Digital Imaging and Communications in Medicine (DICOM)

Supplement 104: DICOM Encapsulation of PDF Documents

Prepared by:

DICOM Standards Committee, Working Group 6

1300 N. 17th Street Suite 1847

Rosslyn, Virginia 22209 USA

VERSION: Final Text

March 22, 2005

Table of Contents

Foreword

This supplement proposes additions to the DICOM Standard to provide for the encapsulation of documents that have been encoded using the Portable Document Format (PDF) so that these documents may be exchanged between various types of equipment using DICOM messages.

This Supplement proposes changes to the following Parts of the DICOM Standard:

PS 3.2 - Conformance

PS 3.3 - Information Object Definitions

PS 3.4 - Service Class Specifications

PS 3.6 - Data Dictionary

PS 3.16 - Content Mapping Resource

Scope and Field of Application

Clinical evidence, measurements and reports are often generated or scanned into a PDF format.

The PDF format is well documented in a publicly available form, requires no license fees for its use, and is supported by tools from multiple vendors as well as from free, public domain sources. In order to exchange and/or handle these documents in an efficient manner in an imaging environment, it is useful to be able to "wrap" these types of documents in a DICOM container, so they can be exchanged as DICOM objects using the DICOM Storage Service and accordingly archived and retrieved.

Some original data objects (scanned documents, forms, waveforms, measurements, and text) could also be exchanged using other, more comprehensive encoding such as the DICOM Secondary Capture, Waveform or SR objects. However, this proposal for PDF encapsulation is no different than using the Secondary Capture (SC) object instead of other alternative objects such as XA or US. Furthermore, some forms of data to be exchanged are not yet standardized.

Therefore, this supplement defines a SOP Class to encapsulate PDF documents into a Composite DICOM SOP Instance, so that it can be exchanged using the appropriate Service Classes and stored and retrieved accordingly.

The PDF format specification is available from: http://partners.adobe.com/asn/tech/pdf/specifications.jsp

Changes to NEMA Standards Publication PS 3.2-2004

Digital Imaging and Communications in Medicine

PART 2 Addendum

Conformance

Item: Add to table A.1-2 categorizing SOP Classes:

The SOP Classes are categorized as follows:

Table A.1-2
UID VALUES

UID Value	UID NAME	Category
…	…	…
1.2.840.10008.5.1.4.1.1.104.1	**Encapsulated PDF Storage SOP Class**	**Transfer**
…	…	…

67

68

69 **Changes to NEMA Standards Publication PS 3.3-2004**

70 **Digital Imaging and Communications in Medicine (DICOM)**

71 **Part 3: Information Object Definitions**

Modify Section 2 Normative References – add new reference.

2 Normative references

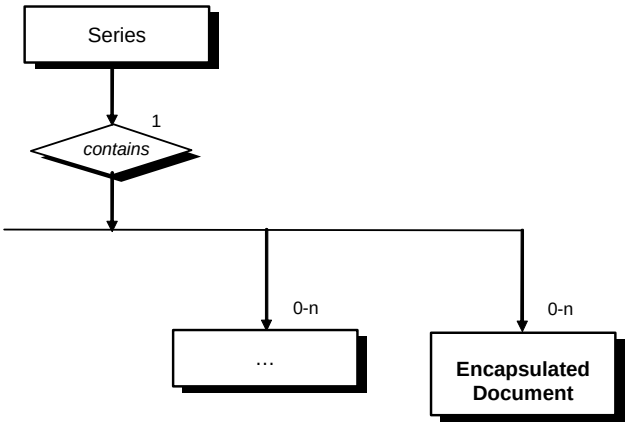
RFC-2046, Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types, November 1996

Modify Section 4 DICOM Symbols and Abbreviations – add new symbol.

The following symbols and abbreviations are used in this Part of the Standard.

