DICOM INTERNATIONAL CONFERENCE & SEMINAR

Oct 9-11, 2010 Rio de Janeiro, Brazil

CBR 10
XXXIX Congresso Brasileiro de Radiologia

DICOM
Digital Imaging and Communications in Medicine

De-identification Revisited

DICOM Supplement 142

David Clunie

CoreLab Partners, Inc.

CoreLab PARTNERS

Use Cases

- **Multi-center Clinical Trial**
  - patients enrolled in a clinical trial and undergoing clinical care
  - consented to have their clinical images submitted for analysis by a third party
  - without revealing their real identity
  - analysis results can be linked to the subject
  - physical characteristics can be used in the analysis (e.g., sex, age, height, weight)
  - limited or broad dissemination (re-use)

Use Cases

- **Teaching File Submission**
  - patients undergoing clinical care
  - have images and clinical data of particular value for teaching or testing students and staff
  - all real identifiers to be removed for privacy
  - limited physical characteristics need to be preserved to interpret the case correctly
  - disseminated broadly, even publicly

Use Cases

- **Remote Equipment Servicing**
  - patients undergoing clinical imaging
  - site staff see quality problems in images
  - remote service staff have no need or right to see real patient identity information
  - given remote access only to images without real identity

Definitions

- **De-identification**
  - removing real patient identifiers
- **Pseudonymization**
  - de-identification and replacement of identifiers with a pseudonym that is unique to the individual and known within a specified context but not linked to the individual in the external world
- **Anonymization**
  - de-identification and further removal or ambiguation of information to reduce the probability of re-identification of the image despite access to other information sources

Adapted from Drug Information Association (DIA) Medical Imaging Standardization Technical Document 1.0 2007/10/10

History

- **DICOM Sup 55 (2002/09/05)**
  - first attempt to standardize a list of attributes that potentially contain identifying information that needs to be removed, and define a "profile"
- **IHE Teaching File & Clinical Trial Export (TCE) Profile (2005/04/22)**
  - specifies use cases, defines actors and transactions to do it, helpful hints based on experience, profile with options (pixel data, remap identifiers (pseudonymization))
- **DICOM Sup 142 (Ballot 2010/08/26)**
  - more comprehensive list of attributes, addresses additional concerns beyond attributes, what attributes to retain for specific use cases, grouped into options

Basic Premises & Conclusion

- **De-identification is hard**
  - choosing what to remove (to protect privacy, reduce risk)
  - and what to keep (to satisfy use case)
  - requires significant expertise
  - technical, statistical, legal
- **Local policy and national regulations**
  - describe requirements in general terms
  - are not image or DICOM-specific
- **Define simple profile and options**
  - easier for ethics committee to understand and agree to
  - simpler and less error-prone for site staff to deploy
  - than individually configuring every attribute manually

Sup 142 Basic Profile

- **Remove/replace *all* attributes at risk**
  - long table of known risky “header” attributes
  - all person names & identifiers (patient & staff)
  - all institution, department, equipment identity
  - all free text comments and descriptions
  - all UIDs
  - all private attributes (since risky if unknown)

Sup 142 Attributes

DICOM℠
Digital Imaging and Communications in Medicine

315

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

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 Chars. Option	Retain Device Ident. Option	Retain Patient Chars. Option	Retain Long. Info Opts	Clean Desc. Option	Clean Struct. Cont. Option	Clean Graph. Option
Accession Number	(0008,0050)	N	Y	Z									
Acquisition Comments	(0018,4000)	Y	N	X							C		
Acquisition Context Sequence	(0040,0555)	N	Y	X								C	
Acquisition Date	(0008,0022)	N	Y	X/Z						K			
Acquisition DateTime	(0008,002A)	N	Y	X/D						K			
Acquisition Device Processing Description	(0018,1400)	N	Y	X/D							C		
Acquisition Protocol Description	(0018,9424)	N	Y	X							C		
Acquisition Time	(0008,0032)	N	Y	X/Z						K			
Actual Human Performers Sequence	(0040,4035)	N	N	X									
Additional Patient's History	(0010,21B0)	N	Y	X							C		
Admission ID	(0038,0010)	N	Y	X									
Admitting Date	(0038,0020)	N	N	X						K			

