**Rationale for Correction:**

Part 5 specifies that the first component group of a PN shall use single byte encoding. However, when using Unicode, this does not allow the encoding of names with non-ASCII characters in the first name component group, since the non-ASCII alphabetic characters require multiple bytes in UTF-8. Whether an alphabetic name is encoded using a single byte character set or using Unicode should not affect its location in the first name component group.

The definition: *Combining characters (e.g., diacritics or vowel marks) separately encoded from base characters shall be considered separate characters for this maximum length* was chosen to be consistent with Unicode and GB18030 definition of character code points.

Note: CP966 proposes additional changes to Annex J.

The term “alphabetic” is used to indicate any non-ideographic writing system, rather than the stricter phrasing used by linguists when describing writing systems.

The code point limitation chosen for the allowed characters when using Unicode UTF-8 includes all of the writing systems for which there are single byte ISO encodings.

**Sections of documents affected**

PS 3.5 Section 6.2; Annexes H, I and J

**Correction Wording:**

### 6.2 Value representation (VR)

... 

**Table 6.2-1**

DICOM VALUE REPRESENTATIONS

<table>
<thead>
<tr>
<th>VR Name</th>
<th>Definition</th>
<th>Character Repertoire</th>
<th>Length of Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN Person Name</td>
<td>A character string encoded using a 5 component convention. … For the purpose of writing names in ideographic characters and in phonetic characters, up to 3 groups of components (see Annexes H, I, and J examples 1 and 2) may be used. The delimiter for component groups shall be the equals character “=” (3DH). The three component groups of components in their order of occurrence are: a single-byte character an alphabetic representation, an ideographic representation, and a phonetic representation. ...</td>
<td>Default Character Repertoire and/or as defined by (0008,0005) excluding Control Characters LF, FF, and CR but allowing Control Character ESC.</td>
<td>64 chars maximum per component group (see NOTE in 6.2)</td>
</tr>
</tbody>
</table>
6.2.1 Ideographic and phonetic characters in Data Elements with VR of PN

Character strings representing person names are encoded using a convention for PN value representations based on component groups with 5 components.

For the purpose of writing names in ideographic characters and in phonetic characters, up to 3 component groups may be used. The delimiter of the component group shall be the equals character “=” (3DH). The three component groups in their order of occurrence are: a single-byte an alphabetic representation, an ideographic representation, and a phonetic representation.

Any component group may be absent, including the first component group. In this case, the person name may start with one or more “=” delimiters. Delimiters are also required for interior null component groups. Trailing null component groups and their delimiters may be omitted.

The first component group (identified by DICOM as “alphabetic”) shall be encoded using a single-byte character encoding from a character set with no Code Extensions. The character set shall be the one specified by the Attribute Specific Character Set (0008,0005), value 1. If Attribute Specific Character Set (0008,0005) is not present, the default Character Repertoire ISO-IR 6 shall be used. ISO 2022 escapes for Code Extension shall not be used in this component group. When Specific Character Set (0008,0005) value 1 specifies a multi-byte character set without Code Extension (i.e., Unicode in UTF-8, or GB18030), the characters of this component group may be encoded with multiple bytes, but shall be drawn from the code points U+0000 through U+1FFF of ISO/IEC 10646.

The length of value value length of each component group is 64 characters maximum, including the delimiter for the component group. Each combining character (e.g., diacritics or vowel marks) shall be considered a separate character for this maximum length, regardless of how an application may display such combining characters (i.e., combined into the glyph for the base character, or rendered separately).

H.3 Example of Person Name Value Representation in the Japanese Language

For the purpose of writing names in ideographic characters and in phonetic characters, up to 3 component groups may be used. The delimiter of the component group shall be the equals character “=” (3DH). The three component groups in their order of occurrence are: a single-byte an alphabetic representation, an ideographic representation, and a phonetic representation.

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 Latin (single-byte alphabetic characters). The three component groups should be written in the order of single-byte alphabetic, ideographic, and phonetic (see Table 6.2-1).
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), or English Latin (single-byte alphabetic characters). The three component groups should be written in the order of single-byte alphabetic, ideographic, and phonetic (see Table 6.2-1).

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), or English Latin (single-byte alphabetic characters). The three component groups should be written in the order of single-byte alphabetic, ideographic, and phonetic (see Table 6.2-1).