



IMAGING DATA ANONYMIZATION

(IN CLINICAL RESEARCH)

Mutiullah Qureshi
Anush Revazyan

24-Mar-2026

AGENDA

2

- Background
- Regulatory Framework
- Metadata Anonymization
- Pixel-Level Controls
- Operational Controls
- Key Takeaways



COLLABORATION



sanofi



Biogen.



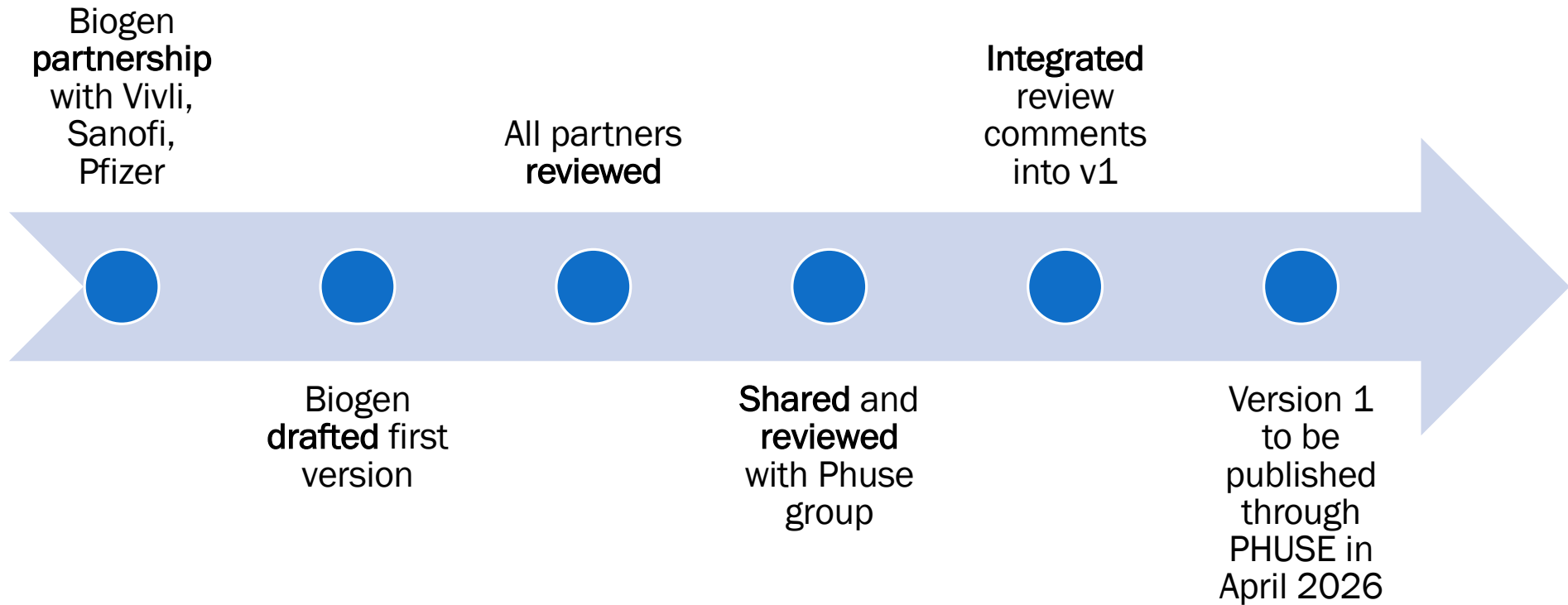
Vivli

CENTER FOR GLOBAL CLINICAL RESEARCH DATA

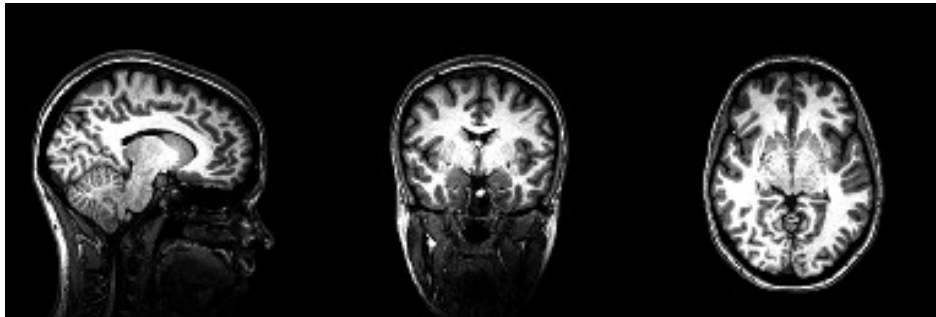


JOURNEY: PHUSE IMAGING ANONYMIZATION

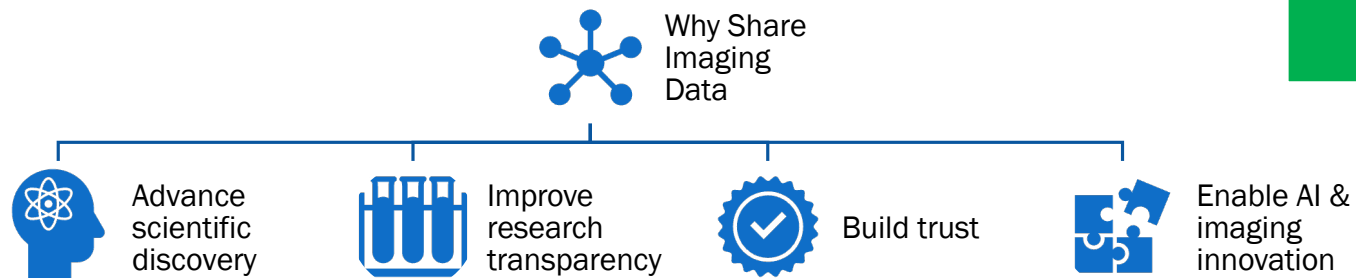
4



BACKGROUND



“Currently Biogen only shares Tabular data. Working with Vivli we plan to extend that to imaging data too.”



Why Anonymization Matters?



