

# THE USER'S GUIDE TO POSDA FOR MEDICAL IMAGE CURATION

FOR POSDA VO.8

# Activity Based Curation

What is Posda?	4
Posda Installation	5
Testing Your Installation	12
Curation Background	13
Data Collections	13
Why use an Activity?	13
Getting Started	13
Collection Preparation	14
DICOM Roots Table Editor	14
Search For Sites or Collections	15
Adding a new Collection	16
Importing Data	
Importing data With the Posda DISK Import tool	
Import Data With DICOM send	
Activity Based Curation	23
Activity Based Curation Layout and Navigation	23
Activity Based Curation Layout and Navigation	23 
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content	
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page	
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries	
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query	
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query	
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen	23 24 25 26 27 28 29 30
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data	23 24 25 26 27 28 29 30 33
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data.	23 24 25 26 27 28 29 30 33 35
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data Associate Imported Data with an Activity Timepoint Import Events By Matching Name	23 24 25 26 27 28 29 30 33 33 35 37
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data Associate Imported Data with an Activity Timepoint Import Events By Matching Name Create Timepoint From All Files in Import Event List	23 24 25 26 27 28 29 30 33 33 35 37 38
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data Associate Imported Data with an Activity Timepoint Import Events By Matching Name Create Timepoint From All Files in Import Event List Using the Inbox	23 24 25 26 27 28 29 30 30 33 35 35 37 38 40
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data Associate Imported Data with an Activity Timepoint Import Events By Matching Name Create Timepoint From All Files in Import Event List Using the Inbox Patient Mapping	23 24 25 26 27 28 29 30 30 33 35 35 37 38 40 40
Activity Based Curation Layout and Navigation Quick Navigation Page Title and Page Content Activity Page Using Queries Searching for a query Running a query The Query Results Screen Creating an Activity and Associating the Data Associate Imported Data with an Activity Timepoint Import Events By Matching Name Create Timepoint From All Files in Import Event List Using the Inbox Patient Mapping Import Patient Mapping	23 24 25 26 27 28 29 30 30 33 33 35 35 37 38 40 40 43

Run Count Checks	52
Check for Duplicate SOPs	54
Run Consistency Check	56
Verify DICOM IOD (Dciodvfy)	57
Visual Review	58
Using Kaleidoscope and Quince	60
Removing Bad Data from activity Timepoint	63
PHI Review - The Posda PHI Scan	64
Check Struct linkage	74
Link RT Data	76
Send to Server	79
Compare Data to Server - Public PHI Scan	82
Closing An Activity	84
Re-opening an Activity	85
Extras: Produce Activity Report, Copy or Consolidate Timepoints, and Other	86
Final Notes	89

## WHAT IS POSDA?

Posda is a suite of medical image curation tools developed by the TCIA team. Originally the "Perl Open Source DICOM Archive" tool, the project has expanded to include tools written in not only Perl, but also Python, Javascript, and more. In addition, while much of the functionality is focused on DICOM file types, many functions are now being designed to support other datatypes. The suite is hosted in a docker container, which contains the main Posda application, the database, and other applications and software tools used in the Posda curation process.

#### WHAT IS TCIA

The Cancer Imaging Archive is a grant initiative that receives imaging collections from researchers, catalogues and de-identifies them, and then hosts the images for further research. Researchers, even those that cannot obtain their own images, now have a great source of image data to use, improving access and reproducibility in the field.

#### WHAT IS CURATION FOR TCIA

For these image collections to be useful, the imaging data must be organized, and contain as much data as possible while avoiding revealing any personal health information(PHI). The data about which kind of cancer is in the image, where it is in the body, what time frame within the treatment the image was taken, how long the cancer had been developing, are all crucial data points for good research. The TCIA team developed a process of curating the collections that is used to organize the collections, remove all personal health information, and leave as much useful data as possible. As part of that process, the Posda curation tools were developed and used by the TCIA.

#### POSDA BEYOND TCIA

Imaging collection curation is a process that many other projects and organizations could find useful. Perhaps you are part of a different project that manages image collections and wishes to make use of the Posda tool suite. Posda was envisioned to be agnostic to the organization using it, so that other research projects can use these tools as part of their own curation process.

#### Want to learn more? Look here:

<u>Posda info on the TCIA Wiki</u> (https://wiki.cancerimagingarchive.net/display/Public/Posda) <u>"Reengineering Workflow for Curation of DICOM Datasets" An academic paper describing</u> <u>Posda</u> (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6261183/)

# POSDA INSTALLATION

Posda is designed to be hosted on a server and accessed via the network by users, rather than being installed locally on each user's machine. You can (and probably should) configure your server to only allow local access. Most users will never have to worry about installation and deployment, they will simply have a URL to access Posda and an account to login. If you are a user looking for usage guidelines, skip this section.

## HOW TO INSTALL POSDA

You can find this section on the web: <u>http://tcia-posda-rh-2.ad.uams.edu/docs/installation/</u>

#### THINGS TO CONSIDER

There are a handful of things you should consider before beginning installation, such as:

- What host OS will you use?
- What type of installation?
- Will storage be separate?
- Will database be separate?

## HOST OS

Posda is distributed as a set of <u>Docker</u> containers, which means that it will technically work on any platform Docker supports (including Windows, MacOS, and Linux). However, we recommend Linux be used for the best compatibility. MacOS is the second-best choice, and we discourage the use of Windows.

Any Linux distribution should work, as long as you can install Docker, though we have tested Posda on the following:

- Ubuntu 16.04, 18.04, 20.04
- RHEL 7
- Alpine

#### TYPE OF INSTALLATION

Posda supports a number of different types of installation. The main options are if separate storage will be used, and/or a separate database server, and/or separate compute nodes (worker nodes). The reasons why you would make these choices are beyond the scope of this documentation, but you should make those choices before beginning installation.

We have included three example configurations in this guide:

- Small Single machine, for a small site, or development, or demonstration
- Medium When separate storage is needed
- Large When separate storage, database, and worker node servers are needed

## INSTALL SIZES / TYPES

Here are three common types of installations, along with a list of which sections you would need to follow for each.

#### SMALL / DEVELOPMENT INSTALLATION

This is the appropriate set of sections you would follow to install Posda on a single machine, such as for development or demonstration purposes, or just an environment where this is all that is required.

- Install Docker
- Clone the Repo
- Configure Common Settings
- Start Posda

#### MEDIUM / SEPARATE STORAGE

- Install Docker
- Clone the Repo
- Connect storage to Host
- Configure storage in docker-compose.yaml
- Configure Common Settings
- Start Posda

#### LARGE / SEPARATE STORAGE, SEPARATE DATABASE, SEPARATE COMPUTE

- Provision database server, install PostgreSQL
- Install Docker
- Clone the Repo
- Connect storage to Host
- Configure storage in docker-compose.yaml
- Configure database
- Configure Common Settings
- Configure Worker Nodes
- Start Posda

#### INSTALL DOCKER

Installing Docker is beyond the scope of this documentation. You can find further information at the <u>Docker Homepage</u>.

Make sure you also install docker-compose. Instructions are available at <u>Install Docker</u> <u>Compose</u>.

### CLONE ONEPOSDA

First you will need to install git on your system. Once you have git, simply clone Posda by typing:

git clone https://code.imphub.org/scm/pt/oneposda.git

You can place this directory anywhere you like, however we recommend placing it

in a location owned by a dedicated user (or root). Any admin working on

this installation will need access to this directory, so it may not make sense

for it to be owned by a normal user.

After it has finished downloading move into the directory and initialize your

configuration (this will also download the docker containers).

cd oneposda ./init

#### **CONFIGURE STORAGE**

The bulk of the storage will be from the "Posda Cache", where all imported and edited DICOM files are stored. There is also the Postgres database data, if you are using the built-in database.

By default both storage locations are managed by Docker, but you may want to configure the location yourself, such as on network attached storage.

The main way you do this is by changing the volume sections within the various services in the docker-compose.yaml file.

For example, if you wanted to place all of the cache (DICOM) files within the host directory /mnt/storage/posda, you would find all lines within docker-compose.yaml that look like:

```
posda_cache_alpine:/home/posda/cache
```

and change them to:

/mnt/storage/posda:/home/posda/cache

Make sure you change all occurrences!

NOTE: You should ensure that the owner:group ID of the chosen location is 1000:2123. This is the UID and GID which Posda runs as inside the containers, and setting it this way will guarantee Posda can write to its storage directory. This is not a hard requirement, so long as the location is <u>writable</u> by UID 1000.

If you additionally wanted to change where the built-in Postgres database stores its data, first choose a location (such as /mnt/storage/database), then change the line that looks like:

pgdata\_alpine:/var/lib/postgresql/data

to instead be:

/mnt/storage/database:/var/lib/postgresql/data

#### CONFIGURE DATABASE

If you have chosen to use a separate database host (instead of using the container), you will need to complete this section.

First, you must remove the db service from the docker-compose.yaml file. Remove these lines:

db:

```
image: postgres:10.1-alpine
restart: always
environment:
    POSTGRES_PASSWORD: example
volumes:
    - pgdata_alpine:/var/lib/postgresql/data
ports:
    - 5433:5432
```

Then, find every occurrence of depends\_on: db and remove it. For example, remove only the – db line from the following:

depends\_on: - db - redis

Remove this everywhere it occurs, otherwise the various containers will fail to start.

Finally, you must edit the database.env file, and set the appropriate values for the database server you have configured. You must use a role which has the ability to create databases. After initial configuration you can remove that right, if you want.

Example:

### CONFIGURE COMMON SETTINGS

Edit the file common.env and set the values. In particular, you should carefully choose POSDA\_EXTERNAL\_HOSTNAME. ;it should be the hostname where users will expect to find the Posda server.

#### **CONFIGURE WORKER NODES**

By default, the docker-compose.yaml defines two Worker Nodes:

- \* High (priority 1)
- \* Low (priority 0)

By default, these will run on the same server as the rest of Posda. You have a number of options for customizing this setup. You can define additional priority levels (numeric), which you can use for more fine-grained association of tasks with nodes (via further configuration that is beyond the scope of this document). You can also choose to run additional nodes, so that additional tasks can execute simultaneously. You can also choose to run these nodes on dedicated systems to greatly expand the available computing resources the Posda system has access to.

You can increase (or decrease) the number of running nodes of each type by adjusting the `replicas` entry in the docker-compose.yaml file for each type of worker node. By default the low (default) priority has 3 replicas, and the high priority has 1.

If you wish to run worker nodes on dedicated servers, you will need to be running separate storage and database servers as well. You will need to copy the `oneposda` repository (including the configuration files) to each host that will run a worker node instance, then modify the docker-compose.yaml (on those hosts only!) to remove all `service`s other than the desired worker node entry. You should then be able to start it using `./manange up -d` just like normal.

#### START POSDA

The first time you start Posda, you will want to start the db and posda containers first, and wait about 30 seconds after each command for initial setup to complete.

The manage script is a simple wrapper around docker-compose, but it configures some things before each run, so it is recommend you use it rather than using docker-compose directly.

WARNING: If you have chosen to have a separate database host, you must skip the first command (because there will be no `db` container).

From the oneposda directory, execute:

./manage up -d db
# Wait 30 seconds for the database to start
./manage up -d posda
# Wait 30 seconds for initial setup to complete
./manage up -d

# **TESTING YOUR INSTALLATION**

#### THINGS TO CONSIDER

- Make sure your preferred browser will allow pop-ups from the posda application.
- Make sure you have a user account in Posda
- Ensure your version of Posda to be tested is installed and running
- Ensure you have the Posda Importer GUI installed on the machine you intend to test

#### IF RUNNING LOCALLY FROM SCRATCH

- locate the oneposda directory
- open a terminal
- type ./init this will update your docker image
- type ./manage up this will bring up your database and applications

#### DOWNLOAD "DICOMTESTSET" DATA

• Found here: <u>https://pathology.cancerimagingarchive.net/download/other/</u> <u>DicomTestSet.tgz</u>

### FOLLOW THE INSTRUCTIONS FOR CURATION

Use this guide to complete a full curation process with the DICOMTestSet data.

## CURATION BACKGROUND

#### DATA COLLECTIONS

Data often comes in to the curation team as a dataset called a Collection. The Principal Investigator and sending Site will work with the curator to give the Collection a proper collection name and other necessary values. For example, in the TCIA system every collection gets a 4 digit collection code that is then associated with the sending site's 4 digit code to create a unique 8 digit ID. In addition, the PI/site can help the curator to choose an appropriate patient mapping system. (See Patient Mapping)

Each collection is unique. Therefore, a real world imaging collection will not always follow the same exact curation steps as another collection. As such, Posda makes many options available at each step.

