Imaging Data Risk Mitigation Considerations

**[Project Name]**

**Document prepared by [NAME OF THE PREPARER] on behalf of the [INSTITUTION / ENTITY] and in discussion with NCI IDC**

**Last updated: 2023-06-02**

**Status: Not started**

# 

[**Introduction** 3](#_Toc136613531)

[**Executive Summary** 4](#_Toc136613532)

[**Scope** 4](#_Toc136613533)

[**Overview of the Types of Imaging Data** 4](#_Toc136613534)

[**Individual Imaging Types and their Risks** 5](#_Toc136613535)

[For each of the following image types: 5](#_Toc136613536)

[Gross Microscopy - Anatomical Pathology - True Color 6](#_Toc136613537)

[**File formats** 6](#_Toc136613538)

[**Metadata elements and areas where PHI could appear** 6](#_Toc136613539)

[Slide Microscopy - Anatomical Pathology - True Color 7](#_Toc136613540)

[**File formats** 7](#_Toc136613541)

[**Metadata elements and areas where PHI could appear** 7](#_Toc136613542)

[Slide Microscopy - Anatomical Pathology - Immunohistochemistry (IHC) 8](#_Toc136613543)

[**File formats** 8](#_Toc136613544)

[**Metadata elements and areas where PHI could appear** 8](#_Toc136613545)

[Multiplexed tissue imaging 9](#_Toc136613546)

[**File formats** 9](#_Toc136613547)

[**Metadata elements and areas where PHI could appear** 9](#_Toc136613548)

[**Mitigations** 10](#_Toc136613549)

[Names 10](#_Toc136613550)

[Addresses 10](#_Toc136613551)

[Telephone and Fax Numbers 10](#_Toc136613552)

[Email addresses 11](#_Toc136613553)

[Person Identifiers 11](#_Toc136613554)

[Vehicle Identifiers 11](#_Toc136613555)

[Device Identifiers 12](#_Toc136613556)

[Internet Identifiers 12](#_Toc136613557)

[Identifiable Features 12](#_Toc136613558)

[Dates and Ages 13](#_Toc136613559)

[Any Other Characteristics 13](#_Toc136613560)

[**Plans if PII or PHI Data is Discovered Retrospectively** 14](#_Toc136613561)

[**Overall Assessment of Risk after Mitigation Steps and Incident Response Plans** 14](#_Toc136613562)

[Abbreviations 15](#_Toc136613563)

# 

# **Introduction**

Submitters of data to the NCI’s Imaging Data Commons (IDC) must ensure that the data have been de-identified with respect to protected health information (PHI) and Personally Identifiable Information (PII).

In this document data submitters to IDC need to describe the practices and processes that have been used to de-identify the images.

This document will be reviewed by the IDC team, subject matter experts, and the NCI security team before data is accepted by IDC.

Potential locations of sensitive data include the following:

* Name of the file, folder, and file system metadata properties (e.g., case/subject/patient ID)
* Metadata within the content of file (structured, in TIFF or OME-TIFF XML elements, or semi-/unstructured, in text elements like SVS ImageDescription)
* Burned into the pixel data
  + annotations added to the image by the acquisition equipment
  + hand-written annotations made on the slide
  + labels attached to the slide (with barcode or text) and scanned
  + recognizable characteristics derivable from the tissue image data

Beyond the HIPAA Privacy Rule “Safe Harbor” 18 types of identifiers that potentially may be present in the data, we recognize the following additional elements that require special handling because they may increase the risk of re-identification:

* Dates (other than DOB and “visit dates”) related to the acquisition of the image or the clinical course of the patient.
* Unique identifiers of entities other than the patient (e.g., identifiers of the equipment, sample, or image) - those can potentially contain dates or other identifying elements, and as unique identifiers are a risk in their own right if they can be matched.

In addition to Protected Health Information (PHI) about the patient (per HIPAA Privacy Rule), the data can potentially contain Personally Identifiable Information (PII) for the personnel involved in the collection of the image/sample. IDC will not accept data that contains PHI or PII.

# **Executive Summary**

Provide 1 - 2 paragraphs summarizing the project and the risk assessment of images to be shared with NCI Imaging Data Commons.

# **Scope**

Briefly describe the project for which the images were generated

# **Overview of the Types of Imaging Data**

Describe what types of imaging data are being considered for submission to the Imaging Data Commons. Break down by site if multiple sites will submit images.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Site** | **Imaging Modality** | **Image Format** | **Count** | **Total Size of Images (Gb)** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Total |  |  |  |  |

# **Individual Imaging Types and their Risks**

## For each of the following image types:

* Describe what file format are the images encoded as (e.g., generic TIFF, OME-TIFF, DICOM, proprietary format such as SVS, …).
* Describe metadata elements and areas of the image where PII or PHI could possibly appear. These may include:
  + Label image
  + Macro image
  + Image path
  + Filename
  + Title
* Describe how the submitting entity is mitigating risks of the imaging type, including
  + the process that is being used,
  + the technologies being used, and
  + the amount of quality control that is being planned or done.