Imaging contains identifiable information



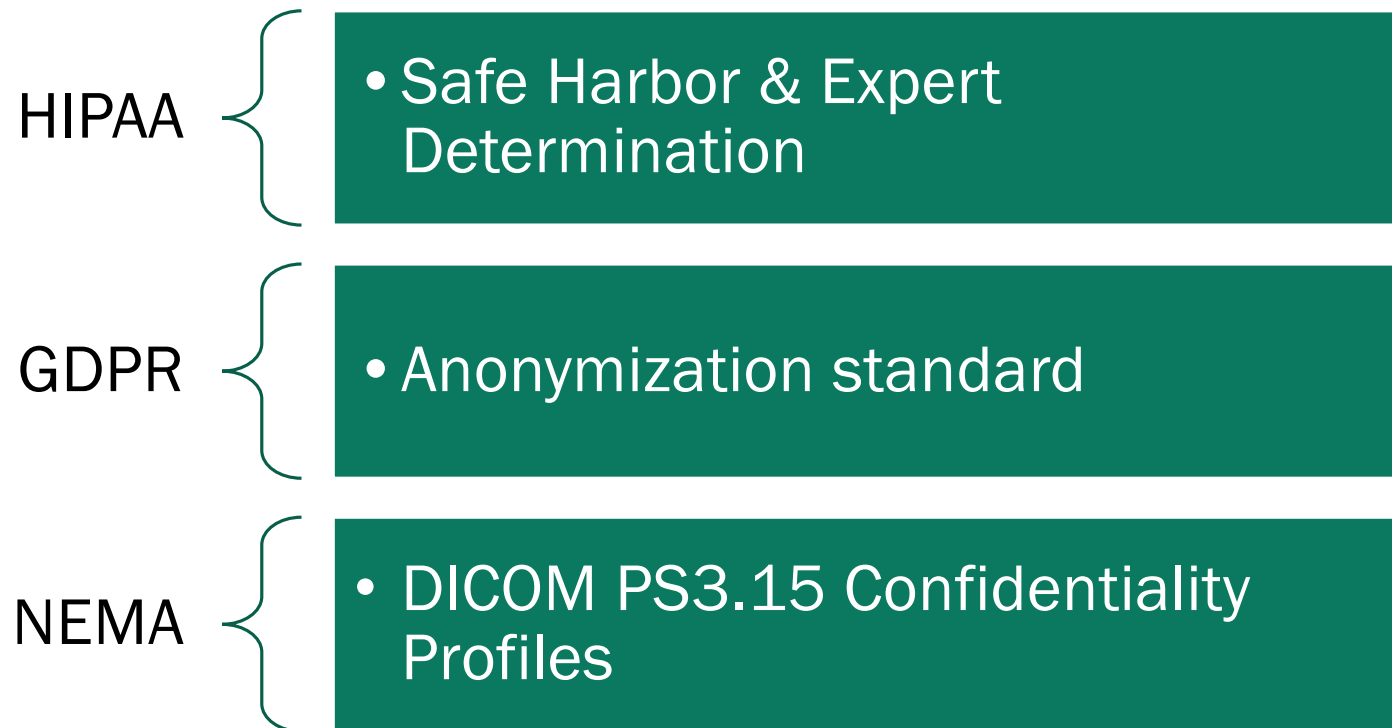
DICOM includes Metadata and Pixel Data



AI increases re-identification risk

Regulatory Foundations

7



8

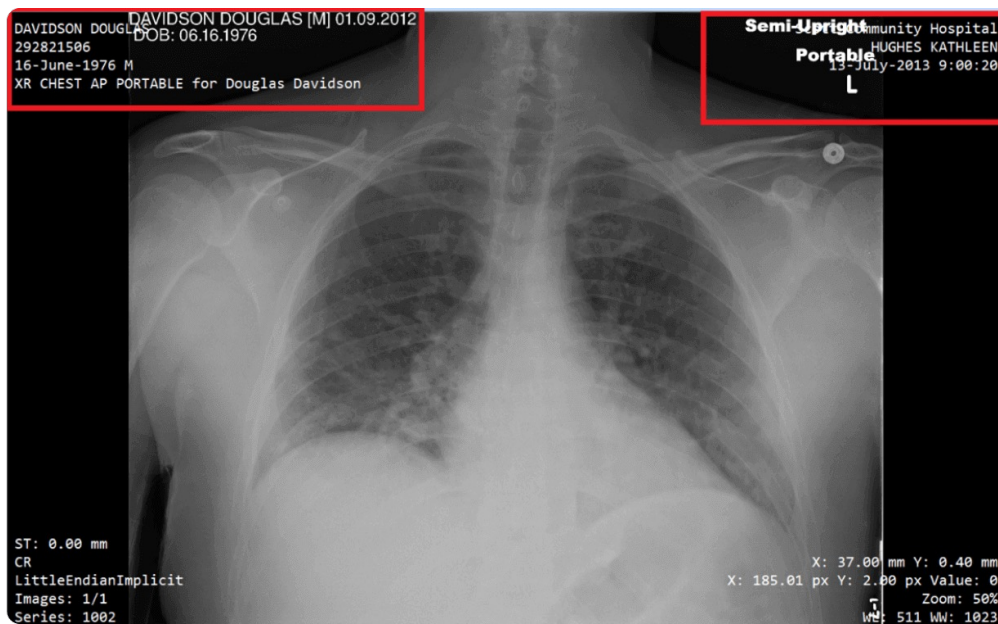
Vendor-specific private tags

(Group,Element)	TAG Description	Value
(0002,0000)	File Meta Information Group Length	184
(0002,0001)	File Meta Information Version	00\01
(0002,0002)	Media Storage SOP Class UID	1.2.840.10008.5.1.4.1.1.4
(0002,0003)	Media Storage SOP Instance UID	1.3.46.670589.11.32643.5.0.8940.2013100111013323381
(0002,0010)	Transfer Syntax UID	1.2.840.10008.1.2.1 (Explicit VR Little Endian)
(0002,0012)	Implementation Class UID	1.2.840.114202.5.2
(0002,0013)	Implementation Version Name	IMS4-6-1-P110
(0008,0005)	Specific Character Set	ISO_IR 192
(0008,0008)	Image Type	ORIGINAL\PRIMARY\M_SE\M\SE
(0008,0012)	Instance Creation Date	20131001
(0008,0013)	Instance Creation Time	113135
(0008,0014)	Instance Creator UID	1.3.46.670589.11.32643.5
(0008,0016)	SOP Class UID	1.2.840.10008.5.1.4.1.1.4
