

2

4

6

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

8

*Supplement 142:  
Clinical Trial De-identification Profiles*

10

12

14

16

18

20 *Prepared by:*

22 **DICOM Standards Committee, Working Group 18 Clinical Trials**

1300 N. 17th Street, Suite 1752

24 Rosslyn, Virginia 22209 USA

26 VERSION: Final Text – 2011/01/25

This supplement is prepared pursuant to work item 2008-04-E

28

## Table of Contents

30	Table of Contents .....	2
	Scope and Field of Application.....	3
32	E.1 BASIC APPLICATION LEVEL CONFIDENTIALITY PROFILES.....	4
	E.1.1 De-Identifier .....	4
34	E.1.2 Re-Identifier .....	30
	E.1.3 Conformance Requirements.....	30
36	E.2 BASIC APPLICATION LEVEL CONFIDENTIALITY PROFILE .....	31
	E.3 BASIC APPLICATION LEVEL CONFIDENTIALITY OPTIONS .....	31
38	E.3.1 Clean Pixel Data Option .....	32
	E.3.2 Clean Recognizable Visual Features Option .....	32
40	E.3.3 Clean Graphics Option .....	33
	E.3.4 Clean Structured Content Option .....	33
42	E.3.5 Clean Descriptors Option .....	34
	E.3.6 Retain Longitudinal Temporal Information Options .....	34
44	E.3.7 Retain Patient Characteristics Option .....	35
	E.3.8 Retain Device Identity Option .....	36
46	E.3.9 Retain UIDs Option .....	36
	E.3.10 Retain Safe Private Option .....	37
48	C.7.1.1 Patient Module.....	38
	C.7.6.1 General Image Module .....	39
50	C.12.1 SOP Common Module .....	40
	CID 7050 De-identification Method .....	41

54

## Scope and Field of Application

56 In clinical trials, images are often acquired during the course of clinical care, in which case the patient's  
57 individually identifiable information needs to be removed to protect the patient's privacy. In addition, there  
58 is often a need to remove other information not directly related to the patient's identity per se, but which  
59 might assist in recovering their identity or bias the image interpretation in some way. Conversely, it is  
60 important to preserve certain specific information for quality control and analysis that is essential to the  
61 conduct of the clinical trial, which might otherwise be removed. Since many clinical trials are conducted  
62 globally, both nationally and locally specific privacy concerns (such as espoused by the EU Directive and  
63 HIPAA rule and individual IRBs and ethics committees) need to be addressed. Data and images acquired  
64 for clinical trials are also often released for secondary re-use, in which case addressing privacy concerns  
65 requires great vigilance. In general, it is impractical to leave the decisions as to what to retain or remove to  
66 the individual sites or trials.

66

67 There are also other scenarios in which de-identification may be required, such as creation of teaching  
68 files, other types of publication, as well as submission of images and associated information to registries,  
69 such as oncology or radiation dose registries.

70

71 The existing confidentiality profile in PS 3.15 lists possible attributes that may cause identity leakage,  
72 without weighing the relative merits of their inclusion or replacement, or describing strategies to prioritize or  
73 selectively replace attribute values.

74

75 WG 18 has determined that it is necessary to add additional confidentiality profiles to the DICOM Standard  
76 that are appropriate to specific types of trials, both to provide instruction for implementers, to assure  
77 compliance, and to provide guidance for sites and trial administrators that has been subject to expert  
78 review. A consensus in this respect has arisen out of work by a joint pharmaceutical and contract research  
79 industry round table convened by the Pharmaceutical Research and Manufacturers of America (PhRMA)  
80 and the Drug Information Association (DIA), with participation by regulators and academics.

81 This document is a Supplement to the DICOM Standard. It is an extension to the following parts of the  
82 published DICOM Standard:

83 PS 3.3 – Information Object Definitions

PS 3.6 – Data Dictionary

85 PS 3.15 – Security and System Management Profiles

PS 3.16 – Content Mapping Resource

88

**Modify PS 3.15 Annex E Attribute Confidentiality Profiles as indicated:**

90 **Annex E           ATTRIBUTE CONFIDENTIALITY PROFILES**

92 **This Annex describes Profiles and Options to address the removal and replacement of Attributes within a DICOM Dataset that may potentially result in leakage of Individually Identifiable Information (III) about the patient or other individuals or organizations involved in acquisition.**

94 **Profiles are provided to address the balance between the removal of information and the need to retain information so that the Datasets remain useful for their intended purpose.**

96 **Options are used in addition to profiles to prevent a combinatorial expansion of different Profiles.**

98 ***Re-factor PS 3.15 Annex E.1 Basic Attribute Confidentiality Profiles to describe common requirements, and reference specific requirements that are relocated elsewhere:***

**E.1                   BASIC APPLICATION LEVEL CONFIDENTIALITY PROFILES**

100 **This Basic** Application Level Confidentiality Profiles ~~addresses~~ the following aspects of security:

— Data Confidentiality at the application ~~level~~layer.

102 Other aspects of security not addressed by ~~this~~these profiles, that may be addressed elsewhere in the standard include:

104 — Confidentiality in other layers of the DICOM model;

— Data Integrity.

106 ~~This~~these Profiles ~~is~~are targeted toward creating a special purpose, de-identified version of an already-existing Data Set. It is not intended to replace the original SOP Instance from which the de-identified SOP Instance is created, nor is it intended to act as the primary representation of clinical Data Sets in image archives. The de-identified SOP Instances are useful, for example, in creating teaching or research files, **performing clinical trials, or submission to registries** where the identity of the patient **and other individuals should be required to be protected**, **In some cases, it is also necessary to provide a means of recovering identity by** ~~but still be accessible to~~ authorized personnel.

114 **Options to profiles are defined for specific applications. These options may specify either that additional Attributes are removed or replaced, or that Attributes that would otherwise be removed or replaced are retained.**

116 **E.1.1           De-Identifier**

118 An Application may claim conformance to ~~the Basic an~~ Application Level Confidentiality Profile **and Options** as a de-identifier if it protects **and retains** all Attributes **as specified in the Profile and Options that might be used by unauthorized entities to identify the patient**. Protection in this context is defined as the following process:

1. The application may create one or more instances of the Encrypted Attributes Data Set and copy Attributes to be protected into the (single) item of the Modified Attributes Sequence (0400,0550) of one or more of the Encrypted Attributes Data Set instances.

124 **Notes:** 1. A complete reconstruction of the original Data Set may not be possible; however, Attributes (e.g. SOP Instance UID) in the Modified Attributes Sequence of an Encrypted Attributes Data Set may refer back to  
126 the original SOP Instance holding the original Data Set.

128 **2. It is not required that the Encrypted Attributes Data Set be created; indeed, there may be**  
130 **circumstances where the Dataset is expected to be archived long enough that any**  
**contemporary encryption technology may be inadequate to provide long term protection against**  
**unauthorized recovery of identification.**

132 **3. Other mechanisms to assist in identity recovery or longitudinal consistency of replaced UIDs**  
**or dates and times are deprecated in favor of the Encrypted Attributes Data Set mechanism that**  
134 **is intended for this purpose. For example, if it is desired to include an encrypted hash of the**  
**Patient's Name, it should not be encoded in a separate private attribute implemented for that**  
136 **purpose, but should be included in the Encrypted Attributes Data Set and encoded using the**  
**standard mechanism. This allows for compatibility between different implementations and**  
138 **provides security based on the quality and control of the encryption keys. Note also, that**  
**unencrypted hashes are considerably less secure and should be avoided, since they are**  
**vulnerable to trivial dictionary based attacks.**

- 140
- 142 2. Each Attribute to be protected shall then either be removed from the dataset, or have its value replaced by a different “replacement value” which does not allow identification of the patient.

144 Note: 1. It is the responsibility of the de-identifier to ensure that this process does not negatively affect the  
146 integrity of the Information Object Definition, i. e. Dummy values may be necessary for Type 1 Attributes  
that are protected but may not be sent with zero length, and are to be stored or exchanged in encrypted  
form by applications that may not be aware of the security mechanism.

148 2. The standard does not mandate the use of any particular dummy value, and indeed it may have some  
150 meaning, for example in a data set that may be used for teaching purposes, where the real patient  
152 identifying information is encrypted for later retrieval, but a meaningful alternative form of identification is  
provided. For example, a dummy Patient's Name (0010,0010) may convey the type of pathology in a  
teaching case. It is the responsibility of the de-identifier **software or human operator** to ensure that  
the dummy values cannot be used to identify the patient.

