

Image DePHI and the DSA: Open Source tools for Histology Image De-Identification

David A Gutman MD PhD
Associate Professor of Pathology Emory University



EMORY
UNIVERSITY
SCHOOL OF
MEDICINE



NATIONAL CANCER INSTITUTE
Informatics Technology for
Cancer Research



EMORY
UNIVERSITY



Disclosures

I own stock in Histowiz Inc, LLC and am a co-founder of SwitchboardMD

The Cancer Digital Slide Archive

CANCER
Digital Slide Archive

Login TCGA Resources Help

CANCER

Slides + - <

Apply Filters Pathology Report Metadata

gdm < > 1/6 > (26 slides)

TCGA-06-0166 < >

TCGA-06-0166-01A-01-B51.2

nationwidechildrens-org_biospecimen_slide_gbm

Key	Value
bcr_patient_uid	4a85a43b-4dbc-4081-b552-df06eef1ad80
bcr_sample_barcode	TCGA-06-0166-01A
bcr_slide_barcode	TCGA-06-0166-01A-01-TS1
bcr_slide_uid	e85993b2-c783-4571-a562-373df206d2ff
image_file_name	TCGA-06-0166-01A-01-TS1.svs
is_derived_from_ffpe	NO
percent_necrosis	50
percent_normal_cells	0
percent_stromal_cells	30
percent_tumor_cells	20
percent tumor nuclei	90

Pathology Reports

TCGA-06-0166.8c5fd571-8b2c-4360-9014-2e9683e8ec5.pdf

MIB-1 PROLIFERATION INDICES: 12% AND 14%.

SEE MICROSCOPIC.

Operation/Specimen: Brain, craniotomy.

GROSS PATHOLOGY:

A. Received fresh, two fragments, 3 x 2 x 0.9 cm. Soft, focally necrotic and hemorrhagic. In total #1-4.

INTRAOPERATIVE CONSULTATION:

1. Brain, imprint: Atypical cells.
2. Brain, squash preparations: Glioma, high histological grade.

B.

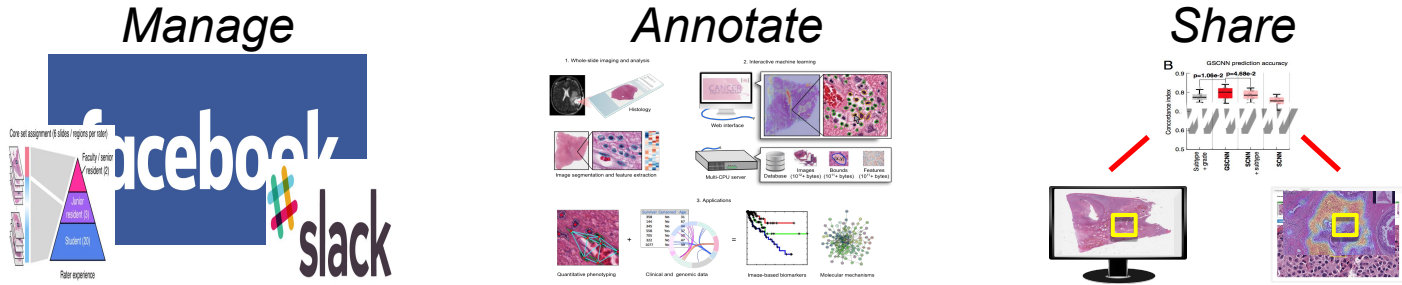
SPECIMEN: 2, brain tumor.
FIXATIVE: Saline.
GENERAL: Three irregularly shaped fragments of tan-red to glistening tan brain tissue, measuring 1.0 x 1.0 x 0.6 cm., 1.5 x 1.2 x .8 cm., and 3.0 x 1.5 x 1.2 cm.
All submitted; 3 - smaller fragment, 4 - intermediate fragment, 5-7 - largest fragment.

SECTIONS:

MICROSCOPIC: A, B. Portions of a neoplastic proliferation of glial cells with an overall astrocytic phenotype. The neoplasm has solid areas and diffusely infiltrates the surrounding brain. The neoplastic cells have nuclear anaplastic features, and there are associated mitotic figures, prominent microvascular cellular proliferation, and areas of necrosis some of them with pseudopallisading.

