THE UNIVERSITY OF CHICAGO | Human Imaging Research Office

CRDW COVID-19 Limited Datamart
Medical Imaging Tip Sheet

The Human Imaging Research Office (HIRO) is collaborating with the Clinical Research Data Warehouse (CRDW) Team to provide medical imaging data for their COVID-19 Limited Datamart. See below for some tips regarding the formatting and use of this image data.

Image Data Format

All images are provided in standard DICOM format and can be opened with any DICOM viewing software. All DICOM image files have the `.dcm` file extension.

Each imaging exam has been zipped using standard zip compression software. Most modern operating systems have the capability to open and decompress zip files automatically. All zip files have the `.zip` file extension.

Imaging Exam Types

The HIRO is currently feeding the Datamart with the following types of imaging exams:

- **Chest radiographs** (aka chest x-rays, including standard and portable exams)
- **Chest CT scans** (including chest/abdomen scans and CTA chest scans)

Data will be updated on a weekly basis. All relevant exams performed on or after a patient's COVID encounter date will be included. In addition, all relevant **chest radiograph exams** performed up to **one year prior** to the COVID encounter date will be included, and all relevant **chest CT scans** performed up to **one month prior** to the COVID encounter date will be included. If additional types of imaging exams are needed, consultation with the CRDW and HIRO will be required.

Folder and File Naming Conventions

Each imaging exam will be compressed into a single zip file. The zip files will be named using the following convention:

- **`SSSS_yyyymmdd-hhmmss_AAAA_ExamDescription.zip`**

Where:

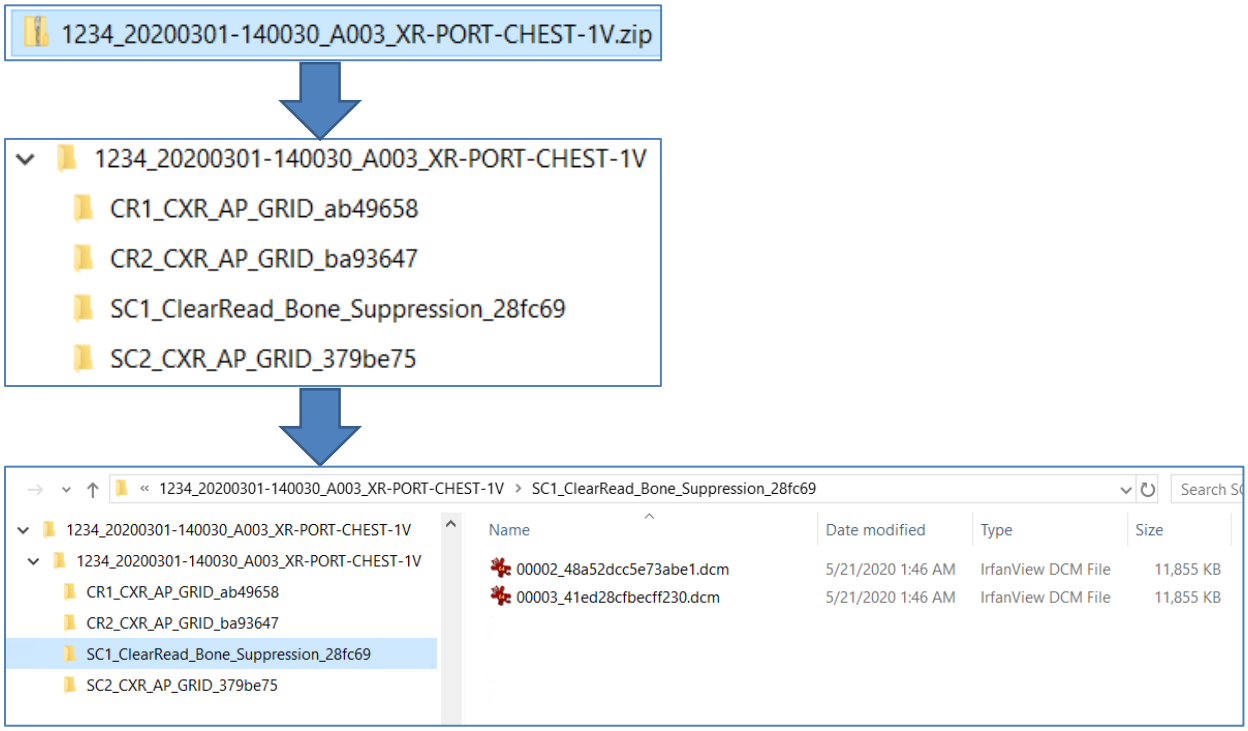
- **`SSSS`** is the CRDW COVID-19 patient ID number.
- **`yyyymmdd-hhmmss`** is the timestamp (date of service) from the exam, expressed as year-month-day-hour-minute-second.
- **`AAAA`** is a unique identifier generated by the HIRO.
- **`ExamDescription`** is a human-readable description of the type of imaging exam based on the EPIC radiology procedure description.