…	…
PDF	Portable Document Format
…	…

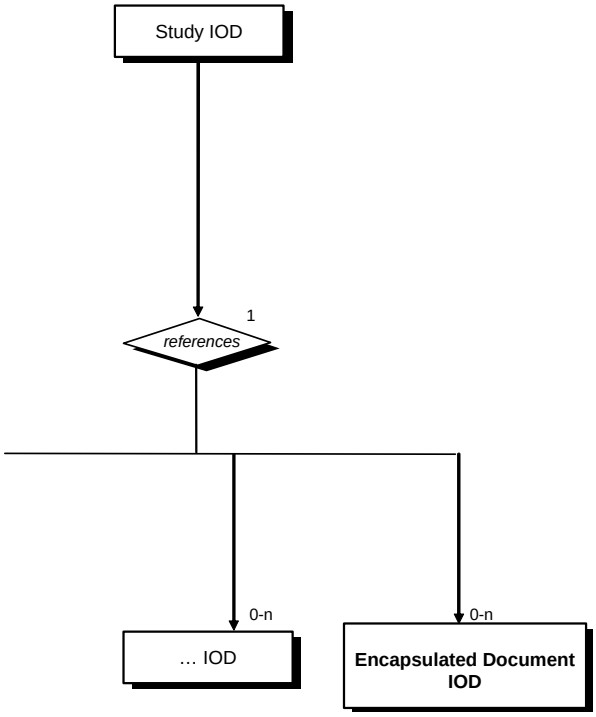
Modify Section 7 DICOM Model of the real world – Add Encapsulated Document to Figure 7-1a



89

90 **Figure 7-1a**
91 **DICOM MODEL OF THE REAL WORLD**

92



93

94

95 **Figure 7-2a**
96 **DICOM INFORMATION MODEL**

97

Modify Section A.1.2 IOD E-R Model – Add Encapsulated Document to Figure A.1-1.

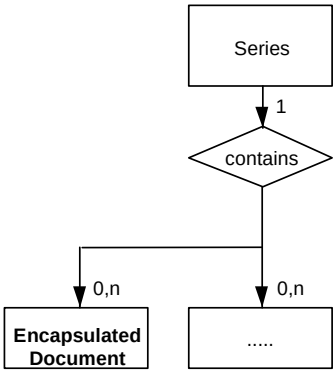


Figure A.1-1
DICOM COMPOSITE INSTANCE IOD INFORMATION MODEL

Modify Section A.1.2 IOD E-R Model – Add description for Encapsulated Document IE

A.1.2.xx Encapsulated Document IE

The Encapsulated Document IE defines the Attributes that describe the content of a non-DICOM formatted document that is encapsulated in a DICOM Attribute. These include Attributes related to document origin, title, and other characteristics. An Encapsulated Document SOP Instance is related to a single Series within a single Study.

Modify Section A.1.4 Overview of the Composite IOD Module Content – Insert Encapsulated PDF

Table A.1-2
COMPOSITE INFORMATION OBJECT MODULES OVERVIEW - NON-IMAGES

IODs Modules	…	Encapsulated PDF	…
Patient		**M**	
Patient Summary			
Specimen Identification		**U**	
Clinical Trial Subject		**U**	

General Study		**M**	
Patient Study		**U**	
Clinical Trial Study		**U**	
Study Content			
Encapsulated Document Series		**M**	
Clinical Trial Series		**U**	
…			
General Equipment		**M**	
SC Equipment		**M**	
…			
Encapsulated Document		**M**	
…			
SOP Common		**M**	

116

117 **Modify Annex A – Insert new section for Encapsulated PDF IOD**

118 **A.YY ENCAPSULATED DOCUMENT INFORMATION OBJECT DEFINITION**

119 **A.YY.1 Encapsulated PDF Information Object Definition**

120 **A.YY.1.1 Encapsulated PDF IOD Description**

121 The Encapsulated PDF Information Object Definition (IOD) describes a PDF document that has been
122 encapsulated within a DICOM information object.

