DICOM Correction Item

Correction Number CP-966	
Log Summary: Specify other Asian name encoding	
Type of Modification Addition	Name of Standard PS 3.5-2009 + CP-964

Rationale for Correction:

While Part 5 specifies alphabetic, ideographic and phonetic component groups of a PN, it is not clear how these should be used for scripts of non-CJK Asian or African languages. An additional section is added to the informative annex on use of Unicode, specifying use of alphabetic and phonetic component groups.

Part 5 Annex J also specifies the use of English (actually Latin), ideographic, and pinyin name representations for Chinese names. However, pinyin is the same thing as Latin representation, and for Chinese names only two component groups are used. Text is clarified

Pinyin is both a phonetic and a Latin alphabet representation. Some Chinese choose a "western" name that they like without regard for phonetic matching. Others choose the pinyin phonetic form of their name. It is extremely rare to have both. Whichever one is chosen, it makes sense to then use it as the Latin version of their name. That is its purpose.

This is different from the Korean, where there may be both a Korean phonetic and a Korean ideographic form for the name.

This CP is written assuming the changes of CP-964 have been accepted as Final Text.

Sections of documents affected

PS 3.5 Annex J

Correction Wording:

Annex J (Informative) Character sets and person name value representation using Unicode UTF-8 and GB18030

The Unicode ~~**3.2**~~ **UTF-8** character set and the GB18030 character set may be used for multiple languages. Some of these languages may also be encoded using other ~~**coding systems**~~ **character sets** that are defined elsewhere in the DICOM standard. ~~**The encoding used for a particular language must be the same**~~ **As Unicode UTF-8 and GB18030 encodings do not allow ISO 2022 character set replacement, these must be used** for all strings in a single SOP Instance. This may have implications for the character set selected for the encoding of the SOP Instance.

J.1 Example of Person Name Value Representation in the Chinese Language Using Unicode

Person names in the Chinese language may be written in ~~**pinyin (phonetic characters),**~~ Hanzi (ideographic characters), **and/**or Latin (alphabetic characters). **The Latin representation may be derived using pinyin or another Romanization method, or may be a chosen "westernized" name.** The ~~**three**~~ **two** component groups should be written in the order of alphabetic, **then** ideographic~~**, and; the**~~ phonetic **component group is typically not used** (see Table 6.2-1). In this example the traditional script is used ~~**and the phonetic component is not being used**~~.

Notes: 1. Some healthcare information systems may encode a "westernized" name with other patient aliases in a separate attribute, e.g., Other Patient Names (0010,1091).

2. Some environments using Chinese language may use the third name component, e.g., for the Yi or Mongolian script, with or without the first name component. This would be similar to the Japanese and Korean name component usage.

In the example below, the Character Set attribute (0008,0005) would contain:
…

J.3 Example of Person Name Value Representation in the Chinese Language Using GB18030

Person names in the Chinese language may be written in ~~**pinyin (phonetic characters),**~~ Hanzi (ideographic characters), **and/**or Latin (alphabetic characters). **The Latin representation may be derived using pinyin or another Romanization method, or may be a chosen "westernized" name.** The ~~**three**~~ **two** component groups should be written in the order of alphabetic, **then** ideographic~~**, and; the** phonetic **component group is typically not used** (see Table 6.2-1). In this example the simplified script is used ~~**and the phonetic component is not being used**~~.

Note: See notes to section J.1.

…

J.5 Person Name Value Representation in Other Languages Using Unicode

Person names in many languages may be written in a local (non-Latin) script, as well as in a transliteration to a Latin script (Romanization). Healthcare information systems in those environments may support one or both name formats. Local scripts may be encoded using Unicode in UTF-8.

For the purpose of exchange in DICOM, there are three typical uses of name component groups using Unicode in UTF-8:

1. **Names in a Latin script may be encoded in the first (alphabetic) component group, and names in a local script (alphabet, abugida, or syllabary) in the third (phonetic) component group (see Table 6.2-1). The second (ideographic) component group is null. This is the preferred use for cross-enterprise or international communication.**

2. **Where the local script historically has a single byte character set defined for Specific Character Set (0008,0005), i.e., Cyrillic, Arabic, Greek, Hebrew, Thai, and the various versions of Latin, only the first name component group might be used. Encoding may be in Unicode in UTF-8, as described in this Annex, as an equivalent for use of that defined single byte character set in the first name component group. (See note 1.)**

3. **Names in the local script may be encoded in the first component group, and names in a Latin script in the third component group, both encoded in Unicode in UTF-8.**

Notes: 1. A previous edition of DICOM required the first name component group to use a single byte character set (see PS3.5-2008). Unicode in UTF-8 may now be used in that component group simply as a matter of a different character set encoding, but with the same application use of that component group.

2. Healthcare information systems will use specific scripts in one, two, or three of the Person Name component groups in accordance with local policy. Conformant DICOM Application Entities that receive name attributes must accept multiple name component groups. An Application Entity that is configurable to allow the use of local script for names in either the first or the third component group, and a transliteration script in the other, would support all these typical representations.

3. The transliteration (from a local script) may be a non-Latin script, e.g., Cyrillic. The same principles apply, and the Cyrillized name might be encoded in the first component group and the local script (which may in fact be a Latin-derived script) in the third component group.