Within each zip file, the exam is divided into a set of sub-folders based upon the imaging series. Each series folder contains the DICOM image files (`.dcm`) associated with the series.

For Example:

The zip file named `1234_20200301-140030_A003_XR-PORT-CHEST-1V.zip` contains the portable chest x-ray exam for Subject 1234 performed on 01-Mar-2020 at 2:00:30PM.

Unzipping the file will create several different folders, each corresponding to a series within the exam. Each of those folders will contain the DICOM images associated with that image series.



Modifications to the DICOM Metadata

The metadata in the DICOM header of each image has been altered to correspond to the COVID-19 Limited Datamart's clinical data. The values of the Patient Name `(0010,0010)` and Patient ID `(0010,0020)` tags have been changed to the corresponding CRDW COVID-19 patient ID number for each patient. All exam timestamps remain intact. The Patient Birthdate `(0010,0030)` tags have been redacted, but the Patient Age `(0010,1010)` tags have been populated. Image and equipment parameter (technique) tags remain intact wherever possible. Most "screen save," 3D reconstruction, and paperwork images have been removed. If there are any specific questions about the DICOM metadata, please feel free to contact the HIRO at hirohelp@bsd.uchicago.edu.

Uchicago.box.com Organization and Uploads

New exams will be uploaded to uchicago.box.com in weekly batches. Each batch will be uploaded into its own folder. The folder naming convention will include the modality and the date of the batch. In addition, the folder will be marked as either "IN PROGRESS" or "COMPLETE" to indicate if an upload is currently underway or if the upload has finished, respectively.

For example, the `CXR-UPDATE-20200523-COMPLETE` folder contains all of the chest radiographs from the 23-May-2020 update. The `CT-UPDATE-20200602-IN-PROGRESS` folder will contain the chest CT scans from the 2-Jun-2020 update, but the data upload is still in progress so you should wait until the name has changed to "complete" before downloading data.

Viewing the Image Data

DICOM images can be viewed using DICOM image viewer software as well as several popular generic image visualization applications. This includes popular scientific visualization applications like ImageJ and Matlab. There are also numerous DICOM viewers freely available on the Internet. Links to a few free DICOM viewing programs for a variety of different operating systems can be found on the HIRO's website at https://hiro.bsd.uchicago.edu/links (see the "Software and Utilities Links" section of this page).

Known Issues with CRDW COVID-19 Patient ID Numbers

Occasionally a patient will receive care at UCM using the incorrect medical record number (MRN). This happens for a variety of reasons, particularly in the Emergency Department: a patient may be admitted and generate data (i.e. radiographs) with one MRN before it is discovered they already exist in the EMR under another MRN. C19 Patient IDs in the COVID-19 datamarts are associated with corresponding MRNs, so an incorrect MRN/Patient ID pair is sometimes assigned to a chest radiograph. Once the correct MRN is identified, additional data for the patient will be associated with the correct MRN/Patient ID pair.

While all of the clinical data in the datamart is routinely regenerated, the image data extracted by the HIRO is only generated once. As a result, if a patient's MRN changes, all of the patient's *clinical* data will be encoded with a new C19 Patient ID during the next datamart regeneration. However, any *image* data that has already been added to the datamart for the patient will retain the "old" C19 Patient ID. This may lead to instances in which some images in the datamart are encoded with C19 Patient IDs that do not have any corresponding clinical data.

