

Established principles of classified image collecting

Image collection

The system collects all the images from the database for each day.

Image collection is limited by Picture Archiving and Communication System (PACS) computed tomography modality.

Range of selected patients is limited by ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision) classification. Chosen lung cancer disease codes are:

- C33 Malignant neoplasm of trachea
- C34 Malignant neoplasm of bronchus and lung
- C34.0 Main bronchus
- C34.1 Upper lobe, bronchus or lung
- C34.2 Middle lobe, bronchus or lung
- C34.3 Lower lobe, bronchus or lung
- C34.8 Overlapping lesion of bronchus and lung
- C34.9 Bronchus or lung, unspecified
- C37 Malignant neoplasm of thymus
- C38 Malignant neoplasm of heart, mediastinum and pleura
- C38.0 Heart
- C38.1 Anterior mediastinum
- C38.2 Posterior mediastinum
- C38.3 Mediastinum, part unspecified
- C38.4 Pleura
- C38.8 Overlapping lesion of heart, mediastinum and pleura
- C39 Malignant neoplasm of other and ill-defined sites in the respiratory system and intrathoracic organs
- C39.0 Upper respiratory tract, part unspecified
- C39.8 Overlapping lesion of respiratory and intrathoracic organs
- C39.9 Ill-defined sites within the respiratory system

Servers used in image collection

1. VUOI range
 1. Image database in PACS server
 2. VUOI database server
 3. IHSIT-CD Application Server
2. VU MIF range
 1. VU MIF database server
 2. WEB server

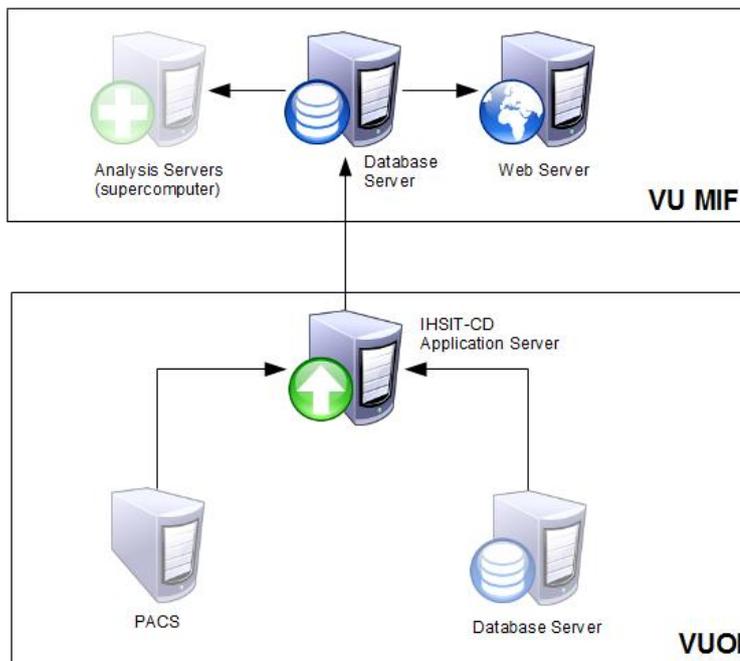


Diagram 1 Servers

Algorithm for image and data collection

Technologies used for image and data collection:

- CloverETL
- DICOM Pixies IO
- MySQL

One iteration designed for one day.

Input data:

1. Study date

Output data:

1. Anonymized DICOM studies with images and metadata
2. Anonymized patient data
3. Anonymized study data
4. Anonymized diagnosis data

Algorithm:

1. Get list of studies by study date from PACS server
2. Get all patients from database server with defined ICD-10 codes
3. Filter studies by selected patients
 - a. Filter key is national identification number
4. Anonymize patient data
5. Anonymize study data
6. Anonymize diagnosis data
7. Download studies by filtered study list
8. Anonymize downloaded studies
9. Save anonymized patients, study and diagnosis data to database
10. Save anonymized patients, study and diagnosis data to files
11. Zip all files and studies