(0008,0018)	SOP Instance UID	1.3.46.670589.11.32643.5.0.8940.2013100111013323381
(0008,0020)	Study Date	20131001
(0008,0021)	Series Date	20131001
(0008,0022)	Acquisition Date	20131001
(0008,0023)	Content Date	20131001
(0008,0030)	Study Time	104458
(0008,0031)	Series Time	105709.620
(0008,0032)	Acquisition Time	105709.62
(0008,0033)	Content Time	105709.62
(0008,0050)	Accession Number	Anonymized
(0008,0060)	Modality	MR
(0008,0070)	Manufacturer	Philips Medical Systems
(0008,0080)	Institution Name	Anonymized
(0008,0090)	Referring Physician Name	Anonymized
(0008,0096)	Referring Physician Identification Sequence	
(FFFE,E000)	Item	
(0040,1101)	Person Identification Code Sequence	
(FFFE,E000)	Item	
(0008,0100)	Code Value	10242
(FFFE,E00D)	Item Delimitation Item	
(FFFE,E0DD)	Sequence Delimitation Item	
(FFFE,E00D)	Item Delimitation Item	

[illegible]

What Makes Imaging Different?

Burned-in annotations



Facial Reconstruction



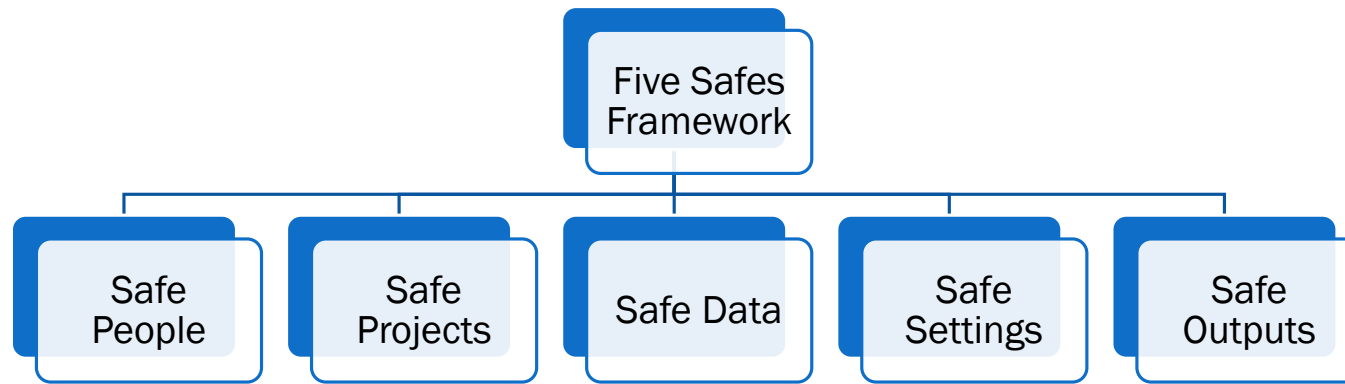
<https://www.sciencedirect.com/science/article/pii/S1053811921001221>