123 **A.YY.1.2 Encapsulated PDF Entity-Relationship Model**

124 The E-R Model in Section A.1.2 of this Part applies to the Encapsulated PDF IOD.

125 **A.YY.1.3 Encapsulated PDF IOD Module Table**

126 Table A.YY.1-1 specifies the Encapsulated PDF IOD Modules.

Table A.YY.1-1
Encapsulated PDF IOD MODULES

IE	Module	Reference	Usage
Patient	Patient	C.7.1.1	M
	Specimen Identification	C.7.1.2	U
	Clinical Trial Subject	C.7.1.3	U
Study	General Study	C.7.2.1	M
	Patient Study	C.7.2.2	U
	Clinical Trial Study	C.7.2.3	U
Series	Encapsulated Document Series	C.YY.1	M
	Clinical Trial Series	C.7.3.2	U
Equipment	General Equipment	C.7.5.1	M
	SC Equipment	C.8.6.1	M
Encapsulated Document	Encapsulated Document	C.ZZ.1	M
	SOP Common	C.12.1	M

A.YY.1.4 Encapsulated PDF IOD content constraints

A.YY.1.4.1 MIME Type of Encapsulated Document

The Enumerated Value of the MIME Type of Encapsulated Document (0042,0012) shall be 'application/pdf'

Modify Annex C – C8.6.1 SC Equipment Module text

This Module describes equipment used to convert images**, documents, and other data** into a DICOM format.

Modify Annex C – Insert new sections for Encapsulated Document Module Tables

C.YY ENCAPSULATED DOCUMENT MODULES

C.YY.1 Encapsulated Document Series Module

Table C.YY-1 defines the Encapsulated Document Series Attributes.

Table C.YY-1
Encapsulated Document Series Module Attributes

Attribute Name	Tag	Type	Attribute Description
Modality	(0008,0060)	1	The modality appropriate for the encapsulated document. This Type definition shall override the definition in the SC Equipment Module.

			See section C.7.3.1.1.1 for Defined Terms.
Series Instance UID	(0020,000E)	1	Unique identifier of the Series.
Series Number	(0020,0011)	1	A number that identifies the Series.
Referenced Performed Procedure Step Sequence	(0008,1111)	3	Uniquely identifies the Performed Procedure Step SOP Instance for which the Series is created. Only a single Item shall be permitted in this sequence. Note: The Performed Procedure Step referred to by this Attribute is the Step during which this Document is generated.
>Referenced SOP Class UID	(0008,1150)	1	Uniquely identifies the referenced SOP Class.
> Referenced SOP Instance UID	(0008,1155)	1	Uniquely identifies the referenced SOP Instance.
Series Description	(0008,103E)	3	User provided description of the Series
Request Attributes Sequence	(0040,0275)	3	Sequence that contains attributes from the Imaging Service Request. The sequence may have one or more Items.
>Requested Procedure ID	(0040,1001)	1C	Identifier that identifies the Requested Procedure in the Imaging Service Request. Required if Sequence Item is present.
>Reason for the Requested Procedure	(0040,1002)	3	Reason for requesting this procedure.
>Reason for Requested Procedure Code Sequence	(0040,100A)	3	Coded Reason for requesting this procedure.
>>Include 'Code Sequence Macro' Table 8.8-1			*No Baseline Context ID is defined.*
>Scheduled Procedure Step ID	(0040,0009)	1C	Identifier that identifies the Scheduled Procedure Step. Required if Sequence Item is present.
>Scheduled Procedure Step Description	(0040,0007)	3	Institution-generated description or classification of the Scheduled Procedure Step to be performed.
>Scheduled Protocol Code Sequence	(0040,0008)	3	Sequence describing the Scheduled Protocol following a specific coding scheme. This sequence contains one or more Items.
>>Include 'Code Sequence Macro' Table 8.8-1			*No Baseline Context ID is defined.*
>>Protocol Context Sequence	(0040,0440)	3	Sequence that specifies the context for the Scheduled Protocol Code Sequence Item. One or more items may be included in this sequence.
>>>Include 'Content Item Macro' Table 10-2			*No Baseline Template is defined.*
>>> Content Item Modifier Sequence	(0040,0441)	3	Sequence that specifies modifiers for a Protocol Context Content Item. One or more items may be included in this sequence. See Section C.4.10.1.
>>>>Include 'Content Item Macro' Table 10-2			*No Baseline Template is defined.*