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 Chars. Option	Retain Device Ident. Option	Retain Patient Chars. Option	Retain Long. Info Opts	Clean Desc. Option	Clean Struct. Cont. Option	Clean Graph. Option
Admitting Diagnoses Code Sequence	(0008,1084)	N	Y	X							C		
Admitting Diagnoses Description	(0008,1080)	N	Y	X							C		
Admitting Time	(0038,0021)	N	N	X						K			
Affected SOP Instance UID	(0000,1000)	N	N	X		K							
Allergies	(0010,2110)	N	N	X					C		C		
Arbitrary	(4000,0010)	Y	N	X									
Author Observer Sequence	(0040,A078)	N	Y	X									
Branch of Service	(0010,1081)	N	N	X									
Cassette ID	(0018,1007)	N	Y	X				K					
Comments on Performed Procedure Step	(0040,0280)	N	Y	X							C		
Concatenation UID	(0020,9161)	N	Y	U		K							
Confidentiality Constraint on Patient Data Description	(0040,3001)	N	N	X									

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 Chars. Option	Retain Device Ident. Option	Retain Patient Chars. Option	Retain Long. Info Opts	Clean Desc. Option	Clean Struct. Cont. Option	Clean Graph. Option
Content Creator's Name	(0070,0084)	N	Y	Z									
Content Creator's Identification Code Sequence	(0070,0086)	N	Y	X									
Content Date	(0008,0023)	N	Y	Z/D						K			
Content Sequence	(0040,A730)	N	Y	X								C	
Content Time	(0008,0033)	N	Y	Z/D						K			
Context Group Extension Creator UID	(0008,010D)	N	Y	U		K							
Contrast Bolus Agent	(0018,0010)	N	Y	Z/D							C		
Contribution Description	(0018,A003)	N	Y	X							C		
Country of Residence	(0010,2150)	N	N	X									
Creator Version UID	(0008,9123)	N	Y	U		K							
Current Patient Location	(0038,0300)	N	N	X									
Curve Data	(50xx,xxxx)	Y	N	X									C
Curve Date	(0008,0025)	Y	Y	X						K			
Curve Time	(0008,0035)	Y	Y	X						K			

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 Chars. Option	Retain Device Ident. Option	Retain Patient Chars. Option	Retain Long. Info Opts	Clean Desc. Option	Clean Struct. Cont. Option	Clean Graph. Option
Custodial Organization Sequence	(0040,A07C)	N	Y	X									
Data Set Trailing Padding	(FFFC,FFFC	N	Y	X									
Derivation Description	(0008,2111)	N	Y	X							C		
Detector ID	(0018,700A)	N	Y	X				K					
Device Serial Number	(0018,1000)	N	Y	X/Z/D				K					
Device UID	(0018,1002)	N	Y	U		K		K					
Digital Signature UID	(0400,0100)	N	Y	X									
Digital Signatures Sequence	(FFFA,FFFA)	N	Y	X									
Dimension Organization UID	(0020,9164)	N	Y	U		K							
Discharge Diagnosis Description	(0038,0040)	Y	N	X							C		
Distribution Address	(4008,011A)	Y	N	X									
Distribution Name	(4008,0119)	Y	N	X									

Remove or Replace

- **Whether to remove or replace**
  - requires preserving integrity of object with respect to DICOM compliance
  - Type 1 – replace with dummy value
  - Type 2 – zero length (empty)
  - Type 3 – remove completely
  - includes recursive handling of sequences

Extended and Retired

- **Standard Extended objects**
  - DICOM allows insertion of standard attributes in images objects that were intended for other purposes
  - these must be removed or replaced as well
  - are listed in the table and identified as such
- **Retired Attributes**
  - no longer described or maintained in standard
  - may be present, may be risky, therefore listed in the table and need to be removed

Two Types of Options

- **Remove more**
  - not in basic profile because too hard
  - *and* usually unnecessary
  - depend on specific type of object
  - non-images
  - specific subject matter (anatomy, modality)
- **Retain more (remove less)**
  - small potential for re-identification (low risk)
  - *and* required for use case

