DICOM Image quality and harmonization in MIDRC

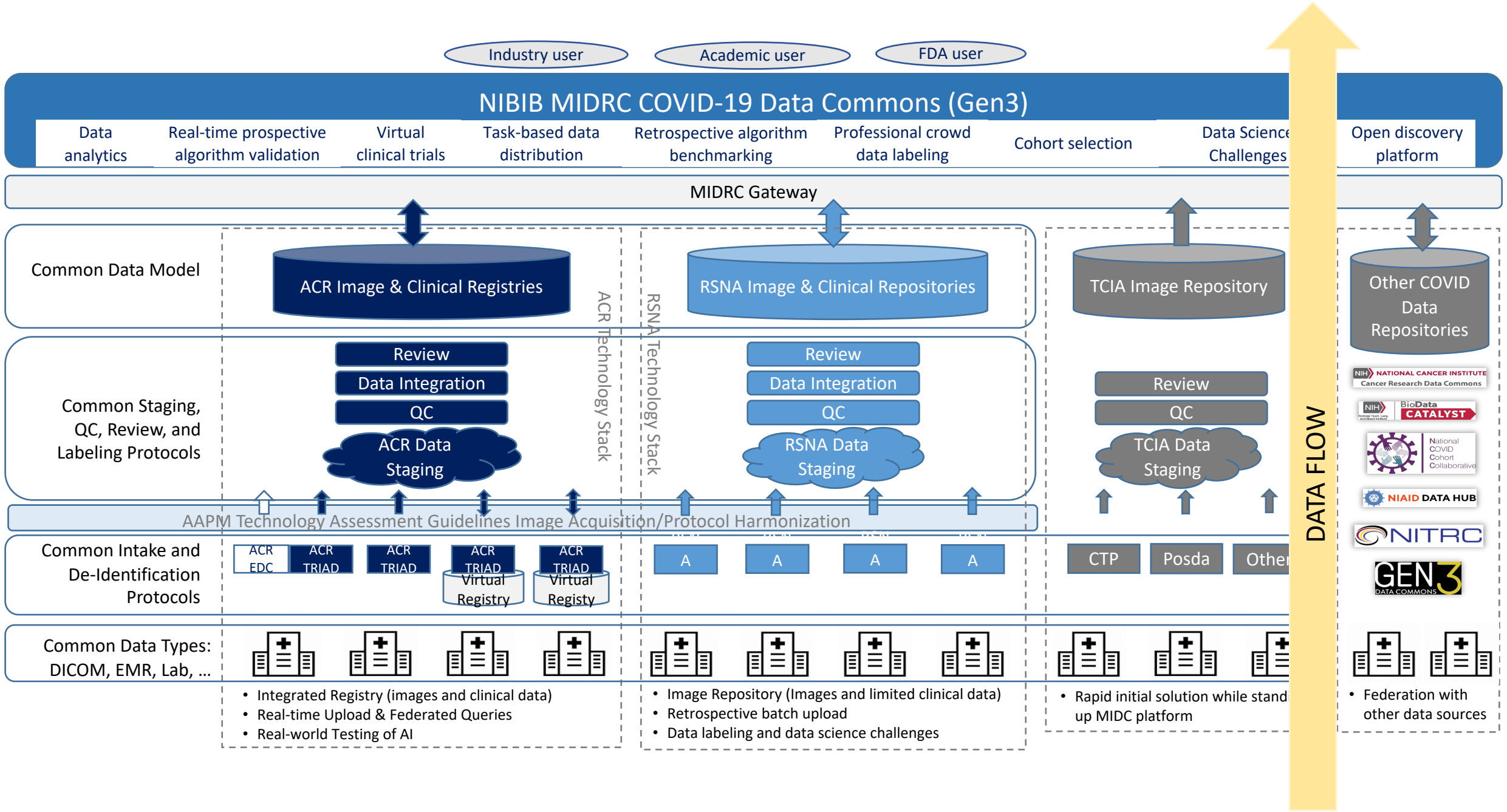
Paul Kinahan, PhD

University of Washington

Technology Development Projects (TDPs)

TDP3: Developing and implementing quality assurance and evaluation procedures for usage across the MIDRC

- Spans from data ingestion through data access and dissemination
- Includes data quality metrics, data provenance, data processing audits paths, and task-directed data distribution
- Procedures are intended to help enable research and translation with MIDRC data
- Developing harmonized data definitions and labeling methods, hosting of data science challenges, and benchmarking of algorithm performance


Industry user
Academic user
FDA user
NIBIB MIDRC COVID-19 Data Commons (Gen3)
Data analytics
Real-time prospective algorithm validation
Virtual clinical trials
Task-based data distribution
Retrospective algorithm benchmarking
Professional crowd data labeling
Cohort selection
Data Science Challenges
Open discovery platform
MIDRC Gateway
Common Data Model
ACR Image & Clinical Registries
RSNA Image & Clinical Repositories
TCIA Image Repository
Other COVID Data Repositories
Common Staging, QC, Review, and Labeling Protocols
Review
Data Integration
QC
ACR Data Staging
RSNA Data Staging
TCIA Data Staging
ACR Technology Stack
RSNA Technology Stack
AAPM Technology Assessment Guidelines Image Acquisition/Protocol Harmonization
Common Intake and De-Identification Protocols
ACR EDC
ACR TRIAD
Virtual Registry
CTP
Posda
Common Data Types: DICOM, EMR, Lab, …
Integrated Registry (images and clinical data)
Real-time Upload & Federated Queries
Real-world Testing of AI
Image Repository (Images and limited clinical data)
Retrospective batch upload
Data labeling and data science challenges
Rapid initial solution while stand up MIDC platform
Federation with other data sources
DATA FLOW

TDP 3 Activities to develop and implement MIDRC quality assurance and evaluation procedures

TDP3a: Develop COVID phantoms (digital and physical)