Performed Procedure Step ID	(0040,0253)	3	User or equipment generated identifier of that part of a Procedure that has been carried out within this step.
Performed Procedure Step Start Date	(0040,0244)	3	Date on which the Performed Procedure Step started.
Performed Procedure Step Start Time	(0040,0245)	3	Time on which the Performed Procedure Step started.
Performed Procedure Step Description	(0040,0254)	3	Institution-generated description or classification of the Procedure Step that was performed.
Performed Protocol Code Sequence	(0040,0260)	3	Sequence describing the Protocol performed for this Procedure Step. One or more Items may be included in this Sequence.
>Include 'Code Sequence Macro' Table 8.8-1			*No Baseline Context ID is defined.*
>Protocol Context Sequence	(0040,0440)	3	Sequence that specifies the context for the Performed Protocol Code Sequence Item. One or more items may be included in this sequence.
>>Include 'Content Item Macro' Table 10-2			*No Baseline Template is defined.*
>> Content Item Modifier Sequence	(0040,0441)	3	Sequence that specifies modifiers for a Protocol Context Content Item. One or more items may be included in this sequence. See Section C.4.10.1.
>>>Include 'Content Item Macro' Table 10-2			*No Baseline Template is defined.*
Comments on the Performed Procedure Step	(0040,0280)	3	User-defined comments on the Performed Procedure Step.

145

146 **C.YY.2 Encapsulated Document Module**

147 Table C.YY-2 defines the Encapsulated Document Attributes.

148 **Table C.YY-2**
149 **Encapsulated Document Module Attributes**

Attribute Name	Tag	Type	Attribute Description
Instance Number	(0020,0013)	1	A number that identifies this SOP Instance. The value shall be unique within a series.
Content Date	(0008,0023)	2	The date the document content creation was started.
Content Time	(0008,0033)	2	The time the document content creation was started.
Acquisition Datetime	(0008,002A)	2	The date and time that the original generation of the data in the document started.
Burned In Annotation	(0028,0301)	1	Indicates whether or not the encapsulated document contains sufficient burned in annotation to identify the patient and date the data was acquired. Enumerated Values: YES NO
Source Instance	(0042,0013)	1C	A sequence that identifies the set of Instances that

Sequence			were used to derive the encapsulated document. One or more Items may be included in this Sequence. Required if derived from one or more DICOM Instances. May be present otherwise.
>Referenced SOP Class UID	(0008,1150)	1	Uniquely identifies the referenced SOP Class.
> Referenced SOP Instance UID	(0008,1155)	1	Uniquely identifies the referenced SOP Instance.
Document Title	(0042,0010)	2	The title of the document. Note: In the case of a PDF encapsulated document, this may be the value of the "Title" entry in the "Document Information Directory" as encoded in the PDF data.
Concept Name Code Sequence	(0040,A043)	2	A coded representation of the document title. Zero or one item may be present.
>Include 'Code Sequence Macro' Table 8.8-1			*Baseline Context Group 7020*
MIME Type of Encapsulated Document	(0042,0012)	1	The type of the encapsulated document stream described using the MIME Media Type (see RFC 2046).
Encapsulated Document	(0042,0011)	1	Encapsulated Document stream, containing a document encoded according to the MIME Type.

Note: One could distinguish four stages in the creation of the Encapsulated Document Object, identified by the following Attributes:

1. Measurement and/or data collection, identified by Acquisition Datetime (0008,002A) in the Encapsulated Document Module.
2. Creation of the original documentation of the data collection, identified by Content Date (0008,0023) and Content Time (0008,0033).
3. Rendering of the original documentation into the format that will be encapsulated, e.g. a PDF document. The rendering time is not captured by any DICOM Attribute, but may be encoded in the rendering.
4. Encapsulation of the rendering into a DICOM Object, identified by Instance Creation Date (0008,0012) and Instance Creation Time (0008,0013) in the SOP Common Module.