SPECIAL STAINS: An immunoperoxidase method for MIB-1 was performed on sections from blocks 1 and 2. In the more cellular areas, a MIB-1 proliferation index of 14% and 12%, respectively, were determined in the more active areas.

Open-source digital pathology: Digital Slide Archive (DSA)

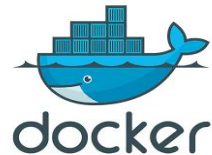
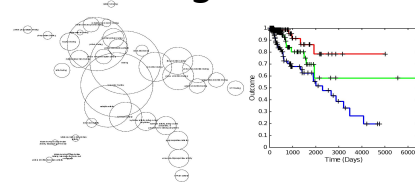
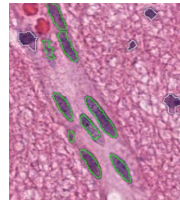


Annotate

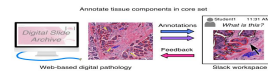
Share

Analyze

Integrate



<http://itcr.nci.nih.gov/>



1U24CA194362-01

DA Gutman *et al*, The Digital Slide Archive: A Software Platform for Management, Integration and Analysis of Histology for Cancer Research, *Cancer Research*, 77(21), pp. 75-78, November 2017

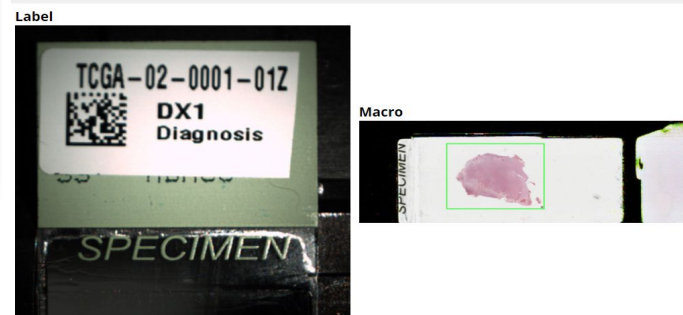
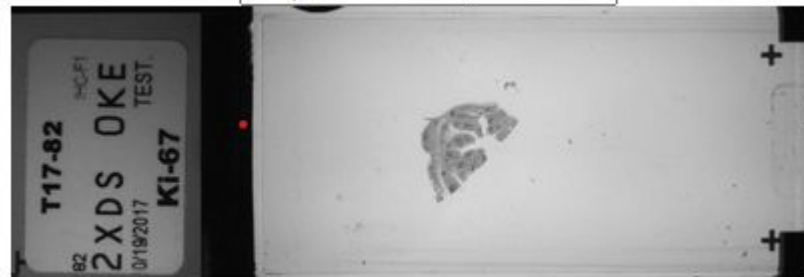
https://digitalslidearchive.github.io/digital_slide_archive/

Protected Health Information in WSI

What & Where?

- For WSI images, if present, there are three places PHI are likely to reside
 - Filename
 - Embedded Image Metadata
 - Embedded Label / Macro Images

Property	Value
aperio.AppMag	20 REDACT <input type="button" value="Keep (do not redact)"/>
aperio.Date	42/14/18 01/01/18
aperio.DisplayColor	0 REDACT <input type="button" value="Keep (do not redact)"/>
aperio.Exposure Scale	0.000001 REDACT <input type="button" value="Keep (do not redact)"/>
aperio.Exposure Time	109 REDACT <input type="button" value="Keep (do not redact)"/>
aperio.Filename	28-A 0579XY112001_03_10
aperio.Focus Offset	-0.000500 REDACT <input type="button" value="Keep (do not redact)"/>