Secure Sharing Context

10



Data shared within Secure Governed Research Environment

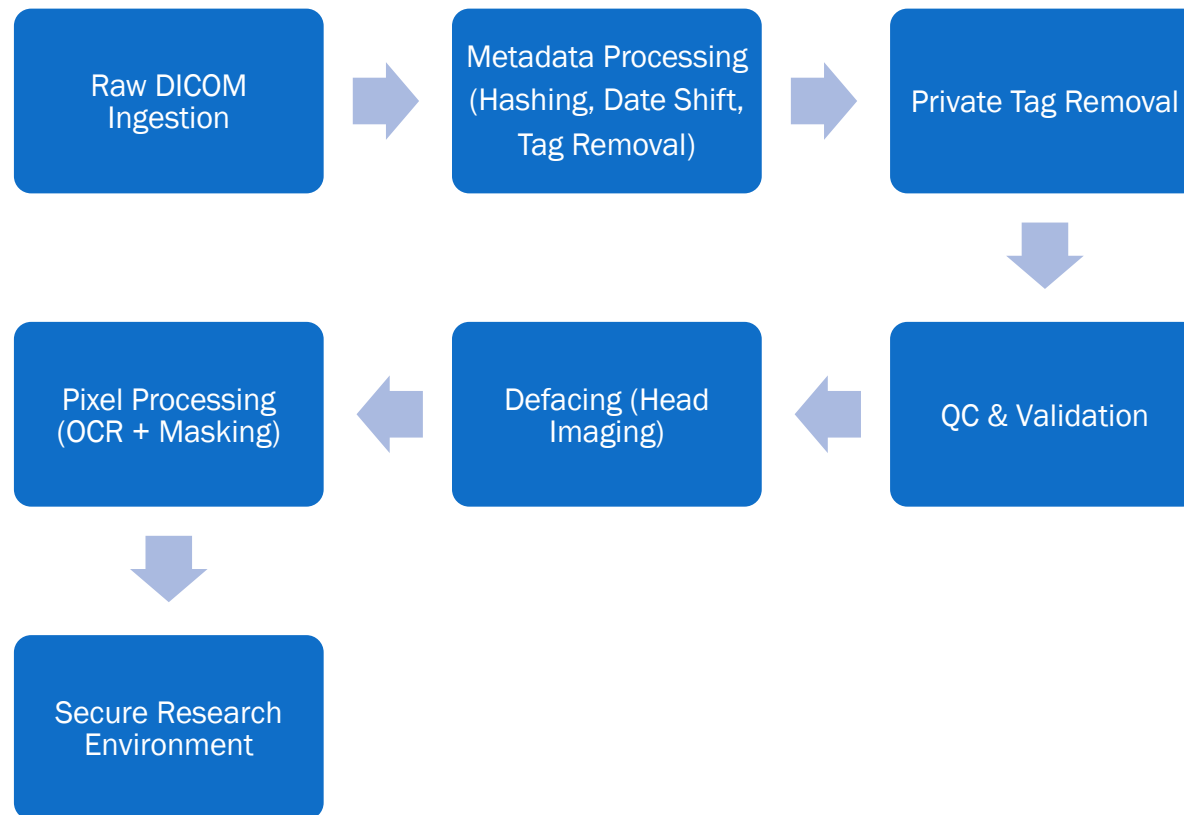


<https://www.gov.uk/data-ethics-guidance/the-five-safes-framework>



Visual Architecture: Imaging Anonymization Workflow

11



DICOM Metadata Risk Areas

12

Patient Name, ID, Birth Date

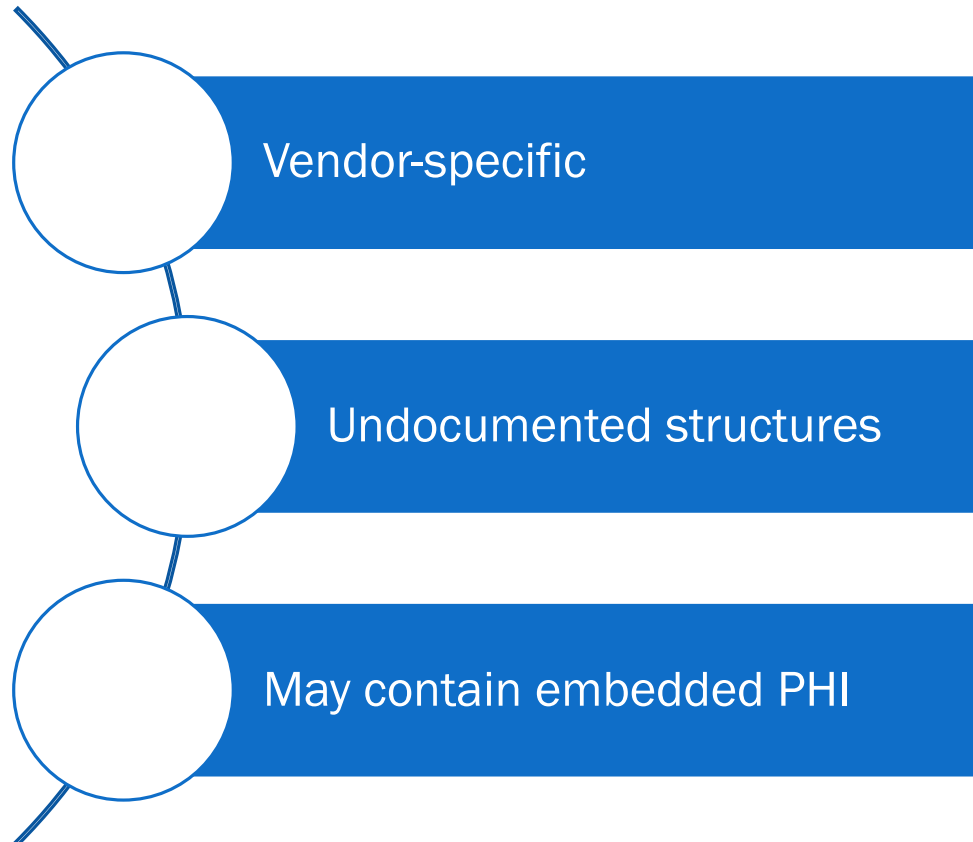
UIDs and Accession Numbers

Institution fields

Private tags

Free-text fields

Private Tags: Hidden Risk



(0029,1010) [Siemens CSA Header]

```
SV10AQAAAAAAQAAAAAABNaXNzaW9uU2l0ZV8wMQAAAAAA  
AAAAAAAAAAAAAAxMjM0NTY3ODkwAAAAAAAAAAAAAAAA  
QWNxVXNlc190YW1lPU9wZXJhdG9yMTIzAAAAAAAAAAAAAA
```

This could contain hidden fields like:

```
AcqUser_Name=Operator123  
SiteName=MissionSite_01
```



