



This programme is funded by the European Union



Latvia, Lithuania and Belarus  
Cross Border Cooperation Programme  
within European Neighborhood and Partnership  
Instrument 2007-2013

This chapter describes lung cancer image importation procedures

## Content

<b>1. Image collection and importation</b>	<b>page 2</b>
<hr/>	
Background information about methods used in image collection and importation	
<b>2. Servers used in image collection and importation</b>	<b>page 3</b>
<hr/>	
Data collecting scheme and servers architecture	
<b>3. Algorithm for image importation and data collection</b>	<b>page 4-5</b>
<hr/>	
Description procedure for data importation from oncology practitioners and annotation	

---

## Image collection and importation

The system collects all the images and medical data from the ambulant database implemented on MySQL and Picture Archiving and Communication System (PACS) for each day. PACS system is a typical picture archiving implementation with archive repository, where all images from medical devices resides.

Image collection is limited by 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 and importation

### 1. NCC (VUOI) range

1. Image database in PACS server (see chapter “Realized functionality of COU”)
2. VUOI database server. Database stores:
  - personal patient information (Name, personal code, ...)
  - diagnosis,
  - gender,
  - patient ID,
  - age,
  - International Classification of Diseases (ICD-10). E.g. C34.2,
  - diagnosis date,
  - all forum information.
3. IHSIT-CD Application Server. Depersonalizes patient data.

### 2. VU MIF range

1. VU MIF database server (MySQL). Stores depersonalized patient data from VUOI DB and medical images with additional data and records. Additional data:
  - Tags for/from images,
  - Oncoexpert system users.
2. WEB server. It's an End-user interface presented via web portal where authorization is implemented via Gmail, Yahoo, Facebook, LinkedIn accounts.

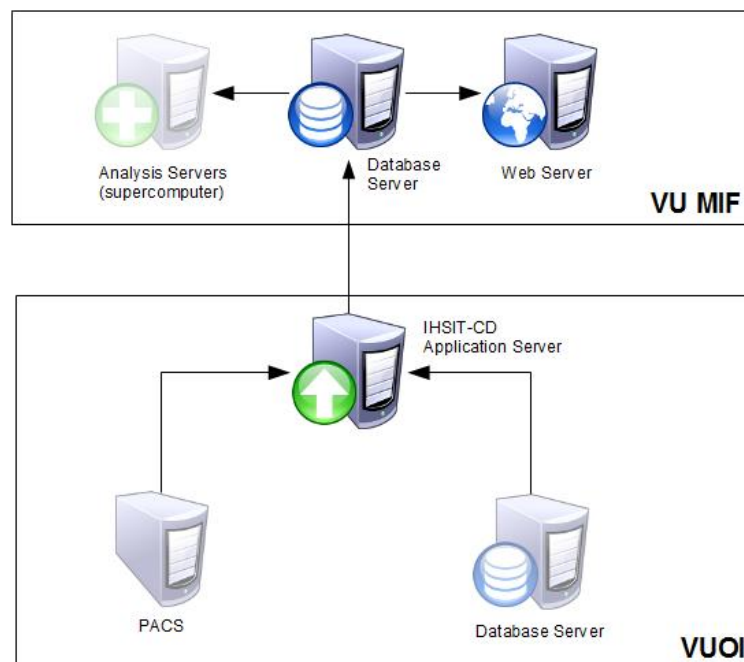


Diagram 1: Servers

---

## Algorithm for image importation and data collection

Description procedure for data importation from oncology practitioners and annotation.  
Technologies used for image and data collection:

- CloverETL. It is an end-to-end data processing and integration solution. Visual tool and IDE enabling data integration design, development, and runtime. Automation, monitoring, and orchestration platform for high performance data integration that includes workflow management and automation.
- DICOM Pixies IO. It is a plug-in that connects the DICOM® standard to the Java™ standard. DICOM is the universal standard for sharing medical imaging resources between heterogeneous and multi-vendor equipment (acquisition device, workstation, storage server, patient management system, etc.)
- MySQL. Open source, relational database management system. Database stores processed medical images.