154 3. It is the responsibility of the de-identifier to ensure the consistency of dummy values for Attributes such  
as Study Instance UID (0020,000D) or Frame of Reference UID (0020,0052) if multiple related SOP  
Instances are protected. **Indeed, all Attributes of every entity about the Instance level should**  
156 **remain consistent for all Instances protected, e.g., Patient ID for the Patient entity, Study ID for**  
**the Study entity, Series Number for the Series entity.**

158 4. ~~This standard does~~ **Some profiles do** not allow selective protection of parts of a Sequence of Items.  
160 If an Attribute to be protected is contained in a Sequence of Items, the complete Sequence of Items **may**  
**needs** to be protected.

162 5. The de-identifier should ensure that **no** identifying information ~~that~~ is burned in to the image pixel data  
**either because the modality does not generate such burned in identification in the first place, or**  
164 **by removing it through the use of the is “blackened” (removed). Clean Pixel Data Option; see**  
**Section E.3. If non-pixel data graphics or overlays contain identification, the de-identifier is**  
166 **required to remove them, or clean them if the Clean Graphics option is supported. See Section**  
**E.4.** The means by which **burned in or graphic** identifying information is located and removed is  
168 outside the scope of this standard.

- 170 3. **Each Attribute specified to be retained shall be retained.** At the discretion of the de-identifier,  
Attributes may be added to the dataset to be protected.

172 Note: As an example, the Attribute Patient's Age (0010,1010) might be introduced as a replacement for  
Patient's Birth Date (0010,0030) if the patient's age is of importance, **and the profile permits it.**

- 174 4. **If used, a**All instances of the Encrypted Attributes Data Set shall be encoded with a DICOM Transfer  
176 Syntax, encrypted, and stored in the dataset to be protected as an Item of the Encrypted Attributes  
Sequence (0400,0500). The encryption shall be done using RSA [RFC 2313] for the key transport of  
178 the content-encryption keys. A de-identifier conforming to this security profile may use either AES or  
Triple-DES for content-encryption. The AES key length may be any length allowed by the RFCs. The

180 Triple-DES key length is 168 bits as defined by ANSI X9.52. Encoding shall be performed according  
to the specifications for RSA Key Transport and Triple DES Content Encryption in RFC-3370 and for  
AES Content Encryption in RFC-3565.

182 Note: 1. Each item of the Encrypted Attributes Sequence (0400,0500) consists of two Attributes, Encrypted  
184 Content Transfer Syntax UID (0400,0510) containing the UID of the Transfer Syntax that was used to  
encode the instance of the Encrypted Attributes Data Set, and Encrypted Content (0400,0520)  
186 containing the block of data resulting from the encryption of the Encrypted Attributes Data Set instance.  
2. RSA key transport of the content-encryption keys is specified as a requirement in the European  
188 Prestandard ENV 13608-2: Health Informatics – Security for healthcare communication – Part 2: Secure  
data objects.

- 190  
192 5. No requirements on the size of the asymmetric key pairs used for RSA key transport are defined in this  
confidentiality scheme. Implementations claiming conformance to the Basic Application Level  
Confidentiality Profile as a de-identifier shall always protect (e.g. encrypt and replace) the SOP  
194 Instance UID (0008,0018) Attribute as well as all references to other SOP Instances, whether  
contained in the main dataset or embedded in an Item of a Sequence of Items, that could potentially  
196 be used by unauthorized entities to identify the patient.

198 Note: In the case of a SOP Instance UID embedded in an item of a sequence, this means that the enclosing  
Attribute in the top-level data set must be encrypted in its entirety.

- 200 6. The attribute Patient Identity Removed (0012,0062) shall be replaced or added to the dataset with a  
value of YES, and **one or more codes from PS 3.16 CID 7050 De-identification Method**  
202 **corresponding to the profile and options used shall be a value inserted in De-identification-**  
**Method (0012,0063) or added to De-identification Method Code Sequence (0012,0064). A text**  
204 **string describing the method used may also be inserted in or added to De-identification**  
**Method (0012,0063), but is not required.**

- 206  
208 7. If the Dataset being de-identified is being stored within a DICOM File, then the File Meta Information  
**including the 128 byte preamble**, if present, shall be replaced with a description of the de-identifying  
application. **Otherwise, there is a risk that identity information may leak through unmodified File**  
210 **Meta Information or preamble.** See PS 3.10.

212 The Attributes listed in Table E.1-1 **for each profile are** contained in Standard IODs, **or may be**  
**contained in Standard Extended IODs typically need to be protected to provide a minimal level of**  
214 **confidentiality from identification.** An implementation claiming conformance to **an the Basic**  
Application Level Confidentiality Profile as a de-identifier shall protect **or retain** all instances of the  
216 Attributes listed in Table E.1-1, whether contained in the main dataset or embedded in an Item of a  
Sequence of Items, **unless the implementation can ensure that the content of these Attributes**  
218 **cannot be used by unauthorized entities to identify the patient. The following action codes are**  
**used in the table:**

- 220 – **D – replace with a non-zero length value that may be a dummy value and consistent with**  
**the VR**  
222 – **Z – replace with a zero length value, or a non-zero length value that may be a dummy value**  
**and consistent with the VR**  
224 – **X – remove**  
– **K – keep (unchanged for non-sequence attributes, cleaned for sequences)**

- 226 – **C – clean, that is replace with values of similar meaning known not to contain identifying information and consistent with the VR**
- 228 – **U – replace with a non-zero length UID that is internally consistent within a set of Instances**
- **Z/D – Z unless D is required to maintain IOD conformance (Type 2 versus Type 1)**
- 230 – **X/Z – X unless Z is required to maintain IOD conformance (Type 3 versus Type 2)**
- **X/D – X unless D is required to maintain IOD conformance (Type 3 versus Type 1)**
- 232 – **X/Z/D – X unless Z or D is required to maintain IOD conformance (Type 3 versus Type 2 versus Type 1)**
- 234 – **X/Z/U\* - X unless Z or replacement of contained instance UIDs (U) is required to maintain IOD conformance (Type 3 versus Type 2 versus Type 1 sequences containing UID references)**
- 236

238 **These action codes are applicable to both Sequence and non-Sequence attributes; in the case of Sequences, the action is applicable to the Sequence and all of its contents. Cleaning a sequence (“C” action) may entail either changing values of attributes within that Sequence when the meaning of the Sequence within the context of its use in the IOD is understood, or recursively applying the profile rules to each Dataset in each Item of the Sequence. Keeping a Sequence (“K” action) requires recursively applying the profile rules to each Dataset in each Item of the Sequence (for example, in order to remap any UIDs contained within that sequence).**

240

242

244

246 **A requirement for an Option, when implemented, overrides any requirement for the underlying Profile.**

- Notes:
- 248 1. The Attributes listed in Table E.1-1 may not be sufficient to guarantee confidentiality of patient identity. In particular, identifying information may be contained in Private Attributes, **new Standard Attributes, Retired Standard Attributes and additional Standard Attributes not present in Standard Composite IODs (as defined in PS 3.3) but used in Standard Extended SOP Classes, Table E.1-1 indicates those Attributes that are used in Standard Composite IODs as well as those Attributes that are Retired. Also included in Table E.1-1 are some Elements that are not normally found in a Dataset, but are used in Commands, Directories and Meta Information Headers, but which could be misused within Private Sequences. Dataset Trailing Padding (FFFC,FFFC), Textual Content Items of Structured Reports, textual annotations of Presentation States, Curves or and Overlays are specifically addressed.** It is the responsibility of the de-identifier to ensure that all identifying information is removed.
- 250
- 252
- 254
- 256
- 258 2. It should be noted that conformance to ~~an the Basic~~ Application Level Confidentiality Profile does not necessarily guarantee confidentiality. **For example, if an attacker already has access to the original images, the Pixel Data could be matched, though the probability and impact of such a threat may be deemed to be negligible. If the Encrypted Attributes Sequence is used, it should be understood that a**Any encryption scheme may be vulnerable to attack. Also, an organization’s Security Policy and Key Management policy are recognized to have a much greater impact on the effectiveness of protection.
- 260
- 262
- 264
- 266 ~~**3. If the image pixel data contains ‘burned in’ identifications, the de-identifier may ‘black’ them out to de-identify the pixel data.**~~
- 268 **34. National and local regulations, which may vary, might require that additional attributes be de-identified, though the Profiles and Options have been designed to be sufficient to satisfy known regulations without compromising the usefulness of the de-identified instances for their intended purpose.**
- 270
- 272 **5. Table E.1-1 is normative, but it is subject to extension as the DICOM Standard evolves and other similar Attributes are added to IODs. De-identifiers may take this extensibility into account, for example, by considering handling all dates and times on the basis of their Value Representation of DT, DA or TM, rather than just those date and time Attributes lists.**
- 274