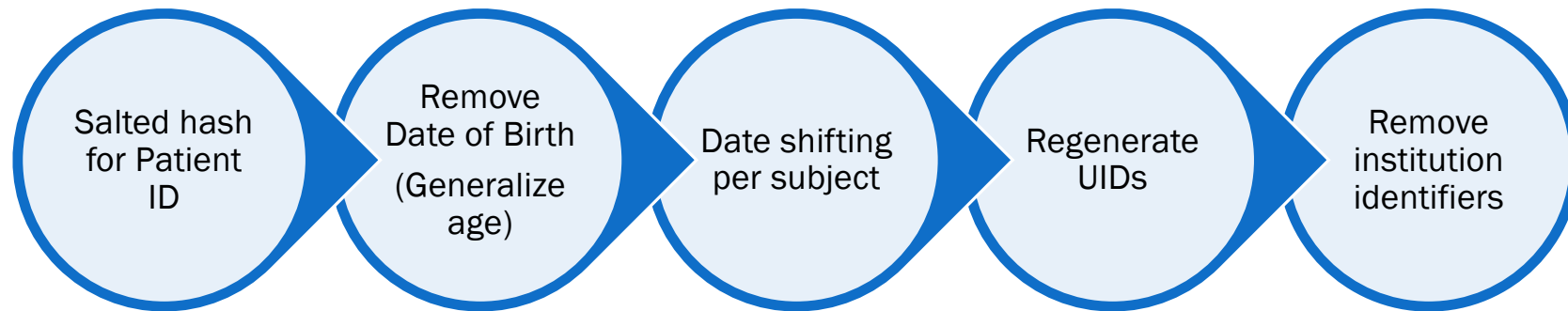
Default: Remove all private tags



Exception: Requires formal review

Core Metadata Transformations

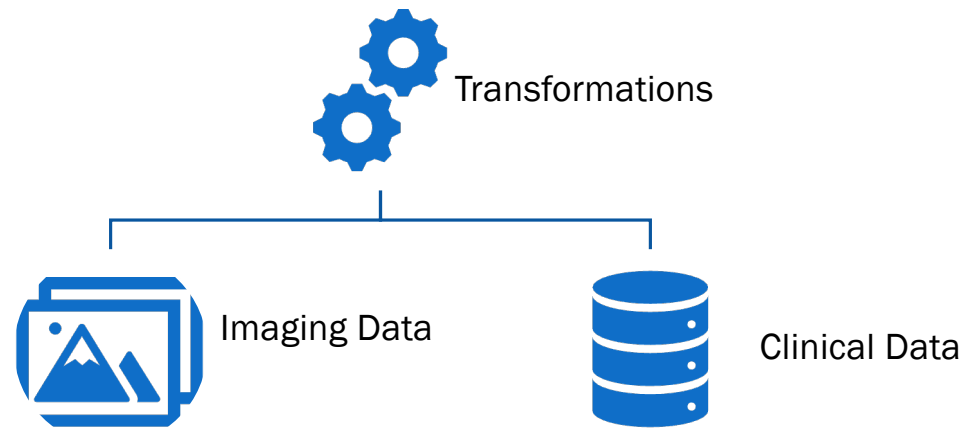
14



(Salt = random secret value added before hashing)

Cross-Dataset Consistency

15



- Consistent date shifting
- Consistent anonymized identifiers

Burned-in Annotations

16

Names, Date of Birth, ID, Institution, etc.

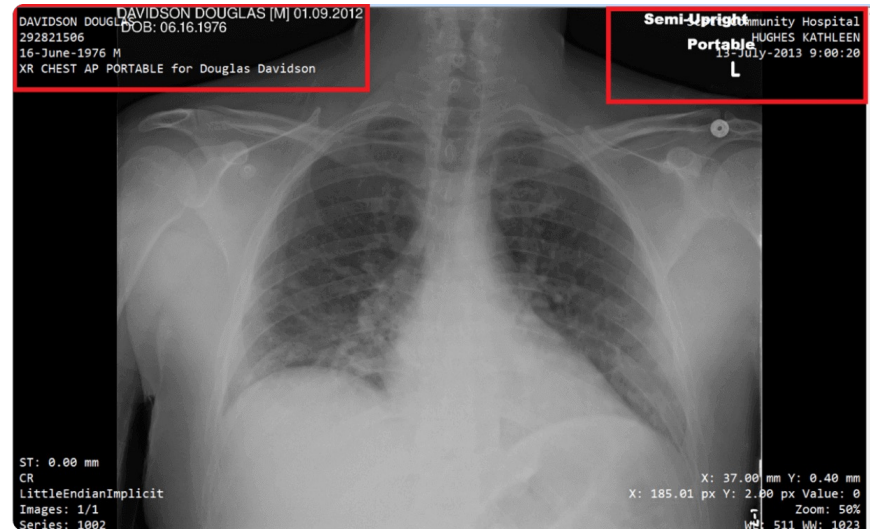
Detection

Handling

OCR or AI
models

Opaque