Example Label and Macro WSI Images

Embedded Image Metadata

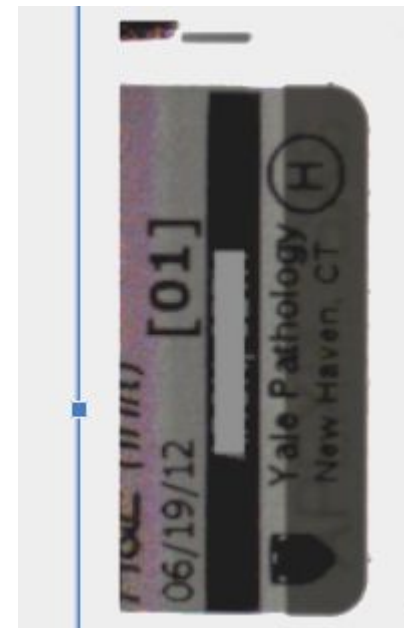
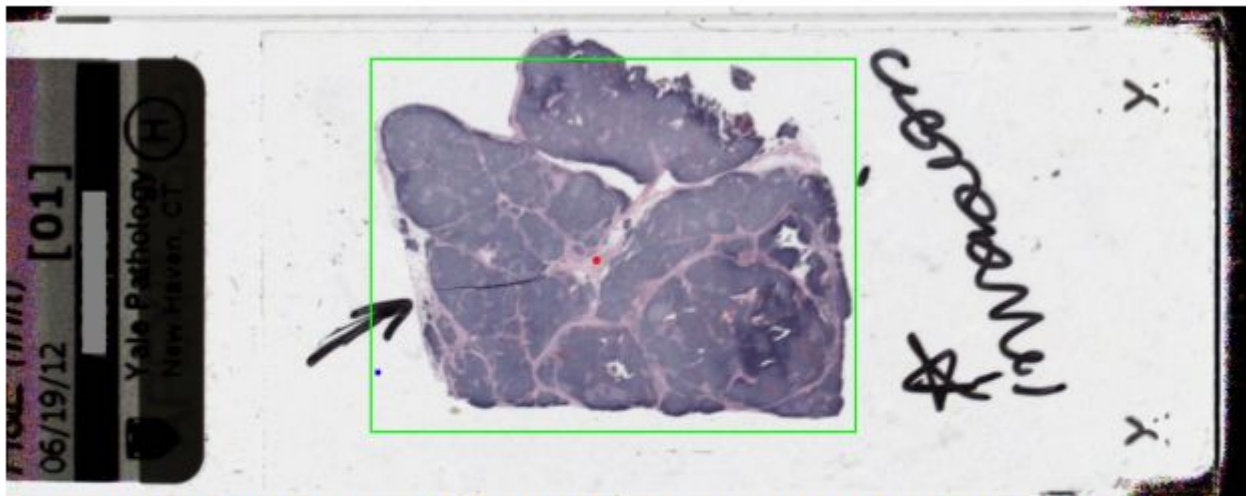


Figure 4: The macro image from TCGA file TCGA-4X-A9FB-01Z-00-DX1.211CC9AA-F721-4D16-8663-68A393223F80.svs. The left side shows part of the label which included the patient name.

This has been manually redacted with a grey bar in this figure. This file was released to the public in 2016.

