



**DATA SCIENCE
INSTITUTE™**

AMERICAN COLLEGE OF RADIOLOGY

Data Privacy and De-identification: How Far Do We Go?

Steve Langer, PHD
Mayo Clinic

Disclaimers:

- Co-chair SiiM Hackathon Committee
- Co-chair SiiM Machine Learning Committee
- Co-Founder FlowSigma
- I am not a lawyer



What is this DICOM thing anyway?



```
(0008,0008) - Image Type  
(0008,0012) - Instance Creation Date  
(0008,0016) - SOP Class UID  
(0008,0018) - SOP Instance UID  
(0008,0020) - Study Date  
(0008,0021) - Series Date  
(0008,0022) - Acquisition Date  
(0008,0023) - Content Date  
(0008,0030) - Study Time
```

```
(0010,0010) - Patient's Name  
(0010,0020) - Patient ID  
(0010,0030) - Patient's Birth Date  
(0010,0040) - Patient's Sex  
(0010,1010) - Patient's Age  
(0010,1020) - Patient's Size  
(0010,1030) - Patient's Weight  
(0010,2110) - Allergies  
(0010,2180) - Occupation  
(0010,2180) - Additional Patient History  
(0018,0010) - Contrast/Bolus Agent  
(0018,0015) - Body Part Examined  
(0018,0060) - KVP  
(0018,1000) - Device Serial Number
```



So if we alias the public tags we are clean?

- well, maybe not
- DICOM allows Private odd numbered groups for the vendor to encode information that is germane to them

```
(0029,E131) - Unknown [#1] [sq][
  (FFFE,E000) - Item [#19]
    (0008,0005) - Specific Character Set
    (0008,0050) - Accession Number
    (0008,0070) - Manufacturer
    (0008,0080) - Institution Name
    (0008,1010) - Station Name
    (0008,1030) - Study Description
    (0008,103E) - Series Description
    (0008,1070) - Operators' Name
    (0008,1090) - Manufacturer's Model Name
    (0010,0010) - Patient's Name
    (0010,0020) - Patient ID
    (0018,1000) - Device Serial Number
    (0018,1020) - Software Version(s)
```



The Challenge(s)

Legal and ethical

- HIPAA/HITECH *
 - Aim; protect patient's PHI
 - Limited Data set; 16 safe harbor tags "removed" **
 - Must either blank, hash, offset or alias

* 45 CFR Parts 160 and 164

** <https://www.federalregister.gov/select-citation/2016/07/30/45-CFR-164.514>

Technical and research

- The "safe call" is to
 - Wipe all private tags
 - Blank, alias, hash or offset the public 16
- We should be safe then right?
 - Well no actually and
 - Some private tags may be needed by analytic



If we wipe the 16 we should be safe right?

- well ... if everyone follows the rules *
- when is an allergy tag not an allergy?

(0010,0010)	- Patient's Name	[PN]	[22]	[1]	:	Public^John^Q^A^A
(0010,0020)	- Patient ID	[LO]	[8]	[1]	:	5347436
(0010,0030)	- Patient's Birth Date	[DA]	[8]	[1]	:	19991012
(0010,0040)	- Patient's Sex	[CS]	[2]	[1]	:	M
(0010,1010)	- Patient's Age	[AS]	[4]	[1]	:	018Y
(0010,1020)	- Patient's Size	[DS]	[0]	[0]	:	
(0010,1030)	- Patient's Weight	[DS]	[0]	[0]	:	
(0010,2110)	- Allergies	[LO]	[0]	[0]	:	Public^John^Q^A^A
(0010,2180)	- Occupation	[SH]	[0]	[0]	:	