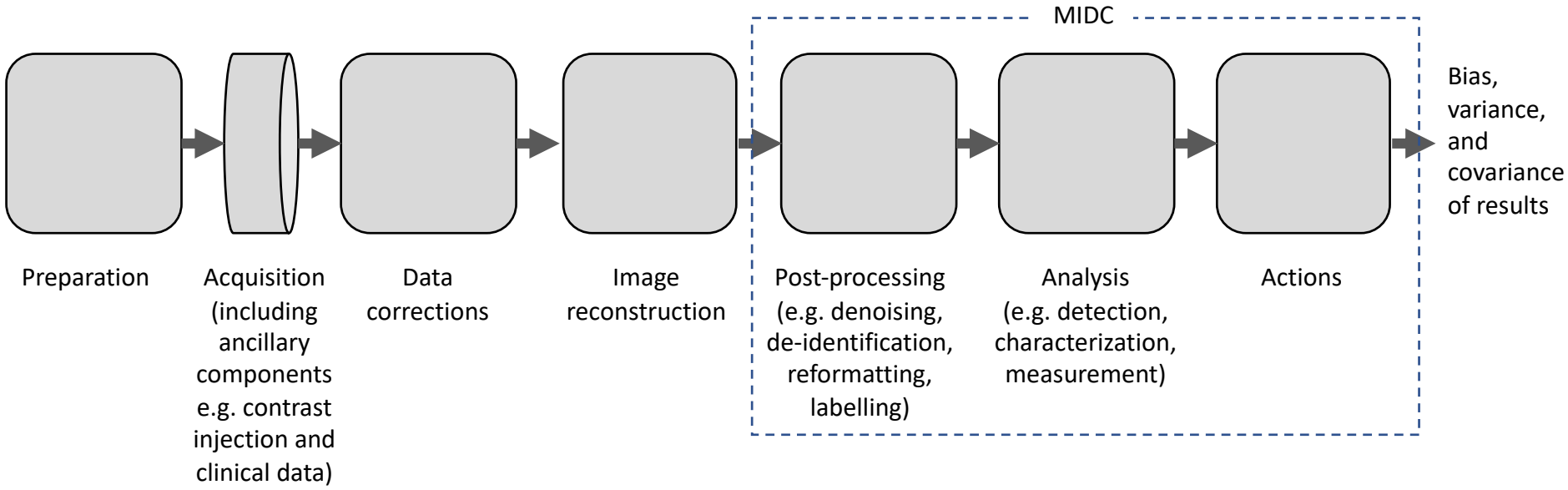
TDP3b: Conduct measurement of CT & CXR image quality & harmonization techniques for image data ingestion

TDP3c: Develop benchmarking methods for technology assessment & clinical tasks in COVID-19 research and translation

TDP3d: Develop task-based distribution methods

TDP3a: Development of digital and physical imaging phantoms (John Boone)

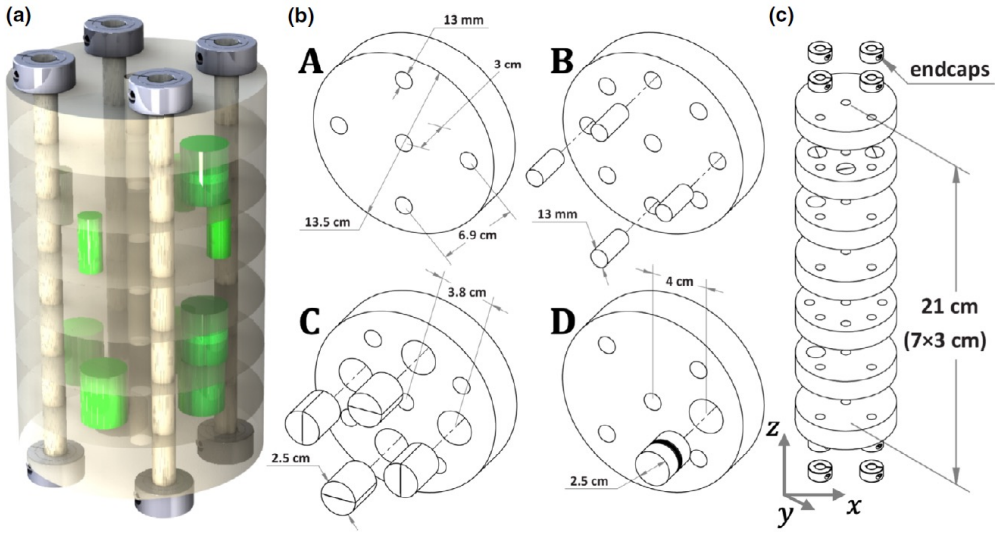
- Evaluation of existing general (ACR) and specific (Corgi) phantoms
- Use of automated image analysis tools
- Developing a digital phantom to assess impact of de-identification methods

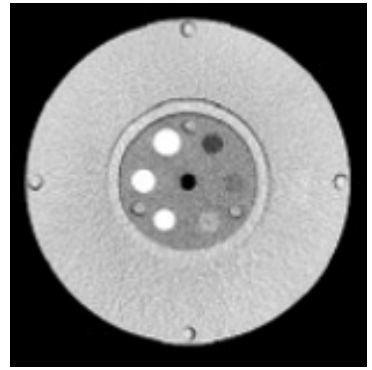


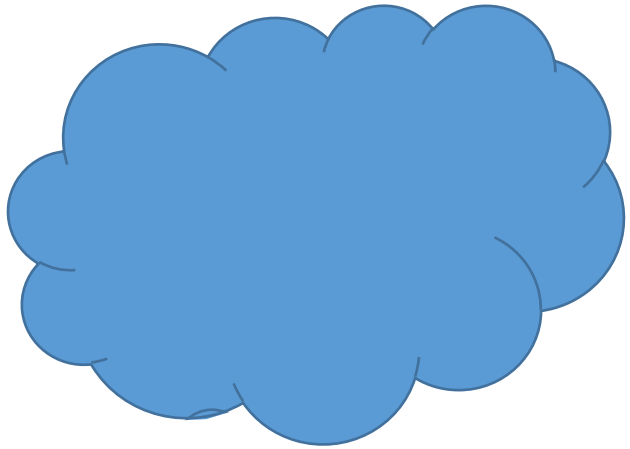
design

fabrication

scan

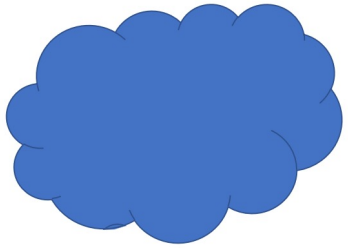






Cloud

CT images

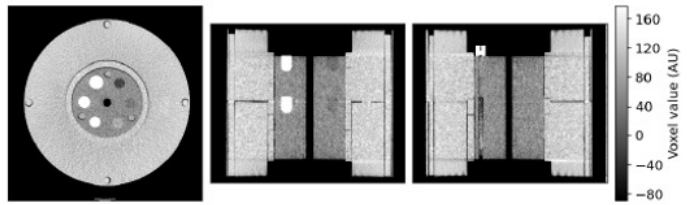


analysis of image quality

Corgi Report

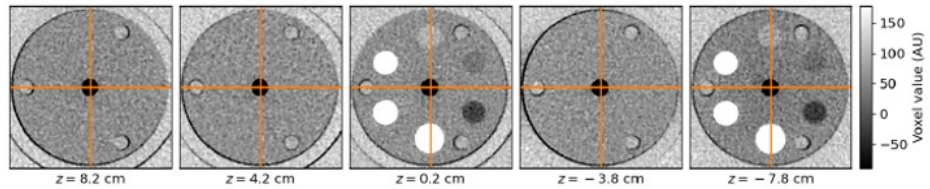
Input Image

TOSHIBA
Aquilion Precision
Study Date: Jul 31, 2020
Study Time: 12:25 PM
Axial FOV: 37 cm
Longitudinal FOV: 31 cm
Tube Voltage: 120 kV
Tube Current: 400 mA
$CTDI_{vol}$: 14.5 mGy