#### WHY USE AN ACTIVITY?

Curation in Posda is done in a workflow where your tasks on a collection are grouped together into an Activity. This allows you better view and understand what tasks have been done on a collection.

When changes are made to the data in the Activity, such as when Edits are imported or data is removed, a new Activity Timepoint is created. Most processes use the Latest Activity Timepoint, but some queries and processes allow you to specify a specific one. In the Activity Timeline you can see each task that created an Activity Timepoint as wells the number of files in those timepoints.

#### GETTING STARTED

Before you can curate your data collection, you need to create its identifiers and then import the data into Posda(you can add more data later). After you have created the collection identifier(Collection Preparation) and have data uploaded (Importing Data), you can begin curating in the main Posda Activity Based Curation application.

# **COLLECTION PREPARATION**

Before you can curate the collection, it must have a unique identifier. This guide will go over the collection creation process used by the TCIA team. Your local team may have their own process for determining these IDs and adding them to the Posda database.

In the TCIA system every collection gets a 4 digit collection code that is then associated with the sending site's 4 digit code to create a unique 8 digit ID. The TCIA team uses the DICOM Roots Table Editor application to add these IDs to the database.

(The examples in this section uses fictional sites and collections.)

### DICOM ROOTS TABLE EDITOR

- Work with your team and site to create a unique collection name and patient mapping plan.
- Go to the landing page for your installation and choose DICOM Roots Table Editor



This opens the DICOM Roots Table Editor application.

	OM Roots V	iewer and Editor						
Search for a sub	mission							
Site Name	-	Search Criteria						
You can change this o	option ^	You can use % as a wildcard						
Collection Name	÷ -	Search Criteria						
You can change this o	option ^	You can use % as a wildcard						
Search Sh	now All Cle	ear						
Add New +								
SITE CODE	SITE NAME	COLLECTION CODE	COLLECTION NAME	PATIENT ID PREFIX	BODY PART	Access Type	BASELINE DATE	DATE SHIFT

CITES OD COLLECTIONS

- In the Search section, look for the collection name, site name, or other data point you think may already exist.
- Add your search criteria.
  - Site Name and Collection are the defaults, but you can click the drop down arrow to change these and search by other parameters, such as Site Code.



SEADCH EOD

- Click Search once your criteria is entered.
- Note the results.
- Click Clear to clear the results.
- Click Show All to see all site + collection combinations in the database.

#### Example

 I want to add a new data collection, but I think the sending site may have sent other collections to us before. I will search for my Site Name to get the exact Site Name and Site Code used previously so that no duplicate sites are added to the system.

Search for a sub	mission	Search Criteria						
Site Name	•	UA%						
You can change this o	ption ^	You can use % as a wildcard						
Collection Name		Search Criteria						
You can change this o	ption ^	You can use % as a wildcard						
Search Sł	low All Cle	par						
Add New +								
SITE CODE	SITE NAME	COLLECTION CODE	COLLECTION NAME	PATIENT ID PREFIX	BODY PART	ACCESS TYPE	BASELINE DATE	DATE SHIFT
4444	UASITE	8909	LUNGDATA	PAT-XXX	LUNG	public	1960-01-01 00:00:00	

#### ADDING A NEW COLLECTION

• Click the "Add New +" button. It appears just under the purple buttons.

You can change	e this option ^	You can use % a	as a wildcard		
Search	Show All	Clear			
Add New +					
SITE CO	DDE SITE	NAME COLLECT	ION CODE	COLLECTION NAME	ŀ

A new data entry form appears above the rows.

	Search Cancel I	Show All Cle	New Collection Code	New Date Shift					
Si	e Code Add This Subr	nission	Site Name	Collection Code	Collection Name	Patient Id Prefix	Body Part	Access Type	Baseline Date (XX-XX-XXXX) Date Shift
	SITE CODE	SITE NAME	COLLECTION CODE	COLLECTION NAME	PATIENT ID PREFIX BODY PART	ACCESS TYPE BASE	INE DATE DATE SHIFT		

There are nine text fields to be completed, but only the first four are required.

- Begin entering your Site Name or Site Code.
  - If the Site exists, the other value will autofill. For example, using the system setup in the above section, if I enter Site Code: 4444, the Site Name: UASITE will automatically populate the Site Name field.
  - If you have a new Site that has not been used before you can enter a Site Name of your choice. For the Site Code you can make one up yourself or with your PI but the best practice is to use the "New Site Code" button. This will populate the code field with a random unused ID number.



• Any new Site will have a highlighted message appear above the Site Code. This is not an error message. It is simply alerting you that this will add a new entry.



- Next enter your Collection Name or Collection Code in the same fashion.
  - If this collection is part of a larger existing Collection, enter one value and the other value will autofill. For example, using the system setup in the above section, if I enter Collection Code: 8909, the Collection Name: LUNGDATA will automatically populate the Collection Name field.
  - If you have a new Collection that has not been used before you can enter a Collection Name of your choice. For the Collection Code you can make one up yourself or with your PI but the best practice is to use the "New Collection Code" button. This will populate the code field with a random unused ID number.

Search	Show All Cle	ear		
Cancel	New Site Code	New Collection Code	New Date Shift	
4444		UASITE	Collection Code	LIVERDATA
Add This S	Submission			

• Any new Collection will have a highlighted message appear above. This is not an error message. It is simply alerting you that this will add a new entry.

This will create a new Collection!	
1440	LIVERDATA

• Enter any other data you have about the Collection including the patient mapping scheme information, and whether this collection requires limited access. These data points are not required at this time, but are saved here for your records.

- If you plan to use the date shift model of patient mapping, the "New Date Shift" button can provide you with a random shift.
- Once your have entered the data, click "Add this Submission"

Search Cancel	Show All Clear New Site Code New Collection Code	New Date Shift					
4444 Add This St	UASITE	This will create a new Collection! 1440	LIVERDATA	Patient Id Prefix	LIVER	public	Baseline Date (XX-XX-XXX: 4438
SITE CO	DDE SITE NAME COLLECTION CODE	COLLECTION NAME PATIENT ID PREFIX	BODY PART ACCESS TYPE	BASELINE DATE D	ATE SHIFT		

• If the identifier combinations were unique with no conflicts, you will receive a success message.

#### Add New + Success! Submission Added!

- If you receive an error, the data will not be submitted. Read the error text and try to resolve the issue and try again. Most often one of the codes or names is already in use. For example, in the above situation, if I tried to add Site Code: 4444 and Site Name: UAsite to a submission, it would error because the Site Code: 4444 is already matched with the Site Name: UASITE.
- Once you receive a success message, your data will have been added to the Posda Database. The identifiers created here will later be mapped to your imported data files in the Patient Mapping step.

## **IMPORTING DATA**

#### IMPORTING DATA WITH THE POSDA DISK IMPORT TOOL

- Locate your data directory.
- Download and install the latest Posda Importer from <a href="https://posda.tools/">https://posda.tools/</a>
- Open the Posda Importer GUI

•	Posda Importer 2.5
	Select Directory
Change	your_posda_environment
Change	your_computer_name
Import Comment:	
	Begin Import

- If this is your first time using the application, or you are using a different database than normal, change your environment with the "Change" buttons. The top is the URL for your posda environment. The second it the computer you are importing data from, most commonly this will be your computer name.
- To start the import process, click "Select Directory"
- Select your data directory.
- Data about the files in the chosen directory should appear in the window (The DicomTestSet has 628 files)

	Colort Directory
	Select Directory
Change	your_posda_environment
Change	your_computer_name
/Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303668
/Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
/Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
/Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
/Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
/Users/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665
lleers/sutecht/	Downloads/DicomTestSet/Patient_031/1.3.6.1.4.1.14519.5.2.1.19591965555960823859303665

Begin Import

- Add an Import Comment. This field is case sensitive. **This will be your key for finding your imported data later so choose a meaningful value!** You may even wish to record the comment in your notes or leave this window open.
- Click "Begin Import"
- Wait for files to complete importing, ( If using the test data it should have 0 errors. Real datasets will sometimes have issues to investigate using the Logs). If you close the application early you will be prompted to resume after reopening.

#### IMPORT DATA WITH DICOM SEND

- Open your data directory in your preferred DICOM editor
- Setup the location for PosdaLocal in your preferences
- Select a subset of files to Send
- Click Send

Example in HOROs on MAC OS with DicomTestSet data:

e 🚽 🛡			_							
					2		INTE ST		<b>1</b>	31
				1 1		19 7:86				T.
loud Dashboard Cloud Report Cloud Sharing Import E	xport Emai	Movie Expor	t Quer	Send And	onymize Burn	Meta-Data Delete	Viewers 2D Viewer ROI	s & Keys 4D	Viewer R	eport
atient name 🗸 🗸 🗸	Rep	Number	of filos:	1201				Modality	ID	
Rsna-2018-033		Number	or mes.	1201			ntrast	MR	0	Mri R
Rsna-2018-032		Select dest	ination:	YOUR LO	OCATION	۵		ОТ	Study	2/3/9
Rsna-2018-032				YOUR LO	OCATION			ОТ	Stu	Ct Ab
Rsna-2018-032								ОТ	Stu€	12/10
⊳Rsna-2018-031		Transfer	Syntax:	Implicit L	little Endian o	only	trast	СТ	0	
Rsna-2018-030	Nun	nber of assoc	iations:	Single as	ssociation	0	-Thigh	CT\PT	0	
⊳Rsna-2018-030						-	ontrast	MR	0	
Rsna-2018-030			Send:	O All im	ages		ontrast	СТ	0	
- Rsna-2018-028				Only	key images		ontrast	СТ	0	
Rsna-2018-027				Only :	Secondary Ca	aptures (SC)	Contrast	MR	0	
Rsna-2018-027			ROIs:	Includ	le ROIs, Com	ments, Reports,	ontrast	СТ	0	
Rsna-2018-026								ОТ	Stu	
- Rsna-2018-026						Cancel	Send	от	Stu3	
Rsna-2018-026								ОТ	Stu2	
- Rsna-2018-026		_	RSNA-	18-026			Unnamed	от	Stu9	
Rsna-2018-025		_	RSNA-	18-025		1572184414852	Ct Abd PelW- Contrast	СТ	0	
⊦Rsna-2018-024		_	RSNA-	18-024		1396157034784	Mr Abdomeo Contrast	MR	0	
- Rsna-2018-024		_	RSNA-	18-024	:	268638063518	Ct Abd PelW- Contrast	СТ	0	
- Rsna-2018-023		_	RSNA-	18-023			Unnamed	от	Stu(	
Rsna-2018-023		_	RSNA-	18-023			Unnamed	ОТ	Stu7	
⊳Rsna-2018-022		_	RSNA-	18-022		16113347327712	Ct Abdomeo Contrast	СТ	0	
Rsna-2018-021		_	RSNA-	18-021	1	9597981416496	Ct Abdomeo Contrast	СТ	0	
- Rsna-2018-020		_	RSNA-	18-020		2952106173864	Mri Abdomen	MR	0	
Rsna-2018-019		_	RSNA-	18-019			Unnamed	ОТ	Stu6	
Dena_2018_010			DONA-	18-010			linnamed	ОТ	Study	
		)							124	
3-Plane Abdomen View size: 256 x 256						A			кыпа-2	018-0,
WL: 6070 WW: 12141										
X: 0 px Y: 0 px Value: 0.00										
						COLUMN TWO IS NOT				
43 Images					-	and the second se				
THE REAL PROPERTY.										
1 4 4 5 2 h										
40 Images R										
		1 × 1	Children -		1					

# ACTIVITY BASED CURATION ACTIVITY BASED CURATION LAYOUT AND NAVIGATION

To open the Activity Based Curation Module, go to the Posda landing page, login, and then hit Launch on the right hand side of the Activity Based Curation row.

Posda.com			
	ActivityBasedCuration	Activity Based Curration Application	Launch

Let's go over the basics of the layout of the Activity Based Curation application.

## **Activity Based Curation**

Activity	
Workflow	Activities
Queries	No Activity Selected Filter:
Inbox	
Upload	Insert a new activity
	Short description
Download	
ShowBackground	Save
Verbose Activity Report	
GetVisualReviewByActivityId	
PublicCollectionCounts	

### QUICK NAVIGATION

On the left is the Quick Navigation menu. By default we launch into the "Activity" screen. Throughout the curation process we will navigate back and forth between the Activity section and other sections.

