

# DICOM Correction Proposal

STATUS	Final Text
Date of Last Update	2012/01/16
Person Assigned	Bas Revet
Submitter Name	Herve HOEHN
Submission date	2011/01/26

Correction Number	CP-1128
Log Summary: Enforce UT CR/LF Part 5 requirement in Part 3	
Type of Modification	Name of Standard
Modification	PS 3.5 2011
<p>Rationale for Correction:</p> <p>Part 5 Section 6.1.3 explicitly mentions that new line shall be represented as a combination CR LF. However, Part 3, Table C.10-5 as defined provides the following description for Unformatted Text Value (0070,0006): <i>“The text value may contain spaces, as well as multiple lines separated by either LF, CR, CR LF or LF CR, but otherwise no format control characters (such as horizontal or vertical tab and form feed) shall be present, even if permitted by the Value Representation of ST.”</i></p> <p>There is therefore a conflict between these two Statements.</p> <p>This CP removes the weakening in Part 3 of the requirement in Part 5.</p>	
<p>Sections of documents affected</p> <p>PS 3.3 Section C.10.5, C.11.18, C.17.3</p>	
Correction Wording:	

*Leave Part 5 unchanged:*

### 6.1.3 CONTROL CHARACTERS

Textual data that is interchanged may require some formatting information. Control Characters are used to indicate formatting, but their use in DICOM is kept to a minimum since some machines may handle them inappropriately. ISO 646:1990 and ISO 6429:1990 define Control Characters. As shown in Table 6.1-1 below, only a subset of four Control Characters from the C0 set shall be used in DICOM for the encoding of Control Characters in text strings.

**Table 6.1-1  
DICOM CONTROL CHARACTERS AND THEIR ENCODING**

Acronym	Name	Coded Value
LF	Line Feed	00/10
FF	Form Feed	00/12
CR	Carriage Return	00/13
ESC	Escape	01/11

The ESC character shall be used only for ISO 2022 character set control sequences, in accordance with section 6.1.2.5.

In text strings (value representation ST, LT, or UT) a new line shall be represented as CR LF.

Note: Some machines (such as UNIX based machines) may interpret LF (00/10) as a new line. In such cases, it is expected that the DICOM format is converted to the correct internal representation for that machine.

...

VR Name	Definition	Character Repertoire	Length of Value
...	...	...	...
UT Unlimited Text	A character string that may contain one or more paragraphs. It may contain the Graphic Character set and the Control Characters, CR, LF, FF, and ESC. It may be padded with trailing spaces, which may be ignored, but leading spaces are considered to be significant. Data Elements with this VR shall not be multi-valued and therefore character code 5CH (the BACKSLASH “\” in ISO-IR 6) may be used.	Default Character Repertoire and/or as defined by (0008,0005).	2 <sup>32</sup> -2  Note: limited only by the size of the maximum unsigned integer representable in a 32 bit VL field minus one, since FFFFFFFFH is reserved.

Fix uses in Part 3 that conflict with Part 5:

**Table C.10-5  
GRAPHIC ANNOTATION MODULE ATTRIBUTES**

Attribute Name	Tag	Type	Attribute Description
...	...	...	...
>>Unformatted Text Value	(0070,0006)	1	Text data which is unformatted and whose manner of display within the defined bounding box or relative to the specified anchor point is implementation dependent. See C.10.5.1.1.  The text value may contain spaces, as well as multiple lines separated by <del>either LF, CR, CR LF or LF CR</del> , but otherwise no format control characters (such as horizontal or vertical tab and form feed) shall be present, even if permitted by the Value Representation of ST.  The text shall be interpreted as specified by Specific Character Set (0008,0005) if present in the SOP Common Module.  Note: The text may contain single or multi-byte characters and use code extension techniques as described in PS 3.5 if permitted by the values of Specific Character Set (0008,0005).
...	...	...	...

**Table C.11.18-1  
Structured Display Annotation Attributes**

Attribute Name	Tag	Type	Attribute Description
...	...	...	...

>Unformatted Text Value	(0070,0006)	1	<p>Text data which is unformatted and whose manner of display within the defined bounding box is implementation dependent. See C.11.18.1.1.</p> <p>The text value may contain spaces, as well as multiple lines separated by <b>either LF, CR, CR LF or LF-CR</b>, but otherwise no format control characters (such as horizontal or vertical tab and form feed) shall be present, even if permitted by the Value Representation of ST.</p> <p>The text shall be interpreted as specified by Specific Character Set (0008,0005) if present in the SOP Instance.</p> <p>Note: The text may contain single or multi-byte characters and use code extension techniques as described in PS 3.5 if permitted by the values of Specific Character Set (0008,0005).</p>
...	...	...	...

**Table C.17-5  
DOCUMENT CONTENT MACRO ATTRIBUTES**

Attribute Name	Tag	Type	Attribute Description
...	...	...	...
Text Value	(0040,A160)	1C	<p>This is the value of the Content Item.</p> <p>Required if Value Type (0040,A040) is TEXT.</p> <p>Text data which is unformatted and whose manner of display is implementation dependent.</p> <p>The text value may contain spaces, as well as multiple lines separated by <b>either LF, CR, CR LF or LF-CR</b>, but otherwise no format control characters (such as horizontal or vertical tab and form feed) shall be present, even if permitted by the Value Representation of UT.</p> <p>The text shall be interpreted as specified by Specific Character Set (0008,0005) if present in the SOP Common Module.</p> <p>Note: The text may contain single or multi-byte characters and use code extension techniques as described in PS 3.5 if permitted by the values of Specific Character Set (0008,0005).</p>
...	...	...	...