- 276 6. The Profiles and Options do not specify whether the design of a de-identifier should be to  
278 remove what is known to be a risk of identity leakage, or to retain only what is known to be safe.  
280 The former approach may fail when the standard is extended, or when a vendor adds  
282 unanticipated standard or private attributes, whilst the latter requires an extensive, if not  
284 complete, comparison of each instance with the Information Object Definitions in PS 3.3 to avoid  
286 discarding required or useful information. Table E.1-1 defines the minimum actions required for  
288 conformance.
- 290 7. De-identification of Private SOP Classes is not defined.
- 292 8. The “C” (clean) action is specified not only for string VRs, but also for Code Sequences, since  
294 the use of private or local codes and non-standard code meanings may potentially cause identity  
296 leakage.
- 298 9. The Digital Signatures Sequences needs to be removed because it contains the certificate of  
300 the signer; theoretically the signature could be verified and the object re-signed by the de-  
302 identifier itself with its own certificate, but this is not required by the Standard.
- 304 10. In general, there are no CS VR Attributes in this table, since it is usually safe to assume that  
306 code strings do not contain identifying information.
- 308 11. In general, there are no Code Sequence Attributes in this table, since it is usually safe to  
310 assume that coded sequence entries, including private codes, do not contain identifying  
312 information. Exceptions are codes for providers and staff.
- 314 12. The Clean Pixel Data and Clean Recognizable Visual Features Options are not listed in this  
table, since they are defined by descriptions of operations on the Pixel Data itself. The Clean  
Pixel Data option may be applied to the Pixel Data within the Icon Image Sequence, or more likely  
the Icon Image Sequence may be recreated entirely once the Pixel Data of the main Dataset has  
been cleaned. The Icon Image Sequence is to be removed when its Pixel Data cannot be cleaned.
13. The Original Attributes Sequence (0400,0561) (which in turn contains the Modified Attributes  
Sequence (0400,0550)) generally needs to be removed, because it may contain unencrypted  
copies of other Attributes that may have been modified (e.g., coerced to use local identifiers and  
names during import of foreign images); an alternative approach would be to selectively modify  
its contents. This is distinct from the use of the Modified Attributes Sequence (0400,0550) within  
the Encrypted Attributes Sequence (0400,0500).
14. Table E.1-1 distinguishes Attributes that are in standard Composite IODs defined in PS 3.3  
from those that are not; some Attributes are defined in PS 3.3 for other IODs, or have a specific  
usage other than in the top level Dataset of a Composite IOD, but are (mis-)used by  
implementers in instances as a Standard Extended SOP Class at other levels than as defined by  
the Standard. Any such Attributes encountered may be removed without compromising the  
conformance of the instance with the standard IOD. For example, Verifying Observer Sequence  
(0040,A073) is only defined in structured report IODs and hence is described in Table E.1-1 as D  
since it is Type 1C; if encountered in an image instance, it should simply be removed (treated as  
X).



316

**Table E.1-1  
Basic Application Level Confidentiality Profile Attributes**

<b>Attribute Name</b>	<b>Tag</b>	<b>Retired (from PS 3.6)</b>	<b>In Std. Comp. IOD (from PS 3.3)</b>	<b>Basic Profile</b>	<b>Retain Safe Private Option</b>	<b>Retain UIDs Option</b>	<b>Retain Device Ident. Option</b>	<b>Retain Patient Chars. Option</b>	<b>Retain Long. Full Dates Option</b>	<b>Retain Long. Modif. Dates Option</b>	<b>Clean Desc. Option</b>	<b>Clean Struct. Cont. Option</b>	<b>Clean Graph. Option</b>
Accession Number	(0008,0050)	<u>N</u>	<u>Y</u>	<u>Z</u>									
<u>Acquisition Comments</u>	<u>(0018,4000)</u>	<u>Y</u>	<u>N</u>	<u>X</u>							<u>C</u>		
<u>Acquisition Context Sequence</u>	<u>(0040,0555)</u>	<u>N</u>	<u>Y</u>	<u>X</u>								<u>C</u>	
<u>Acquisition Date</u>	<u>(0008,0022)</u>	<u>N</u>	<u>Y</u>	<u>X/Z</u>					<u>K</u>	<u>C</u>			
<u>Acquisition DateTime</u>	<u>(0008,002A)</u>	<u>N</u>	<u>Y</u>	<u>X/D</u>					<u>K</u>	<u>C</u>			
<u>Acquisition Device Processing Description</u>	<u>(0018,1400)</u>	<u>N</u>	<u>Y</u>	<u>X/D</u>							<u>C</u>		
<u>Acquisition Protocol Description</u>	<u>(0018,9424)</u>	<u>N</u>	<u>Y</u>	<u>X</u>							<u>C</u>		
<u>Acquisition Time</u>	<u>(0008,0032)</u>	<u>N</u>	<u>Y</u>	<u>X/Z</u>					<u>K</u>	<u>C</u>			
<u>Actual Human Performers Sequence</u>	<u>(0040,4035)</u>	<u>N</u>	<u>N</u>	<u>X</u>									
Additional Patient's History	(0010,21B0)	<u>N</u>	<u>Y</u>	<u>X</u>							<u>C</u>		
<u>Admission ID</u>	<u>(0038,0010)</u>	<u>N</u>	<u>Y</u>	<u>X</u>									
<u>Admitting Date</u>	<u>(0038,0020)</u>	<u>N</u>	<u>N</u>	<u>X</u>					<u>K</u>	<u>C</u>			



Supplement 142 – Clinical Trials De-identification

<u>Attribute Name</u>	<u>Tag</u>	<u>Retired</u> (from <u>PS 3.6</u> )	<u>In Std.</u> <u>Comp.</u> <u>IOD</u> (from <u>PS 3.3</u> )	<u>Basic</u> <u>Profile</u>	<u>Retain</u> <u>Safe</u> <u>Private</u> <u>Option</u>	<u>Retain</u> <u>UIDs</u> <u>Option</u>	<u>Retain</u> <u>Device</u> <u>Ident.</u> <u>Option</u>	<u>Retain</u> <u>Patient</u> <u>Chars</u> <u>Option</u>	<u>Retain</u> <u>Long.</u> <u>Full</u> <u>Dates</u> <u>Option</u>	<u>Retain</u> <u>Long.</u> <u>Modif.</u> <u>Dates</u> <u>Option</u>	<u>Clean</u> <u>Desc.</u> <u>Option</u>	<u>Clean</u> <u>Struct.</u> <u>Cont.</u> <u>Option</u>	<u>Clean</u> <u>Graph.</u> <u>Option</u>
-----------------------	------------	--	--	--------------------------------	---	---	--	--	---	---	---	---	--

<u>Content Creator's Name</u>	(0070,0084)	<u>N</u>	<u>Y</u>	<u>Z</u>									
<u>Content Creator's Identification Code Sequence</u>	(0070,0086)	<u>N</u>	<u>Y</u>	<u>X</u>									
<u>Content Date</u>	(0008,0023)	<u>N</u>	<u>Y</u>	<u>Z/D</u>					<u>K</u>	<u>C</u>			
<u>Content Sequence</u>	(0040,A730)	<u>N</u>	<u>Y</u>	<u>X</u>								<u>C</u>	
<u>Content Time</u>	(0008,0033)	<u>N</u>	<u>Y</u>	<u>Z/D</u>					<u>K</u>	<u>C</u>			
<u>Context Group Extension Creator UID</u>	(0008,010D)	<u>N</u>	<u>Y</u>	<u>U</u>		<u>K</u>							
<u>Contrast Bolus Agent</u>	(0018,0010)	<u>N</u>	<u>Y</u>	<u>Z/D</u>							<u>C</u>		
<u>Contribution Description</u>	(0018,A003)	<u>N</u>	<u>Y</u>	<u>X</u>							<u>C</u>		
<u>Country of Residence</u>	(0010,2150)	<u>N</u>	<u>N</u>	<u>X</u>									
<u>Creator Version UID</u>	(0008,9123)	<u>N</u>	<u>Y</u>	<u>U</u>		<u>K</u>							
<u>Current Patient Location</u>	(0038,0300)	<u>N</u>	<u>N</u>	<u>X</u>									
<u>Curve Data</u>	(50xx,xxxx)	<u>Y</u>	<u>N</u>	<u>X</u>									<u>C</u>
<u>Curve Date</u>	(0008,0025)	<u>Y</u>	<u>Y</u>	<u>X</u>					<u>K</u>	<u>C</u>			
<u>Curve Time</u>	(0008,0035)	<u>Y</u>	<u>Y</u>	<u>X</u>					<u>K</u>	<u>C</u>			