- Activity When no activity is selected, an insert or select option appears. Once an activity is chosen, this section shows a timeline of all the tasks completed on that activity.
- Workflow A menu to launch important queries and operations.
- Queries A screen to search and run data queries. <u>Click here for details.</u>
- Inbox Where notifications on completed tasks and their results are viewed.
- Upload Some steps will require downloading, editing, and re-uploading spreadsheets. This is where you will upload them and run the processes.
- Download Downloading directories. Advanced use only.
- ShowBackground Some steps will cause queries to run in the background. Here you can view information about those queries. A popup window will appear listing recent queries including those currently running. Buttons will allow you to view the logs and error output for any query.
- Quick Access Queries some popular queries are listed in the quick navigation menu as well for easy access
  - Verbose Activity Report
  - GetVisualReviewByActivityId
  - PublicCollectionCounts

#### PAGE TITLE AND PAGE CONTENT

The title of the current page is normally at the top and below it is the Page Content. The Quick Navigation Menu on the left allows you to change pages (see previous section). As you change pages throughout the Posda System the title and content displayed will change.

### ACTIVITY PAGE

The Activity Page allows you to select a current Activity, that can be used throughout the system. The first time you visit the page you can add a new activity, or select an existing one. Once an activity is selected the page will immediately change. Notice that the selected Activity is now displayed in the title. The "Choose Another Activity" button will take you back to the previous view, where you can select a different existing Activity, or create a new one.

Once an activity is selected this page shows the Activity Timeline for that activity. This timeline will show all the tasks that have been run on the selected activity. <u>More details</u> <u>about the Activity TImeline are discussed in the Using the Inbox section.</u>

Activ Is thi	Activit ity 1: Test Activity rd party: Yes O No	Choose	Another Activity	ura	atic	on (	Page 1: Tes	Title St Activ	has up /ity)	dated	
id	operation	start	duration	ol	tp id	cmp fr	to	tp files	view	user	command
			Page Co	onter	nt ha	is upo	lated				

#### USING QUERIES

This page allows you to run saved queries on the database. Most queries only display data. Additional buttons may appear dynamically to the left side of the screen when you have options to interact with the resulting data.



You can use the above tools at the top of the page to change your view.

- Recent: will have queries you used earlier this session and can be very useful for running things that you need often.
- Search: can be used to find a query by name, argument, returned item, and more.
- Workflow: combined with the drop down will bring up queries associated with each workflow step.

If you want to know more about searching for a query read the next section. Otherwise skip to <u>Running a Query</u>.

#### SEARCHING FOR A QUERY

• On the Queries page, select Search

#### Activity Based Curation (1: Path)

	Args containing:	Columns containing:	Query matching:	Name matching:	
✓ - recent o - search o - workflow					clear

- You have multiple fields for searching for your query.
  - Args containing search for a query that takes this value as input (arguments).
  - Columns containing search for a query that gives this value as output.
  - Query matching Advanced Use. Search for a query by the actual text within it.
  - Name Matching search for the query by name
- Enter your conditions in the boxes. This is case sensitive, and needs exact spelling.
- Hit Enter

◯ - recent <b>⊙</b> - search ◯ - workflow	Args containing:	Columns containing: Q	uery matching:	Name matching: EventsByMatchingName	
Searched queries (3 rows) (name like '%ImportEventsByMatchir	ngName%')				
name		params	columns returned		make query
ImportEventsByMatchingName		import_comment_like	import_event_id, impo num_images	ort_comment, import_time, import_close_time,	foreground background
ImportEventsByMatchingNameAndTyp	00	import_comment_like, import_type_like	import_event_id, import_event_id, import_event_images	ort_type, import_comment, import_time, 3	foreground background
ImportEventsByMatchingNameAndTyp	peWithCollectionSitePatientId	import_comment_like, import_type_like, from, to	import_event_id, impo duration, collection, si	ort_type, import_comment, import_time, ite, patient, dicom_file_type, num_images	foreground

- The queries that matched our search term are returned. At the top you can see the row count and the details of the search you ran.
- Clicking foreground will run the query immediately and display it on the screen. Sometimes the results will have buttons for further interaction.
- Clicking background will run the query behind the scenes and send a notification with the results as a report.

#### RUNNING A QUERY

• Find the query you want to run, and click **foreground**.



- The page updates to display:
  - The query name
  - A list of everything it will return
  - Input boxes for any parameters the query needs
  - The full text of the query.

query       Clear       Query name: ApilmportEvents         Columns returned: import_event_id, import_comment, import_time, import_close_time, num_images         Arguments:							
import_comment_like							
select							
<pre>import_event_id, import_comment, import_time, import_close_time, count(distinct file_id) as num_images</pre>							
import_event natural join file_import							
<pre>where import_comment like ? and import_type = 'posda-api import' group by import_event_id, import_comment, import_time, import_close_time</pre>							

- Enter your parameters into the input boxes. The wildcard '%' can be added to the end of a word to search for anything starting with that word. It can also be used alone to search for 'all', but use it carefully in large systems with a lot of data. Case matters.
- Click the blue **query button**. The results will be displayed.

Examples:

import_comment_like Path%	Returns data where the import comment starts with "Path"
import_comment_like %Path	Returns data where the import comment ends with "Path"
import_comment_like %	Returns all data! Be careful if used without other parameters

#### THE QUERY RESULTS SCREEN

Let's go over a few of the features of this page before we move on.

Activity Query Name and input Fi		Filter options	No filter currently defined		
Workflow	Full Results for: Impo	rtEventsByMatchingName(P%)	• Unfiltered rows: 2		
Queries	First row: 0 Show	v: 30 pg-dn pg-up	Paging options		
Inbox	import_event_id	import_comment	import_time	import_close_time	num_images
Upload	1	PATH518	2021-05-18 20:31:24.646227+00	2021-05-18 20:33:13.39193+	00 6
General Query Actions	41	PathologyMixed	2021-05-28 16:54:23.772849+00	2021-05-28 16:57:39.005398	+00 12
Unselect	Query Results				
Edit Filter					
Clear Filter					
Dismiss					
Refresh					
Rerun					
Download					
Chain					
Specific Query Actions					
Create Timepoint From DICOM Files in Import Event List					
Create Timepoint From All Files in Import Event List					
Download					

#### **Activity Based Curation**

- Query name and input At the top, we see which query the results below are for, including the parameters we chose. If your results are unexpected, look here to see if your parameters had a mistake.
- Resulting Data rows Skipping to the bottom for a moment, you can see the resulting data rows(Highlighted in the blue box in the image above). Depending on the query the returned columns and data will differ.
- Action buttons within the resulting data rows Some queries will have buttons in the data rows. These buttons typically allow you to view or act on a single point of the returned data.
- Filter Options- If you look to the right of the query name you see a set of radio buttons. Unfiltered Rows: # and Filtered Rows: #. By default you are likely on the Unfiltered rows. Sometimes the parameters for the query did not narrow down your results to what you needed. For instance, in the above query, if someone else also did an import today and named it "PediatricCollection" my query would have returned that data too! To narrow down my results even more I can click the Edit Filter button on the left and create a filter. After inputing the parameters and clicking "set filter" the page will update. The radio buttons for Unfiltered and Filtered rows will now show differing numbers. Now, I can click the Filtered rows button to view and use only the results that meet my criteria. If you change your mind, you can Clear the feature with the button on the left.

• Paging Options - The First row currently shown on this page is listed, along with the number of rows per page. These can be edited to move quickly through the pages. There are also Pg-up and Pg-dn buttons for navigating between multiple pages of result data.

Query General Actions - these buttons appear on the left whenever you are in the queries screen.

- Unselect button This will "unselect" the current query, and return us to the first Query section page to choose a different query.
- Edit Filter Set a filter. See Filter Options above for more details.
- Clear Filter Clear the filter. See Filter Options above for more details.
- Dismiss button This will dismiss the current query, and return us to the search query page to choose a different query. This will also remove the query from your "queries used this session" history.
- Refresh button- This will refresh our results. If you performed tasks in a different section, someone else changed the data, or something outside the program changed the data between now and when you last ran this query you can use this button to refresh the data and see the most up to date results without having to go back to any earlier pages or reenter any parameters.
- Re-run button This lets you re-run the same query with new parameters.
- Download buttons Download the filtered or unfiltered result sets as a csv file.

Specific Query Actions - Some queries have unique actions listed here.

• Action buttons on the left - Buttons here will perform operations using the entire result set as input. (The filtered result set if filtered rows is selected). Different queries will have different actions available. For example, the pictured query has two buttons. Each opens a popup to begin a commonly used process, with the input set to the results of your query.

#### Example:

Activity				No filter cur
Workflow	Full Results for	ImportEventsByMatchingName(P%)	O Unfiltered rows: 2 ○ Filtered rows: 2	
Queries	First row: 0	Show: 30 pg-dn pg-up		
Inbox	import_event_id	import_comment	import_time	i
Upload	1	PATH518	2021-05-18 20:31:24.646227+00	2
	41	PathologyMixed	2021-05-28 16:54:23.772849+00	2
Unselect		-		
Information fr	om •••		Process Operation Popup	
		ocalhost:64616/465485435379975/Refre	esh?obj_path=Dblf/TableLevelPopup_2	110%
Clear Filter	Posda.c	om Process Operation Popup		Signed in
Dismiss				
Refresh			Start	
Rerun			Cancel	
Download			Help	
Create Timepoint From DICOM Files in Import Event List Create Timepoint From All Files in Import Event List Click the button to open popu	Parameters • activ	s: ity_id :		
Download	<n< td=""><td>one&gt;</td><td></td><td></td></n<>	one>		
ShowBackground	• com	ment :		
Verbase Activity Papart	• notif	v:		
	ad	min		
PublicCollectionCounts	Expanded	Command:		
	CreateAd	ctivityTimepointFromImportEvent	IdAll.pl bkgrnd_id? <none> "" ad</none>	min
	2 lines to s	upply as input:		
	1 41			

#### ACTIVITY DASED CUTATION

#### CREATING AN ACTIVITY AND ASSOCIATING THE DATA



The initial steps of curation are to import the data, create an activity, and then tie them together.

First, make sure the data has been properly imported (See the importing data section). You will want to remember some information that identifies this import from previous ones such as the import comment, or the exact time of the import.

Next, we will create the Activity that we can associate with our imported data. Open Posda, and navigate to the Activity Module.

As soon as you launch the module it should place you in the Activity section. (If not click Activity on the lefthand Quick Navigation menu)

Begin by choosing a meaningful "name" for your new Activity.

- Create a New Activity by entering the name in the input text box
- Then Click Save

#### Insert a new activity



Nothing changed?

Actually, your new activity was created and added to the list! However, the program does not automatically select the activity. If you want to start working on the Activity now, you need to select it from the list.

Just above the Insert section, you should see an Activities drop down. Open this and choose your activity. The page will automatically change. (If you have several activities you can use the filter on the right to filter your activity list)

## Activities



The page updates to the timeline view. Since no tasks have been completed on this activity yet, the timeline is blank.

	Activity Based Curation (1: Test Activity)										
Activ Is thi	ity 1: Test Activity rd party: Yes 🔿 No 💽	Choose	Another Activity								
id	operation	start	duration	ol	tp	cmp		tp	view	user	command
					id	fr	to	files			

#### ASSOCIATE IMPORTED DATA WITH AN ACTIVITY TIMEPOINT

#### • In the Quick Navigation menu select Workflow.

Create and Manage Activity Timepoints
Patient Mapping
Initial Anonymization
Run Count Checks
Check for Duplicate SOPs
Run Consistency Check
Verify DICOM IOD (Dciodvfy)
Visual Review
PHI Review
Check Struct Linkage
Link RT Data
Send to Server (NBIA)
Compare data to Server (NBIA)
Produce Activity Report
Copy or Consolidate Timepoints
Other

- Click the top menu item. Create and Manage Activity Timepoints.
- Click Suggested Queries for Series or Suggested Queries for Import Events.

Create and Manage Activity Timepoints
Curation workflow tasks performed on a collection are grouped together into an Activity Timepoint to allow for better management and analysis. This step is tying the data to the Activity.
Note: You must first Import Data into Posda and create an Activity!
Possible Operations:
• Find files In Timepoint Not In Public
Possible Queries:
Suggested Queries for Creating a Timepoint from Import Event IDs
Suggested Queries for Creating a Timepoint from Date Range for CTP data

- Suggested Queries for Creating a Timepoint from Import Events: This will give you data that can be grouped by import events. This is the most frequently used import query.
- Suggested Queries for Creating a Timepoint from Date Range for CTP data: This is useful for certain data that comes in as small chunks over long time periods.

The Queries page should now be visible.

By accessing this page from the Workflow, you are automatically set to see the queries associated with this workflow item.



Notice that workflow is selected along with the workflow step in the dropdown.

To learn more read the section on basic usage of the <u>Queries</u> page. The pages below will detail the use of the **ImportEventsByMatchingName** query. However the steps for other queries will be similar.
### IMPORT EVENTS BY MATCHING NAME

This query can be used to single out find the data we imported earlier.

If you are working in a production system, you are likely looking for a *specific import* out of *many* imports into the system. You'll need to know either the import\_comment or import\_id for your specific import. If you used the Posda Import GUI tool, the import\_comment is the name you gave the import.

