

DICOM Change Proposal Form

Tracking Information - Administrator Use Only	
Change Proposal Number:	CP-107
STATUS:	Voting Packet Oct. 1998
Date of Last Update:	31 Aug 1998
Person Assigned:	L. Tarbox

Submission Abstract: Add support for UNICODE (ISO 10646) character sets	
Submitter Name: Lawrence Tarbox	Submission date: 8 May 1997
Organization: Siemens	Type of Change Proposal: Addition
Name of Document: Part 3: Information Object Definitions, Part 5: Data Structures and Encoding	Version Number: PS 3.3-1996, PS 3.5-1996
<p>Rationale for change: The use of Unicode (now standardized via ISO10646) is becoming more prevalent. Support is built into some operating systems. In addition, Unicode covers a much broader base of character sets than currently covered by the sets listed in Part 3. In order to better support Unicode, DICOM should allow it as one of its supported character sets. This is not intended to replace national character sets defined by specific countries, but rather to provide support for multinational systems, particularly where countries have not yet recommended national character sets for use in DICOM.</p>	

Sections of document affected/ Suggest Wording of Change:

Part 3 C.12.1.1.2 Specific Character Set

Add the following entry to Table C.12-2

Character Set Description	Defined Term	ISO registration number	Number of characters	Code element	Character Set
ISO 10646 UTF-8	ISO_IR 196	ISO-IR 196	128	G1	ISO 10646 characters encoded as bytes according to UTF-8
		ISO-IR 196	94	G0	ISO 10646 character code values from 00 to 7F as per UTF-8.

Note: The ISO 10646 code values from 00 to 7F encoded via UTF-8 are essentially equivalent to the DICOM default Character Repertoire (ISO-IR 6).

Add the following text:

When the value ISO_IR 196 (i.e., ISO 10646 UTF-8) is used for Specific Character Set, then only one value of Specific Character Set shall be present.

Note: Since ISO 10646 UTF-8 maps to most character sets, there is no need for ISO escape sequences to switch character sets.

Part 5 Section 6

Add the following to the end of section 6.1 “Support of character repertoires”:

In addition, DICOM supports ISO 10646 in the UTF-8 encoding as a character set.

Note Unicode can be considered a profile of ISO 10646.

Modify the last sentence of section 6.1.1 “Representation Of Encoded Character Values” to read:

Only some Control Characters from the C0 set are used in DICOM (see Section 6.1.3), and characters from the C1 set shall not be used unless ISO 10646 encoded as UTF-8 is selected as the Character Repertoire. When ISO 10646 is selected as the current Character Repertoire, values from the C1 set are permitted in the character byte stream as defined by the UTF-8 encoding rules.

Modify the last bullet item of section 6.1.2.5.1 “Assumed Initial States” to read:

- The primary set of Control Characters shall always be designated as the C0 code element and this shall be invoked in the CL area of the code table. The C1 code element shall not be used unless ISO 10646 encoded as UTF-8 is selected as the Character Repertoire. When ISO 10646 is selected as the current Character Repertoire, values from the CR area of the code table are permitted in the character byte stream as defined by the UTF-8 encoding rules.