Modify Annex F Basic Directory Information Object Definition (Normative)– add new item to tables.

Table F.3-3
DIRECTORY INFORMATION MODULE

Attribute Name	Tag	Type	Attribute Description
…	…	…	…
>Directory Record Type	(0004,1430)	1C	Defines a specialized type of Directory Record by reference to its position in the Media Storage Directory Information Model (see Section F.4). Required if the Directory Record Sequence (0004,1220) is not zero length.

			Enumerated Values (see Section F.5): PATIENT STUDY SERIES IMAGE OVERLAY MODALITY LUT VOI LUT CURVE TOPIC VISIT RESULTS INTERPRETATION STUDY COMPONENT STORED PRINT RT DOSE RT STRUCTURE SET RT PLAN RT TREAT RECORD PRESENTATION WAVEFORM SR DOCUMENT KEY OBJECT DOC SPECTROSCOPY RAW DATA REGISTRATION FIDUCIAL **ENCAP DOC** PRIVATE = Privately defined record hierarchy position. Type shall be defined by Private Record UID (0004,1432). MRDR = Special Directory Record which allows indirect reference to a File by multiple Directory Records. Instead of directly referencing a File by its Referenced File ID (0004,1500), a Directory Record of any of the Types define above (except MRDR) may reference a Multi-Referenced File Directory Record which in turn will reference the File by its File ID. Note: Enumerated Values PRINT QUEUE, FILM SESSION, FILM BOX, and IMAGE BOX were previously defined in DICOM for this Attribute. They are now retired. See PS3.3-1998.

Table F.4-1
RELATIONSHIP BETWEEN DIRECTORY RECORDS

Directory Record Type	Section	Directory Record Types which may be included in the next lower-level directory Entity
(Root Directory Entity)	—	PATIENT, TOPIC, PRIVATE
PATIENT	F.5.1	STUDY, PRIVATE
STUDY	F.5.2	SERIES, VISIT, RESULTS, STUDY COMPONENT PRIVATE
SERIES	F.5.3	IMAGE, OVERLAY, MODALITY LUT, VOI LUT, CURVE, STORED PRINT, RT DOSE, RT STRUCTURE SET, RT PLAN, RT TREAT RECORD, PRESENTATION, WAVEFORM, SR DOCUMENT, KEY OBJECT DOC, SPECTROSCOPY, RAW DATA, REGISTRATION, FIDUCIAL, PRIVATE, **ENCAP DOC**
ENCAP DOC	**F.5.31**	**PRIVATE**
TOPIC	F.5.9	STUDY, SERIES, IMAGE, OVERLAY, MODALITY LUT, VOI LUT, CURVE, STORED PRINT, RT DOSE, RT STRUCTURE SET, RT PLAN, RT TREAT RECORD, PRESENTATION, WAVEFORM, SR DOCUMENT, KEY OBJECT DOC, SPECTROSCOPY, RAW DATA, REGISTRATION, FIDUCIAL, PRIVATE, **ENCAP DOC**

Modify Annex F Basic Directory Information Object Definition (Normative)– add new section.

F.5.31 Encapsulated Document directory record definition

The Directory Record is based on the specification of Section F.3. It is identified by a Directory Record Type of Value "ENCAP DOC." Table F.5-31 lists the set of keys with their associated Types for such a Directory Record Type. The description of these keys may be found in the Modules related to the Encapsulated Document IE of the Encapsulated PDF IOD. This Directory Record shall be used to reference an Encapsulated PDF SOP Instance. This type of Directory Record may reference a Lower-Level Directory Entity that includes one or more Directory Records as defined in Table F.4-2.