• You will see a place to enter your parameter. Enter the import\_comment you chose earlier. You can also use the wildcard character '%' at the end to match anything starts with the text entered. For example the search below for "Path%" will find the import comment "PathTest". (A wildcard alone will match all results, which may be slow in a large system).

Query name: ImportEventsByMatchingName         Query clear         Columns returned: import_event_id, import_comment, import_time, import_close_time, num_images
Arguments:
import_comment_like Path%
<pre>select   import_event_id, import_comment, import_time, import_close_time, count(distinct file_id) as   num_images</pre>
<pre>from   import_event natural join file_import   where     import comment like ?</pre>
<pre>group by import_event_id, import_comment, import_time, import_close_time</pre>

• Click the blue **query button**. The query will run and then the results will be displayed.

<i>Full Results</i> <i>for:</i> ImportEvent	sByMatchingName(	Path%)	<ul> <li>Unfiltered rows: 1</li> </ul>	<ul> <li>Filtered</li> <li>rows: 1</li> </ul>	No filter currently define		
First row: 0	Show: 30	pg-dn	pg-up				
import_event_id	import_comment	import_ti	me	import_cle	ose_time	num_images	
126	PathTest	2021-03-2 19:49:06.8	24 874465+00	2021-03-2 19:51:53.6	4 29413+00	6	

## CREATE TIMEPOINT FROM ALL FILES IN IMPORT EVENT LIST

Now that we have our results, and have filtered them if necessary, we can see the exact set of imported files we want to associate with an activity.

- Along with the query results, a new set of buttons has appeared on the left, under the quick navigation menu (If the button is not visible, refresh the page. A quick refresh option is to click pg-up)
- Click the Create Timepoint From All Files in Import Event List

Create Timepoint From All Files in Import Event List

Create Timepoint From DICOM Files in Import Event List

- A new window will appear. Click over to that window if it does not automatically take focus. The Process popup shows us details on the process we are about to run including:
  - Start button Start the process.
  - Cancel button Cancel the process entirely.
  - Help button get more information. Operation The name of operation we are performing.
  - Command Exactly which perl file is being called and what parameters it expects.
  - Parameters Our input. Defaults are sometimes filled in automatically.
  - Expanded Command. Similar to command above, but with the parameters filled in to match the input provided.

- *#* lines to supply as input More inputs. These are the datapoints from the query we ran that are being passed to this operation. This process requires parameters to be specified. Default options are already filled in.
- In the Comment section add a meaningful comment, such as why you need to run this step.
- If necessary, change which user should be notified when the process completes. This should default to your account, which is normally correct.

Posda.com	Process Operation Popup	Signed in as admin Close
		Start
		Cancel
		Help
Operation: Create Command: Create Input line format:	ActivityTimepointFromImportEventl ActivityTimepointFromImportEvent <import_event_id></import_event_id>	d IdAll.pl bkgrnd_id? <activity_id> "<comment>" <notify></notify></comment></activity_id>
Parameters: • activity_id :		
3		
• comment :		
import of	documentation test data	
• notify : admin		
Expanded Comma	and:	
CreateActivit	yTimepointFromImportEventIdA	<pre>ll.pl <?bkgrnd_id?> 3 "import of documentation test data" admi</pre>
1 lines to supply a	is input:	
126		

- Click Start.
- The popup page will refresh with its current status
- Once the process popup says something like "Successfully queued background process" or "response\_complete" you may close the popup window.
- You should receive a notification, continue following along with the next section.

# USING THE INBOX

When you return to the main window, it is likely that your Inbox in the Quick Navigation menu will be red. This means you have a new notification, likely a message that a process has completed. If you are ever running a process in the background and it has not completed by the time you expected it to complete, click ShowBackground to view the status of all processes currently running in the background. For example, this may be needed if you realize there is an error in the parameters that has caused the process to get stuck.

Activity
Workflow
Queries
Inbox (1)
Upload

• Click Inbox

#### Message Inbox for admin

Message ID	Operation	Status	Created Date
1	CreateActivityTimepointFromSeriesList	entered	2019-12-05 20:05:39.159284

Here you can see all the messages you have received but not yet dismissed.

• Click on the ID button for the newest message pertaining to CreateActivityTimepointFromImportEventId in order to open the notification.

Status	Date Entered		Date Dismissed						
read	2021-04-28 14:11:29.695968								
Background pro Creating time Activity Time Created tables Preparing repo Background pro Total time ela Report 'Conde Report 'Timepo	Background process /home/posda/posdatools/Posda/bin/CreateActivityTimepointFromImportEventIdAll.pl Creating timepoint from named import for 6:at Wed Apr 28 09:11:24 CDT 2021 Activity Timepoint Id: 7 Created tables in 0 seconds. Preparing reports. Background process ended at: 2021-04-28T14:11:24 Total time elapsed: 0 Report 'Condensed Timepoint Report': http://localhost/papi/file/39/9e8113ac-a82b-11eb-83d8-3a09999 Report 'Timepoint Report': http://localhost/papi/file/40/9f9f709e-a82b-11eb-83d8-3a099980257								
Dismiss this mes	sage Forward this message								
This message h	This message has already been filed.								
Recent operations on this message									
What	When	Who	How						

What	When	Who	How
message read	2021-04-29 18:12:26.02339	admin	Posda::Inbox
entered	2021-04-28 14:11:29.707688	nobody	Posda::BackgroundProcess

This notification display tell you when the notice arrive, when it was last read, and the results of the process that sent the message, including information on whether the process succeeded or failed.

A successful process response may include downloadable data spreadsheets for edit, or a report with extra details about the process and its effects on the data. A failed process response may have a useful error message that can help you find the problem.

There are two or three Action buttons available.

- Dismiss this message remove the notice, disregard it
- File this message button Filed messages are archived and those relating to an activity can be viewed from the Activity timeline. Activity related notifications will often file themselves, so instead of the button you will see the green notice shown in the image above.
  - If you are filing the message manually, make sure you have the correct activity selected, so that the message appears on the proper activity timeline.
- Forward this message send a copy of this message to another user.

- If you see the green "This message has already been filed." notice, click on the Dismiss button. Otherwise click "File this message" and then click yes on the confirmation.
- Return to Activity with the Quick Navigation menu on the left.

The timeline now has its first entry. We can see which process was run, which user ran it, when it was run, how many files it affected, and we can view the filed notification. If you click email, a popup appears with the filed notification results, and options to download the message.

Activity 4: PathologyTest Is third party: Yes O No O	• 34: Creat	eActivityTimep	ointFr	romIm	oortEven	tld - Created activity_tim	epoint 5	with 6 files dismiss
id operation	start	duration o	ol tp	cmp	tp	view	user	command
			id	fr t	o files			
34 CreateActivityTimepointFromImportEventId	2021-04-27 17:03:32.48	00:00:0.10	) 5	0 (	6	email resp input	admin	CreateActivityTimepointFromImportEventIdAll.pl 4 "A sample pathology collection" admin

You can see in the center of the table the Activity Timepoint ID and the number of files in the Timepoint (In the above example it is Activity Timepoint 5 and there are 6 files). Any task that edits or removes data will create a new timepoint. Most processes will use the most recent timepoint automatically. However when troubleshooting you may want to examine older timepoints, so it is helpful to remember you can see the Activity Timepoint IDs here. You can verify that an edit task behaved properly by examining the file count and checking that it matches your expectations. You can also select 2 timepoints with the small radio buttons and hit the "cmp" button to launch the Compare Timepoints operation.

## PATIENT MAPPING

This step can be skipped if the data is already mapped, such as when it was sent from CTP (A tool some sites use that sends and partially de-identifies data).

When removing PHI many data-points are mapped rather than deleted. For example, while we do not want to keep the patient's name or hospital ID, we need to keep the fact that three different image sets refer to this singular patient. This means a patient mapping system will be chosen by the curator and PI/site team to map every patient to a new ID and to shift the identifying dates to new dates with the same time gaps between them.

Once created and properly formatted the patient mapping can be loaded into the Posda Application.

The following fields are required in the Import Patient Mapping CSV spreadsheet.

- from\_patient\_id: This is the value of the patient id in the original image header. This must be entered **EXACTLY** as it is in the image header (upper and lower case characters matter). This can be upper or lower case alphanumeric values.
- to\_patient\_id: This is the new patient id after the mapping is completed. This can be upper or lower case alphanumeric values.
- to\_patient\_name: Usually when you de-identify data, you use the same value for the patient name as is used for the to\_patient\_id.
- collection\_name: This is the collection of data that the data will be assigned to. If the data is being submitted to TCIA, TCIA staff will work with the site to identify the correct Collection name. There is a corresponding collection\_code that gets assigned behind the scene for data submitted to TCIA. The process for adding these codes to Posda are covered in the <u>Collection Preparation</u> section.
- site\_name: This is the site name submitting the data. If the data is being submitted to TCIA, TCIA staff will work with the site to identify an appropriate site name (often a shortened name). There is a corresponding site\_code that gets assigned behind the scene for data submitted to TCIA. The process for adding these codes to Posda are covered in the <u>Collection Preparation</u> section.
- diagnosis\_date: yyyy-mm-dd format. If your data is being de-identified using a baseline date, then this value is required. Date and Date-Time fields in the DICOM image headers

will be de-identified by normalizing them to a baseline date (see below) and then shifting them by the number of days between the original Study Date and the Date of Diagnosis. If your data is not be de-identified using a baseline date, then this field should be empty or contain the value "<undef>",which means undefined.

- baseline\_date: yyyy-mm-dd format. If your data is being de-identified using a baseline date, then this date is required. It represents the date by which all date and date-time fields in the image header will be shifted. If the data is being submitted to TCIA, TCIA staff will work with you to identify the baseline date to use if requested. If your data is not be de-identified using a baseline date, then this field empty or contain the value "<undef>",which means undefined.
- uid\_root: This is the DICOM Unique Identifier root that all the DICOM UID fields will be hashed to so that the original DICOM UID values are not visible in the de-identified image header. If the data is being submitted to TCIA, TCIA staff will provide the site with the correct UID root to use.
- batch\_number: This field is reserved for future use and should contain <undef>
- date\_shift: "-####" days format. If your data is not being de-identified using a baseline date, then this value should contain the number of days that the date and date-time fields will be shifted. The "-" identifies that the dates are shifted back in time, the # identifies the number of days. If the data is being submitted to TCIA, TCIA staff will work with the site to identify the correct date\_shift value to use. If your data is being de-identified using a baseline date, then this field should contain <undef>
- Operation: This field must contain the value ImportPatientMapping exactly as it is spelled here (with the same upper and lower case characters)
- notify: The username of the Posda user who should be notified when the operation completes. This is usually the same Posda user submitting the file.
- description: A short explanation of what you are doing such as "Import Patient Mapping for collection and site"

Example files:

sample using diagnosis and baseline dates

from_patient_id	to_patient_id	to_patient_name	collection_name	site_name	diagnosis_date	baseline_date	uid_root	batch_number	date_shift	Operation	notify	description
123456789	ACNS123_P01	ACNS123_P01	ACNS123	UAMS	<1/23/2010>	<1/1/1950>	1.3.6.1.4.1.14519.5.2.1	<undef></undef>	<undef></undef>	ImportPatientMapping	admin	Import patient mapping table for ACNS123
234567890	ACNS123_P02	ACNS123_P02	ACNS123	UAMS	<6/25/2010>	<1/1/1950>	1.3.6.1.4.1.14519.5.2.1	<undef></undef>	<undef></undef>			

sample using date shift

from_patient_id	to_patient_name	to_patient_id	collection_name	site_name	diagnosis_date	baseline_date	uid_root	batch_number	date_shift	Operation	notify	description
DBT-P00015	Breast-Cancer-Screening-DBT-P00015	Breast-Cancer-Screening-DBT-P00015	Breast-Cancer-Screening-DBT	Duke	<undef></undef>	<undef></undef>	1.3.6.1.4.1.14519.5.2.1	<undef></undef>	<-2000 days>	ImportPatier	admin	Import patient mapping table for Duke Breast DBT
DBT-P00016	Breast-Cancer-Screening-DBT-P00016	Breast-Cancer-Screening-DBT-P00016	Breast-Cancer-Screening-DBT	Duke	<undef></undef>	<undef></undef>	1.3.6.1.4.1.14519.5.2.1	<undef></undef>	<-2000 days>			

### IMPORT PATIENT MAPPING

You should have an activity tied to your imported collection at this point. We will next perform an operation that will ask us to upload a spreadsheet. The spreadsheet must be in the correct format to be used.

- Login and launch the Activity Based Curation Module , if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title)
- Click Workflow.
- Click Create Patient Mappings for Timepoint.
- Click the Import Patient Mapping button.