CloverETL algorithm for data collecting and anonymizing:

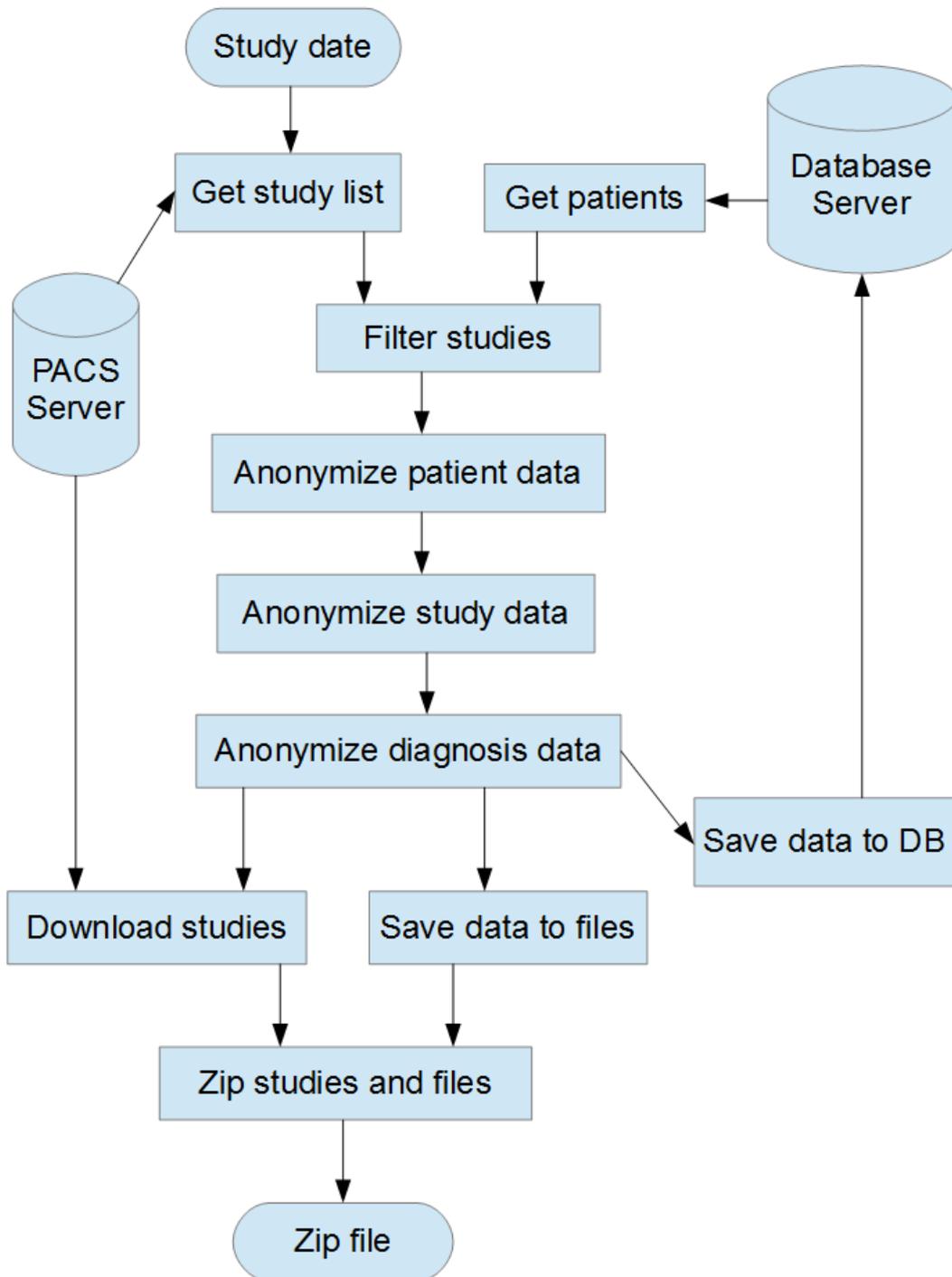


Diagram 2 Algorithm