<b>Attribute Name</b>	<b>Tag</b>	<b>Retired (from PS 3.6)</b>	<b>In Std. Comp. IOD (from PS 3.3)</b>	<b>Basic Profile</b>	<b>Retain Safe Private Option</b>	<b>Retain UIDs Option</b>	<b>Retain Device Ident. Option</b>	<b>Retain Patient Chars Option</b>	<b>Retain Long. Full Dates Option</b>	<b>Retain Long. Modif. Dates Option</b>	<b>Clean Desc. Option</b>	<b>Clean Struct. Cont. Option</b>	<b>Clean Graph. Option</b>
-----------------------	------------	--------------------------------------	--	--------------------------	---	-----------------------------------	--	--	---	---	-----------------------------------	---	------------------------------------

<b>Plate ID</b>	(0018,1004)	<b>N</b>	<b>Y</b>	<b>X</b>			<b>K</b>						
<b>Pre-Medication</b>	(0040,0012)	<b>N</b>	<b>N</b>	<b>X</b>				<b>C</b>					
<b>Pregnancy Status</b>	(0010,21C0)	<b>N</b>	<b>N</b>	<b>X</b>				<b>K</b>					
<b>Private attributes</b>	( <i>gggg,eeee</i> <i>where gggg</i> <i>is odd</i> )	<b>N</b>	<b>N</b>	<b>X</b>	<b>C</b>								
<b>Protocol Name</b>	(0018,1030)	<b>N</b>	<b>Y</b>	<b>X/D</b>							<b>C</b>		
<b>Reason for Imaging Service Request</b>	(0040,2001)	<b>Y</b>	<b>N</b>	<b>X</b>							<b>C</b>		
<b>Reason for Study</b>	(0032,1030)	<b>Y</b>	<b>N</b>	<b>X</b>							<b>C</b>		
<b>Referenced Digital Signature Sequence</b>	(0400,0402)	<b>N</b>	<b>Y</b>	<b>X</b>									
<b>Referenced Frame of Reference UID</b>	(3006,0024)	<b>N</b>	<b>Y</b>	<b>U</b>			<b>K</b>						
<b>Referenced General Purpose Scheduled Procedure Step Transaction UID</b>	(0040,4023)	<b>N</b>	<b>N</b>	<b>U</b>			<b>K</b>						
<b>Referenced Image Sequence</b>	(0008,1140)	<b>N</b>	<b>Y</b>	<b>X/Z/U*</b>			<b>K</b>						









Supplement 142 – Clinical Trials De-identification

<u>Attribute Name</u>	<u>Tag</u>	<u>Retired</u> <u>(from</u> <u>PS 3.6)</u>	<u>In Std.</u> <u>Comp.</u> <u>IOD</u> <u>(from</u> <u>PS 3.3)</u>	<u>Basic</u> <u>Profile</u>	<u>Retain</u> <u>Safe</u> <u>Private</u> <u>Option</u>	<u>Retain</u> <u>UIDs</u> <u>Option</u>	<u>Retain</u> <u>Device</u> <u>Ident.</u> <u>Option</u>	<u>Retain</u> <u>Patient</u> <u>Chars</u> <u>Option</u>	<u>Retain</u> <u>Long.</u> <u>Full</u> <u>Dates</u> <u>Option</u>	<u>Retain</u> <u>Long.</u> <u>Modif.</u> <u>Dates</u> <u>Option</u>	<u>Clean</u> <u>Desc.</u> <u>Option</u>	<u>Clean</u> <u>Struct.</u> <u>Cont.</u> <u>Option</u>	<u>Clean</u> <u>Graph.</u> <u>Option</u>
-----------------------	------------	--	--	--------------------------------	---	---	--	--	---	---	---	---	--

<u>Scheduled</u> <u>Performing</u> <u>Physician</u> <u>Identification</u> <u>Sequence</u>	<u>(0040,000B)</u>	<u>N</u>	<u>N</u>	<u>X</u>									
<u>Scheduled</u> <u>Performing</u> <u>Physician</u> <u>Name</u>	<u>(0040,0006)</u>	<u>N</u>	<u>N</u>	<u>X</u>									
<u>Scheduled</u> <u>Procedure</u> <u>Step</u> <u>End</u> <u>Date</u>	<u>(0040,0004)</u>	<u>N</u>	<u>N</u>	<u>X</u>					<u>K</u>	<u>C</u>			
<u>Scheduled</u> <u>Procedure</u> <u>Step</u> <u>End</u> <u>Time</u>	<u>(0040,0005)</u>	<u>N</u>	<u>N</u>	<u>X</u>					<u>K</u>	<u>C</u>			
<u>Scheduled</u> <u>Procedure</u> <u>Step</u> <u>Description</u>	<u>(0040,0007)</u>	<u>N</u>	<u>Y</u>	<u>X</u>							<u>C</u>		
<u>Scheduled</u> <u>Procedure</u> <u>Step</u> <u>Location</u>	<u>(0040,0011)</u>	<u>N</u>	<u>N</u>	<u>X</u>			<u>K</u>						
<u>Scheduled</u> <u>Procedure</u> <u>Step</u> <u>Start</u> <u>Date</u>	<u>(0040,0002)</u>	<u>N</u>	<u>N</u>	<u>X</u>					<u>K</u>	<u>C</u>			
<u>Scheduled</u> <u>Procedure</u> <u>Step</u> <u>Start</u> <u>Time</u>	<u>(0040,0003)</u>	<u>N</u>	<u>N</u>	<u>X</u>					<u>K</u>	<u>C</u>			

Supplement 142 – Clinical Trials De-identification

<b>Attribute Name</b>	<b>Tag</b>	<b>Retired (from PS 3.6)</b>	<b>In Std. Comp. IOD (from PS 3.3)</b>	<b>Basic Profile</b>	<b>Retain Safe Private Option</b>	<b>Retain UIDs Option</b>	<b>Retain Device Ident. Option</b>	<b>Retain Patient Chars Option</b>	<b>Retain Long. Full Dates Option</b>	<b>Retain Long. Modif. Dates Option</b>	<b>Clean Desc. Option</b>	<b>Clean Struct. Cont. Option</b>	<b>Clean Graph. Option</b>
-----------------------	------------	--------------------------------------	--	--------------------------	---	-----------------------------------	--	--	---	---	-----------------------------------	---	------------------------------------

<b>Scheduled Station AE Title</b>	<b>(0040,0001)</b>	<b>N</b>	<b>N</b>	<b>X</b>			<b>K</b>						
<b>Scheduled Station Geographic Location Code Sequence</b>	<b>(0040,4027)</b>	<b>N</b>	<b>N</b>	<b>X</b>			<b>K</b>						
<b>Scheduled Station Name</b>	<b>(0040,0010)</b>	<b>N</b>	<b>N</b>	<b>X</b>			<b>K</b>						
<b>Scheduled Station Name Code Sequence</b>	<b>(0040,4025)</b>	<b>N</b>	<b>N</b>	<b>X</b>			<b>K</b>						
<b>Scheduled Study Location</b>	<b>(0032,1020)</b>	<b>Y</b>	<b>N</b>	<b>X</b>			<b>K</b>						
<b>Scheduled Study Location AE Title</b>	<b>(0032,1021)</b>	<b>Y</b>	<b>N</b>	<b>X</b>			<b>K</b>						
<b>Series Date</b>	<b>(0008,0021)</b>	<b>N</b>	<b>Y</b>	<b>X/D</b>					<b>K</b>	<b>C</b>			
Series Description	(0008,103E)	<b>N</b>	<b>Y</b>	<b>X</b>							<b>C</b>		
Series Instance UID	(0020,000E)	<b>N</b>	<b>Y</b>	<b>U</b>		<b>K</b>							
<b>Series Time</b>	<b>(0008,0031)</b>	<b>N</b>	<b>Y</b>	<b>X/D</b>					<b>K</b>	<b>C</b>			
<b>Service Episode Description</b>	<b>(0038,0062)</b>	<b>N</b>	<b>Y</b>	<b>X</b>							<b>C</b>		
<b>Service Episode ID</b>	<b>(0038,0060)</b>	<b>N</b>	<b>Y</b>	<b>X</b>									
<b>Smoking Status</b>	<b>(0010,21A0)</b>	<b>N</b>	<b>N</b>	<b>X</b>				<b>K</b>					
SOP Instance UID	(0008,0018)	<b>N</b>	<b>Y</b>	<b>U</b>		<b>K</b>							



Supplement 142 – Clinical Trials De-identification

Attribute Name	Tag	Retired (from PS 3.6)	In Std. Comp. IOD (from PS 3.3)	Basic Profile	Retain Safe Private Option	Retain UIDs Option	Retain Device Ident. Option	Retain Patient Chars Option	Retain Long. Full Dates Option	Retain Long. Modif. Dates Option	Clean Desc. Option	Clean Struct. Cont. Option	Clean Graph. Option
----------------	-----	-----------------------------	---	------------------	-------------------------------------	--------------------------	--------------------------------------	--------------------------------------	--	--	--------------------------	-------------------------------------	---------------------------

