De-identification Revisited

DICOM Supplement 142

David Clunie
CoreLab Partners, Inc.
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*
• **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
• **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**

### Table 2.5.1

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Tag</th>
<th>Relevant Profiles</th>
<th>De-Identification Level</th>
<th>De-Identification Profile Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accession Number</td>
<td>0008,0050</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Accession Code</td>
<td>0008,0051</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Accession Time</td>
<td>0008,0020</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Accession Time</td>
<td>0008,0020</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Additional Patient's History</td>
<td>0010,0011</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Additional Code</td>
<td>0008,0020</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

### Table 2.5.2

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Tag</th>
<th>Relevant Profiles</th>
<th>De-Identification Level</th>
<th>De-Identification Profile Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contract Creation</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Creation 2nd</td>
<td>0010,0001</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Type</td>
<td>0008,0010</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Address</td>
<td>0010,0010</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Status</td>
<td>0008,0010</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Description</td>
<td>0010,0010</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Number</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contract Date</td>
<td>0008,0000</td>
<td>Y</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>Contract Time</td>
<td>0008,0000</td>
<td>Y</td>
<td>N</td>
<td>X</td>
</tr>
</tbody>
</table>

### Table 2.5.3

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Tag</th>
<th>Relevant Profiles</th>
<th>De-Identification Level</th>
<th>De-Identification Profile Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contractor</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contractor's Identification Code</td>
<td>0010,0001</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Contractor's Identification Code 2nd</td>
<td>0010,0002</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

### Table 2.5.4

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Tag</th>
<th>Relevant Profiles</th>
<th>De-Identification Level</th>
<th>De-Identification Profile Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participants' Address</td>
<td>0010,0001</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participants' Name</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participants' Type</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participants' Gender</td>
<td>0010,0000</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Participants' Date</td>
<td>0008,0000</td>
<td>Y</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>Participants' Time</td>
<td>0008,0000</td>
<td>Y</td>
<td>N</td>
<td>X</td>
</tr>
</tbody>
</table>

**DICOM**

Digital Imaging and Communications in Medicine
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
• **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
• **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
• **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
  – Pixelmed DicomCleaner
    • turnkey – gives users choices like Sup 142 options