The process popup appears. Unlike other process in the system, this popup asks for a formatted spreadsheet rather than only text parameters.

Start
Cancel
Help
Operation: ImportPatientMapping Command: ImportPatientMapping.pl bkgrnd_id? <notify> Input line format: <from_patient_id>&amp;<to_patient_id>&amp;<collection_name>&amp;<site_name>&amp;<batch_number>&amp; <date_shift>&amp;<diagnosis_date>&amp;<baseline_date>&amp;<uid_root></uid_root></baseline_date></diagnosis_date></date_shift></batch_number></site_name></collection_name></to_patient_id></from_patient_id></notify>
Parameters: • notify : admin
Expanded Command: ImportPatientMapping.pl bkgrnd_id? admin

Upload a spreadsheet matching the above input format.

Drop files here to upload No files have been uploaded Working with the coordinating site, you should be able to determine the patient mapping choices such as how to mask the patient's name and how to shift the dates. Once you have this information, enter it into a csv spreadsheet in the format described in the previous section. The Operation Column should be "ImportPatientMapping".

- Upload the Spreadsheet into the popup window. You can drag the file into the box or click the box to open a file chooser window.
- The page will refresh and a new Chain button will appear. (If the page does not refresh automatically, you can open and close the help dialog to force a page refresh)

This Process Requires a Spreadsheet									
Jpload a spreadsheet matching the above input format.									
		Drop fi	iles here to upload						
Files Uploaded									
file_name	size	mime_type	description Dismiss	file_id					
DicomTestPatientMap.csv	471	text/csv chain	UTF-8 Unicode (with BOM) text, with C	694					

• Click the chain button. The page will refresh(It may take a moment). Instead of an upload box the page should now list the input lines the application read from your spreadsheet.

Operation: ImportPatientMapping Command: ImportPatientMapping.pl bkgrnd_id? <notify> Input line format: <from_patient_id>&amp;<to_patient_id>&amp;<to_patient_name>&amp;<collection_name>&amp;<site_name>&amp;<batch_number> <date_shift>&amp;<diagnosis_date>&amp;<baseline_date>&amp;<uid_root></uid_root></baseline_date></diagnosis_date></date_shift></batch_number></site_name></collection_name></to_patient_name></to_patient_id></from_patient_id></notify>
Parameters:
notify :
admin
Expanded Command:
ImportPatientMapping.pl bkgrnd_id? admin
2 lines to supply as input:
&Pat_938&Pat_938&NewColl&UAMS&&&&<1960-01-01>&1.3.6.1.4.1.14519.5.2.1.999.1001 &Pat_522&Pat_522&NewColl&UAMS&&&&<1960-01-01>&1.3.6.1.4.1.14519.5.2.1.999.1001

• Start the process.

- Once the screen updates to "Successfully queued background process" it is safe to close the popup.
- Return to the main window.

When the process is complete, a notification should appear in the Inbox, and Inbox will be red (If you are unfamiliar with Posda notifications see the Using the Inbox section of this guide).

- Click Inbox.
- Click the ID of the notification for SuggestPatientMapping to open the notification.
- Click Dismiss (Be sure you have the correct activity selected).
- Click Yes on the confirmation.

At this point the import is complete. If you would like to verify the import's success you can run the GetPatientMappingByCollection query.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title)
- Click Queries.
- Click Search.
- In the "Name matching" box enter GetPatientMappingByCollection.
- Click foreground.
- Enter in your Collection name exactly as it appeared in the uploaded spreadsheet.
- You should see the resulting data, verifying that your collection was mapped successfully.

# INITIAL ANONYMIZATION

Once the mapping is in place, we can run the initial anonymization step. Again, this step can often be skipped if the data is already initially de-identified, such as when it was sent from CTP. (CTP is a tool some sites use that sends and partially de-identifies data. CTP stands for Clinical Trial Processor and comes from The Radiological Society of North America(RSNA) Medical Imaging Resource Center(MIRC) )

- Click Workflow on the Quick Navigation menu on the left.
- Click Initial Anonymization.
- Click Produce Initial Anonymizer for Timepoint
- Input the collection and site. These need to be the *exact* collection name and *exact* site name.
- Click Start.
- Close the popup when the message appears.
- View and dismiss the Inbox notification.
- Click Activity to view the timeline.
- Click Email.
- Click on the hyperlink within the message to download the spreadsheet you created with the subprocess.
- Save the file when prompted.
- Navigate to the file on your machine(default locations are often Documents or Downloads).
- Open the file with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).

series_instance_uid	ор	tag	val1	val2	Operation	activity_id	edit_description	notify
2.25.26694550284351243382345180032248872228.91.200					BackgroundEditTp	1	Initial Anonymzer Edits	admin
2.25.26694550284351243382345180032248872228.361.300								
2.25.26694550284351243382345180032248872228.2.100								
	set_tag	<(0013,"CTP",10)>	<newcollection1></newcollection1>	$\Leftrightarrow$				
	set_tag	<(0013,"CTP",11)>	<newcollection1></newcollection1>	<>				
	set_tag	<(0013, "CTP", 12)>	<uams></uams>	<>				
	set_tag	<(0013, "CTP", 13)>	<9991001>	<>				
	set_tag	<patientid></patientid>	<pat_180></pat_180>	$\Leftrightarrow$				
	set_tag	<patientname></patientname>	<pat_180></pat_180>	$\Leftrightarrow$				

#### downloaded\_file\_10(1)

This report shows the changes your mapping will have on the data. Use this report to verify that your mapping will be applied correctly.

- Carefully review the spreadsheet. If the changes seem incorrect you may want to re-run the Patient Mapping steps.
- Save the sheet as a .csv file. Close the file.

2	Operation	activity_id	edit_description	notify
	BackgroundEditTp	1	Initial Anonymzer Edits	admin

- Return to your Posda window.
- Click Upload in the Quick Navigation Menu on the left.
- Drag your saved file into the box, or click on the box to bring up a file selection window and choose your file. Wait for the upload to complete.
- Click chain in the file list once it appears.
- Review the Process Popup . Make sure the input lines and parameters match the expected operation.
- Click Start.
- Click Activity in the Quick Navigation Menu on the left to return to the timeline.
- Look at the status updates on the top right. Note the process updates its status text as it runs. The Initial anonymization process will edit every file in the dataset.



• Wait for the process to complete. If you have not already, now may be a good time to dismiss older status messages for completed tasks that you no longer need to see. Note: This only dismisses the message text, it will not cancel any processes so dismissing too many is only a minor inconvenience, you will still receive a notice in the inbox on final completion.



- When the process finishes, you will receive an inbox notification. View and dismiss the notice.
- Return to Activity, click email to open the filed notice.



- At the bottom of the notice are 2 reports. Download and open these in your preferred spreadsheet program. Be careful to look out for formatting issues, you may need to expand the columns or change other settings. Some users have found that they have different preferred programs for different reports(Excel, Numbers, Libre Office).
- Review the reports. These summarize the edits that were performed on the data based on your input. Close the reports after review.
- If you are <u>not</u> satisfied with the Edits, Click "Reject Edits and Delete Temporary Files". You may need to go back to previous steps to get the data ready for acceptance.
- If you <u>are</u> satisfied with the Edits, Click "Accept Edits, Import and Delete Temporary Files".



- A new popup appears. Review the parameters.
- Click Start.
- When it says "Entering Background" it is safe to close both popups and return to the main window.
- On the main window click Activity. Notice the Status message is updating. Wait for it to finish. You will receive a notification.
- View and dismiss the Inbox message.
- You may view the report that was produced by clicking email in the timeline, and clicking the hyperlink.
- You can also review the Activity Timeline and check that number of files in the latest activity timepoints is as expected after the change. See <u>Activity Timeline details</u>.

### **RUN COUNT CHECKS**

This is an optional step to verify the number of files sent so that we can confirm with the sending site that everything arrived as expected.

- Login and launch the Activity Based Curation module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title)
- Click Workflow in the Quick Navigation Menu on the left.
- Select "Run Count Checks"
- Click "Suggested Queries for Count Checks"
- Choose the query that's name looks most relevant to your data. Click foreground.
- Input the parameters and click query.
- Review the results and confirm with the sending site that all file counts make sense. If files, series, or patients are missing from the expectation, the site may need to resend the data.

See the example results on the next page.

For example, I have run CountsByCollection with the input "NewCollection1"

My results show 2 patients, each with 3 series. One patient has 3 series of OT modality. The other has 3 series of CT modality. I can hit view on the series uid column to see the pixel data for the series.

If I were to contact the site and the sender agreed this was the expected number of series I could move on. If instead the sender said there should have been a third patient with 3 more series, I would know that data was missing and steps would need to be retaken.

Results fo	countsByCollection(N	ewCollectic	on1) 👩 Unfil	tered rows: 6	No	filter currently defined		
First row: 0	Show: 30	pg-dn	pg-up					
patient_id	image_type	modality	study_date	study_description	series_description	n study_instance_uid		series_instance_uid
Pat_180	DERIVED\SECONDARY	ОТ	2018-10-26			1.3.6.1.4.1.14519.5.2.1.999.1001	.2098227556369894215385542055890	1.3.6.1.4.1.14519.5.2.1.999.100 view
Pat_180	DERIVED\SECONDARY	ОТ	2018-10-26			1.3.6.1.4.1.14519.5.2.1.999.1001	1.2459310009885137221581717153216	1.3.6.1.4.1.14519.5.2.1.999.100 view
Pat_180	DERIVED\SECONDARY	ОТ	2018-10-26			1.3.6.1.4.1.14519.5.2.1.999.1001	1.8450695175676197734378786085699	1.3.6.1.4.1.14519.5.2.1.999.100 view
Pat_181	ORIGINAL\PRIMARY \LOCALIZER	СТ	1992-05-04	CT CHEST W/ CONTRAST	SCOUTS	1.3.6.1.4.1.14519.5.2.1.999.1001	1.1721577734805282481271389278283	1.3.6.1.4.1.14519.5.2.1.999.100 view
Pat_181	ORIGINAL\PRIMARY \AXIAL	СТ	1992-05-04	CT CHEST W/ CONTRAST	NON-CONTRAST	1.3.6.1.4.1.14519.5.2.1.999.1001	1.1721577734805282481271389278283	1.3.6.1.4.1.14519.5.2.1.999.100 view
Pat_181	ORIGINAL\SECONDARY \AXIAL	СТ	1992-05-04	CT CHEST W/ CONTRAST	AXIAL CAP/KIDNEYS	1.3.6.1.4.1.14519.5.2.1.999.1001	1.1721577734805282481271389278283	1.3.6.1.4.1.14519.5.2.1.999.100 view

# CHECK FOR DUPLICATE SOPS

SOPs (Service Object Pairs), are a type of identifier in DICOM. If you have a duplicate SOP that means you have a duplicate image. This process builds a report to alert you to duplicated data or data where multiple entities are using the same identifiers. If duplicates are found, you may need to work with the site sending the data to delete unnecessary or broken files or to choose which file to keep from a duplicate pair.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title)
- Click Workflow in the Quick Navigation Menu on the left.
- Click Check for Duplicate SOPs.
- Click Compare Duplicate Sops in Timepoint.
- Switch focus to the Process Popup window if it does not switch automatically.
- Change the default parameters if necessary (Your Activity, and who should be notified).
- Click Start.
- It is safe to close the popup once it says the background worker has been queued.
- View and dismiss the inbox message. (See using the Inbox)
- Click Activity in the Quick Navigation menu to view the timeline.
- Click email to view the notification.
- Click the hyperlink labeled Difference Report.

```
Background process /home/posda/posdatools/Posda/bin/CompareDupSopsInTp.pl begun at 2021-09-21T19:3
Background process ended at: 2021-09-21T19:30:09
Total time elapsed: 1
Report 'DifferenceReport': http://localhost/papi/file/203/561046f6-1b12-11ec-a383-fc42b35093d2
```

- Save the file when prompted.
- Navigate to the file on your machine(default locations are often Documents or Downloads).

• Open the file with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).



- This report shows you how many SOPs are duplicated and what differences appear in the different copies. In the example, the age differs, some dates are changed, and some information was abbreviated. This could happen if the sending site sends a file, then resends it after doing some attempts at de-identification on their end. Now that you know what differs between these files of the same patient, you can better choose which to keep and which to remove from your dataset. Some choices may require asking for more information from the sending site. (Another useful query report is **DupSopsInTimepointWithSeriesFileIdAndLoadTimes** which is also linked in the workflow step).
- Using these reports you can analyze duplicate data or data where multiple entities are using the same identifiers, and then choose which files to remove or replace. This may require communicating with the sending site. A commonly used way to remove the unneeded files is with the **CopyPriorTimepointExcludingFiles** function. Using the reports you have analyzed, you can list the FileIds to be removed in a properly formatted csv file, and use the operation to copy all the files except the duplicates into a new timepoint. (See the <u>Extras</u> section for more information).
- If there are inconsistencies other than older versions of files that can be removed, you may also need to use **BackgroundEditTp** to edit the data. More details on this function are covered in the <u>PHI Review Section</u>.