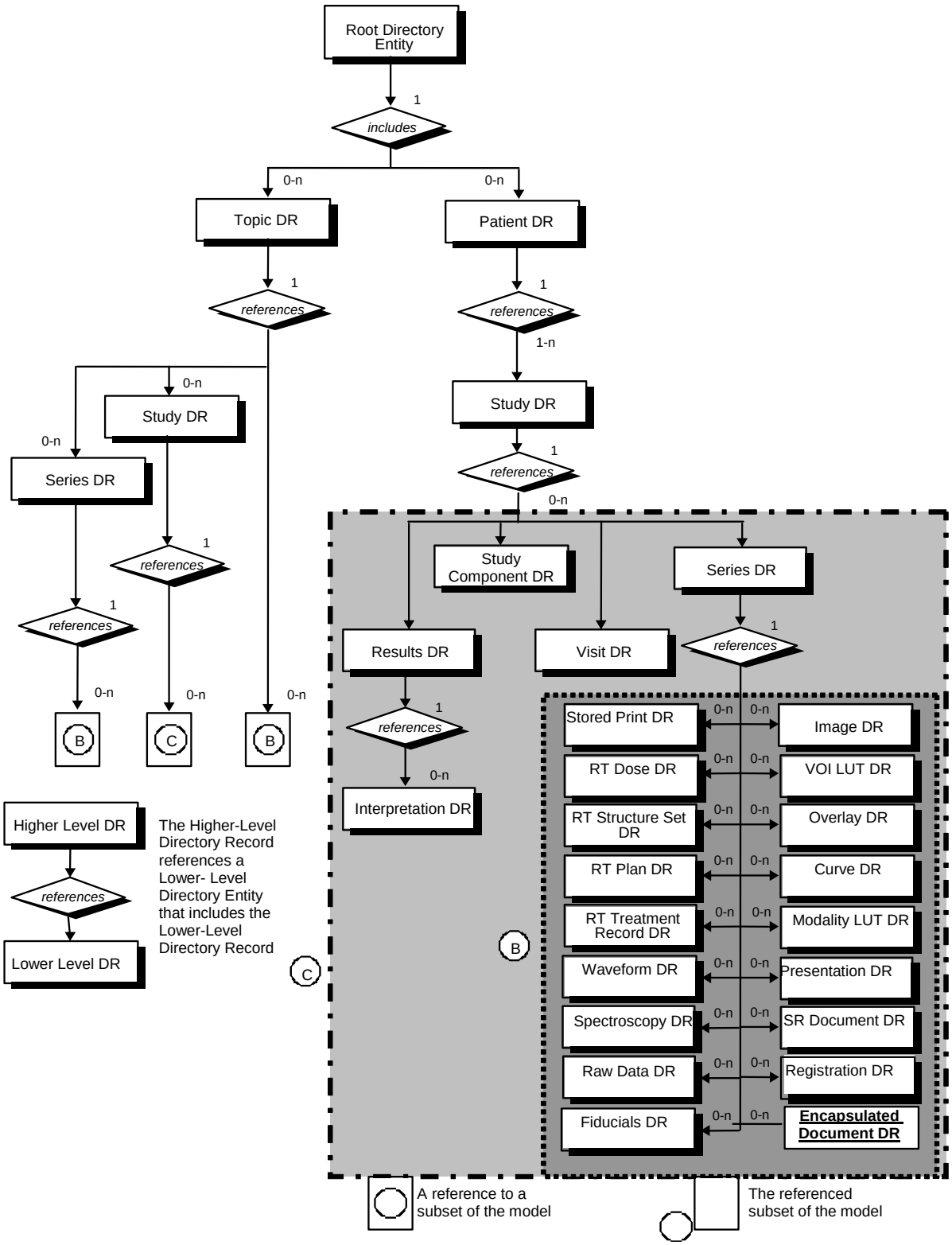
Note: Other Encapsulated Document SOP Classes may be added to the standard in the future and these will likely be referenced by this directory record. Therefore, the MIME Type should be checked rather than assuming that the referenced file contains PDF.