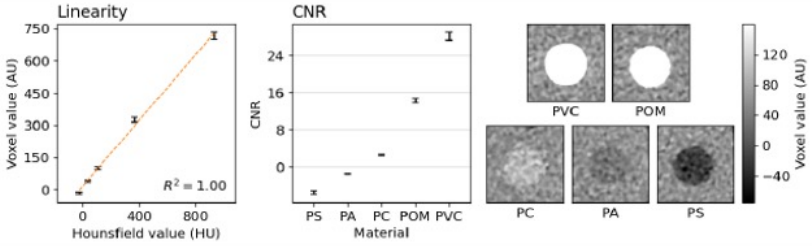
Corgi Modules

Module count: 5
Phantom *x,y* offset: 0.5 cm
Phantom *z* tilt: 0.1°



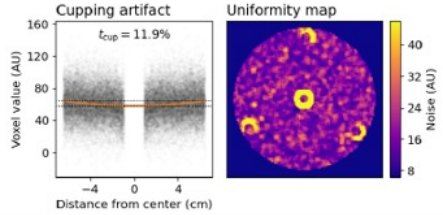
Sensitometry

CNR for PVC: 28.1
CNR for POM: 14.3
CNR for PC: 2.5
CNR for PA: -1.5
CNR for PS: -5.6
Linearity (R^2): *1.00*



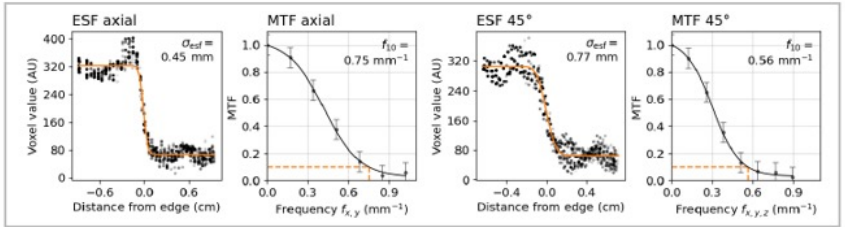
Uniformity

Mean value: 60.3 AU
t_{cup}: 11.9%



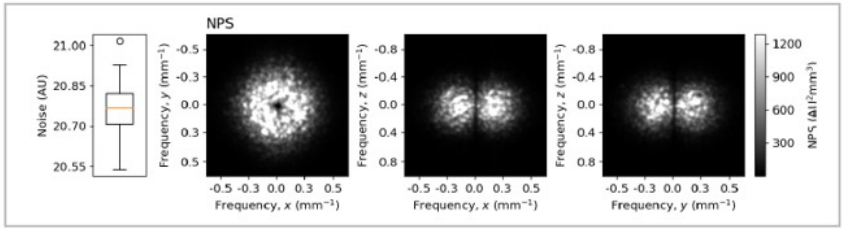
Spatial Resolution

σ_{ESF} axial: 0.45 mm
σ_{ESF} 45°: 0.77 mm
f_{10} axial: 0.75 mm^{-1}
f_{10} 45°: 0.56 mm^{-1}



