

1	Status	Final Text
2	Date of Last Update	2015/03/18
3	Person Assigned	David Clunie
4		mailto:dclunie@dclunie.com
5	Submitter Name	David Clunie
6		mailto:dclunie@dclunie.com
7	Submission Date	2014/07/29

8	Correction Number CP-1425	
9	Log Summary: Tab character should be permitted be in text VRs (ST, LT, UT)	
10	Name of Standard	
11	PS3.5 2015a	
12	Rationale for Correction:	
13	Text VRs explicitly permit several control characters but not TAB, which is surprising given that they are intended to be (primitively)	
14	"formatted". Since increasingly such VRs are used to encode XML text, this is a significant limitation and an unnecessary restriction.	
15	Also, the Character Repertoire column for these and other VRs needs to be cleaned up to be consistent with any restrictions specified	
16	in the Definition column, and the general definition of control characters needs to be updated to mention TAB, and the behavior when	
17	ISO 2022 escape sequences are used needs to be addressed (in a manner consistent with IETF RFC 1554 ISO-2022-JP-2).	
18	Correction Wording:	

Amend DICOM PS3.5 as follows:

6.1.2.5.3 Requirements

The character set specified by value 1 of the Attribute Specific Character Set (0008,0005), or the default character repertoire if value 1 is missing, shall be active at the beginning of each textual Data Element value, and at the beginning of each line (i.e., after a CR and/or LF) or page (i.e., after an FF).

If within a textual value a character set other than the one specified in value 1 of the Attribute Specific Character Set (0008,0005), or the default character repertoire if value 1 is missing, has been invoked, the character set specified in the value 1, or the default character repertoire if value 1 is missing, shall be active in the following instances:

- before the end of line (i.e., before the CR and/or LF)
- before the end of a page (i.e., before the FF)
- **before any other Control Character other than ESC (e.g., before any TAB)**
- before the end of a Data Element value (e.g., before the 05/12 character code that separates multiple textual Data Element Values - 05/12 corresponds to "\" (BACKSLASH) in the case of default repertoire IR-6 or "¥" (YEN SIGN) in the case of IR-14).
- before the "^" and "=" delimiters separating name components and name component groups in Data Elements with a VR of PN.

If within a textual value a character set other than the one specified in value 1 of the Attribute Specific Character Set (0008,0005), or the default character repertoire if value 1 is missing, is used, the Escape Sequence of this character set must be inserted explicitly in the following instances:

- before the first use of the character set in the line
- before the first use of the character set in the page
- before the first use of the character set in the Data Element value
- before the first use of the character set in the name component and name component group in Data Element with a VR of PN

Note

These requirements allow an application to skip lines, values, or components in a textual data element and start the new line with a defined character set without the need to track the character set changes in the text skipped. A similar restriction appears in the RFCs describing the use of multi-byte character sets over the Internet. An Escape Sequence switching to the value 1 or default Specific Character Set is not needed within a line, value, or component if no Code Extensions are present. Nor is a switch needed to the value 1 or default Specific Character Set if this character set has only the G0 Code Element defined, and the G0 Code Element is still active.

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 five** 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
<u>TAB</u>	<u>Horizontal Tab</u>	<u>00/09</u>

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

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

Note

1. 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.
2. **In previous editions of the standard (see PS3.5 2015a), the TAB character was not listed as a Control Character.**

Table 6.2-1. DICOM Value Representations

VR Name	Definition	Character Repertoire	Length of Value
AE Application Entity	A string of characters that identifies an Application Entity with leading and trailing spaces (20H) being non-significant. A value consisting solely of spaces shall not be used.	Default Character Repertoire excluding character code 5CH (the BACKSLASH "\" in ISO-IR 6), and all control characters LF, FF, CR and ESC .	16 bytes maximum
LO Long String	A character string that may be padded with leading and/or trailing spaces. The character code 5CH (the BACKSLASH "\" in ISO-IR 6) shall not be present, as it is used as the delimiter between values in multiple valued data elements. The string shall not have Control Characters except for ESC.	Default Character Repertoire and/or as defined by (0008,0005) excluding character code 5CH (the BACKSLASH "\" in ISO-IR 6), and all Control Characters except ESC when used for ISO 2022 escape sequences.	64 chars maximum (see ???)
LT Long Text	A character string that may contain one or more paragraphs. It may contain the Graphic Character set and the Control Characters, TAB (see Note 3) , 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) excluding Control Characters except TAB, LF, FF, CR (and ESC when used for ISO 2022 escape sequences).	10240 chars maximum (see ???)