masking (not
blurring)

Exclude images
if unsafe to
redact



Facial Re-identification Risk

17



MRI/CT faces can be reconstructed



Considered biometric data (GDPR)



Defacing required for head imaging

Defacing strategy

18



Use a well-tested
tool (e.g.,
mri_reface)

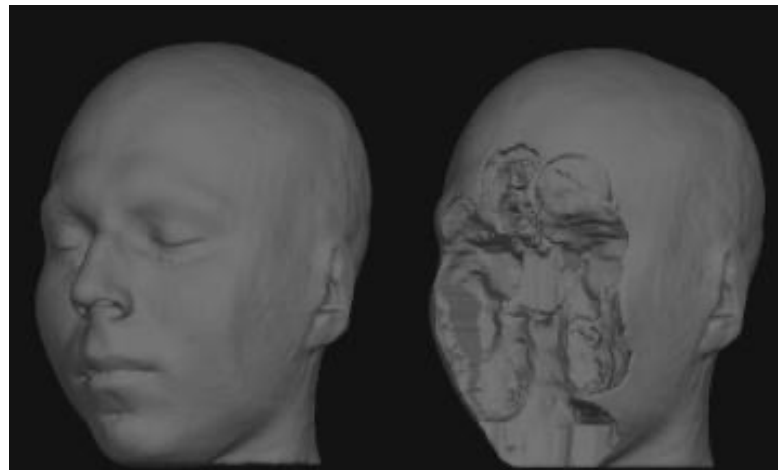


Preserve brain
structure



QC / review results

QC Approach:
Automated validation of
DICOM pixel integrity
combined with risk-based
manual sampling and
exception-driven review.