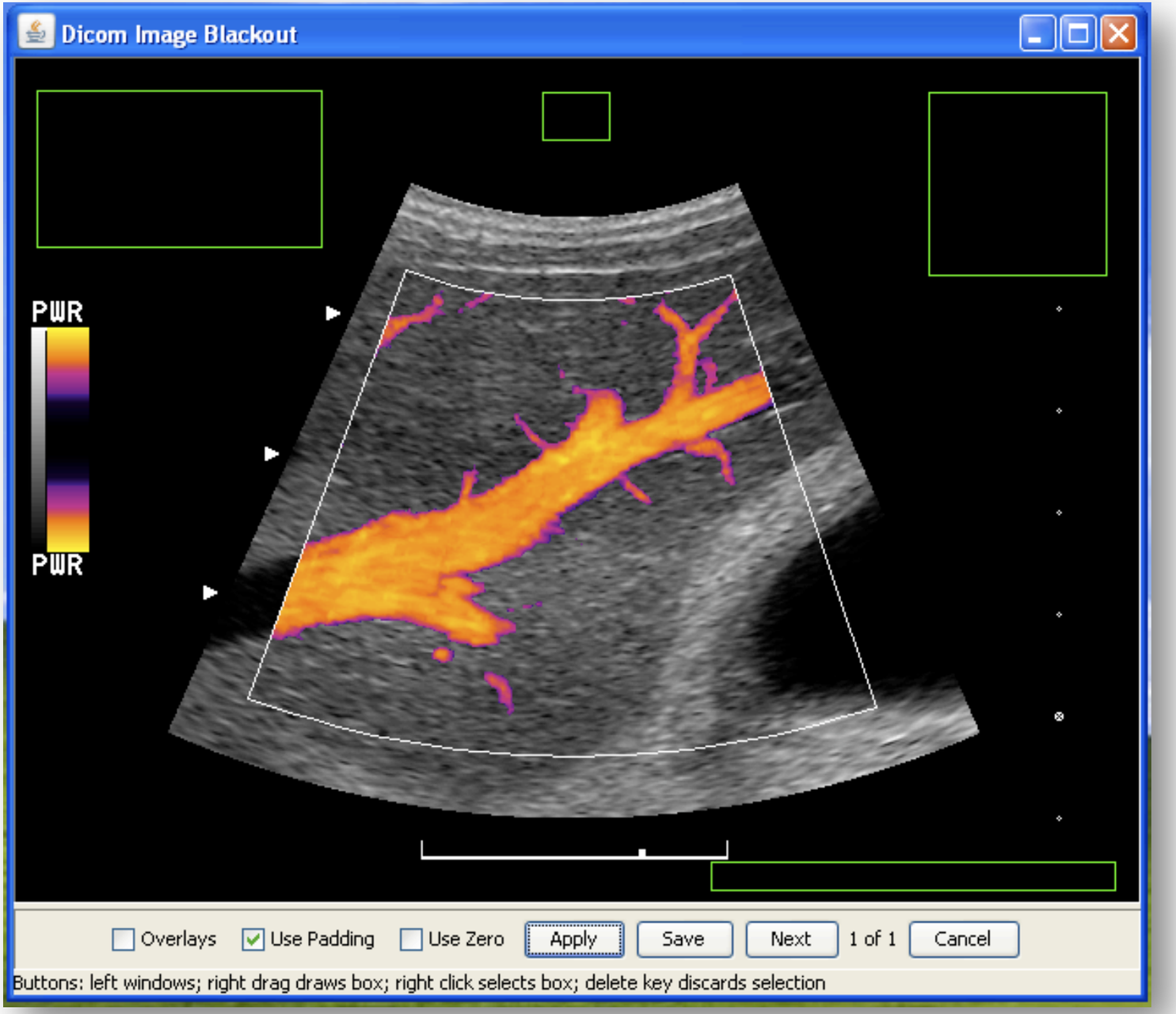
Options Summary

- **Remove more**
  - Clean Pixel Data Option
  - Clean Recognizable Visual Features Option
  - Clean Graphics Option
  - Clean Structured Content Option
  - Clean Descriptors Option
- **Retain more (remove less)**
  - Retain Longitudinal Option
  - Retain Patient Characteristics Option
  - Retain Device Information Option
  - Retain UIDs
  - Retain Safe Private Option