4X	Yale University	Thymoma	NCH
----	-----------------	---------	-----

Embedded Image Metadata

```
TIFF Directory at offset 0xbbc777fe (3150411774)
Subfile Type: (0 = 0x0)
Image Width: 109559 Image Length: 84381 Image Depth: 1
Tile Width: 240 Tile Length: 240
Bits/Sample: 8
Compression Scheme: JPEG
Photometric Interpretation: RGB color
YCbCr Subsampling: 2, 2
Samples/Pixel: 3
Planar Configuration: single image plane
ImageDescription: Aperio Image Library v12.0.15
111760x84481 [0,100 109559x84381] (240x240) JPEG/RGB Q=70|AppMag = 40|StripeWidth = 2032|ScanScope ID = APERIONP|Filename = E20-72_8 TAU|Date = 08/24/20|Time = 22:27:25|Time Zone = GMT-04:00|User = 81e564c9-09ad-4e80-95ce-ec8daa7330cd|MPP = 0.2519|Left = 17.393621|Top = 22.502415|LineCameraSkew = 0.000365|LineAreaXOffset = -0.014056|LineAreaYOffset = 0.007910|Focus Offset = 0.000000|ImageID = 1030008|Exposure Time = 45|Exposure Scale = 0.000001|DisplayColor = 0|SessonMode = NR|OriginalWidth = 111760|OriginalHeight = 84481|ICC Profile = AT2
ICC Profile: <present>, 1687824 bytes
JPEG Tables: (574 bytes)
TIFF Directory at offset 0xbbf713d0 (3153531856)
Subfile Type: (0 = 0x0)
Image Width: 997 Image Length: 768 Image Depth: 1
Bits/Sample: 8
Compression Scheme: JPEG
Photometric Interpretation: RGB color
YCbCr Subsampling: 2, 2
Samples/Pixel: 3
Rows/Strip: 16
Planar Configuration: single image plane
ImageDescription: Aperio Image Library v12.0.15
109559x84381 -> 997x768 - |AppMag = 40|StripeWidth = 2032|ScanScope ID = APERIONP|Filename = E20-72_8 TAU|Date = 08/24/20|Time = 22:27:25|Time Zone = GMT-04:00|User = 81e564c9-09ad-4e80-95ce-ec8daa7330cd|MPP = 0.2519|Left = 17.393621|Top = 22.502415|LineCameraSkew = 0.000365|LineAreaXOffset = -0.014056|LineAreaYOffset = 0.007910|Focus Offset = 0.000000|ImageID = 1030008|Exposure Time = 45|Exposure Scale = 0.000001|DisplayColor = 0|SessonMode = NR|OriginalWidth = 111760|OriginalHeight = 84481|ICC Profile = AT2
JPEG Tables: (289 bytes)
```

Results of tiffinfo on an Aperio Image provides access to some of the internal TIFF metadata.

```
ImageDescription: Aperio Image Library v12.0.15
109559x84381 -> 997x768 - |AppMag = 40|StripeWidth = 2032|ScanScope ID = APERIONP|Filename = E20-72_8 TAU|Date = 08/24/20
0.2519|Left = 17.393621|Top = 22.502415|LineCameraSkew = 0.000365|LineAreaXOffset = -0.014056|LineAreaYOffset = 0.007910|
DisplayColor = 0|SessonMode = NR|OriginalWidth = 111760|OriginalHeight = 84481|ICC Profile = AT2
```

As an example, the ImageDescription tag may contain the original filename, it may contain specific dates or other PHI.

DSA Plugin for DeID*

WSI DeID

WSI DeID

-
- Approved ▶
- AvailableToProcess ▶
- Original ▶
- Quarantined ▶
- Redacted ▶
- Rejected ▶
- Reports ▶
- Schema ▶
- Unfiled ▶

Step 1: Load / Copy over image set for De-ID

- User supplies CSV file or uploads metadata to the DSA via API

```
Metadata
{
  "ASSAY": "H&E",
  "BLOCK": "BR0011",
  "CASE": "TestProject2",
  "INDEX": "9",
  "ImageID": "SA",
  "OutputFileName": "TCGA-12-0654.S9.DEID.svs",
  "PROJECT": "HP0600-001",
  "REPOSITORY": "DCEG",
  "STUDY": "MR-0600",
  "SampleID": "TCGA-12"
}
deidUpload }
```

1 0 (3) ↑

Redact Checked Import Find label text i ↕ 🔒 📁

Public

*Under Active Development

<https://github.com/dgutman/nci-dsa-deid>

AVOID CLICKS AT ALL COSTS!



Metadata Mapping... The Hard Part

- Slide label and image name contain primary mechanism to map slide → patient
- Metadata is validated against a custom JSON Schema
 - System won't allow redaction if validation fails
- Metadata can also be encoded into bar code



More Text To Stare At

Schema 1.0 →

```
"$schema": "http://json-schema.org/draft-07/schema",
"$id": "http://example.com/example.json",
"type": "object",
"title": "Working schema for NCI CBIIT deidentification workflow validation",
"description": "The schema to validate the DeID Upload CSV file for the NCI CBIIT DeID Project.",
"default": {},
"required": ["InputFileName", "SampleID", "OutputFileName"],
"properties": {
  "PatientID": {
    "type": "string",
    "pattern": "^[a-zA-Z0-9]",
    "description": "The patient ID has to start with a letter or number."
  },
  "SampleID": {
    "type": "string"
  },
  "REPOSITORY": {
    "type": "string"
  },
  "STUDY": {
    "type": "string"
  },
  "PROJECT": {
    "type": "string"
  },
  "CASE": {
    "type": "string"
  },
  "BLOCK": {
    "type": "string"
  },
  "ASSAY": {
    "type": "string"
  },
  "INDEX": {
    "type": "string"
  }
}
```