<b>Text String</b>	(2030,0020)	<u>N</u>	<u>N</u>	<u>X</u>									
<b>Timezone Offset From UTC</b>	(0008,0201)	<u>N</u>	<u>Y</u>	<u>X</u>					<u>K</u>	<u>C</u>			
<b>Topic Author</b>	(0088,0910)	<u>Y</u>	<u>N</u>	<u>X</u>									
<b>Topic Key Words</b>	(0088,0912)	<u>Y</u>	<u>N</u>	<u>X</u>									
<b>Topic Subject</b>	(0088,0906)	<u>Y</u>	<u>N</u>	<u>X</u>									
<b>Topic Title</b>	(0088,0904)	<u>Y</u>	<u>N</u>	<u>X</u>									
<b>Transaction UID</b>	(0008,1195)	<u>N</u>	<u>N</u>	<u>U</u>		<u>K</u>							
UID	(0040,A124)	<u>N</u>	<u>Y</u>	<u>U</u>									
<b>Verifying Observer Identification Code Sequence</b>	(0040,A088)	<u>N</u>	<u>Y</u>	<u>Z</u>									
<b>Verifying Observer Name</b>	(0040,A075)	<u>N</u>	<u>Y</u>	<u>D</u>									
<b>Verifying Observer Sequence</b>	(0040,A073)	<u>N</u>	<u>Y</u>	<u>D</u>									
<b>Verifying Organization</b>	(0040,A027)	<u>N</u>	<u>Y</u>	<u>X</u>									
<b>Visit Comments</b>	(0038,4000)	<u>N</u>	<u>N</u>	<u>X</u>							<u>C</u>		

320 **E.1.2 Re-Identifier**

322 An Application may claim conformance to ~~an the Basic~~ Application Level Confidentiality Profile as a re-  
324 Encrypted Attributes Sequence (0400,0500) of the SOP instance are available. Removal of protection in  
this context is defined as the following process:

- 326 1. The application shall decrypt, using its recipient key, one instance of the Encrypted Content  
328 (0400,0520) Attribute within the Encrypted Attributes Sequence (0400,0500) and decode the resulting  
330 block of bytes into a DICOM dataset using the Transfer Syntax specified in the Encrypted Content  
Transfer Syntax UID (0400,0510). Re-identifiers claiming conformance to this profile shall be capable  
of decrypting the Encrypted Content using either AES or Triple-DES in all possible key lengths  
specified in this profile.

332 Note: If the application is able to decode more than one instance of the Encrypted Content (0400,0520)  
334 Attribute within the Encrypted Attributes Sequence (0400,0500), it is at the discretion of the application  
to choose any one of them.

- 336 2. The application shall move all Attributes contained in the single item of the Modified Attributes  
338 Sequence (0400,0550) of the decoded dataset into the main dataset, replacing “dummy value”  
Attributes that may be present in the main dataset, and remove the Modified Attributes Sequence  
(0400,0550).

340 Notes: 1. Re-identification does not imply a complete reconstruction of the original SOP Instance, since it is not  
342 required that all Attributes being protected be part of the Encrypted Attributes Data Set. If the original  
UIDs are part of the Encrypted Attributes Data Set, they might be usable to gain access to the original,  
unprotected SOP Instance.

344 2. The presence of an encrypted data set that cannot be decrypted indicates that some or all of the  
attribute values in the message may not be real (they are dummies). Therefore, the recipient must not  
assume that any value in the message is diagnostically relevant.

- 346 3. The attribute Patient Identity Removed (0012,0062) shall be replaced or added to the dataset with a  
348 value of NO and De-identification Method (0012,0063) and De-identification Method Code Sequence  
(0012,0064) shall be removed.

**E.1.3 Conformance Requirements**

350 The Conformance Statement of an application that claims conformance to ~~an the Basic~~ Application Level  
Confidentiality Profile shall describe:

- 352 — which Attributes are removed during protection;  
— which Attributes are replaced by dummy values and how the dummy values are generated;  
354 — which Attributes are included in Encrypted Attributes Data Sets for later re-identification, and  
any pertinent details about how keys are selected for performing the encryption;  
356 — ~~the scope across which whether or not~~ the application is able to ensure **referential** integrity  
of **dummy replacement** values for references such as SOP Instance UID, Frame of Reference  
358 UID, etc. if multiple SOP instances are protected (**e.g., across multiple Studies, consistent  
replacement if the same Study processed more than once, etc.**);  
360 — which Attributes and Attribute values are inserted during protection of a SOP instance;  
— which Transfer Syntaxes are supported for encoding/decoding of the Encrypted Attributes Data  
362 Set;  
— which Confidentiality Schemes are supported;

- 364 — which **Options** are supported;  
— any additional restrictions (e. g. key sizes for public keys).

366

**Add new section to PS 3.15 Annex E Attribute Confidentiality Profiles and Options:**

368 **E.2 BASIC APPLICATION LEVEL CONFIDENTIALITY PROFILE**

370 This profile is intended for use in clinical trials, and other scenarios in which de-identification may be required, such as creation of teaching files, other types of publication, as well as submission of images and associated information to registries, such as oncology or radiation dose registries.

372 This Basic Application Level Confidentiality Profile defines an extremely conservative approach that removes all information related to:

- 374 – the identity and demographic characteristics of the patient  
– the identity of any responsible parties or family members  
376 – the identity of any personnel involved in the procedure  
– the identity of the organizations involved in ordering or performing the procedure  
378 – additional information that could be used to match instances if given access to the originals, such as UIDs, dates and times  
380 – private attributes

when that information is present in the non-Pixel Data Attributes, including graphics or overlays, as described in Table E.1-1.

384 **Note:** Unless the Clean Pixel Data Option is also specified, this profile does not address information burned-in to the pixels.

386 The Attribute Longitudinal Temporal Information Modified (0028,0303) shall be added to the Dataset with a value of “REMOVED” if none of the Retain Longitudinal Temporal Information Options is applied.

**E.3 BASIC APPLICATION LEVEL CONFIDENTIALITY OPTIONS**

388 Various options are defined to be applicable to the Basic Application Level Confidentiality Profile. Some of these options require removal of additional information, and some of these options require retention of  
390 information that would otherwise be removed.

The following options are defined that require removal of additional information:

- 392 – Clean Pixel Data Option  
– Clean Recognizable Visual Features Option  
394 – Clean Graphics Option  
– Clean Structured Content Option  
396 – Clean Descriptors Option

398 The following options are defined that require retention of information that would otherwise be removed but which is needed for specific uses:

- 400 – Retain Longitudinal Temporal Information with Full Dates Option  
– Retain Longitudinal Temporal Information with Modified Dates Option

- 402 – Retain Patient Characteristics Option
- Retain Device Information Option
- 404 – Retain UIDs
- Retain Safe Private Option

406

### **E.3.1 Clean Pixel Data Option**

408 When this Option is specified in addition to an Application Level Confidentiality Profile, any information  
burned in to the Pixel Data (7FE0,0010) corresponding to the Attribute information specified to be removed  
410 by the Profile and any other Options specified shall also be removed, as described in Table E.1-1.

This may require intervention of or approval by a human operator.

412 The Attribute Burned In Annotation (0028,0301) shall be added to the Dataset with a value of “NO”.

- 414 Notes:
- 416 1. This capability is called out as a specific option, since it may be extremely burdensome in practice to implement and is unnecessary for the vast majority of modalities that do not burn in such annotation in the first place. For example, CT images do not normally contain such burned in annotation, whereas Ultrasound images routinely do.
  - 418 2. Though image processing and optical character recognition techniques can be used to detect the presence of and location of burned in text, and matching against known identifying information can be applied, deciding whether or not that text is identifying information or some other type of information may be non-trivial. Compliance with this option requires that identifying information is removed, regardless of how that is achieved. It is not required that information specified to be retained in the non-pixel data by other Options (e.g., physical characteristics, dates or descriptors) also be retained burned-in to the pixel data. Thus the most conservative approach of removing any and all burned in text would be compliant. This may involve sacrificing additional potentially useful information such as localizer posting and manual graphic annotations.
  - 420 3. The stored pixel values are to be changed (blacked out); it is not sufficient to superimpose an overlay or graphic annotation or shutter to obscure the pixel data values, since those may not be ignored by the receiving system.
  - 422 4. This option is intended to apply to the Pixel Data (7FE0,0010) Attribute that occurs in the top level Dataset of an Image Storage SOP Instance. The other standard use of Pixel Data (7FE0,0010) is within Icon Image Sequence (0088,0200), which is already described in Table E.1-1 and the accompanying note as requiring removal. This option does not require the ability to manually or automatically process the pixel values of Pixel Data (7FE0,0010) occurring in any other location than the top level dataset, but it does not prohibit it. Pixel Data (7FE0,0010) occurring within private Attributes will be removed because such Attributes will not be known to be safe.