The HIRO will routinely update a "mapping" file named `C19_HIRO_LDS_REMAP_PUB-LIST.xlsx` that will allow you to associate a patient's "old" C19 Patient ID with their current C19 Patient ID. If you are working with any images that are encoded with a C19 Patient ID that does not appear to have any corresponding clinical data in the datamart, please check this mapping file and see if the patient has been assigned a new C19 Patient ID.

Known Issues with Pediatric Radiographs

Due to a quirk in the manner in which images associated with pediatric chest radiographs are often stored in the hospital PACS, some exams may appear in the datamart twice. Pediatric exams can be identified via the exam description – the description will contain a 'P' after the modality code (for example, `XR P PORT CHEST 1 VIEW`).

The images associated with these "duplicate" exams will be identical and they will have identical timestamps, but the "unique HIRO identifier" may differ. We suggest utilizing the timestamp as well as the names of the individual image sub-folders and files to help identify these duplicate exams so you can discard them if necessary.

Known Issues with Some CT Scan Timestamps

Due to a quirk in the manner in which some CT scans are performed, in rare cases images from a single exam may appear to be split into two or more exams. For example, you may encounter two zip files, `234567_20200508-105300_A002_CT-CHEST-WO.zip` and `234567_20200508-110200_A002_CT-CHEST-WO.zip`, which appear to be two different scans. However, as demonstrated by the timestamps, the scans are mere minutes apart (10:53AM vs. 11:02AM). Both scans also have the same "unique HIRO identifier" (`A002`). In this case, the images in both zip files are part of the same scan. We are working to correct this issue

for future batches, but you may encounter it in existing data. If you see multiple CT scans from a single patient with timestamps that are only minutes apart **and** have the same unique HIRO identifier, it is likely that these "scans" are part of a single exam.

Known Issues with Some Duplicate Exams

We have noted that there have been some duplicate chest x-rays across some batches of data. These exams will share the same timestamp as the patient's previous exam but will have a different "unique HIRO identifier." Upon inspection, however, it is obvious that the images in the "new" exam are identical to the images in the previous exam. For example, the 23-May-2020 datamart update may include an exam labeled `5678_20200401-121530_A004_XR-CHEST-PA-LATERAL.zip`, and the 18-Jun-2020 datamart update may include the similarly labeled exam `5678_20200401-121530_A005_XR-CHEST-PA-LATERAL.zip`. Reviewing the images will reveal that these exams are actually identical; the latter exam is a duplicate that was pulled again by mistake.

We believe we have resolved this error in most of the previous batches and we do not believe it will recur in any future batches, but you may encounter it in existing data. To help more easily identify potential duplicates, the following criteria may be helpful:

- The duplicates do not appear to occur *within* a single update/batch. They generally occur *across* batches. So as noted in the example above, you should suspect a duplicate if the new case closely matches the timestamp of a case from a previous batch.
- When you unzip a suspected duplicate case, you will note that although it has a different unique HIRO identifier (`Axxx`), the sub-folder and image filenames are identical to the original case. Identically named sub-folder and image files indicate that the new case is a duplicate.
- Finally, if the value of the SOP Instance UID tag (`0008,0018`) in the DICOM header of the "new" image is identical to the value in the image of the suspected original case, the image (and therefore the case) is a duplicate.
- Note that it is *not* safe to assume that a new chest radiograph exam with a timestamp identical to a previous exam is a duplicate. There are many reasons why a patient may have multiple different chest radiographs with identical timestamps. As such, you should use the criteria above to determine if an exam is a duplicate.

Please note that this issue is separate and distinct from the "pediatric radiographs" issue above.

Questions?

If you have any questions regarding this data, please feel free to reach out to the HIRO's general questions email (hirohelp@bsd.uchicago.edu) or the HIRO's Technical Director (Nick Gruszauskas, ngruszauskas@bsd.uchicago.edu).