# RUN CONSISTENCY CHECK

This steps verifies that the DICOM metadata is consistent throughout the series and studies. For example, the "Series Description" should be the same for all images in the same series.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title)
- Click Workflow.
- Click Run Consistency Check.
- Click the Check Consistency button.
- Switch focus to the new window if it does not take focus automatically.
- Edit the parameters if necessary.
- Click Start.
- Once it say "Going to the background" close the popup window.
- View and dismiss the inbox notification.
- Return to Activity, and click email on the current task.
- A report is generated if there are inconsistencies. If any inconsistencies are found, edits will be needed to correct them. Work with the sending site to determine where the error originates and work back through the previous steps as necessary.
- Editing files requires using the **BackgroundEditTp** process. More details on this function are covered in the <u>PHI Review Section</u>.

# VERIFY DICOM IOD (DCIODVFY)

This step verifies the DICOM files meet the DICOM standard.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click Verify DICOM IOD (Dciodvfy)
- Click the Dciodvfy button.
- Switch focus to the new window if it does not take focus automatically.
- Fill in the type parameter.
  - type Type of scan:
    - "one\_per\_series" scan one file in series (most common choice)
    - "all\_per\_series" scan all files in series
    - "per\_sop" sops are SOP instances one file per SOP
- Click Start.
- Once the message confirms the worker node is queued, close the popup window.
- View and dismiss the inbox notification.
- Click Activity in the Quick Navigation menu on the left.
- On the Activity timeline, click email.
- Click the hyperlink and download the report. This report may require you to expand the columns to view.
- If problems are found work with the site to edit the data as necessary.

### VISUAL REVIEW

In this step the pixel data will be examined to ensure no PHI is contained. To do this we will schedule and manage the review in the Activity module, and use another Posda tool called Kaleidoscope to do the actual review.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click Visual Review.
- Click the Schedule Visual Review button.
- Switch focus to the new window if it does not take focus automatically.
- Click Start.
- Once the message confirms the worker node is queued, close the popup window.
- View and dismiss the inbox notification.
- Click Activity in the Quick Navigation menu on the left.
- Notice on the upper right that the status message for VisualReview from Timepoint is updating.



- Wait until the status text says "ReadyToReview 100%". If there are any errors, these will need to be addressed before moving on.
- Click Workflow.
- Click Visual Review.
- Click the Suggested Queries for Visual Review button.
- Click foreground next to VisualReviewScanInstances.

- No parameters are needed. Click query.
- Your scheduled review should appear in the results.

#### Activity Based Curation (1: Test Activity)

Results for: VisualReviewScanInstances() OII filtered rows: 1 OFiltered rows: 1				No filter currently defined			
First row: 0	Show: 30 pg-dn	pg-up					
chain	id reason	who_by	num_series	when_scheduled	num_done	num_equiv	sched_finish_time
Details	1 Activity Id: 1	admin	6	2019-12-12 19:43:55.138747	6	6	2019-12-12 19:44:07.527009

- Click the Details button where it appears in the chain column. This is "chaining" these results into the another query.
- Click query, this will run VisualReviewStatusById with the ID from the Visual Review we were just reviewing. The results appear separated by DICOM modality.

<b>Results for:</b> Vis	sualReviewStatusById(1)	Unfiltered rows: 2	No filter currently d	efined		
First row: 0	Show: 30 pg-dr	n pg-up				
chain	id	processing_status	review_status	dicom_file_type	num_equiv	num_series
Details	1 review	ReadyToReview		CT Image Storage	4	3
Details	1 review	ReadyToReview		Secondary Capture Image Storage	3	3

In this example we see that there are 3 series of CT data, and 3 series of Secondary Capture data. These have been divided into 4 and 3 Image Equivalency Classes respectively. IECs are a subset of a series intended to separate out scouts and other set of images in different planes.

• Click Review on one of the data sets to launch Kaleidoscope.

chain	id	processing_status	r
Details	1 review	ReadyToReview	
Details	1 review	ReadyToReview	

### USING KALEIDOSCOPE AND QUINCE

This is Kaleidoscope. In the screenshot above, you can see that IEC 1 has 44 images, of Patient 007, examining the kidney.

The 3 images are projections created by stacking those 44 frames and looking at the brightest/average/and darkest pixel which allow you to search for PHI across many frames at once. This is a major improvement upon the alternative process, of viewing every frame individually.

Above the projections are the Review (Green, Red, Yellow) and Navigation (Blue) buttons.

To see the whole IEC, one frame at a time, Click the **Open in Quince** button on the IEC row. To see the whole series, one frame at a time, click the corresponding button on the Series Instance UID row. This opens the lightweight Quince DICOM Viewer.



You can play the frames as a video, scroll manually, adjust the window and level manually or with presets, and more. Anytime Kaleidoscope does not provide enough info to check for PHI launch Quince to review further.

- Review the open projection, open Quince if you have need for further detail.
- If there is no PHI in the pixel data, click the green Good button.
- If there is PHI visible, click the red Bad button.

- The yellow buttons are optional ( teams may use these to mark as "one review completed out of two", or a "go back and review a second time").
- This automatically moves you forward in the review set. Review the next IEC.
- Review all the IECs in this set. When no more IECs display, close the Kaleidoscope window.
- Return to the main Posda window where your query results are still displayed. Click Refresh on the left menu.
- The query runs again using the parameters you provided last time. Review the changes to the results.

The processing status has changed from ReadyToReview to Reviewed for the data that you reviewed. The review status lists whether the set was reviewed as good or bad.

- Visually Review all the data marking each IEC as good or bad.
  - If the Visual Review projections have errored, you can click on the details box and use Quince to review.
- At this point your team may have multiple curators in charge of review. It is even possible to require a set number of curators to label an IEC as good before it receives the good label. Once all Series have been labeled good or bad, the bad sets must be removed.

### REMOVING BAD DATA FROM ACTIVITY TIMEPOINT

- Refresh the results of the VisualReviewStatusById query. (Use the recents option on the queries page if you cannot easily get to the results)
- At this point all rows should have a processing status of Reviewed and a Review status of either Good or Bad.
- Click Details on a Bad row.
- Click query, this will run VisualReviewStatusDetails.
- Click pg-up to refresh the page.
- A new buttons should appear on the left.

Remove Bad Sops from Timepoint

- If these files were marked Bad by mistake you can use Change Review Status to correct the review.
- To remove bad files click Remove Bad Sops from Timepoint.
- Verify the input is correct.
- Click Start.
- The response should inform you of the number of files expected to be hidden. Close the popup.
- View and dismiss the Inbox notification.
- On the Activity Timeline you should see that the number of files has changed.
   Note: Bad files will still appear in the Visual Reviewer, even after they have been removed from the timepoint.

### PHI REVIEW - THE POSDA PHI SCAN

This process will provides multiple reports showing the metadata in the files that must be manually scanned for remaining PHI.

There are several steps to a standard PHI review:

- Schedule a PHI Scan to create the PHI reports.
- Review the generated PHI Reports.
- Populate the "Edit Skeleton" spreadsheet with the elements in need of editing.
- Upload that sheet as input for the Propose Edits process.
- Review the results of the Proposed Edits.
- If satisfactory upload the result sheet as input for the next process Background Edits.
- Review the results from Background Edits. These reports explain the final changes that will be done to the dataset.
- Accept or Reject the edits.

Now that we have covered the big picture, let's go through the process in detail.

#### Schedule a PHI Scan to create the PHI reports:

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click PHI Review.
- Click the Schedule PHI Review button.
- Switch focus to the new window if it does not take focus automatically.
- Click Start.

- Close the popup once the worker is queued. This process may take some time to complete.
- View and dismiss the Inbox notification once the process has completed. This process may take some time to complete.
- Click Activity to view the timeline.

#### Review the generated PHI Reports.

• Click Email next to the completed Schedule PHI Review task.

```
Background process /home/posda/posdatools/Posda/bin/SchedulePhiReviewFromActivityTimepoint.pl begun
Starting PHI scan: "First Pass Scan activity timepoint for activity 4 (admin)"
6 series to scan
Created scan id: 2 in 166 seconds
Creating "SimplePublicPhiReportSelectedVrWithMetaquotes" report.
Creating "SimplePhiReportAllRelevantPrivateOnlyWithMetaQuotes" report.
Background process ended at: 2020-08-12T19:42:55
Total time elapsed: 172
Report 'Edit Skeleton': http://localhost/papi/file/29/05283b9a-dcd4-11ea-9f6a-a7e5dec73e88
Report 'Selected Private': http://localhost/papi/file/30/05bdd7fe-dcd4-11ea-9f6a-a7e5dec73e88
Report 'Selected Public VR': http://localhost/papi/file/31/06543b72-dcd4-11ea-9f6a-a7e5dec73e88
```

- Click on the hyperlinks within the message to download the spreadsheets you created with the subprocess. The top link is the Edit Skeleton spreadsheet that will be modified, and the other two are the reports to review for PHI.
- Save the files when prompted.
- Navigate to the files on your machine(default locations are often Documents or Downloads).
- Open the files with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).
- Carefully review the spreadsheets.
  - The "Selected Private" report is a report of unique values in the Private DICOM Tags that are likely to contain PHI; not all DICOM tags are displayed in the report.
  - The "Selected Public VR" report is a report of unique values in the Public DICOM Tags that are likely to contain PHI; not all DICOM tags are displayed in this report.

#### Example "Selected Private" report :

element	vr	q_value	description	disp	num_series
<(0013,"CTP",10)>	LO	<	Project Name	k	24
<(0013,"CTP",11)>	LO	<	Trial Name	k	24
<(0013,"CTP",12)>	LO	<	Site Name	k	24
<(0013,"CTP",13)>	LO	<	Site ID	k	24

Example snippet from "Selected Public VR" report :

element	vr	q_value	description	disp	num_series
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		2
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		2
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		2
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0010,1010)>	AS		Patient's Age		1
<(0008,0060)>	CS		Modality		24
<(0010,0040)>	CS		Patient's Sex		24
<(0012,0062)>	CS		Patient Identity Removed		24
<(0028,0303)>	CS		Longitudinal Temporal Information Modified		24
<(0040,a040)>	CS		Value Type		24
<(0040,a050)>	CS		Continuity Of Content		24
<(0040,a491)>	CS		Completion Flag		24
<(0040,a493)>	CS		Verification Flag		24
<(0040,a504)[<0>](0008,0105)>	CS		Content Template Sequence:Mapping Resource		24
<(0040,a504)[<0>](0040,db00)>	CS		Content Template Sequence:Template Identifier		24
<(0040,a730)[<0>](0040,a010)>	CS		Content Sequence:Relationship Type		24
<(0040,a730)[<0>](0040,a010)>	CS	>	Content Sequence:Relationship Type		24

- The element column contains the DICOM tag.
- The VR is the Value Representation, a DICOM coding for the type of data.
- The q\_value is the value seen in a tag in one or more files in the dataset. (These are redacted in the example images.)
- The description is what the value represents. Disp refers to Disposition and is related to how the tag is handled in the apply dispositions steps. Private Tags will normally have a disp of 'k'.
- Num\_series is the number of Series in the dataset that have this q\_value in this tag.

In the example you can see that the (0010,1010) tag is the DICOM tag for the Patient's Age. The q\_values then are the ages of the patients. Most patients have different ages. However all 24 series have the same value for Modality, tag (0008,0060).

#### Populate the "Edit Skeleton" spreadsheet with the elements in need of editing.

element	vr	q_value	edit_description	disp	num_series	p_op	q_arg1	q_arg2	Operation	activity_id	scan_id	notify	sep_char
									ProposeEditsTp	4	2	admin	%

- Any row that contains PHI in either report should have a row created for it in the Edit Skeleton spreadsheet.
- You can copy and paste from the Selected Public or Selected Private\_VR report to the Edit Skeleton spreadsheet (do not paste over the Operation, activity\_id, scan\_id, notify, or sep\_char fields in row 1). This will complete the first 6 columns.
- For the next 3 columns you will be choosing an operation to run anywhere the element + value pair is found. P\_op is the operation name, q\_arg1 is the first argument value, and q\_arg2 is the second argument value or '<>' if only one argument is needed.
  - Edit operations are listed below in the format p\_op(<q\_arg1>,<q\_arg2>):
    - shift\_date(<shift\_count>)
    - shift\_date\_by\_year(<shift\_count>)
    - copy\_date\_from\_tag\_to\_dt(<from\_tag>)
    - copy\_from\_tag(<from\_tag>)
    - delete\_tag()
    - set\_tag(<value>)
    - substitute(<existing\_value>, <new\_value>) (\*exact match)
    - string\_replace(<old\_text>, <new\_text>) (\*replaces part of a string)
    - empty\_tag()
    - short\_hash()
    - hash\_unhashed\_uid(<uid\_root>)
    - date\_difference(<ref\_tag>, <date>)
- <u>The sep\_char needs to be repeated on every row</u>, but the Operation, activity\_id, scan\_id, and notify only need to be in the first data row.