436

### **E.3.2 Clean Recognizable Visual Features Option**

438 When this Option is specified in addition to an Application Level Confidentiality Profile, if there is sufficient visual information within the Pixel Data of a set of instances to allow an individual to be recognized from  
440 the instances themselves or a reconstruction of a set of instances, then sufficient removal or distortion of the Pixel Data shall be applied to prevent recognition.

442 This may require intervention of or approval by a human operator.

444 The Attribute Recognizable Visual Features (0028,0302) shall be added to the Dataset with a value of “NO”.

- 446 Notes:
1. This capability is called out as a specific option, since it may be extremely burdensome in practice to implement and is unnecessary for the vast majority of anatomic sites and modalities.



448 2. In the case of full-face photographs, the risk of visual identification is obvious, and numerous  
techniques are well established for de-identification, such as applying black rectangles over the eyes,  
etc.

450 3. In the case of high-resolution cross-sectional imaging of the entire head and neck, it has been  
452 suggested that a 3D volume or surface rendering of the pixel data may be sufficient to allow identification  
(or matching against a constrained subset of individuals) under some circumstances.

454 4. Application of this option may render the pixel data unusable for the purpose for which it has been  
collected, and hence its use may require a compromise between de-identification and utility based on  
456 obtaining appropriate ethical approval and informed consent. Consider for example, the case of dental  
images.

### 458 **E.3.3 Clean Graphics Option**

Instances of various Standard and Standard Extended SOP Classes, including Images, Presentation  
460 States and other Composite SOP Instances, may contain identification information encoded as graphics,  
text annotations or overlays. This does not include information contained in Structured Report SOP  
462 Classes.

When this Option is specified in addition to an Application Level Confidentiality Profile, any information  
464 encoded in graphics, text annotations or overlays corresponding to the Attribute information specified to be  
removed by the Profile and any other Options specified shall also be removed, as described in Table E.1-  
466 1.

This may require intervention of a human operator.

- 468 Notes: 1. This capability is called out as a specific option, since it may be more practical to simply remove all  
such graphics, text annotations or overlays (as required by the profile without this option).  
470 2. As with burned-in pixel data annotation, deciding whether or not text is identifying information or some  
472 other type of information may be non-trivial. It is not required that information specified to be retained in  
the non-pixel data by other Options (e.g., physical characteristics, dates or descriptors) also be retained  
in graphics, text annotations or overlays.

### 474 **E.3.4 Clean Structured Content Option**

476 Instances of Structured Report SOP Classes may contain identifiable information in a Content Sequence  
(0040,A730) encoded in Content Items. Instances of other SOP Classes may contain structured content  
478 encoded in a similar manner in the Acquisition Context Sequence (0040,0555) or Specimen Preparation  
Sequence (0040,0610).

480 When this Option is specified in addition to an Application Level Confidentiality Profile, any information  
encoded in SR Content Items or Acquisition Context or Specimen Preparation Sequence Items  
482 corresponding to the Attribute information specified to be removed by the Profile and any other Options  
specified shall also be removed.

- 484 Notes: 1. For example, the “observer” responsible for a diagnostic imaging report may be explicitly identified in  
Observation Content related Content Items in an SR.  
486 2. A de-identifier that does not implement this option creates significant risk when attempting to de-  
identity a Structured Report unless it is only used to de-identify instances that are known to have no  
488 identifying information in the Content Sequence.

490 **E.3.5 Clean Descriptors Option**

492 Even though many Attributes are defined in the DICOM Standard for specific purposes, such as to  
describe a Study or a Series, those that contain plain text over which an operator has control may contain  
unstructured information that includes identities.

494 When this Option is specified in addition to an Application Level Confidentiality Profile, any information that  
is embedded in text or string Attributes corresponding to the Attribute information specified to be removed  
496 by the Profile and any other Options specified shall also be removed, as described in Table E.1-1.

Notes: 1. For example, an operator may include a person's name or a patient's demographics or physical  
498 characteristics in the Study Description (0008,1030), perhaps because their modality user interface does  
not provide other fields or because other systems do not display them. E.g., the description might contain  
500 "CT chest abdomen pelvis – 55F Dr. Smith".

2. One approach to cleaning such text strings without human intervention is to extract and retain only  
502 values known to be useful and safe and discard all others. For example, in the string "CT chest abdomen  
pelvis – 55F Dr. Smith" are found in Study Description (0008,1030), then it would be feasible to detect  
504 and retain "CT chest abdomen pelvis" and discard the remainder. In an international setting, this may  
require an extensive dictionary of words that are safe to retain, e.g., to detect "Buik" for abdomen in  
506 Dutch or "λεκάνη" for pelvis in Greek. Another possibility is to extract such information and attempt to  
code the information in other Attributes (if otherwise absent or empty) such as Anatomic Region  
508 Sequence (0008,2218). However, the possibility of string values being both identifying and descriptive in  
different uses needs to be considered, e.g. "Dr. Hand" or "M. Genou".

3. Table E.1-1 calls out specific Attributes known to be at risk, but an implementer may want to consider  
510 any attribute that could potential contain character data, though this Option does not require that this be  
done. For example, all SH, LO, ST, LT and UT Value Representations could perhaps be misused. Code  
512 strings, CS, are not generally at risk, but a check against known Defined Terms and Enumerated Values  
could be performed. Though extremely unusual, it is conceivable that even a DS or IS string could be  
514 misused, and a check could be made that only legal numeric characters were used. Any PN Attribute is  
obviously at risk. The OB VR is discussed in the Retain Safe Private Option.

3. This Option specifies what needs to be removed, not what needs to be retained. Depending on the  
518 application, it may be desirable to retain some information, such as technique description, but discard  
other information, such as diagnosis, for example because it may bias the interpretation in a clinical trial.  
520 For example, one approach is to remove all description and comment attributes except Series  
Description (0008,103E), since this Attribute rarely contains identifying or diagnosis information yet is  
522 typically a reliable source of useful information about the acquisition technique populated automatically  
from modality device protocols, though it still could be cleaned as described in Note 2.

4. It should be recognized that if any descriptor contains information about a particularly unusual  
524 procedure or condition, then in conjunction with other demographic information it might reduce the  
number of possible individuals that could be the imaging subject. However, this is to some extent true  
526 also if the condition or other unusual physical features are obvious from visual examination of the images  
themselves. E.g., how many conjoined twins born in a particular month in Philadelphia might there be?  
528

530 The manner of cleaning shall be described in the Conformance Statement.

**E.3.6 Retain Longitudinal Temporal Information Options**

532 Dates and times are recognized as having a potential for leakage of identity because they constrain the  
number of possible individuals that could be the imaging subject, though only if there is access to other  
534 information about the individuals concerned to match it against.

However, there are applications that require dates and times to be present to able to fulfill the objective.  
536 This is particularly true in therapeutic clinical trials in which the objective is to measure change in an  
outcome measure over time. Further, it is often necessary to correlate information from images with  
538 information from other sources, such as clinical and laboratory data, and dates and times need to be  
consistent.

540 Two options are specified to address these requirements:

- Retain Longitudinal Temporal Information With Full Dates Option
- 542 – Retain Longitudinal Temporal Information With Modified Dates Option

544 When the Retain Longitudinal Temporal Information With Full Dates Option is specified in addition to an  
Application Level Confidentiality Profile, any dates and times present in the Attributes shall be retained, as  
546 described in Table E.1-1. The Attribute Longitudinal Temporal Information Modified (0028,0303) shall be  
added to the Dataset with a value of “UNMODIFIED”.

548 When the Retain Longitudinal Temporal Information With Modified Dates Option is specified in addition to  
an Application Level Confidentiality Profile, any dates and times present in the Attributes listed in Table  
550 E.1-1 shall be modified. The modification of the dates and times shall be performed in a manner that:

- aggregates or transforms dates so as to reduce the possibility of matching for re-identification
- 552 – preserves the gross longitudinal temporal relationships between images obtained on different  
dates to the extent necessary for the application
- 554 – preserves the fine temporal relationships between images and real-world events to the extent  
necessary for analysis of the images for the application

556 The Attribute Longitudinal Temporal Information Modified (0028,0303) shall be added to the Dataset with a  
value of “MODIFIED”.