Clean Pixel Data Option

- **Text identifiers in the "picture" (pixel data)**
  - secondary capture
    - screen shots (e.g., analysis result screens)
    - video
    - scanned film or paper prints
    - scanned documents (requests or reports)
  - ultrasound (historically was video capture)
  - angiography or fluoroscopy (occasionally)
- **Clean Pixel Data option requires removal**
  - manual
  - automatic (desirable, hard, may remove other stuff)

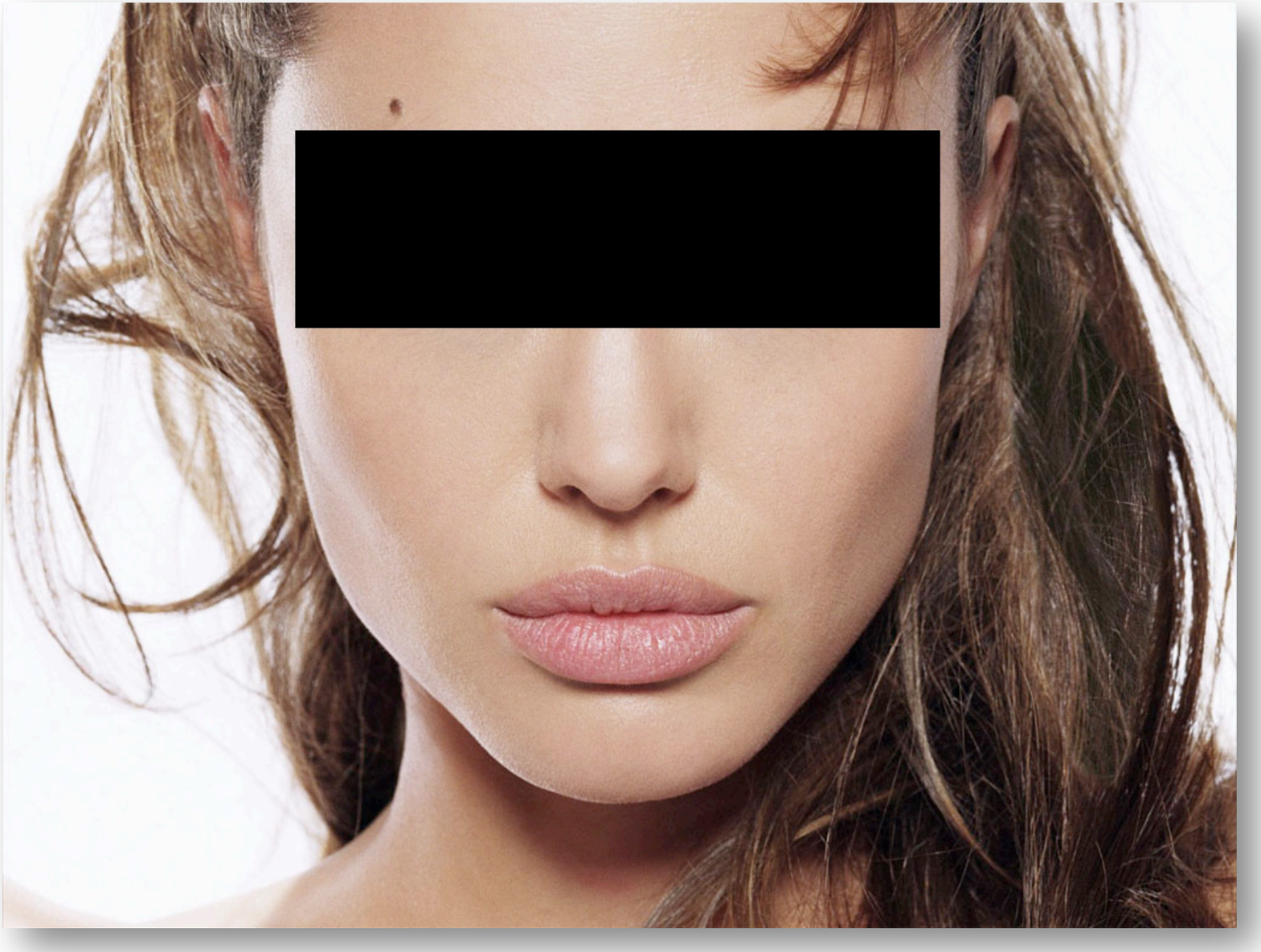
Clean Pixel Data Option



Clean Recognizable Visual Features Option

- **Visible Light**
  - photographs of faces
  - traditionally blacked out in publications
- **Cross-sectional thin slice CT or MR**
  - theoretically can reconstruct a “face”
  - arguable whether these are “recognizable” (Chen J et al. SIIM 2007)
  - can add noise to facial region to disrupt
  - renders images useless for some purposes

