10.1148/radiol.2021203957

# Appendix E1

## Participating Sites

The process of assembling a single data set derived from multiple disparate institutions presented several challenges in the data procurement process. Broadly, these can be divided into refinements for data discovery, curation, and anonymization. The data imbalance problem was principally related to variations in data discovery methods employed at the contributing sites given the inclusion criteria were chiefly driven by a request for COVID-19 positive patient imaging data. While RSNA made tools available for many important steps, identifying and retrieving data relied on site-specific methods for data extraction from their respective archives; the demographic and case proportions varied by site. The number of cases per institution was 28/19/58/15 in collection 1a, 30/30/30/30 in collection 1b, and 76/433/491 in collection 1c.

### Koç University Hospital, Turkey

A search was performed in the institutional PACS (Centricity, GE Healthcare, Chicago, IL) for chest CT studies with the "Clinical Indication" field containing the keywords "COVID" or "COVID-19." The corresponding reports and the RT-PCR result of each patient were reviewed by a radiologist. Only patients with positive RT-PCR results were included. Axial soft tissue window images were manually exported from the PACS as DICOM images and de-identified using DICOM Anonymizer (RSNA, Oak Brook, IL). A manual review of images was performed to ensure successful de-identification.

### UC San Francisco, USA

A clinical data warehouse was created using Clarity (Epic Systems Corporation, Madison, WI), tracking every SARS-CoV-2 test administered in the UCSF Health System. Queries based on these data were used to identify all PUI and assign COVID status. Chest CT and CXR studies for PUI were extracted from the institutional PACS (Agfa IMPAX, Agfa Health care, Mortsel, Belgium) and de-identified using the same profile employed by DICOM Anonymizer (RSNA, Oak Brook, IL). Images were manually reviewed for PHI prior to release.

### Unity Health Toronto, Canada

Nuance mPower (Nuance Communications, Burlington, MA) was used to search the institutional radiology information system (syngo, Siemens Medical Solutions USA Inc., Malvern, PA) for chest radiograph and CT scan reports containing the keyword "COVID" performed between February 1 and June 4, 2020. The RT-PCR status of each patient identified by this search was manually reviewed in the electronic medical record (Cerner Soarian Clinicals, Cerner Corporation, North Kansas City, MO). The chest radiographs and CT scan images of patients with a positive RT-PCR test were downloaded from the institutional PACS (Carestream PACS, Carestream Health, Rochester, NY, U.S.) using DICOM Anonymizer (RSNA, Oak Brook, IL). A custom written Python script was used to extract axial soft tissue window images. The DICOM Anonymizer software tool was also used to perform image de-identification. A manual review of all the images was performed to ensure that PHI was not inadvertently included in the data set.

10.1148/radiol.2021203957

### Universidade de São Paulo, Brazil

Electronic medical records were searched for all patients with RT-PCR positive status. All chest CTs and radiographs from these patients were manually downloaded from PACS (Synapse, Fujifilm corporation, Tokyo, Japan) due to ambiguity in the Study Description. A custom Python script was written to extract axial soft tissue kernel series of ≤ 3.0 mm slice thickness from the CTs. Conversion of thinner thicknesses to 3.0 mm was performed using an in-house Python script. De-identification was using the standard script in DICOM Anonymizer (RSNA, Oak Brook, IL). Manual review was done to ensure no PHI was disclosed. Images were uploaded to md.ai through a web browser.