Example completed Edit Skeleton:

element	vr	q_value	edit_description	disp	num_series	p_op	q_arg1	q_arg2	Operation	activity_id	scan_id	notify	sep_char
<(0008,1030)>	LO	<ct lungs="" of=""></ct>	Study Description		1	set_tag	<lung></lung>	<>	ProposeEditsTp	1	2	admin	%
<(0008,1030)>	LO	<us 984-934=""></us>	Study Description		1	string_replace	< 984-934>	< 988-888>					%
<(0010,1010)>	AS	<112Y>	Patient's Age		2	delete_tag	<>	<>					%

Take a look at the values in this example.

This skeleton is proposing that every file where the Study Description tag (0008, 1030) has the value "CT of Lungs" we will replace that value with "LUNG". We expect only one series in our dataset will have a match and be changed. It is also proposing that every file where that tag instead has the value <US 984-934> we will look at that string value and replace the "984-934" section with "988-888". We expect only one series in our dataset will have a match and be changed. Finally we want to completely delete the tag value pair wherever the patient's age (0010, 1010) was set to 140 years. We expect 2 series to be changed from that command. (At this point we do not know if either of the planned edits to different tags are for the same series. The next process will tell us more)

- Complete your edit skeleton with an operation to perform for every PHI element found.
- Save the Edit Skeleton spreadsheet as a .csv file. Close the file.
- Return to your Posda window.

### Upload that sheet as input for the Propose Edits process.

- Click Upload in the Quick Navigation Menu on the left.
- Drag your saved file into the box, or click on the box to bring up a file selection window and choose your file. Wait for the upload to complete.
- Click chain in the file list once it appears.
- Review the Process Popup. Make sure the input lines and parameters match the expected operation.
- Click Start. This process will look through the data for the tag + value pairs you input in the Edit Skelton and find which series contain them. It will provide an output file with the series to be changed, and the changes you suggested.
- Close the popup when safe to do so.

- When the process finishes, you will receive an inbox notification. View and dismiss the notice.
- Return to Activity, click email to open the filed notice.

```
Background process /home/posda/posdatools/Posda/bin/ProposeEditsTp.pl begun at 2021-09-23T17:49:
Starting simple look up of Series with PHI
Scan_id: 2
Retrieved 1 for:
        element: (0008,1030)
        value: CT of Lungs
Retrieved 1 for:
        element: (0008,1030)
        value: US 984-934
Retrieved 2 for:
        element: (0010,1010)
        value: 112Y
3 distinct edit groups found
Command group: delete_tag|(0010,1010)||
Edit: delete_tag|(0010,1010)||
Command group: set_tag|(0008,1030)|LUNG|
Edit: set_tag|(0008,1030)|LUNG|
Command group: string_replace|(0008,1030)| MRI| TEST
Edit: string_replace (0008,1030) | MRI | TEST
finished scan
duration 0 seconds
Background process ended at: 2021-09-23T17:49:04
Total time elapsed: 0
Report 'EditSpreadsheet': http://localhost/papi/file/205/8c7c3b20-1c96-11ec-9edc-0b43b35093d2
```

- The results should appear similar to the image above. For every row you submitted, a number of series should have been found matching the expected num\_series in the last step.
  - If you see any places where the result says "Retrieved 0" then that row of data found no matching files with the tag + value pair you were looking to edit. Since you would not have tried to change data that did not exist, there is likely something wrong in the Edit Skeleton. Go back to that step.

```
Retrieved <u>0</u> for: Problem
element: (0008,1030)
value: US 984-934
```

- Click the hyperlink to download the EditSpreadsheet.
- Save the files when prompted.
- Navigate to the files on your machine(default locations are often Documents or Downloads).

• Open the files with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).

#### Review the results of the Proposed Edits.

• Carefully review the spreadsheet. The Proposed Edits process took the Edit skeleton and found every series that contained a tag+value pair that you suggested needed to be changed. This report shows you its findings.

series_instance_uid	num_files	ор	tag	val1	val2	Operation	edit_description	notify	activity_id
1.3.6.1.4.1.14519.5.2.1.9999.1001.772340686308580778543863085807	134					BackgroundEditTp	From PHI Scan: 2	admin	1
1.3.6.1.4.1.14519.5.2.1.9999.1001.228812415447949757555061193306	135								
		delete_tag	<(0010,1010)>	<>	<>				
1.3.6.1.4.1.14519.5.2.1.9999.1001.254198654664922839290857382007	567								
		set_tag	<(0008,1030)>	<lung></lung>	<>				
1.3.6.1.4.1.14519.5.2.1.9999.1001.763591437674261091741576749864	29								
		string_replace	<(0008,1030)>	< 984-934>	< 988-888>				

• For each potential change you can see the list of series that will be affected. In this example you can see that 2 series will have tag (0010, 1010) deleted. Only 1 series will receive the next change. Only one series will receive the final change. In this case no series has multiple changes, but that often can occur. If we look back at our Edit Skeleton and PHI report we can see that the num\_series lines up with the number of series that were found.

#### If satisfactory upload the result sheet as input for the next process Background Edits.

- Click Upload in the Quick Navigation Menu on the left.
- Drag the saved EditSpreadsheet into the box, or click on the box to bring up a file selection window and choose your file. Wait for the upload to complete.
- Click chain in the file list once it appears.
- Review the Process Popup. Make sure the input lines and parameters match the expected operation.
- Click Start. This process will determine the full extent of changes needed, going through each series one at a time.
- When the process finishes, you will receive an inbox notification. View and dismiss the notice.

Review the results from Background Edits. These reports explain the final changes that will be done to the dataset.

• Return to Activity, click email to open the filed notice.



- Click on the hyperlinks within the message to download the spreadsheets.
- Save the files when prompted.
- Navigate to the files on your machine(default locations are often Documents or Downloads).
- Open the files with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).

#### Example EditDifferences:

Short Report	Long Report	short_file_id	long_file_id	num_files
Only in from file:	Only in from file:	620	621	2
(0010,1010)	(0010,1010) : "112Y"			
Elements changed:	Elements changed:	622	623	1
(0008,1030):	(0008,1030): "US 984-934" => "US 988-888"			
-	Elements changed:	622	624	1
	(0008,1030): "CT of Lungs" => "LUNG"			

Example ShortEditDifferences: (Same data as EditDifferences but with fewer columns so you can focus on examining the file and tag information)

downloaded_file_210				
Report	short file id	nu		

Short Report	short_file_id	num_files
Only in from file:	620	2
(0010,1010)		
Elements changed:	622	2
(0008,1030):		

- This report shows the changes you are being asked to accept in detail. The "From" file refers to the file before the change. The "To" file refers to the file after the change. In this example you can see how for the Delete operation the tag value pair only exists in the before file(from file), since we have deleted it. Other changes we see are in both files, as a change in value.
- Look at the report in detail. Are all the changes you requested listed? Do you see any unexpected results?

Accept or Reject the edits.


- If your reports show that the requested changes were not properly found:
  - Click the Reject Edits button.
  - Switch focus to the new window if it does not take focus automatically.
  - Click Start.
  - Close the popup once the worker is queued. This process may take some time to complete.
  - View and dismiss the Inbox notification once the process has completed.
  - Go back through the previous steps and try to discover what caused the problem. Could there have been a typo in the Edit Skeleton? Could you have run the file on the wrong activity?
- Once you are confident that the reports confirm that all the changes you requested in your Edit Skeleton have been found and planned out properly:
  - Click the Accept Edits button.
  - Switch focus to the new window if it does not take focus automatically.
  - Click Start to begin the ImportEdits process. This process will finally apply the Edits that you have chosen and verified.
  - Close the popup once the worker is queued. This process may take some time to complete.
  - View and dismiss the Inbox notification once the process has completed.
  - Your Activity will have a new latest Activity Timepoint with files that have your specified edits.

## CHECK STRUCT LINKAGE

DICOM images can have extra files that include structures, especially ROIs(Region of Interest) contours related to Radiation Therapy(RT). The Check Struct Linkage process scans the files and verifies that the Structure data was found and successfully parsed.

Not all collections will include Radiation Therapy structure data, so this step will be skipped for some collections.

This step verifies that the data necessary for properly linking RTSTRUCTS with the images is present within the dataset.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click Check Struct Linkage.
- Click the Check Structure Set Linkage button.
- Switch focus to the new window if it does not take focus automatically.
- Click Start.
- Close the popup when it is safe to do so.
- View and dismiss the Inbox notification.
- Click Activity to view the timeline.
- Click Email.
- Click on the hyperlinks within the message to download the spreadsheets you created with the subprocess.
- Save the files when prompted.
- Navigate to the files on your machine (default locations are often Documents or Downloads folders).
- Open the files with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).

• This report shows if the Structure data was found and successfully parsed.

Example:

Collection	ite	Patient	File Id	Number F	Rois Sops in V	olui So	ops Linked To Roi	Internal Linkages Consistent	Files Linked	i Volume Files In Posda	Volume Files In Public	Roi Files In Posda	Roi Files In Public	Frame of Re	Unlinked Closed Planar Contours	Points Within Volume
			92491653	1	145 3	211 CL	LOSED_PLANAR:	Ok	Ok	Found: 211; Not found: 0	Found: 0; Not found: 211	Found: CLOSED_PLANAR = 211	Not Found: CLOSED_PLANAR = 211	Matches	None	Not Currently Implemented
			92491244	1	174 :	147 CL	LOSED_PLANAR:	Ok	Ok	Found: 147; Not found: 0	Found: 0; Not found: 147	Found: CLOSED_PLANAR = 147	Not Found: CLOSED_PLANAR = 147	Matches	None	Not Currently Implemented
			92491679		87 :	150 CL	LOSED_PLANAR:	Ok	Ok	Found: 150; Not found: 0	Found: 0; Not found: 150	Found: CLOSED_PLANAR = 150	Not Found: CLOSED_PLANAR = 150	Matches	None	Not Currently Implemented

- Look closely at the "Volume Files in Posda" column. We want to see the "Not Found: 0". If the value is something other than 0 there is an issue in the data.
- In the "ROI Files in Posda" column, find the Found: "CLOSED\_PLANAR = number" It should match the Found number in "Volume Files in Posda." If the numbers do not match, there is an issue.
- Check that the values are as expected. If your data has a problem, you will need to troubleshoot this issue before continuing to the Link RT Data step.

# LINK RT DATA

DICOM images can have extra files that include structures, especially ROIs(Region of Interest) contours related to Radiation Therapy(RT). The Link RT Data process links the ROI to the corresponding pixel data. Once linked, the images are viewed in Quince to confirm that the ROIs are properly aligned over the pixel data.

Not all collections will include Radiation Therapy structure data, so this step will be skipped for some collections.

The Check Struct Linkage steps verify that the necessary data exists. The Link RT Data creates the linkages.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click Link RT Data. Make sure you have done the previous step "Check Structure Linkages" first in order to resolve any "unfound" files in that report.
- Click the Link RT Data.
- Switch focus to the new window if it does not take focus automatically.
- Click Start.
- Close the popup when it is safe to do so.
- View and dismiss the Inbox notification.
- Click Workflow.
- Click Link RT Data.
- Click Series Linked to RTStructs.
- Click foreground.
- Click Query. Wait for the page to populate.
- The results will have a view button on the Series Instance UID column. Click View.

#### • Click Open in Quince



- You can click on the image and use the mouse wheel to enlarge it.
- Region of Interest(ROI)s can be displayed in Quince by toggling on the ROI button. Each ROI is named and displayed as seen in the picture above. The number of ROIs will vary between series and may range from one to multiples.
- Click on the ROI name to enable/disable the drawing of that specific ROI.
- Click the Transport button ( the car/bus) next to any ROI name, and Quince will skip to the first frame of the image that includes that ROI.
- Review the ROIs making sure the contour lines overlay a portion of the image that is not sitting outside the pixel data. If an ROI is outside the image, you will need to troubleshoot this issue before moving on to the next step.
- Continue to click through each frame of the series to view each ROI.

- You can also use the Enable/Disable and Play buttons, but the ROIs appear quickly and their location may be difficult to discern.
- When finished with the first series, return to the query and click the next View button then follow the steps above until all series are complete.