Clean Recognizable Visual Features Option



Clean Recognizable Visual Features Option

MRI Defacer - http://www.nitrc.org/projects/mri_deface/

Clean Graphics Option

- **"Header" may contain graphics**
  - overlays
  - curves
  - graphics in presentation states
  - presentation state mechanisms used in images (standard extended)
- **Basic profile requires complete removal**
  - may discard useful info (lesions, measurements)
- **Clean Graphics option**
  - selective "cleaning" (manual or automatic)

Clean Structured Content Option

- **DICOM Structured Reports**
  - tree of content items in sequences
  - identifying information depends on coded concepts defined in DICOM PS 3.16
  - beyond the scope of Sup 142 to enumerate
- **Basic profile**
  - addresses only the “header” and not the tree
- **Clean Structured Content option**
  - commitment to clean the tree as necessary

Clean Descriptors Option

- **"Header" may contain free text**
  - comments and descriptions
  - patient, study, series, image, protocol
  - copied from work list (relatively safe)
  - entered by operator (very dangerous)
- **Basic profile requires complete removal**
  - may discard useful info (procedure, anatomy)
- **Clean Descriptors option**
  - selective "cleaning" (manual or automatic)

Clean Descriptors Option

- **Example – Study Description**
  - "CT chest abdomen pelvis – 55F Dr. Smith"
  - retain only "CT chest abdomen pelvis"
  - extract SNOMED codes for anatomic region
- **Example – Multiple Language support**
  - "Buik" for abdomen in Dutch
  - "λεκάνη" for pelvis in Greek
- **Example – person names are keywords**
  - "Dr. Hand" or "M. Genou"

Retain Longitudinal Options

- **"Header" contains many dates & times**
  - constrain the number of possible individuals that could be the subject
- **Basic profile**
  - requires removal
- **Retain Longitudinal options**
  - Full Dates – just keep them
  - Modified Dates – adjust them consistently

Retain Patient Characteristics Option

- **Information about the patient**
  - as distinct from name, medical record number
  - e.g., sex, age, height, weight
  - critical for PET SUV, DEXA, MRI measures of body composition (normalized to body size)
- **Basic profile**
  - requires removal
- **Retain Patient Characteristics option**
  - keep them

Retain Device Options

- **Scanner identification & characteristics**
  - characteristics – important when a particular class of scanner is required (e.g., Acme 3T)
  - identification – important when a particular scanner has been qualified (e.g., by phantom)
- **Basic profile**
  - requires removal
- **Retain Device options**
  - Retain Device Characteristics Option
  - Retain Device Identity Option

Retain UIDs Option

- **Unique Identifiers (UIDs)**
  - patients do not have unique identifiers
  - but studies, series, instances and other entities do
  - all cross-references between objects are by UIDs
  - replacement jeopardizes audit trail, repeated submission duplicate detection, long term consistency
- **Basic profile requires**
  - replacement of all UIDs
  - such that they are “internally consistent with a set”
- **Retain UIDs option**
  - just keep them without change

Retain Safe Private Option

- **Private Attributes**
  - are vendor proprietary & often undocumented
  - could contain anything
  - some contain vital information
  - e.g., Philips Private SUV Scale Factor
- **Basic profile**
  - requires removal of all private attributes
- **Retain Safe Private option**
  - keep those known to be safe
  - a partial list of these will be maintained in PS 3.15

Implementations

- **Currently Sup 142 is out for ballot**
- **Prototype implementations of concepts**
  - MIRC Clinical Trial Processor (CTP)
    - highly configurable – now has Sup 142 templates
    - http://mircwiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor
  - PixelMed DicomCleaner
    - turnkey – gives users choices like Sup 142 options http://www.dclunie.com/pixelmed/software/webstart/DicomCleanerUsage.html