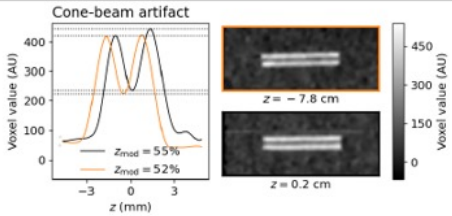
Noise

Mean noise: 20.8 AU



Cone-Beam Artifact

z_{mod} **at 0.2 cm:** 55%
z_{mod} **at -7.8 cm:** 52%

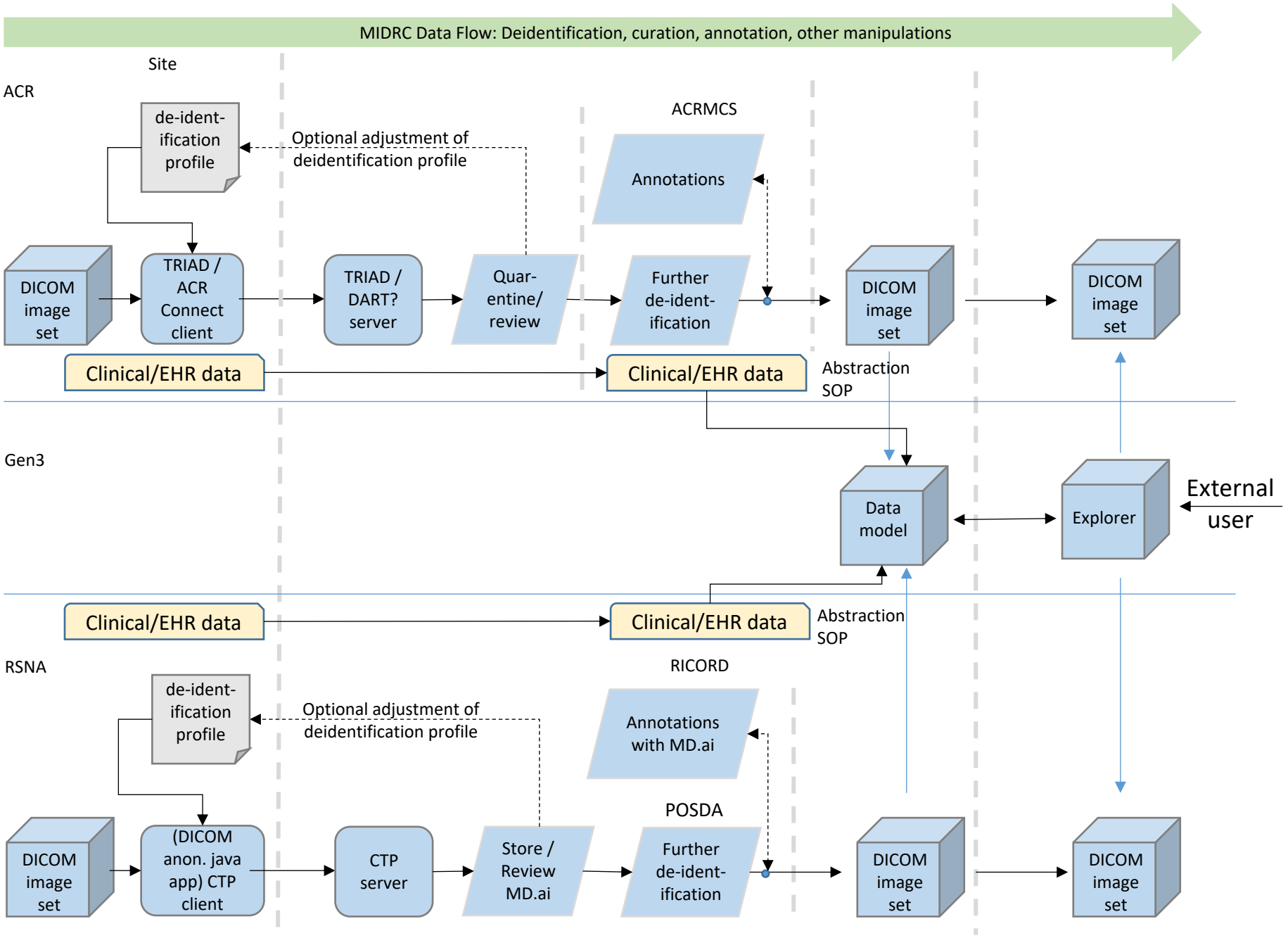


TDP3b: Assessment of image quality on ingestion into MIDRC

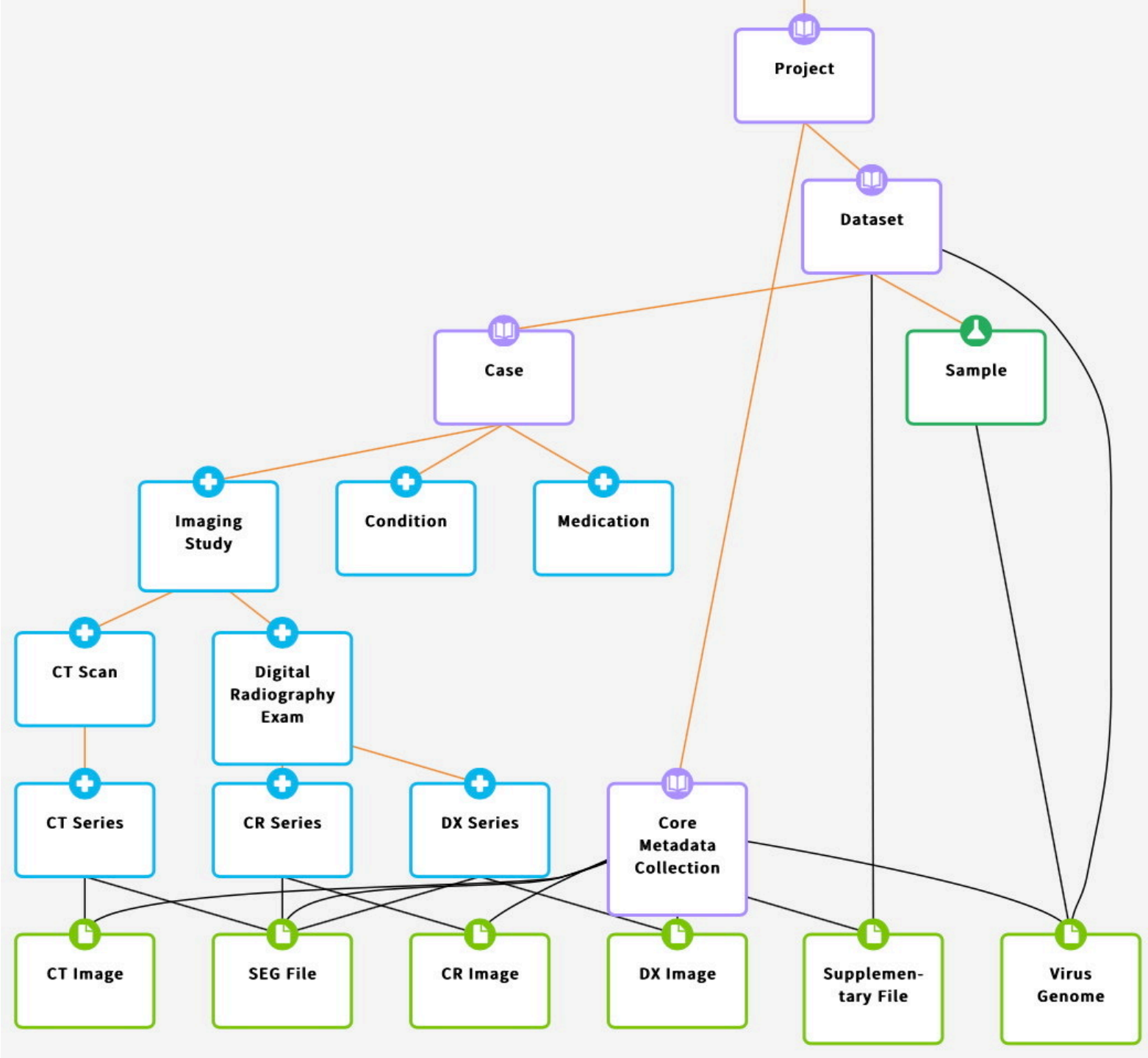
Two main components:

1. Assessment of the images themselves where there are manual and semi-automated approaches available to search for and assess known structures and artifacts (e.g. from implants)
2. Use of the image meta-data that is stored in the DICOM image
    - An exploratory component is to combine parameters derived from the images with the meta-data to general image quality metrics

Efforts on evaluating image data quality early in the ingestion process are driving the efforts of the Data Quality and Harmonization SC


MIDRC Data Flow: Deidentification, curation, annotation, other manipulations
Site
ACR
de-ident-ification profile
Optional adjustment of deidentification profile
ACRMCS
Annotations
DICOM image set
TRIAD / ACR Connect client
TRIAD / DART? server
Quar-entine/ review
Further de-ident-ification
DICOM image set
DICOM image set
Clinical/EHR data
Clinical/EHR data
Abstraction SOP
Gen3
Data model
Explorer
External user
Clinical/EHR data
Clinical/EHR data
Abstraction SOP
RSNA
RICORD
de-ident-ification profile
Optional adjustment of deidentification profile
Annotations with MD.ai
POSDA
DICOM image set
(DICOM anon. java app) CTP client
CTP server
Store / Review MD.ai
Further de-ident-ification
DICOM image set
DICOM image set

Explorer data model for cohort selection



MIDRC Approved Properties [WIP]

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

A1:D1 | The following list includes the variables (properties) at the IMAGE SERIES (aka Exam) level. Users will be able to search on these data for all IMAGE SERIES. These data are defined as, "Information related to a CR/DX/CT imaging series to which one or more computed