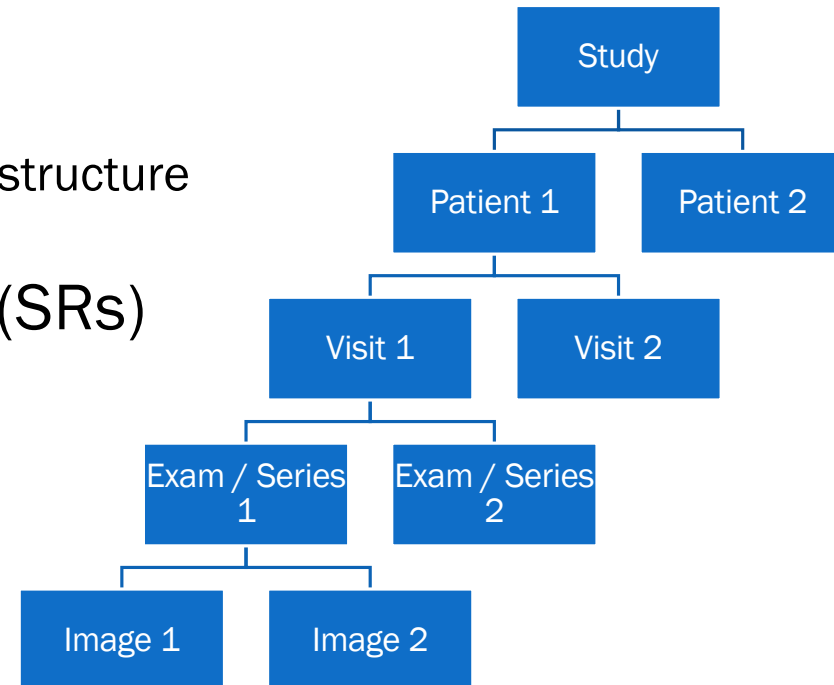


https://www.na-mic.org/wiki/File:Defacing_figure.jpg

Other Risk areas

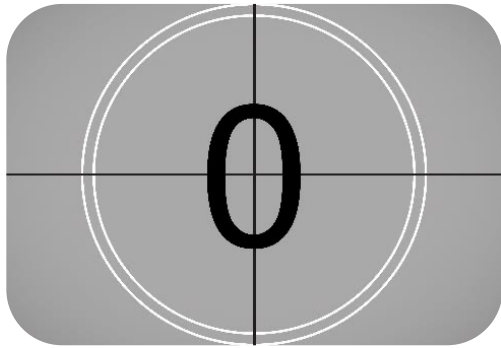
19

- Folder names and file names
 - Standardized anonymized directory structure
- Embedded PDFs and reports (SRs)



Re-identification Risk Philosophy

20



Risk cannot be
zero



Reduce to 'not
reasonably likely'



Controlled secure
environment

Operational Deliverables

21



ReadMe File
(Transformation
log)



Exclusion log



Dataset
Inventory

Key Takeaways

22



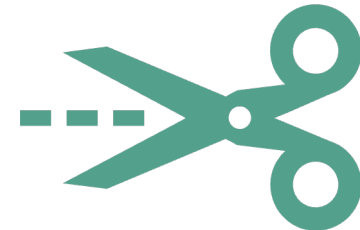
Dual-layer
anonymization:
Metadata + Pixels



Private tags are high
risk by default



Sync Imaging and
Tabular data
transformation



Burnt-in text
redaction



Defacing for head
imaging



Governance
environment +
technical controls



Balance privacy with
scientific utility

Thank You!

Questions