#### SEND TO SERVER

Your instance should have a server where the final data that has been fully de-identified is released for use (For many of our current users this is NBIA, often called PUBLIC). This process will apply the Background Dispositions and then send out your now clean data. **Ensure your dataset is ready before beginning this process**!!

(Dispositions are general rules about how DICOM private tags are handled for repository data. Each tag that has special rules has them applied in this step as part of the server sending process. The possible dispositions are 'k' keep, 'd' delete, 'h' hash, 'na' not applicable, 'o' offset in date format, and 'oi' offset in integer format. Most tags will already have the disposition set in the database.)

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click Send to Server (NBIA).
- Recall during Patient Mapping whether your collection used a baseline date or a dateoffset (non-baseline). Click the appropriate Apply Background Dispositions to Timepoint.
  - Example with baseline date:

Start
Cancel
Help
<pre>Dperation: BackgroundPrivateDispositionsTpBaseline Command: BackgroundPrivateDispositionsTpBaseline.pl <?bkgrnd_id?> <activity_id> "<notify>" "<skip_dispositions>" " <upd_nbia>" "<dir>"</dir></upd_nbia></skip_dispositions></notify></activity_id></pre>
<ul> <li>Parameters:</li> <li>activity_id :</li> </ul>
633
• dir :
notify :
admin
skip_dispositions :
0
• upd_nbia :
1
Expanded Command:

Start	Operation: BackgroundPrivateDispositionsTp Command: BackgroundPrivateDispositionsTp pl_c2bkgrad_id2_captivity_ids_" (uid_roots " " coffeets " coffeets " " coffeets" " coffeets " coffeets " coffeets " coffeets " coffeets " coffeets" " coffeets " coffeets" coffeets
Cancel	<ul> <li>Command: BackgroundPrivateDispositionsTp.pi &lt; /pkgrid_id /&gt; <activity_id> <uld_root> <onset> <notify> <skip_dispositions> <upd_nbia>" "<dir>"</dir></upd_nbia></skip_dispositions></notify></onset></uld_root></activity_id></li> </ul>
Help	
	Parameters:
	activity_id :
	648
	• dir :
	notify :
	admin
	offset :
	-9500
	skip_dispositions :
	0
	uid_root :
	upd_nbia :
	1
	Expanded Command:
	BackgroundPrivateDispositionsTp.pl bkgrnd_id? 648 """""""""""""""""""""""""""""""""""

- Switch focus to the new window if it does not take focus automatically. Complete the required information.
- Click Start. Depending on several factors, such as the size of the data, this may take some time to complete.
- Close the popup when it is safe to do so.
- View and dismiss the Inbox notification.
- Click Activity to view the timeline.
- Click Email.
- Click on the hyperlink within the message to view the Public Copy Status Report(Example image on following page). It will open in a new tab. This report will let you know if the data was successfully sent.
  - Examine the "Count of files" values. True is the number of files that transferred successfully. False is the number of files that produced an error. Most errors will cause the file not to transfer. The Both value should be the sum of the True and False values, if it is not there may be a problem that needs troubleshooting. If there were any failures, then a second section appears below the count section showing

the error message and details for the first five failures (Since many times multiple files will fail for the same reason only a small sample number is shown). You can use the error information to troubleshoot why the transfers failed, so that you can resolve them in a later send.

#### **Public Copy Report for ID #75555**

#### **Overall status**

Count of files	Sucessful?
120	False
167900	True
168020	Both

#### First 5 files that failed

file_id	<u>55555555</u>
collection	NewCollection1
site	UAMS
site_id	99991001
filename	/server-place/directory/example/45/99/34248234ds3423.dcm
third_party_analysis_url	None
error_message	(400, b'Error in ExampleOperation — Exception in TestExampleOperation java.lang.Exception: There would be some error information here — java.lang.Exception: Continued error text)

## COMPARE DATA TO SERVER - PUBLIC PHI SCAN

This process will compare your data in Posda to the Release Server. This is commonly used right after the previous process to confirm that all of the data sent to the release server properly. TCIA uses the NBIA server as its release/public server. This set of instructions is specific to that process, but the process should have similarities for other Release servers.

- Login and launch the Activity Module if you are not already there.
- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Workflow.
- Click Compare data to Server (NBIA).
- Click Find Files in Tp, not in Public.
- Switch focus to the new window if it does not take focus automatically.
- Click Start.
- Close the popup when it is safe to do so.
- View and dismiss the Inbox notification.
- Click Activity to view the timeline.
- Click Email.
- If files were found that are in this Timepoint that are not in PUBLIC, they may have failed to send and need to be resent. If all files are in PUBLIC, then continue these instructions. Otherwise you will need to troubleshoot and retry Send to Server.
- Click Workflow.
- Click Compare data to Server (NBIA).
- Click Public Phi Scan Based on Current TP by Activity.
- Switch focus to the new window if it does not take focus automatically.
- Click Start.

- Close the popup when it is safe to do so.
- View and dismiss the Inbox notification.
- Click Activity to view the timeline.
- Click Email.
- Click on the hyperlinks within the message to download the spreadsheets you created with the subprocess.
- Save the files when prompted.
- Navigate to the files on your machine(default locations are often Documents or Downloads).
- Open the files with your preferred spreadsheet editor(Excel/Numbers/LibreOffice/etc).
- This report will let you analyze the metadata just as you did in the PHI Review step. Double check that no metadata contains PHI. If you find PHI, you will need to go back to your timeline. Find the most recent PHI Review. Download the edit skeleton and use your data from this step to complete the Edit spreadsheet. Rerun the PHI Review process starting from the Upload Proposed Operation step. Then rerun the Send to Server process.
- If you use NBIA, you will next do a Visual Spot check within the NBIA system, as a final verification that the data is correct.

Once you have completed these processes you should be finished with this dataset! The deidentification of your dataset is complete, the data is available on the release server, and you have confirmed the release data is correct and PHI free!

The last step remaining is to close your activity. This optional step will keep your Posda instance more manageable.

## CLOSING AN ACTIVITY

The Activity dropdown used to select an activity will be quickly overwhelmed if you leave your activities open. Closing an activity will hide it from the dropdown and other visuals. However, a closed activity can be opened again at any time if you need it again later.

Login and launch the Activity Module if you are not already there.

- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Queries.
- Search for CloseActivity.

Activity Based Curation								
◯ - recent	Args containing:	Columns containing:	Query matching:	Name matching: Close				
Searched queries (8 rows) (name like '%Close%')								
name	par	ams	columns returned	make query				
CloseActivity	acti	vity_id		foreground				
				background				

- Click foreground.
- Input the ID of the Activity you wish to close.
- Click query.
- Your activity is now Closed.

## **RE-OPENING AN ACTIVITY**

In the previous step we learned how to close an activity. If you ever need to work on an activity again after closing it, you can re-open the activity by following the steps below.

Login and launch the Activity Module if you are not already there.

- Select your Activity if it is not already selected. (Name of Activity shows in the title).
- Click Queries.
- Search for ListClosedActivities.

<ul> <li>○ - active ○ - recent ○ - search ○ - workflow</li> </ul>	Args	containing:	Columns containing:	Query matching:	Name matching:	clear
Searched queries (2 rows) (name like '%ListClo%')						
name	params	columns retu	irned		make query	
ListClosedActivities		activity_id, br	ief_description, when_	_created, who_created	, foreground	
		_010000		background		

- Click foreground.
- Find the activity you wish to re-open. You may want to use the Edit Filter option. See the <u>Query Results Screen section</u> for more details on using filters.

Full Results for: ListCle	osedActivities	s()	No fi	lter currently	defined
First row: 0 Show	: 30	pg-dn pg-up			
chain	activity_id	brief_description	when_created	who_created	when_closed
re-open Details	1	SR Breast Data Test	2021-09-01 21:24:52.84507+00	admin	2021-11-05 13:00:00.924279+00

- Note the Activity ID.
- Click re-open in the row with the correct activity.
- Input the Activity ID. Click query.
- Your Activity is now Open.

## EXTRAS: PRODUCE ACTIVITY REPORT, COPY OR CONSOLIDATE TIMEPOINTS, AND OTHER

The remaining options on the workflow page are not required steps in the curation process, but offer convenient access to useful operations and queries that are more situational.

#### **Produce Activity Report:**

• **Produce Condensed Activity Timepoint Report** - This creates a summary report that provides information for each patient\_id in this activity.

key	value							
script	/home/posda/posdatools/Posda/bin/CondensedActivityTimepointReport.pl							
activity_id	4							
activity_timepoint_id	9							
activity_timepoint_comment	New Timepoint for ImportedEdits 17							
activity_timepoint_date	2020-08-12 19:13:21.437398							
when	Thu Aug 13 11:25:51 CDT 2020							
who	admin							
collection	site	patient_id	num_studies	num_series	num_modalities	num_sop_classes	num_sops	num_files
NewCollection1	UAMS	Pat_308	3	3	1	1	454	454
NewCollection1	UAMS	Pat_920	1	3	1	1	174	174

• Activity Report Queries - A collection of useful queries for examining your data. Some queries take the Activity ID as input, while others will take the Activity Timepoint ID.

Suggested	Queries	for	Activity	Timepoint	Reports
-----------	---------	-----	----------	-----------	---------

name	params	columns returned	make query
VerboseActivityTimepointReport	activity_id	collection, site, patient_id, patient_age, study_instance_uid, study_date,	foreground
		study_description, series_instance_uid, series_date, series_description,	background
		dicom_file_type, modality, num_files	
FilesSeriesNumFIlesAndSopsVisibilityInTimepoint	activity_timepoint_id	patient_id, study_instance_uid, series instance uid, collection, site.	foreground
		visibility, modality, type, num_files,	background
		num_sops	
TimepointCreationReport	activity_timepoint_id	collection, site, patient, num_studies, num_series, num_sop_classes,	foreground
		num_modalities, num_sops, num_files	background

#### Copy and Consolidate Timepoints:

Occasionally, you will need to restore files to an older version or incorporate new files into an activity timepoint. Below are operations that allow you to modify the files in the activity based on activity timepoints.

- **Copy Prior Timepoint** Make a new timepoint with the same files as another. The copied timepoint does not have to be the latest timepoint. This can be very useful for recovering a timepoint after the latest timepoint has had a problem, or to back out edits that were imported.
- **Copy Prior Timepoint Excluding Files** Similar to Copy Prior Timepoint while also allowing the removal of files no longer needed in the activity timepoint. Requires a formatted spreadsheet as input. Columns: [file\_id, Operation, notify, description]
- **Consolidate Timepoints** Merge timepoints. This is useful for situations when a set of files from a particular collection were edited in another activity and you want to merge them into the activity containing the rest of the collection. Requires a formatted spreadsheet as input. Columns: [ activity\_timepoint\_id , Operation, notify, description ]
- **Consolidate Activities** Merge Activities. Requires a formatted spreadsheet as input. Columns: [ activity\_id , Operation, notify, description ]

#### Other:

- Make a Downloadable Directory A way to share files for analysis outside of Posda. Can be helpful for troubleshooting errors. Requires activity\_id, notify\_id, and sub\_dir. In sub\_dir box, type in the desired directory name such as CollectionNameActivityID. There should be no spaces. Once the directory is created, click on Download in the lefthand menu and select the directory name entered previously from the pulldown. Only share files with other internal users!
- Import Downloadable Directory Import a cached directory.
- **Copy SOP Class and SOP instance from Meta-header** Used to fix certain malformed DICOM files from a specific collection.
- Generate List of Weekly Uploads By File Type useful for gathering usage statistics for your instance of Posda.

- Make Worksheet For Dispositions Needed Creates reports for Dispositions Needed. Clicking on the Make Worksheet button will result in an email with a spreadsheet which contains private tags that need to have dispositions defined in the database. The possible dispositions are 'k' keep, 'd' delete, 'h' hash, 'na' not applicable, 'o' offset in date format, and 'oi' offset in integer format. You may find it easier to review the spreadsheet in another program. Add the appropriate value in the "disp" column (k,d, h,o, oi) of the CSV spreadsheet. The disposition is always lowercase. Delete the value column from the spreadsheet then save. Upload the spreadsheet using the Upload button in the lefthand menu. Click on the chain option when it appears and run the process.
- Create Private Tag VR and Disposition Reports Creates a report showing the dispositions assigned to each Private tag.
- Create New Timepoint from Old Timepoint ID a troubleshooting operation sometimes needed when fixing issues with an activity timepoint.

Other new features and developer use only features are sometimes added to this section before being moved to more relevant places.

## FINAL NOTES

This document and the application it describes are updated regularly. You may find new features in your application that have not yet been added to the documentation. Official releases of this document and the Posda application are available at <a href="https://posda.com/">https://posda.com/</a>. If you have questions or need assistance please contact <a href="https://posda.com/">help@cancerimagingarchive.net</a>.