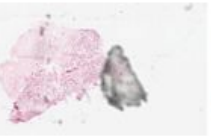



```
InputFileName, SampleID, REPOSITORY, STUDY, PROJECT, CASE, BLOCK, ASSAY, INDEX, ImageID, OutputFileName
TCGA-06-5408-01A-01-TS1.68665795-aac7-4a6d-9a14-c306bc941b26.svs, TCGA-06, DCEG, MR-0600, HP0600-001, TestProject, BR6011, H&E, 1, SA, TCGA-06.S1.DEID.svs
TCGA-06-5412-01A-01-TS1.4e47da5e-588a-4cc1-8c10-a701764b1f7c.svs, TCGA-06, DCEG, MR-0600, HP0600-001, TestProject, BR6011, H&E, 2, SA, TCGA-06.S2.DEID.svs
TCGA-06-5414-01A-01-TS1.660d19f8-1237-47f0-8d5d-71037df4ed67.svs, TCGA-06, DCEG, MR-0600, HP0600-001, TestProject, BR6011, H&E, 3, SA, TCGA-06.S3.DEID.svs
TCGA-06-5416-01A-01-TS1.1749a247-c33d-4a06-88f2-4deef3c5d982.svs, TCGA-06, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 4, SA, TCGA-06.S4.DEID.svs
TCGA-06-5417-01A-01-TS1.a84b8f46-a57d-469d-a65e-4aa5f45da5f3.svs, TCGA-06, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 5, SA, TCGA-06.S5.DEID.svs
TCGA-06-5418-01A-01-TS1.6600b787-bac7-4ad1-8711-f27bae721e7a.svs, TCGA-06, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 6, SA, TCGA-06.S6.DEID.svs
TCGA-08-0509-01A-01-TS1.27f6ae4c-e445-4e2e-a94b-31d449ba69c9.svs, TCGA-08, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 7, SA, TCGA-08.S7.DEID.svs
TCGA-12-0620-01Z-00-DX1.5d4cc50f-9b78-4b8b-8e8e-1c20ae415d71.svs, TCGA-12, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 8, SA, TCGA-12.S8.DEID.svs
TCGA-12-0654-01C-01-BS1.52cba7d5-7130-48c1-a31e-349f8b12822b.svs, TCGA-12, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 9, SA, TCGA-12-0654.S9.DEID.svs
TCGA-12-0820-01Z-00-DX1.43DEB9E2-D419-4263-8188-20FEBB070AAF.svs, TCGA-12, DCEG, MR-0600, HP0600-001, TestProject2, BR6011, H&E, 10, SA, TCGA-12.S10.DEID.svs
```

Select Cases for Redaction

WSI DeID / AvailableToProcess / Patient2

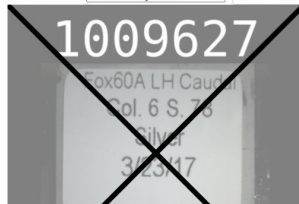


Redact Checked

macro	thumbnail	Item	aperio.DSR ID	aperio.Date	aperio.Filename	aperio.Filtered
		 Patient2_1.4_WSIDEID.svs (1)   11.34 MB	ap1546-dsr REDACT <input type="text" value="Keep (do not redact)"/>	03/24/11 01/01/11	25675 Patient2_1.4_WSIDEID	
		 Patient2_4.2_WSIDEID.svs   9.961 MB	path-mclen001ms REDACT <input type="text" value="Keep (do not redact)"/>	12/02/09 01/01/09	213 Patient2_4.2_WSIDEID	
		 Patient2_4.6_WSIDEID.svs   7.566 MB	AP1258-DSR REDACT <input type="text" value="Keep (do not redact)"/>	12/18/08 01/01/08	7191 Patient2_4.6_WSIDEID	5 REDACT <input type="text" value="Keep (do not redact)"/>