One iteration designed for one day.

Input data:

Study date

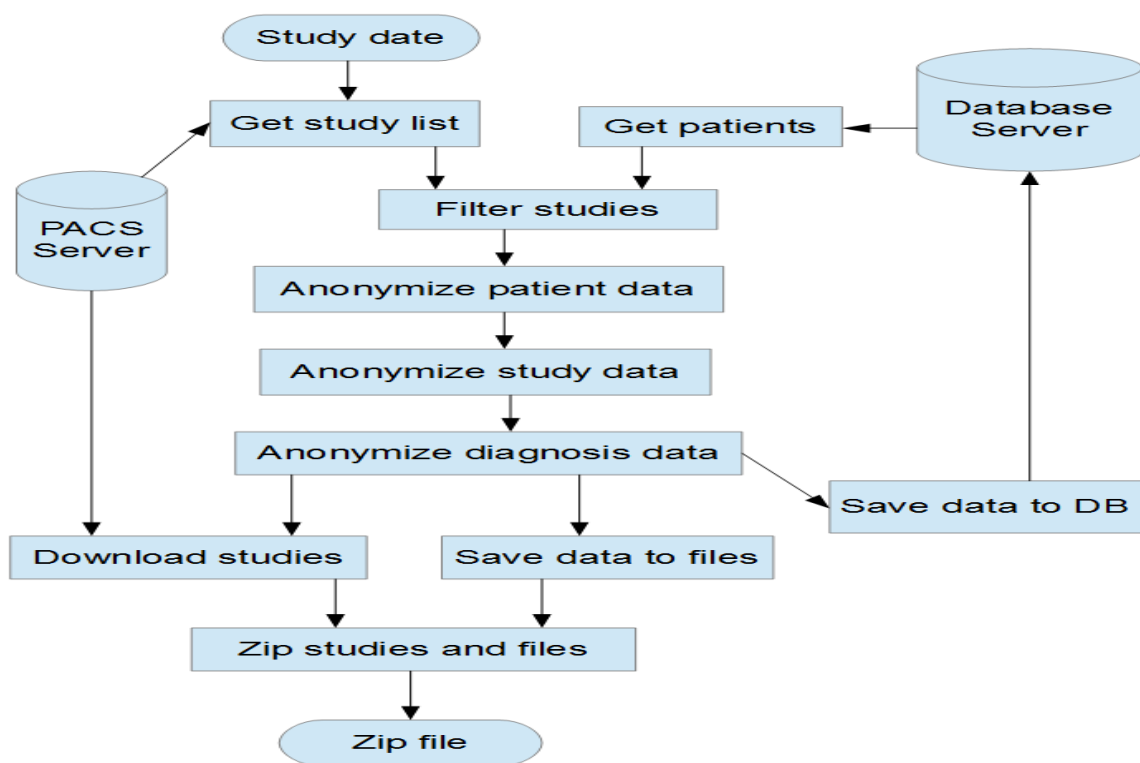
Output data:

- Anonymized DICOM studies with images and metadata (lots of technical information about image)
- Anonymized patient data (gender, patient ID)
- Anonymized study data (patient ID, patient age at the time of the study, Dicom image ID, study ID)
- Anonymized diagnosis data (ICD-10 codes, other technical ID's)

Algorithm:

- Get list of studies by study date from PACS server
- Get all patients from database server with defined ICD-10 codes
- Filter studies by selected patients
  - Filter key is national identification number
- Anonymize patient data
- Anonymize study data
- Anonymize diagnosis data
- Download studies by filtered study list
- Anonymize downloaded studies
- Save anonymized patients, study and diagnosis data to database
- Save anonymized patients, study and diagnosis data to files
- Zip all files and studies

CloverETL algorithm for data collecting and anonymizing:



**Diagram 2: Algorithm**