	A	B	C	D	E	F	G	H	I	J	K	L
1	The following list includes the variables (properties) at the **IMAGE SERIES** (aka Exam) level. Users will be able to search on these data for all **IMAGE SERIES.** These data are defined as, "Information related to a CR/DX/CT imaging series to which one or more computed radiographs/digital radiographs (x-rays)/a stack of CT slices belong." *Note: While we understand most of this information is provided within each Image File, we want to limit the amount of searching done at that level as this could cause performance issues. Ideally, we want to eliminate anything that is completely or heavily redundant for each image. In otherwords, these properties are going to be "rolled up" from the image files and some of these properties are arrays because more than one value might be relevant for a single series.*				Applicable to Modality							
2	**Node**	**Property**	**Description**	**Data type**	MRI	CT	CR	DX	Include as UI filter	Needs lookup table (**draft completed**)	Required	DICOM
3	Series	series_description	A general description of or additional information about the imaging series.	string	yes	yes	yes	yes	yes	yes	no	yes
4	Series	series_uid	The identification code or scan number of the imaging series.	string	yes	yes	yes	yes	no	no	no	yes
5	Series	acquisition_type	(0018, 9302) Acquisition Type	array, enum [SEQUENCED I SPIRAL I CONSTANT_ANGLE I STATIONARY I FREE]	no	yes	no	no	yes	??	no	yes
6	Series	contrast_bolus_agent_number	(0018, 0010) Contrast/Bolus Agent	string	yes	yes	yes	yes	yes	??	no	yes
7	Series	convolution_kernel	(0018, 1210) Convolution Kernel	string	no	yes	no	no	yes	??	no	yes
8	Series	detector_type	(0018, 7004) Detector Type	enum [DIRECT I SCINTLLATOR I STORAGE I FILM]	no	no	no	yes	yes	??	no	yes
9	Series	exposure_modulation_type	(0018, 9323) Exposure Modulation Type	string	no	yes	no	no	yes	??	no	yes
10	Series	imager_pixel_spacing	(0018, 1164) Imager Pixel Spacing	number	no	no	no	yes	yes	no	no	yes
11	Series	image_type	(0008, 0008) Image Type	enum Varies by modality	yes	yes	yes	yes	yes	??	no	yes
12	Series	lossy_image_compression	(0028, 2110) Lossy Image Compression	enum [00 I 01]	yes	yes	yes	yes	yes	no	no	yes
13	Series	magnetic_field_strength	(0018, 0087) Magnetic Field Strength	string	yes	no	no	no	yes	??	no	yes
14	Series	manufacturer	(0008, 0070) Manufacturer	string	yes	yes	yes	yes	yes	**yes**	no	yes
15	Series	manufacturer_model_name	(0008, 1090) Manufacturer's Model Name	string	yes	yes	yes	yes	yes	??	no	yes
16	Series	modality	(0008, 0060) Modality	enum (long list of enums linked here: https://dicom.innolitics.com/ciods/ct-image/gen	yes	yes	yes	yes	yes	??	no	yes
17	Series	pixel_spacing	(0028, 0030) Pixel Spacing	number	yes	yes	yes	yes	yes	no	no	yes
18	Series	scanning_sequence	(0018, 0020) Scanning Sequence	string	yes	no	no	no	yes	??	no	yes
19	Series	slice_thickness	(0018, 0050) Slice Thickness	number	yes	yes	no	no	yes	no	no	yes
20	Series	spacing_between_slices	(0018, 0088) Spacing Between Slices	number	yes	yes	no	no	yes	no	no	yes
21	Series	spatial_resolution	(0018, 1050) Spatial Resolution	number	yes	yes	yes	yes	yes	no	no	yes
22	Series	view_position	(0018, 5101) View Position	enum [AP I PA I LL I RL I RLD I LLD I RLO I LLO]	no	no	yes	yes	yes	??	no	yes
23	Series	data_format	Format of the data files.	enum (DICOM I NIfTI)	yes	yes	yes	yes	yes	no	yes	no
24	ADD annotations generated by the MIDRC DQH group... need a detailed list from DQH. (connect with Adam Flanders)											

MIDRC Data Harmonization Mapping (Lookup Table)

