DCMTK - Bug #1123

Private Creator tags require special handling in character set conversion

2024-05-31 14:37 - Marco Eichelberg

Status:	Closed		**Start date:**	2024-05-31
Priority:	Normal		**Due date:**	
Assignee:	Marco Eichelberg		**% Done:**	100%
Category:	Library and Apps		**Estimated time:**	2:00 hours
Target version:	3.7.1+			
Module:	dcmdata		**Compiler:**	
Operating System:				

Description

Private Creator elements (gggg,0010-00FF) have a Value Representation of "LO". Normally, "LO" is a value representation that permits extended character sets. However, DICOM Part 3, Section 7.8.1 states:

> The Private Creator Data Elements shall only contain characters from the Default Character Repertoire and not an Extended or Replacement Character Repertoire, even though the LO VR is one that is affected by the Specific Character Set (0008,0005).

There is one special case that is probably not handled correctly in DCMTK: When the source dataset is encoded in JIS X 0201 (ISO_IR 13), then character 0x7E is not defined as "~", but as an Overline "¯". When converting to UTF-8 (ISO_IR 192) or any other character set, however, 0x7E must be mapped to "~" and not to Overline, since only characters from the Default Character Repertoire are permitted. This special handling only applies to Private Creator elements.

This is the second,special case required for ISO_IR 13. The handling of the backslash character / Yen character as a separator was already fixed in commit #da5370947.

Source: E-mail discussion with Mathieu Malaterre, Jouke Numan and David Clunie concerning the upcoming DICOM CP 2396.

History

#1 - 2024-06-30 12:21 - Marco Eichelberg

- Status changed from New to Closed

- Assignee set to Marco Eichelberg

- % Done changed from 0 to 100

- Estimated time set to 2:00 h

Closed by commit #65d4f8617.