Table F.5-31
Encapsulated Document KEYS

Key	Tag	Type	Attribute Description
Content Date	(0008,0023)	2	The date the content creation started.
Content Time	(0008,0033)	2	The time the content creation started.
Instance Number	(0020,0013)	1	A number that identifies this instance
Document Title	(0042,0010)	2	The title of the document.
Concept Name Code Sequence	(0040,A043)	2	A coded representation of the document title. Zero or one item may be present.
>Include 'Code Sequence Macro' Table 8.8-1		1	*Baseline Context Group 7020*
MIME Type of Encapsulated Document	(0042,0012)	1	The type of the encapsulated document stream described using the MIME Media Type (see RFC 2046).
Any other Attribute of the Encapsulated Document Module except Encapsulated Document (0042,0011)		3	

Modify figure F.4-1 BASIC DIRECTORY IOD INFORMATION MODEL – add new item.

189


Root Directory Entity
1
includes
0-n
0-n
Topic DR
Patient DR
1
references
1
references
1-n
0-n
Study DR
Study DR
0-n
Series DR
1
references
1
references
1
references
0-n
0-n
0-n
B
C
B
0-n
Study Component DR
Series DR
Results DR
Visit DR
1
references
1
references
0-n
Interpretation DR
Stored Print DR
Image DR
RT Dose DR
VOI LUT DR
RT Structure Set DR
Overlay DR
RT Plan DR
Curve DR
RT Treatment Record DR
Modality LUT DR
Waveform DR
Presentation DR
Spectroscopy DR
SR Document DR
Raw Data DR
Registration DR
Fiducials DR
Encapsulated Document DR
Higher Level DR
references
Lower Level DR
The Higher-Level Directory Record references a Lower- Level Directory Entity that includes the Lower-Level Directory Record
A reference to a subset of the model
The referenced subset of the model

190

191

Changes to NEMA Standards Publication PS 3.4-2004

Digital Imaging and Communications in Medicine (DICOM)

Part 4: Service Class Specifications

Modify Annex B.5 STANDARD SOP CLASSES – add new item.

B.5 STANDARD SOP CLASSES

Table B.5-1
STANDARD SOP CLASSES

SOP Class Name	SOP Class UID	IOD (See PS 3.3)
…		
Encapsulated PDF Storage	**1.2.840.10008.5.1.4.1.1.104.1**	**Encapsulated PDF IOD**
…		

Modify Annex I.4 MEDIA STORAGE STANDARD SOP CLASSES – add new item.

I.4 MEDIA STANDARD STORAGE SOP CLASSES

Table I.4-1
Media Storage Standard SOP Classes

SOP Class Name	SOP Class UID	IOD Specification
…		
Encapsulated PDF Storage	**1.2.840.10008.5.1.4.1.1.104.1**	**Encapsulated PDF IOD**
…		

206

207 # Changes to NEMA Standards Publication PS 3.6-2004

208 # Digital Imaging and Communications in Medicine (DICOM)

209 # Part 6: Data Dictionary

210

Modify PS3.6 Section 6 Registry of DICOM data elements – add new items.

Tag	Name	VR	VM	
…				
(0042,0010)	**Document Title**	**ST**	**1**	
(0042,0011)	**Encapsulated Document**	**OB**	**1**	
(0042,0012)	**MIME Type of Encapsulated Document**	**LO**	**1**	
(0042,0013)	**Source Instance Sequence**	**SQ**	**1**	
…				

Modify PS3.6 Annex A Registry of DICOM unique identifiers (UID) – add new items.

UID Value	UID NAME	UID TYPE	Part
…			
1.2.840.10008.5.1.4.1.1.104.1	**Encapsulated PDF Storage**	**SOP Class**	**PS 3.4**
…			

219

220 # Changes to NEMA Standards Publication PS 3.16-2004

221 # Digital Imaging and Communications in Medicine (DICOM)

222 # Part 16 Addendum

223

Add to PS3.16 Annex B

CID 7020 Document Titles

Context Group ID 7020 comprises all document names (i.e., terms with Scale “DOC”) within the HIPAA Attachments class of the LOINC coding scheme. The Coding Scheme Designator shall be LN.

Note: 1. A subset of this Context Group directly applicable to imaging reports is in CID 7000.

2. The LOINC coding scheme can be found at: http://www.regenstrief.org/loinc