- 558 Notes:
1. Aggregation of dates may be performed by various means such as setting all dates to the first day of  
560 the month, all months to the first month of the year, etc., depending on the precision required for the  
application.
  - 562 2. It is possible to modify all dates and times to dummy values by shifting them relative to an arbitrary  
epoch, and hence retain the precise longitudinal temporal relationships amongst a set of studies, when  
either de-identification of the entire set is performed at the same time, or some sort of mapping or  
564 database is kept to repeat this process on separate occasions.
  - 566 3. Transformation of dates and times should be considered together, in order to address studies that  
span midnight.
  - 568 4. Any transformation of times should be performed in such a manner as to not disrupt computations  
needed for analysis, such as comparison of start of injection time to the acquisition time for PET SUV, or  
extraction of time-intensity values from dynamic contrast enhanced studies.

570 The manner of date modification shall be described in the Conformance Statement.

### **E.3.7 Retain Patient Characteristics Option**

572 Physical characteristics of the patient, which are descriptive rather than identifying information per se, are  
recognized as having a potential for leakage of identity because they constrain the number of possible  
574 individuals that could be the imaging subject, though only if there is access to other information about the  
individuals concerned to match it against.

576 However, there are applications that require such physical characteristics in order to perform the  
computations necessary to analyze the images to fulfill the objective. One such class of applications is  
578 those that are related to metabolic measures, such as computation of PET Standard Uptake Values (SUV)  
or DEXA or MRI measures of body composition, which are based on body weight, body surface area or  
580 lean body mass.

When this Option is specified in addition to an Application Level Confidentiality Profile, information about  
582 age, sex, height and weight and other characteristics present in the Attributes shall be retained, as  
described in Table E.1-1.

584 The manner of cleaning of retained attributes shall be described in the Conformance Statement.

### **E.3.8 Retain Device Identity Option**

586 Information about the identity of the device that was used to perform the acquisition is recognized as  
587 having a potential for leakage of identity because it may constrain the number of possible individuals that  
588 could be the imaging subject, though only if there is access to other information about the individuals  
concerned to match it against.

590 However, there are applications that require such device information to perform the analysis or  
591 interpretation. The type of correction for spatial or other inhomogeneity may require knowledge of the  
592 specific device serial number. Confirmation that specific devices that have been previously qualified (e.g.,  
593 with phantoms) may be required. Further, there may be a need to maintain a record of the device used for  
594 regulatory or registry purposes, yet the acquisition site may not maintain an adequate electronic audit trail.

When this Option is specified in addition to an Application Level Confidentiality Profile, information about  
596 the identity of the device in the Attributes shall be retained, as described in Table E.1-1.

### **E.3.9 Retain UIDs Option**

598 Though individuals do not have unique identifiers themselves, studies, series, instances and other entities  
599 in the DICOM model are assigned globally unique UIDs. Whilst these UIDs cannot be mapped directly to  
600 an individual out of context, given access to the original images, or to a database of the original images  
containing the UIDs, it would be possible to recover the individual's identity.

602 However, there are applications that require the ability to maintain an audit trail back to the original images  
603 and though there are other mechanisms they may not scale well or be reliably implemented. This Option is  
604 provided for use when it is judged that the risk of gaining access to the original information via the UIDs is  
small relative to the benefit of retaining them.

606 When this Option is specified in addition to an Application Level Confidentiality Profile, UIDs shall be  
retained, as described in Table E.1-1.

608 Notes: 1. A UID of a DICOM entity is not the same as a unique identifier of an individual, such as would be  
proscribed by some privacy regulations.  
610 2. UIDs are generated using a hierarchical scheme of "roots", which may be traceable by a  
611 knowledgeable person back to the original assignee of the root, typically the device manufacturer, but  
612 sometimes the organization using the device.  
614 3. When evaluating the risk of matching UIDs with the original images or PACS database, one should  
615 consider that even if the UIDs are changed, the pixel data itself presents a similar risk. Specifically, the  
616 pixel data of the de-identified image can be matched against the pixel data of the original image. Such  
617 matching can be greatly accelerated by comparing pre-computed hash values of the pixel data. Removal  
618 of burned-in identification may change the pixel data but then matching against a sub-region of the pixel  
619 data is almost certainly possible (e.g., the central region of an image). Even addition of noise to an image  
620 is not sufficient to prevent re-identification since statistical matching techniques can be used. Ultimately,  
621 if any useable pixel data is retained during de-identification, then re-identification is nearly always  
622 possible if one has access to the original images. Ergo, replacement of UIDs should not give rise to a  
false confidence that the images have been more thoroughly de-identified than if the UIDs are retained.  
624 4. Regardless of this option, implementers should take care not to remove UIDs that are structural and  
defined by the standard as opposed to those that are instance-related. E.g., one would never remove or  
replace the SOP Class UID for de-identification purposes.  
626 5. The Implementation Class UID (0002,0012) is not included in the list of UID attributes to be retained,  
627 since it is part of the File Meta Information (see PS 3.10), which is entirely replaced whenever a file is  
628 stored or modified during de-identification. See E.1.1.

630 **E.3.10 Retain Safe Private Option**

632 By definition, Private Attributes contain proprietary information, in many cases the nature of which is known only to the vendor and not publicly documented.

634 However, some Private Attributes may be necessary for the desired application. For example, specific technique information such as CT helical span pitch, or pixel value transformation, such as PET SUV rescale factors, may only be available in Private Attributes since the information is either not defined in 636 Standard Attributes, or was added to the DICOM Standard after the acquisition device was manufactured.

638 When this Option is specified in addition to an Application Level Confidentiality Profile, Private Attributes that are known by the de-identifier to be safe from identity leakage shall be retained, together with the Private Creator IDs that are required to fully define the retained Private Attributes; all other Private 640 Attributes shall be removed.

When this Option is not specified, all Private Attributes shall be removed, as described in Table E.1-1.

642 Notes: 1. A sample list of Private Attributes thought to be safe is provided here. Vendors do not guarantee them to be safe, and do not commit to sending them in any particular software version (including future 644 products).

Data Element	Private Creator	VR	VM	Meaning
(7053,xx00)	Philips PET Private Group	DS	1	SUV Factor – Multiplying stored pixel values by Rescale Slope then this factor results in SUVbw in g/l
(7053,xx09)	Philips PET Private Group	DS	1	Activity Concentration Factor – Multiplying stored pixel values by Rescale Slope then this factor results in MBq/ml.
(00E1,xx21)	ELSCINT1	DS	1	DLP
(01E1,xx26)	ELSCINT1	CS	1	Phantom Type
(01E1,xx50)	ELSCINT1	DS	1	Acquisition Duration
(01F1,xx01)	ELSCINT1	CS	1	Acquisition Type
(01F1,xx07)	ELSCINT1	DS	1	Table Velocity
(01F1,xx26)	ELSCINT1	DS	1	Pitch
(01F1,xx27)	ELSCINT1	DS	1	Rotation Time
(0019,xx23)	GEMS_ACQU_01	DS	1	Table Speed [mm/rotation]
(0019,xx24)	GEMS_ACQU_01	DS	1	Mid Scan Time [sec]
(0019,xx27)	GEMS_ACQU_01	DS	1	Rotation Speed (Gantry Period)
(0043,xx27)	GEMS_PARM_01	SH	1	Scan Pitch Ratio in the form "n.nnn:1"
(0045,xx01)	GEMS_HELIOS_01	SS	1	Number of Macro Rows in Detector
(0045,xx02)	GEMS_HELIOS_01	FL	1	Macro width at ISO Center
(0903,xx10)	GEIIS PACS	US	1	Reject Image Flag
(0903,xx11)	GEIIS PACS	US	1	Significant Flag
(0903,xx12)	GEIIS PACS	US	1	Confidential Flag
(2001,xx03)	Philips Imaging DD 001	FL	1	Diffusion B-Factor
(2001,xx04)	Philips Imaging DD 001	CS	1	Diffusion Direction
(0019,xx0C)	SIEMENS MR HEADER	IS	1	B Value

(0019,xx0D)	SIEMENS MR HEADER	CS	1	Diffusion Directionality
(0019,xx0E)	SIEMENS MR HEADER	FD	3	Diffusion Gradient Direction
(0019,xx27)	SIEMENS MR HEADER	FD	6	B Matrix
(0043,xx39)	GEMS_PARM_01	IS	4	1 <sup>st</sup> value is B Value

646

648

650

652

654

656

2. One approach to retaining Private Attributes safely, either when the VR is encoded explicitly or known from a data dictionary (such as may be derived from published DICOM Conformance Statements or previously encountered instances, perhaps by adaptively extending the data dictionary as new explicit VR instances are received), is to retain those Attributes that are numeric only. For example, one might retain US, SS, UL, SS, FL and FD binary values, and IS and DS string values that contain only valid numeric characters. One might assume that other string Value Representations are unsafe in the absence of definite confirmation from the vendor to the contrary; code strings (CS) may be an exception. Bulk binary data in OB Value representations is particularly unsafe, and may often contain entire proprietary format headers in binary or text or XML form that includes the patient's name and other identifying information.