Associated Images

Label **redact** Redact Area



Macro **redact**



Thumbnail **redact**

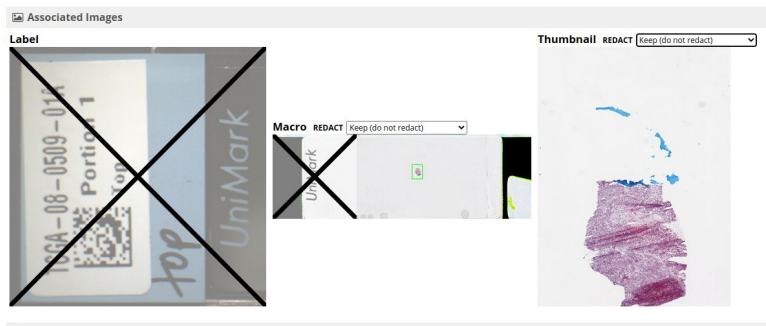


Redact: With Default Profile

WSI DeID / AvailableToProcess / TCGA-08

Redact Checked Import Find label text i [upload] [lock] [share]

name	aperio.ImageID	aperio.Originalheight	aperio.ScanScope ID	aperio.Time	aperio.Title	aperio.User	tiff.ResolutionUnit
2626	8081	SS1258	09:25:54	none	81cbf3e8-d31e-4c48-bb52- TCGA- 08.S7.DEID.svs	bb47e6745daa	Inch
REDACT	Keep (do not redact)	REDACT	Keep (do not redact)	REDACT	Keep (do not redact)	REDACT	Keep (do not redact)



Review Redacted Results

WSI DeID

WSI DeID / Redacted / Case12



Approve Checked



WSI	label	macro	thumbnail	Item	aperio.CustomField.BLOCK	aperio.CustomField.CASE	aperio.CustomField.INDEX	aperio.CustomField.IN
-----	-------	-------	-----------	------	--------------------------	-------------------------	--------------------------	-----------------------

	<p>Case12_1_6_WSIDEID</p>  			<p> Case12_1_6_WSIDEID.svs 1 11.04 MB</p>  		Case3	6	TCGA-19-2623-01B-01- TS1.f7894961-4e2c-4b1 442d1933fef9.svs
--	--	---	---	--	--	-------	---	---

	<p>Case12_3_4_WSIDEID</p>  			<p> Case12_3_4_WSIDEID.svs 3 9.825 MB</p>  				
--	--	---	---	--	--	--	--	--

 **Files & links**

 Case12_1_6_WSIDEID.svs   11.04 MB

 **WSI DeID Workflow**

06-5412-01A-01-
:47da5e-588a-4c
64b1f7c.svs

Quarantine

Approve

< Previous

Export Results To Target

Export Recent

Export All

WSI	label	macro	thumbnail	Item	aperio.CustomField.BLOCK	aperio.CustomField.CASE	aperio.CustomField.
-----	-------	-------	-----------	------	--------------------------	-------------------------	---------------------

	<p>Patient2_1_4_WSIDEID</p> 		
--	---------------------------------	--	--

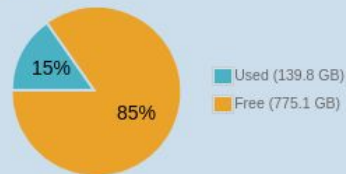
	<p>Patient2_4_2_WSIDEID</p> 		
--	---------------------------------	--	--

Assetstores

[View all past Imports](#)

Below is a list of all of the assetstores available to the server. The one set as "current" is the one that uploaded files will be written to.