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

E11:G11 | CT Chest

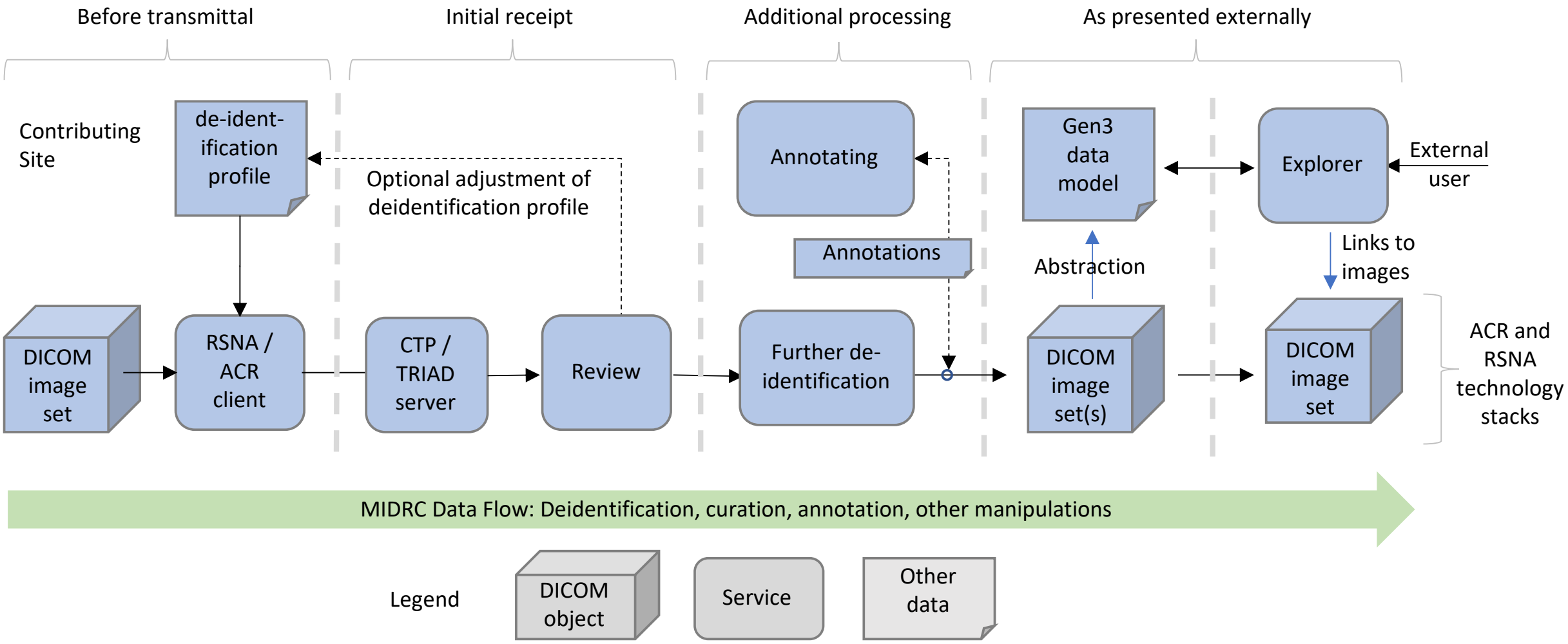
	A	B	C	D	E	F	G	H
1						Mapping to standardized lexicon		
2	***node***	**property**	**raw_value (submitted by contributors)**	**raw_count**	**proposed mapped midrc_value (RadLex Playb**	**Coding Scheme**	**Code**	**difference**
7	imaging_study	study_description	XR CHEST 1 VIEW AP	477	XR CHEST 1V	RadLex Playbook	RPID2502	TRUE
8	imaging_study	study_description	CXR	432	XR CHEST* (*non-RLPB)			TRUE
9	imaging_study	study_description	XR CHEST AP PORTABLE	204	XR CHEST 1V	RadLex Playbook	RPID2502	TRUE
10	imaging_study	study_description	CHEST 1V	76	XR CHEST 1V	RadLex Playbook	RPID2502	TRUE
11	imaging_study	study_description	CT CHEST	58	CT Chest	LOINC	24627-2	
12	imaging_study	study_description	THORAX PE	34	CT CHEST W EMBOLUS	RadLex Playbook	RPID147	TRUE
13	imaging_study	study_description	XR CHEST AP ONLY	19	XR CHEST 1V	RadLex Playbook	RPID2502	TRUE
14	imaging_study	study_description	CT PE CHEST	17	CT CHEST W EMBOLUS	RadLex Playbook	RPID147	TRUE
15	imaging_study	study_description	THORAX/ABDOMEN/PELVIS + CONT	14	CT CHEST/ABD/PELV W	RadLex Playbook	?	TRUE
16	imaging_study	study_description	XR CHEST 2 VIEWS PA AND LATERAL	13	XR CHEST 2V	RadLex Playbook	RPID2503	TRUE
17	imaging_study	study_description	CT CHEST PULMONARY EMBOLISM (CTPE)	13	CT CHEST W EMBOLUS	RadLex Playbook	RPID147	TRUE
18	imaging_study	study_description	Non-contrast Chest CT	11	CT CHEST WO	RadLex Playbook	RPID16	TRUE
19	imaging_study	study_description	Low-dose Chest CT	7	CT CHEST WO SCREEN	RadLex Playbook	RPID6002	TRUE
20	imaging_study	study_description	THORAX PE + LOW DOSE	5	CT CHEST W EMBOLUS	RadLex Playbook	RPID147	TRUE
21	imaging_study	study_description	CT CHEST ABDOMEN PELVIS W	5	CT CHEST/ABD/PELV W		?	TRUE
22	imaging_study	study_description	XR CHEST PA AND LATERAL	5	XR CHEST 2V	RadLex Playbook	RPID2503	TRUE
23	imaging_study	study_description	THORAX + CONT	3	CT CHEST W	RadLex Playbook	RPID18	TRUE
24	imaging_study	study_description	CT ANGIOGRAM CHEST	3	CT CHEST ANGIO W/WO	RadLex Playbook	RPID6	TRUE
25	imaging_study	study_description	XR ACUTE ABDOMINAL SERIES W PA CHEST POF	2	XR ABD SUPINE/ERECT+CHEST	RadLex Playbook	RPID6044	TRUE
26	imaging_study	study_description	XR CHEST PA ONLY	2	XR CHEST 1V PA	RadLex Playbook	RPID6035	TRUE
27	imaging_study	study_description	CARDIAC CORONARY CTA	1	CT HEART CORONARY ANGIO W	RadLex Playbook	RPID6003	TRUE
28	imaging_study	study_description	CT CHEST HIGH RESOLUTION	1	CT CHEST WO LUNG PARENCHYMA	RadLex Playbook	RPID6001	TRUE
29	imaging_study	study_description	CT PE	1	CT CHEST W EMBOLUS	RadLex Playbook	RPID147	TRUE
30	imaging_study	study_description	THORAX LOW DOSE	1	CT CHEST WO SCREEN	RadLex Playbook	RPID6002	TRUE

