 of PS 3.5.

I.3 EXAMPLE OF LONG TEXT VALUE REPRESENTATION IN THE KOREAN LANGUAGE WITHOUT EXPLICIT ESCAPE SEQUENCES BETWEEN CHARACTER SETS

Hangul (ISO IR 149) and ASCII (ISO 646) character sets can be used intermingled without explicit escape sequences between them. The Hangul character set ISO IR 149 is invoked to the G1 area, so this invocation doesn't affect the G0 area to which the ASCII character set has been invoked. The following is an example of a Long Text value representation which includes ASCII and Hangul character set.

```
(0008,0005) \ISO 2022 IR 149
```

| The 1st line includes 한글. |
| The 2nd line includes 한글, too. |
| The 3rd line. |

Encoded String:
```
ESC 02/04 02/09 04/03 The 1st line includes 한글. 
ESC 02/04 02/09 04/03 The 2nd line includes 한글, too. 
The 3rd line. 
```

Once having invoked the ISO IR 149 character set to G1 area by the escape sequence in the head of line, one can use Hangul and ASCII intermixed in that line.
Table I-1.
CHARACTER SETS AND ESCAPE SEQUENCES USED IN THE EXAMPLES

<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>Character Set: Purpose of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>First: Single-byte</td>
<td>Value 1: none</td>
<td>ISO-IR 6</td>
<td></td>
<td>GL</td>
<td>ISO 646:</td>
</tr>
</tbody>
</table>