658

The safe private attributes that are retained shall be described in the Conformance Statement.

660

**Add context group for de-identification method to PS 3.3 C.7.1.1 General Image Module as indicated:**

662

**C.7.1.1 Patient Module**

664 ...

666

**Table C.7-1  
PATIENT MODULE ATTRIBUTES**

Attribute Name	Tag	Type	Attribute Description
...	...	...	...
Patient Identity Removed	(0012,0062)	3	The true identity of the patient has been removed from the Attributes and the Pixel Data Enumerated Values: YES NO
De-identification Method	(0012,0063)	1C	A description or label of the mechanism or method use to remove the patient's identity. May be multi-valued if successive de-identification steps have been performed. Notes: 1. This may be used to describe the extent or thoroughness of the de-identification, for example whether or not the de-identification is for a "Limited Data Set" (as per HIPAA Privacy Rule). 2. The characteristics of the de-

			<p>identifying equipment and/or the responsible operator of that equipment may be recorded as an additional item of the Contributing Equipment Sequence (0018,A001) in the SOP Common Module. De-identifying equipment may use a Purpose of Reference of (109104,DCM,“De-identifying Equipment”).</p> <p>Required if Patient Identity Removed (0012,0062) is present and has a value of YES and De-identification Method Code Sequence (0012,0064) is not present. <b>May be present otherwise.</b></p>
De-identification Method Code Sequence	(0012,0064)	1C	<p>A code describing the mechanism or method use to remove the patient’s identity. One or more Items shall be present. Multiple items are used if successive de-identification steps have been performed <b>or to describe options of a defined profile.</b></p> <p>Required if Patient Identity Removed (0012,0062) is present and has a value of YES and De-identification Method (0012,0063) is not present. <b>May be present otherwise.</b></p>
>Include Code Sequence Macro Table 8.8-1			<b><del>No Baseline Context ID is defined</del>Defined CID 7050.</b>

668 **Add to PS 3.3 C.7.6.1 General Image Module as indicated (and also make same addition to all IODs that do not use the General Image Module but use Burned In Annotation in a specific module):**

670 **C.7.6.1 General Image Module**

...

672

**Table C.7-9  
GENERAL IMAGE MODULE ATTRIBUTES**

Attribute Name	Tag	Type	Attribute Description
...	...	...	...
Burned In Annotation	(0028,0301)	3	<p>Indicates whether or not image contains sufficient burned in annotation to identify the patient and date the image was acquired.</p> <p>Enumerated Values: YES NO</p>

			If this Attribute is absent, then the image may or may not contain burned in annotation.
<b>Recognizable Visual Features</b>	<b>(0028,0302)</b>	<b>3</b>	<b>Indicates whether or not the image contains sufficiently recognizable visual features to allow the image or a reconstruction from a set of images to identify the patient.</b> <b>Enumerated Values:</b> <b>YES</b> <b>NO</b> <b>If this Attribute is absent, then the image may or may not contain recognizable visual features.</b>
Lossy Image Compression	(0028,2110)	3	Specifies whether an Image has undergone lossy compression. Enumerated Values: 00 = Image has NOT been subjected to lossy compression. 01 = Image has been subjected to lossy compression. See C.7.6.1.1.5

674

**Add to PS 3.3 C.12.1 SOP Common Module as indicated:**

676 **C.12.1 SOP Common Module**

...

678

**Table C.12-1  
SOP COMMON MODULE ATTRIBUTES**

<b>Attribute Name</b>	<b>Tag</b>	<b>Type</b>	<b>Attribute Description</b>
<b>Longitudinal Temporal Information Modified</b>	<b>(0028,0303)</b>	<b>3</b>	<b>Indicates whether or not the date and time attributes in the instance have been modified during de-identification.</b> <b>Enumerated Values:</b> <b>UNMODIFIED</b> <b>MODIFIED</b> <b>REMOVED</b> <b>See PS 3.15.</b>

680

682 **Add new data elements to PS 3.6 as indicated:**



(0028,0301)	Burned In Annotation	BurnedInAnnotation	CS	1
<b>(0028,0302)</b>	<b>Recognizable Visual Features</b>	<b>RecognizableVisualFeatures</b>	<b>CS</b>	<b>1</b>
<b>(0028,0303)</b>	<b>Longitudinal Temporal Information Modified</b>	<b>LongitudinalTemporalInformationModified</b>	<b>CS</b>	<b>1</b>

684

**Add context group UIDs to PS 3.6 as indicated:**

686

**Table A-3  
CONTEXT GROUP UID VALUES**

Context UID	Context Identifier	Context Group Name
<u>1.2.840.10008.6.1.925</u>	<u>7050</u>	<u>De-identification Method</u>

688

**Add context groups to PS 3.16 as indicated:**

690

**CID 7050 De-identification Method**

692

**Context ID 7050**

**De-identification Method**

694

**Type: Extensible Version: 20110123**

Coding Scheme Designator (0008,0102)	Code Value (0008,0100)	Code Meaning (0008,0104)
DCM	113100	Basic Application Confidentiality Profile
DCM	113101	Clean Pixel Data Option
DCM	113102	Clean Recognizable Visual Features Option
DCM	113103	Clean Graphics Option
DCM	113104	Clean Structured Content Option
DCM	113105	Clean Descriptors Option
DCM	113106	Retain Longitudinal Temporal Information With Full Dates Option
DCM	113107	Retain Longitudinal Temporal Information With Modified Dates Option
DCM	113108	Retain Patient Characteristics Option
DCM	113109	Retain Device Identity Option
DCM	113110	Retain UIDs Option

<b>Coding Scheme Designator (0008,0102)</b>	<b>Code Value (0008,0100)</b>	<b>Code Meaning (0008,0104)</b>
DCM	113111	Retain Safe Private Option

696 **Add code definitions to PS 3.16 Annex D as indicated:**

...

113100	Basic Application Confidentiality Profile	De-identification using a profile defined in PS 3.15 that requires removing all information related to the identity and demographic characteristics of the patient, any responsible parties or family members, any personnel involved in the procedure, the organizations involved in ordering or performing the procedure, additional information that could be used to match instances if given access to the originals, such as UIDs, dates and times, and private attributes, when that information is present in the non-Pixel Data Attributes, including graphics or overlays	
113101	Clean Pixel Data Option	Additional de-identification according to an option defined in PS 3.15 that requires any information burned in to the Pixel Data corresponding to the Attribute information specified to be removed by the Profile and any other Options specified also be removed.	
113102	Clean Recognizable Visual Features Option	Additional de-identification according to an option defined in PS 3.15 that requires that sufficient removal or distortion of the Pixel Data shall be applied to prevent recognition of an individual from the instances themselves or a reconstruction of a set of instances.	
113103	Clean Graphics Option	Additional de-identification according to an option defined in PS 3.15 that requires that any	

		information encoded in graphics, text annotations or overlays corresponding to the Attribute information specified to be removed by the Profile and any other Options specified also be removed.	
113104	Clean Structured Content Option	Additional de-identification according to an option defined in PS 3.15 that requires that any information encoded in SR Content Items or Acquisition Context Sequence Items corresponding to the Attribute information specified to be removed by the Profile and any other Options specified also be removed.	
113105	Clean Descriptors Option	Additional de-identification according to an option defined in PS 3.15 that requires that any information that is embedded in text or string Attributes corresponding to the Attribute information specified to be removed by the Profile and any other Options specified also be removed.	
113106	Retain Longitudinal Temporal Information With Full Dates Option	Retention of information that would otherwise be removed during de-identification according to an option defined in PS 3.15 that requires that any dates and times be retained,	
113107	Retain Longitudinal Temporal Information With Modified Dates Option	Retention of information that would otherwise be removed during de-identification according to an option defined in PS 3.15 that requires that any dates and times be modified in a manner that preserves temporal relationships. E.g., Study Date and Time.	
113108	Retain Patient Characteristics Option	Retention of information that would otherwise be removed during de-identification according to an option defined in PS 3.15 that requires that any physical characteristics of the patient, which are descriptive rather than identifying information per se, be	

		retained. E.g., Patient's Age, Sex, Size (height) and Weight.	
113109	Retain Device Identity Option	Retention of information that would otherwise be removed during de-identification according to an option defined in PS 3.15 that requires that any information that identifies a device be retained. E.g., Device Serial Number.	
113110	Retain UIDs Option	Retention of information that would otherwise be removed during de-identification according to an option defined in PS 3.15 that requires that UIDs be retained. E.g., SOP Instance UID.	
113111	Retain Safe Private Option	Retention of information that would otherwise be removed during de-identification according to an option defined in PS 3.15 that requires that private attributes that are known not to contain identity information be retained. E.g., private SUV scale factor.	