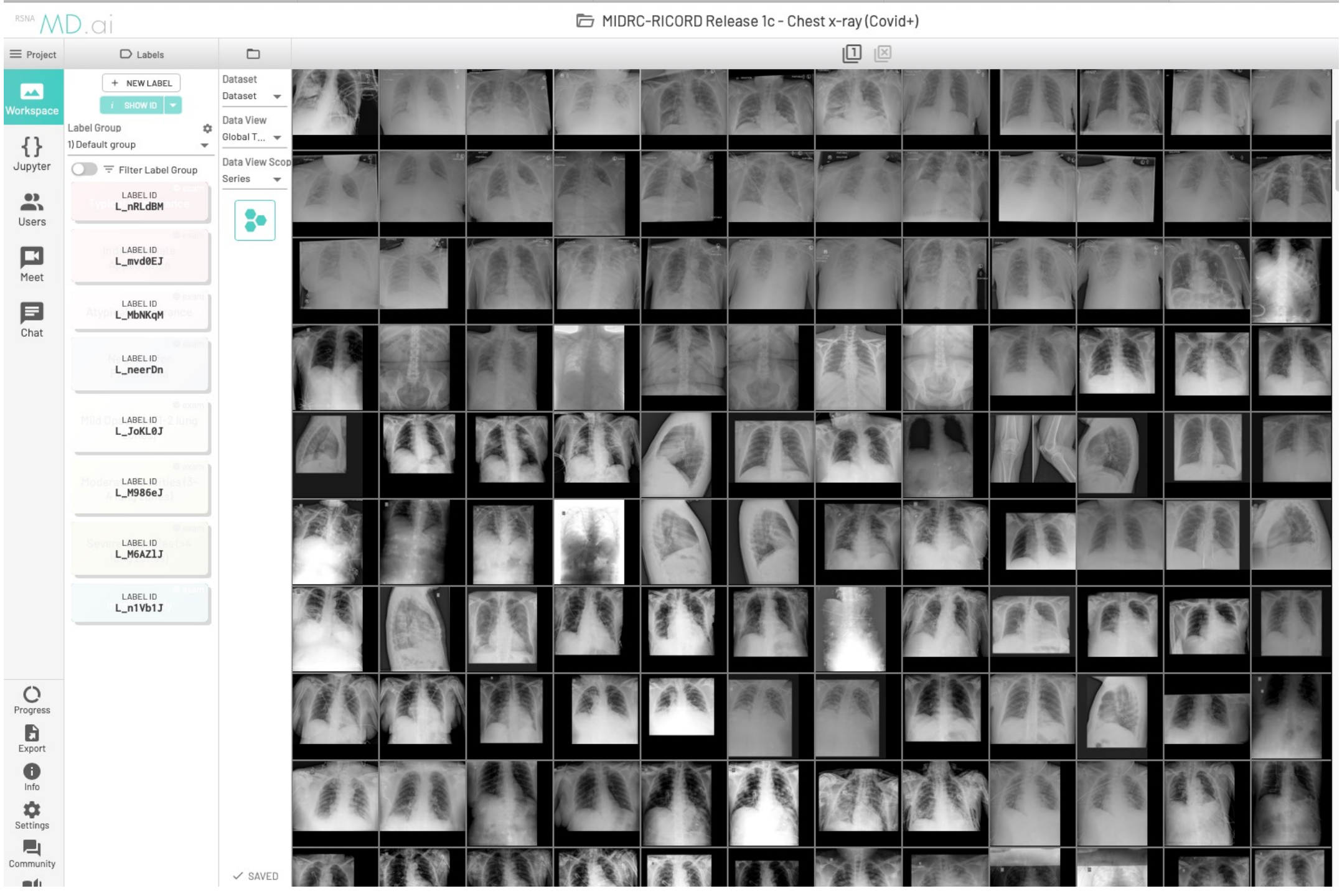
TDP 3b Data Quality Activities

The MIDRC will use data management methods harmonized across all participating organizations at three critical stages:

1. intake, including curation, de-identification, abstraction, and quality assessment
2. annotation and labelling of imaging and other data using semi-automated approaches
3. distributed access and query methods

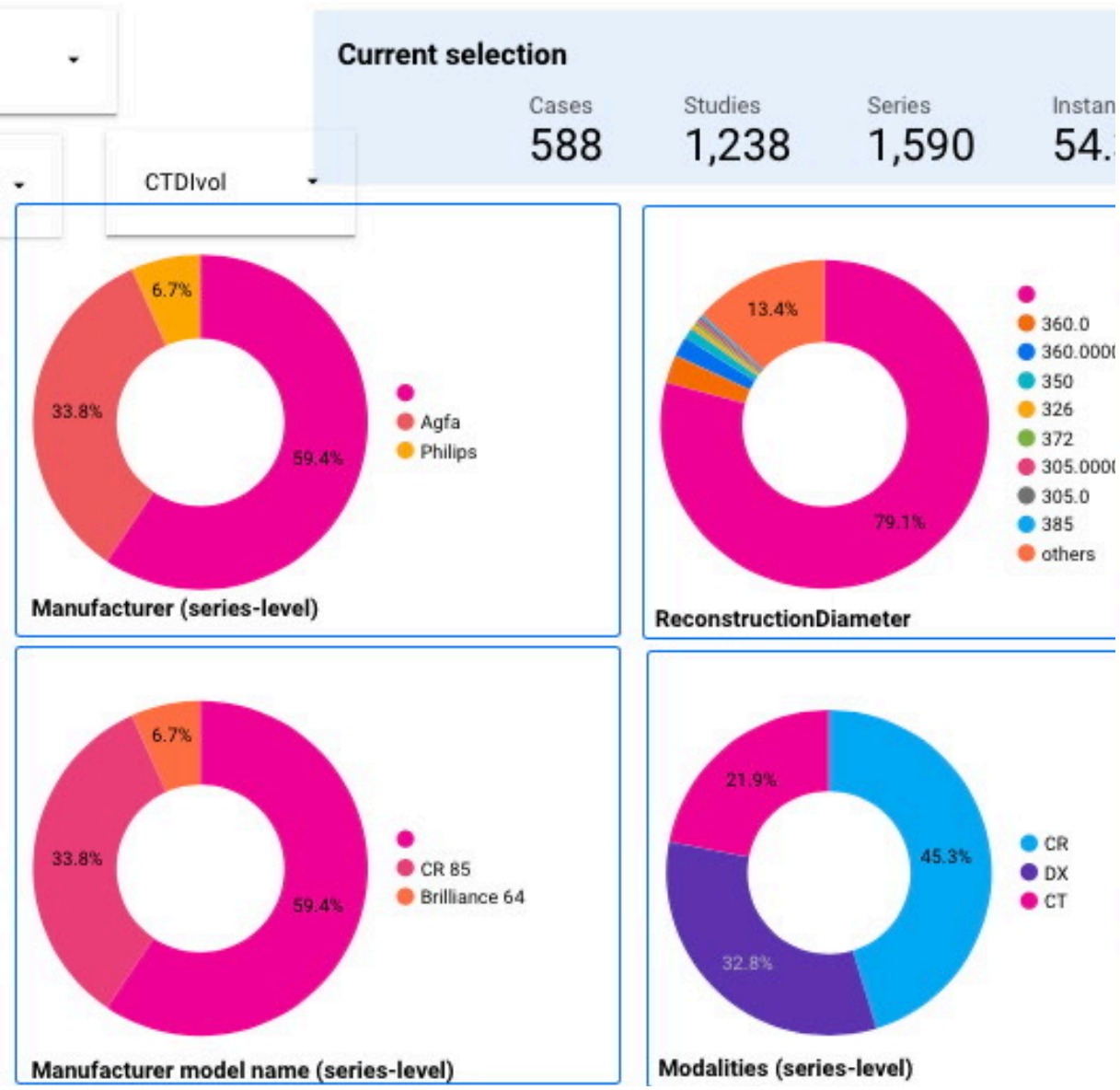
Schematic of the DICOM image data quality SOP

