**RT-HaND\_I Strategy Summary for maintenance and growth of the database: Post-EPIC Cohort Data Curation**

This document lists a brief overview of the strategy to build prospective patients into the RT-HaND\_I. The prospective cohort comprises of patients seen by the HNC oncology team after 4/10/23, or those seen before but who did not complete their radiotherapy course until after 4/10/23. For full inclusion/exclusion criteria please see the full clinical dataset build SOP document (RT-HaND- 015 RT-HaND\_C Retrospective Database Build SOP).

**Step 1: Prospective Patient Identification**

Prospective patients are identified in line with the procedure described in RT-HanD 019a: RT-HaND\_C Strategy Summary for maintenance and growth of the database: Prospective Cohort Data Curation.

Patients returning for further treatment and already existing on the retrospective cohort list can be considered as retrospective patients for the purposes of imaging and radiotherapy data ingestion. This is because they will have already had all of their historic imaging updated.

**Step 2: Ingesting from PACS**

To update the lake for prospective patients, PACS should be queried for all studies available for these patients. The descriptions should be used to select the studies for ingestion. There is a list of agreed *include* and *exclude* studies to use to initially filter the list. This can be found in 012a: Imaging Study Descriptions known Include vs Exclude criteria. It is likely that not all study descriptions will be included on either the *include/exclude* list. This is due to updates in scanning protocols and new machines. A clinician should be consulted regarding the new study descriptions and the *include/exclude* lists updated accordingly. With imaging studies to be ingested filtered out, CSC can then ingest via REST-API with access to the PACS-XNAT pipeline given with respect to clinical priorities.

The basic flowchart for this querying process is shown in Figure 1.

RT-HaND- 011 Transfer of Imaging Data SOP details how the data availability document should be updated accordingly after ingestion.

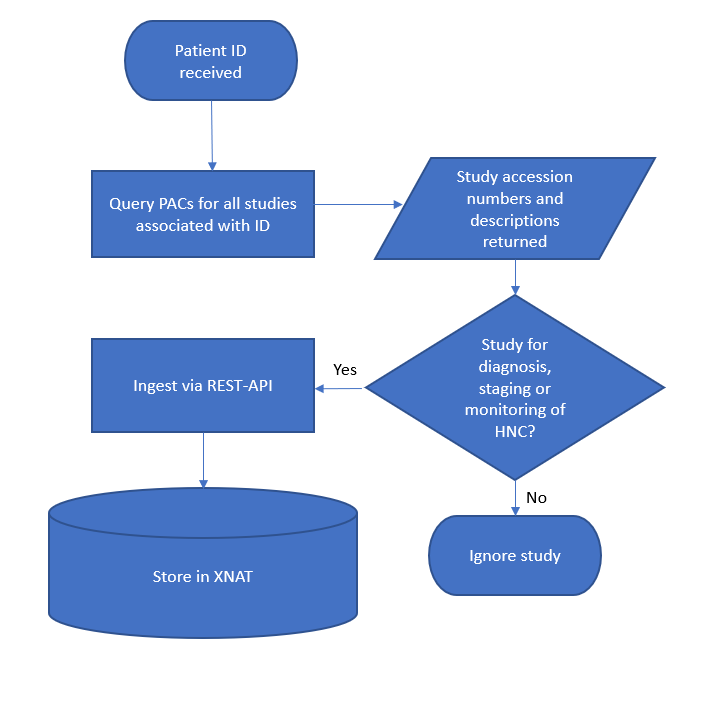


Figure 1: PACS ingestion flowchart

**Step 2: Ingesting from Aria**

To update the lake for previous retrospective patients, Aria should be queried for all studies. The studies should then be named according to RT-HaND-013 Transfer of Radiotherapy Data SOP and ingested into XNAT.

In the ingestion of the retrospective cohort most patients were treated in Mosaiq and so details on their prescription, completed fractions and radiotherapy start and end dates were taken from EDW and stored within RT-HaND\_C in EDW as this is treatment information. The patients treated through Aria had this data manually compiled.

Currently CSC are trying to connect Aria’s database containing this prescription and treatment information to a similar warehousing system so that these radiotherapy details do not need to be complied manually. The database containing these radiotherapy details is called “AURA”. Eclipse scripting is another potential method that could locate a lot of these pieces of data to minimise manual data curation in future.

**Step 3: Updating the “retrospective patient list”**

For the purposes of RT-HaND\_I, after the initial ingestion of these prospective patients into RT-HaND\_I these patients can now be considered “retrospective” patients. in that the next time the PACS and Aria databases are queried, there is no need to query all available data for these patients, only imaging and radiotherapy sessions after the date of the last ingestion and providing the patients are still alive. This will prevent data duplication within XNAT. XNAT does flag to the user if ingested scans are identical to previous scans (name and UID) and these studies can be deleted.