Name: Assetstore (Current assetstore)
Unique ID: 63bda50af50b104ce17506de
Type: Filesystem
Root path: /assetstore
File creation permissions (octal): 600
Capacity: 775.1 GB free of 914.8 GB total



Edit

Delete

[Import data](#)

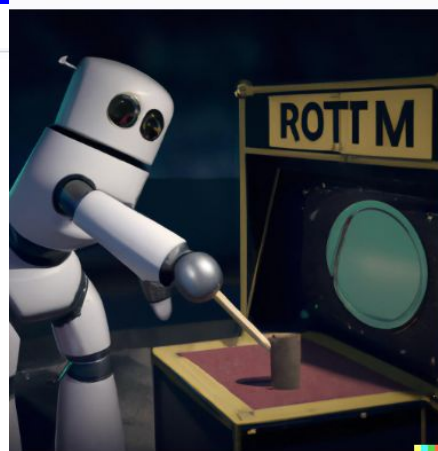
[View past Imports](#)

Image Format Whack-a-Mole

Running on a recent .svs fails with default rules due to unexpected keys:

- CalibrationAverageBlue
- CalibrationAverageGreen
- CalibrationAverageRed
- ScannerType
- SessionMode
- Scan Warning

This slide was generated using a Leica GT450. If necessary, we can find a shareable instance.



```
$ ./imagedephi-macOS gui
[2023-05-17 14:34:26 -0500] [44941] [INFO] Running on http://127.0.0.1:65305 (CTRL + C to quit)
Server is running at http://127.0.0.1:65305/ .
Redacting 017a8f83-6be0-63e4-5fe6-125901b4e243_4_5_173735.svs...
Redaction could not be performed for 017a8f83-6be0-63e4-5fe6-125901b4e243_4_5_173735.svs.
The following keys were found in Aperio ImageDescription strings and could not be redacted given the current se
Missing key (Aperio ImageDescription): CalibrationAverageGreen
Missing key (Aperio ImageDescription): ScannerType
Missing key (Aperio ImageDescription): SessionMode
Missing key (Aperio ImageDescription): Scan Warning
Missing key (Aperio ImageDescription): CalibrationAverageRed
Missing key (Aperio ImageDescription): CalibrationAverageBlue
```

Stand Alone App

ImageDePHI

ImageDePHI is an application to redact personal data (PHI) from whole slide images (WSIs).

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, under Contract No. 75N91022C00033

Goal: Standalone application that works on Linux, Windows & MacIntosh

Format Support: SVS, NDPI, Philips TIFF, iSyntax**

ImageDePHI: Basic GUI or command line

Directory Selector

Input Directory

/ > home > dagutman > data
> import

- TCGA-3C-AALJ-01A-03-TSC.272FA991-8382-409A-A00A-9A3BAA6EE041.svs
- PatientC.Slide2.svs
- TCGA-3C-AALJ-01Z-00-DX1.777C0957-255A-42F0-9EEB-A3606BCF0C96.svs
- PatientA.Slide3.svs
- SampleA.svs
- PatientB.Slide2.svs
- TCGA-02-0006-01B-01-BS1.e426e878-9367-482d-8cd9-2f389aa96dfe.svs
- SampleB.svs
- PatientA.Slide2.svs
- TCGA-02-0006-01B-01-TS2.15b581bb-b760-46b4-b8ae-79c1fcd17d67.svs

[4 More Images](#)

SELECT CURRENT DIRECTORIES

Output Directory

/ > home > dagutman > data
> export

- Case1
- Case2
- Case3

```
dagutman@dagutman-XPS-9315:~/Downloads$ ./imagedephi-Linux --help
```

```
Usage: imagedephi-Linux [OPTIONS] COMMAND [ARGS]...
```

```
Redact microscopy whole slide images.
```

Options:

```
--version           Show the version and exit.  
-r, --override-rules FILENAME  User-defined rules to override defaults.  
--help             Show this message and exit.
```

Commands:

```
gui  Launch a web-based GUI.  
plan Print the redaction plan for images.  
run  Perform the redaction of images.
```

```
{"message": "You chose this input directory: /home/dagutman/data/import and this output directory: /home/dagutman/data/export"}
```

Set it and forget it....