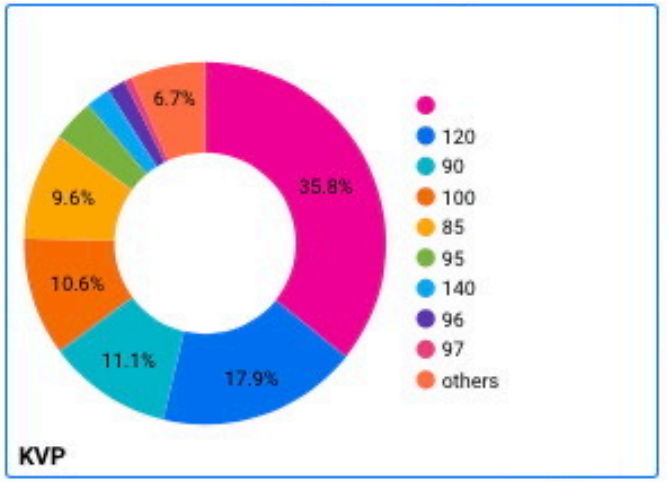


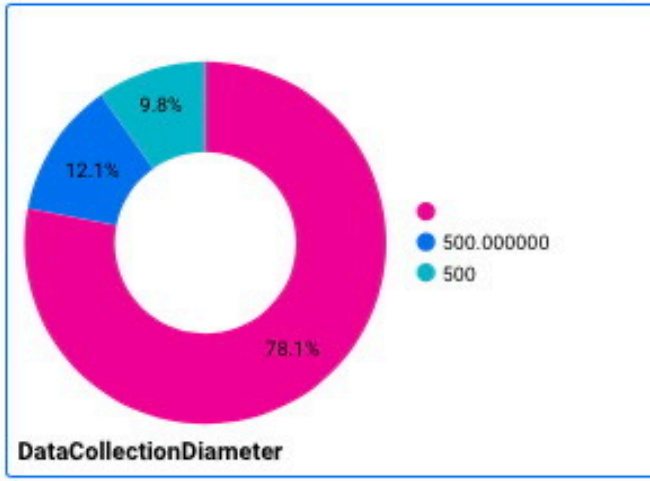

RSNA MD.ai
MIDRC-RICORD Release 1c - Chest x-ray (Covid+)
Project
Labels
Workspace
Jupyter
Users
Meet
Chat
NEW LABEL
SHOW ID
Label Group
1) Default group
Filter Label Group
LABEL ID
L_nRLdBM
LABEL ID
L_mvd0EJ
LABEL ID
L_MbNKqM
LABEL ID
L_neerDn
LABEL ID
L_JoKL0J
LABEL ID
L_M986eJ
LABEL ID
L_M6AZ1J
LABEL ID
L_n1Vb1J
Dataset
Dataset
Data View
Global T...
Data View Scop
Series
SAVED
Progress
Export
Info
Settings
Community

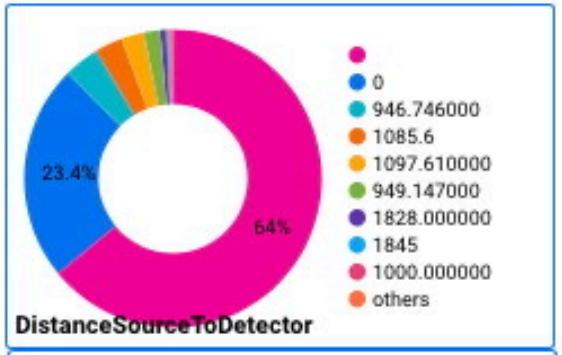

RSNA MD.ai
MIDRC-RICORD Release 1b - Chest CT (Covid-)
Project
Labels
Workspace
Jupyter
Users
Meet
Chat
Progress
Export
Info
Settings
NEW LABEL
LABEL CONTROLS
Label Group
Default group
Filter Label Group
No labels currently in this label group.
Dataset
Dataset
Data View
Global T...
Data View Scop
Exam

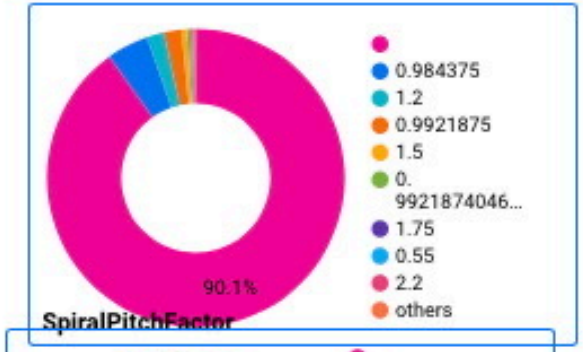
Cohort selection characteristics

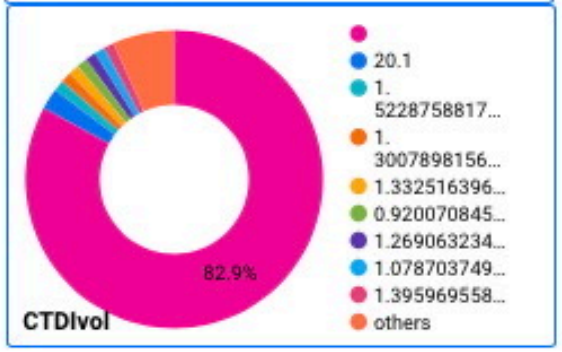


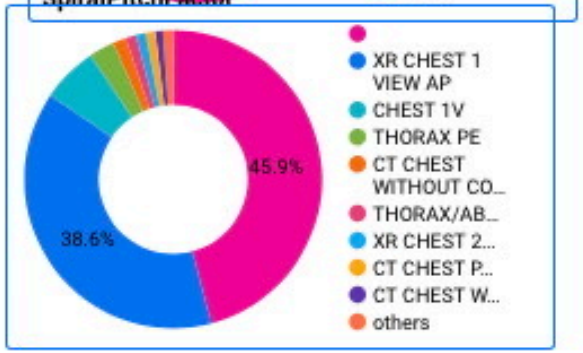












Added DICOM Image Quality Metrics – Under Consideration

- mobile_xray_unit (assuming we can figure out a way to do this)
- missing_slices/images (can we do this automatically?)
- image_artifacts (similarly can we do this automatically?)
- image_modified (e.g. altering burned-in information)
- incorrect_header_information (e.g. wrong body_part_examined NOT SURE (won't we remove these cases?)
- MIDRC_acceptable_image_quality (assuming we can figure out how to do this automatically on a large scale)
- phantom_image
- MIDRC_high_noise_image (assuming we can figure out a way to do this automatically)
- MIDRC_low_resolution_image (assuming we can figure out a way to do this automatically)
- study_description_mapped
- manufacturer_mapped
- body_part_examined_mapped
- annotations_provided
- deidentification_profile_used
- CT slice thickness categories (thin, medium, thick)
- CT dose categories (ultra low dose (CTDIvol<1mGy), low dose (1mGy< CTDIvol< 3mGy), modest dose (3 mGy<CTDIvol< 10 mGy), diagnostic dose (10mGy< CTDIvol).

Synergies between AAPM TDP 3 and CRPs 9-12

TDP 3a *Develop COVID phantoms (Boone)*

- Physical and digital phantoms to assess quality

TDP 3b *Measure and Harmonize image quality (Kinahan)*

- Assessment of images and meta data

TDP 3c *Benchmarking methods (McNitt-Gray)*

- Clarification paper on COVID-19 tasked-based performance metrics
- Online decision tree for performance metrics; benchmarking

TDP 3d *Task-based distribution methods (Myers)*

- Task-based distribution methods for Challenges and industry
- Initiated MIDRC Bias & diversity working group
- Assessing diversity in incoming data
- Developing sequestering algorithm

CRP 9 *AI for detection and diagnosis (Armato)*

- Grand Challenges
- Development / testing of AI diagostic methods

CRP 10 *Visualization & Explainability of AI (Giger)*

- "Lessons learned"; AI vs. AI med imaging vs. AI COVID-19 imaging
- "Explainability & Interpretability" paper
- Development / testing of AI prognostic methods

CRP 11 *Image-based biomarkers for radiogenomics (Chen)*

- Enclaves for integrating and interoperability of multi-modality (multi-omics) data sources
- Containerizing techniques (how to) as MIDRC connects with other COVID-19 registries

CRP 12 *Data quality, provenance, and harmonization (Fedorov)*

- Introduction of active tracking methods w DICOM