* Data has been offset or aliased in this example



And sometimes the Private tags are needed

- the vendor needs the below, but what is really in those encrypted tags?

```
(0863,0010) - Private creator  
(0863,1010) - Unknown  
(0863,1023) - Unknown  
(0863,1026) - Unknown  
(0863,1027) - Unknown  
(0863,1028) - Unknown  
(0863,1032) - Unknown  
(0863,1033) - Unknown  
(0863,1034) - Unknown  
(0863,1035) - Unknown  
(0863,1036) - Unknown  
(0863,1037) - Unknown  
(0863,1038) - Unknown  
(0863,1039) - Unknown
```

```
[LO] [ 22] [ 1]: Biospace Med : EOS Tag  
[UN] [ 4] [ 1]: 0 [ 03 00 00 00 ]  
[UN] [ 4] [ 1]: 0 [ 01 00 00 00 ]  
[UN] [ 4] [ 1]: 0 [ 01 00 00 00 ]  
[UN] [ 4] [ 1]: 0 [ 01 00 00 00 ]  
[UN] [ 4] [ 1]: 0 [ 01 00 00 00 ]  
[UN] [ 4] [ 1]: [ 00 00 00 00 ]  
[UN] [ 4] [ 1]: ä [ E3 06 00 00 ]  
[UN] [ 4] [ 1]: [ 00 80 65 44 ]  
[UN] [ 4] [ 1]: 0 [ 01 00 00 00 ]  
[UN] [ 4] [ 1]: 0 [ 15 00 00 00 ]  
[UN] [ 4] [ 1]: 0 [ 04 00 00 00 ]  
[UN] [ 4] [ 1]: - [ 96 00 00 00 ]  
[UN] [ 4] [ 1]: - [ 96 00 00 00 ]
```



Enough tags, what about just images?

- well, yes, what about them?
- In this image the yellow words are in DICOM tags – and can be manipulated (aliased, hashed, offset or blanked)
- But the lead “R” marker is **in** the image. It can only be blacked out. This is a destructive operation that cannot be undone

Technical data (left):

- Slice 1/1
- Size: 1216x2859x16 (6,953,088 bytes)
- 45296VV / 23840L = 1192.46488
- DX
- Anatomy: LSPINE
- Uncompressed
- MONOCHROME2
- Pixel Range: (8, 65280) - 65272#
- Slope: 1 / Intercept: 0
- Unsigned Short [0008, ff00] - 16 bits
- TLHC(0.00, 0.00, 0.00)
- 0V0V0V0V0 - \ Front/Back= /
- 70.11 FPS
- 0 Overlays
- File Offset: 0 (0x00000000)
- Image Offset: 10,520 (0x00002918)

Image (center):

- Yellow text: "R STANDING"
- Bottom text: "DX LUMBAR SPINE 4+ VIEWS WLSPLAT"

Right side information:

- 197253
- Patient Name: F 075Y
- DOB: 19441003 - 75 Years
- Series: 2
- 04/02/2019
- 13:05:41
- Read 0.4637 sec
- 0.00 Bytes per sec
- Total [1] 0.4637, Avg: 0.4637
- Parse 0.0670 sec
- Total [1] 0.0670, Avg: 0.0670
- Series: 0.5308, Img: 0.5308
- Pixel X/Y: 0.148000/0.148000
- FOV (X/Y): 179.97/423.13
- Station: YsioMax-27145



Are we done yet?

- Well, no actually
- We have learned that we must
 - Process the 16 safe harbor tags
 - Scan the rest of the Public tags for “illegal DICOM uses” that still have PHI
 - Look at vendor required Private tags for PHI
 - And check the images for burned in PHI
- But wait --- there’s more ...



NIH [10.1111/j.1601-6343.2008.00425.x](https://doi.org/10.1111/j.1601-6343.2008.00425.x)



<https://techreviewpro.com/best-facial-recognition-search-engines-search-faces-online-9335/>

Contents [hide]

[5 Best Facial Recognition Search Engines to Search Person's Faces Online](#)

[1. Google Face Search](#)

[2. Betaface](#)

[3. PicWiser](#)

[4. Pictriv](#)

[5. TwinsOrNot.net](#)

[The Pro Review From TechReviewPro](#)

[Don't Miss:](#)

5 Best Facial Recognition Search Engines to Search Person's Faces Online



Data Privacy and De-identification: How Far Do We Go?

Questions ?