Case1



REDACTED_PatientA.Slide2.
svs

Case2



REDACTED_PatientA.Slide3.
svs

Case3



REDACTED_PatientA.Slide4.
svs

DeID Export Job 20221222
181324.xlsx



REDACTED_PatientB.Slide1.
svs

REDACTED_PatientA.Slide1.
svs



REDACTED_PatientB.Slide2.
svs



REDACTED_PatientC.Slide2.
svs



REDACTED_SampleA.svs



REDACTED_SampleB.svs



REDACTED_TCGA-02-0006-
01B-01-BS1.e426e878-9367-
482d-8cd9-2f389aa96dfe.svs



REDACTED_TCGA-02-0006-
01B-01-TS2.15b581bb-b760-
46b4-b8ae-79c1fcd17d67....

What did it do?

```
TRL + C to quit)
Server is running at http://127.0.0.1:41011/ .
Redacting TCGA-3C-AALJ-01A-03-TSC.272FA991-8382-409A-A00A-9A3BAA6EE041.svs...
Redacting PatientC.Slide2.svs...
Redacting TCGA-3C-AALJ-01Z-00-DX1.777C0957-255A-42F0-9EEB-A3606BCF0C96.svs...
Redacting PatientA.Slide3.svs...
Redacting SampleA.svs...
Redacting PatientB.Slide2.svs...
Redacting TCGA-02-0006-01B-01-BS1.e426e878-9367-482d-8cd9-2f389aa96dfe.svs...
Redacting SampleB.svs...
Redacting PatientA.Slide2.svs...
Redacting TCGA-02-0006-01B-01-TS2.15b581bb-b760-46b4-b8ae-79c1fcd17d67.svs...
```

```
Tiff Tag 32997 - ImageDepth: keep (base)
Tiff Tag 273 - StripOffsets: keep (base)
Tiff Tag 278 - RowsPerStrip: keep (base)
Tiff Tag 279 - StripByteCounts: keep (base)
Tiff Tag 317 - Predictor: keep (base)
SVS Image Description - AppMag: keep (base)
SVS Image Description - StripeWidth: keep (base)
SVS Image Description - ScanScope ID: delete (base)
SVS Image Description - Filename: delete (base)
SVS Image Description - Title: delete (base)
SVS Image Description - Date: delete (base)
SVS Image Description - Time: delete (base)
SVS Image Description - User: delete (base)
SVS Image Description - MPP: keep (base)
SVS Image Description - Left: keep (base)
SVS Image Description - Top: keep (base)
SVS Image Description - LineCameraSkew: keep (base)
SVS Image Description - LineAreaXOffset: keep (base)
SVS Image Description - LineAreaYOffset: keep (base)
SVS Image Description - DSR ID: delete (base)
SVS Image Description - ImageID: delete (base)
SVS Image Description - OriginalWidth: keep (base)
SVS Image Description - OriginalHeight: keep (base)
The redaction plan is comprehensive.
```

Thanks

Lee AD Cooper (Northwestern)

David Manthey (Kitware)

Juan Carlos Vizcarra (Emory)

Alison Van Dyke (NCI)

Scott Lawrence (NCI)

Questions: Email me at dgutman@emory.edu or dagutman@gmail.com

Supported by NCI ITCR U24CA194362, U01-AG061357-04

and NCI SBIR Contract No. 75N91022C00033

Label

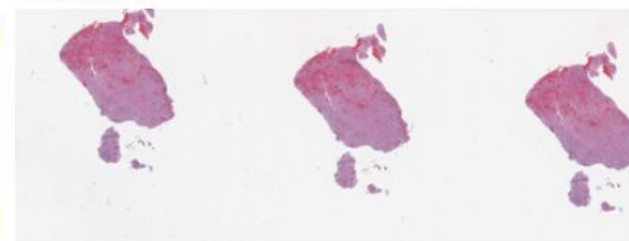
Patient3_3.5_WSIDEID



Macro



Thumbnail



Recent export task completed. 3 images exported. See the [Excel report](#) for more details.

Export Recent Export All ⓘ ⬆️ 🔒 🗑️

tomField.CASE aperio.CustomField.INDEX aperio.CustomField.InputFi

i Info

335.6 MB (351938845 bytes)

Created on May 23, 2023 at 9:56:39

Updated on May 23, 2023 at 9:56:39

Unique ID: 646cc6170ebbc64be7b1a48

Image Viewer

Clear Area

Other PHI/PII

Reject

Keep (do not reject)