VR Name	Definition	Character Repertoire	Length of Value
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 PN Person Name	<p>A character string encoded using a 5 component convention. The character code 5CH (the BACKSLASH "\" in ISO-IR 6) shall not be present, as it is used as the delimiter between values in multiple valued data elements. The string may be padded with trailing spaces. For human use, the five components in their order of occurrence are: family name complex, given name complex, middle name, name prefix, name suffix.</p> <p>Note</p> <p>HL7 prohibits leading spaces within a component; DICOM allows leading and trailing spaces and considers them insignificant.</p> <p>Any of the five components may be an empty string. The component delimiter shall be the caret "^" character (5EH). Delimiters are required for interior null components. Trailing null components and their delimiters may be omitted. Multiple entries are permitted in each component and are encoded as natural text strings, in the format preferred by the named person.</p> <p>For veterinary use, the first two of the five components in their order of occurrence are: responsible party family name or responsible organization name, patient name. The remaining components are not used and shall not be present.</p> <p>This group of five components is referred to as a Person Name component group.</p> <p>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) 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: an alphabetic representation, an ideographic representation, and a phonetic representation.</p> <p>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 required for interior null component groups. Trailing null component groups and their delimiters may be omitted.</p> <p>Precise semantics are defined for each component group. See ???.</p> <p>For examples and notes, see ???.</p>	Default Character Repertoire and/or as defined by (0008,0005) excluding character code 5CH (the BACKSLASH "\" in ISO-IR 6), and all Control Characters LF, FF, and CR but allowing Control Characters except ESC when used for ISO 2022 escape sequences.	64 chars maximum per component group (see ???)
49 50 51 52 53 54 55 56 SH Short String	<p>A character string that may be padded with leading and/or trailing spaces. The character code 05CH (the BACKSLASH "\" in ISO-IR 6) shall not be present, as it is used as the delimiter between values for multiple data elements. The string shall not have Control Characters except ESC.</p>	Default Character Repertoire and/or as defined by (0008,0005) excluding character code 5CH (the BACKSLASH "\" in ISO-IR 6), and all Control Characters except ESC when used for ISO 2022 escape sequences.	16 chars maximum (see ???)

VR Name	Definition	Character Repertoire	Length of Value
ST Short Text	A character string that may contain one or more paragraphs. It may contain the Graphic Character set and the Control Characters, TAB (see Note 3) , 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) excluding Control Characters except TAB, LF, FF, CR (and ESC when used for ISO 2022 escape sequences).	1024 chars maximum (see ???)
UC Unlimited Characters	A character string that may be of unlimited length that may be padded with trailing spaces. The character code 5CH (the BACKSLASH "\" in ISO-IR 6) shall not be present, as it is used as the delimiter between values in multiple valued data elements. The string shall not have Control Characters except for ESC.	Default Character Repertoire and/or as defined by (0008,0005) excluding character code 5CH (the BACKSLASH "\" in ISO-IR 6), and all Control Characters except ESC when used for ISO 2022 escape sequences.	2 ³² -2 bytes maximum See Note 2
UT Unlimited Text	A character string that may contain one or more paragraphs. It may contain the Graphic Character set and the Control Characters, TAB (see Note 3) , 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) excluding Control Characters except TAB, LF, FF, CR (and ESC when used for ISO 2022 escape sequences).	2 ³² -2 bytes maximum See Note 2

Note

1. For attributes that were present in ACR-NEMA 1.0 and 2.0 and that have been retired, the specifications of Value Representation and Value Multiplicity provided are recommendations for the purpose of interpreting their values in objects created in accordance with earlier versions of this standard. These recommendations are suggested as most appropriate for a particular attribute; however, there is no guarantee that historical objects will not violate some requirements or specified VR and/or VM.
2. The length of the value of the UT VR is limited only by the size of the maximum unsigned integer representable in a 32 bit VL field minus one, since FFFFFFFFH is reserved.
3. **In previous editions of the standard (see PS3.5 2015a), the TAB character was not listed as permitted for the ST, LT and UT VRs. It has been added for the convenience of formatting and the encoding of XML text.**