Where appropriate, provide pointers to existing documentation

Providing example screenshots of how images are de-identified is beneficial.

Please fill out the mitigation section below for each imaging type. If abbreviations are used, please add these to the table at the end of this document.

## Gross Microscopy - Anatomical Pathology - True Color

This includes visible light images of gross specimens to localize where samples have been obtained.

#### **File formats**

#### **Metadata elements and areas where PHI could appear**

## Slide Microscopy - Anatomical Pathology - True Color

This includes visible light images (e.g., bright-field), frozen section or formalin-fixed paraffin-embedded (FFPE), stained with visible light stains such as, but not limited to, H&E. Includes whole slide images (may be tiled, pyramidal), sub-regions and derived (processed, registered, re-sampled) images.

#### **File formats**

#### **Metadata elements and areas where PHI could appear**

## Slide Microscopy - Anatomical Pathology - Immunohistochemistry (IHC)

This includes visible light images and single-channel fluorescence images. Includes whole slide images (may be tiled, pyramidal), sub-regions and derived (processed, registered, re-sampled) images.

#### **File formats**

#### **Metadata elements and areas where PHI could appear**

## Multiplexed tissue imaging

This includes multiplex images derived from multiplex techniques. Includes whole slide images (may be tiled, pyramidal), sub-regions and derived (processed, registered, re-sampled) images, regardless of the specific specimen preparation and acquisition technique (e.g., CyCIF, etc.).

#### **File formats**

#### **Metadata elements and areas where PHI could appear**

Note: labels of slides, and how they are de-identified should be discussed.

# **Mitigations**

For each of the following types of information, whether it be present in structured or unstructured form in metadata or data, including burned into the pixel data, describe the mitigation(s) planned, including the processes applied by the submitting sites and the processes used by the DCC to confirm.

Note: The fields below come from the list of the HIPAA De-identification Standards Safe Harbor Method: [HIPAA Privacy Rule 18 elements](https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard).

## Names

For all image types, how are names being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Addresses

For all image types, how are addresses being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Telephone and Fax Numbers

For all image types, how are telephone and fax being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Email addresses

For all image types, how are email addresses being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Person Identifiers

For all image types, how are identifiers of persons, including but not limited to, social security numbers (SSN), medical record numbers (MRN), health plan beneficiary numbers, account numbers, certificate, or license numbers, being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Vehicle Identifiers

For all image types, how are identifiers of vehicles, including but not limited to, license plate numbers and serial numbers, being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Device Identifiers

For all image types, how are identifiers of devices, including but not limited to, device identifiers and serial numbers, being detected, and if necessary, remediated?

This includes identifiers of scanning and image processing devices.

*[Note that this is a conflict of re-identification risk versus provenance, and arguably identifiers of machines within a research lab do not pose a significant re-identification risk; IDC will discuss details of this with you if required.]*

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Internet Identifiers

For all image types, how are web URLs and IP addresses being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **True positives:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Identifiable Features

For all image types, how are identifiable features such as fingerprints, voice prints, photographic images (including but not limited to the face) being detected, and if necessary, remediated?

This may include addressing the uniqueness of the nature of the data itself (e.g., very rare molecular profile).

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

## Dates and Ages

For all image types, how are actual dates, including, but not limited to, birthdate, admission date, discharge date, date of death, being detected, and if necessary, remediated?

|  |  |
| --- | --- |
| **Detection:** |  |
| **Remediation:** |  |
| **Value present:** | **Likelihood of value being present [X]/5.** |
| **False positives:** | **Likelihood value being falsely detected [X]/5.** |

For all image types, how are ages, including exact age if over 89 (per HIPAA Privacy Rule), being detected, and if necessary, remediated?

## Any Other Characteristics

For all image types, how are any other characteristics that could uniquely identify the individual being detected, and if necessary, remediated?

# **Plans if PII or PHI Data is Discovered Retrospectively**

Describe how the submitting entity would respond if submitting member sites, DCC, or another team such as IDC discovered possible PII or PHI data (i.e., define the self- and external incident response plans):

* What steps would you take to mitigate that risk?
* Who would be involved?
* With what timelines?

# **Overall Assessment of Risk after Mitigation Steps and Incident Response Plans**

Describe, and quantify or categorize (e.g., very high, high, moderate, low, very low), the overall risk that the submitting entity perceives for the images, after they have gone through mitigation steps and an incident response plan has been put in place.

## Abbreviations

DCC – Data Coordination Center

DOB – date of birth

FFPE – formalin-fixed and paraffin embedded

H&E – Hematoxylin and Eosin

HIPAA – Health Insurance Portability and Accountability Act

IDC – Imaging Data Commons

IHC – Immunohistochemistry

IP – Internet Protocol

MRN – Medical Record Number

NCI – National Cancer Institute

PII – Personal Identifiable Information

PHI – Protected Health Information

SSN – Social Security Number

URL – Unique Resource Locator