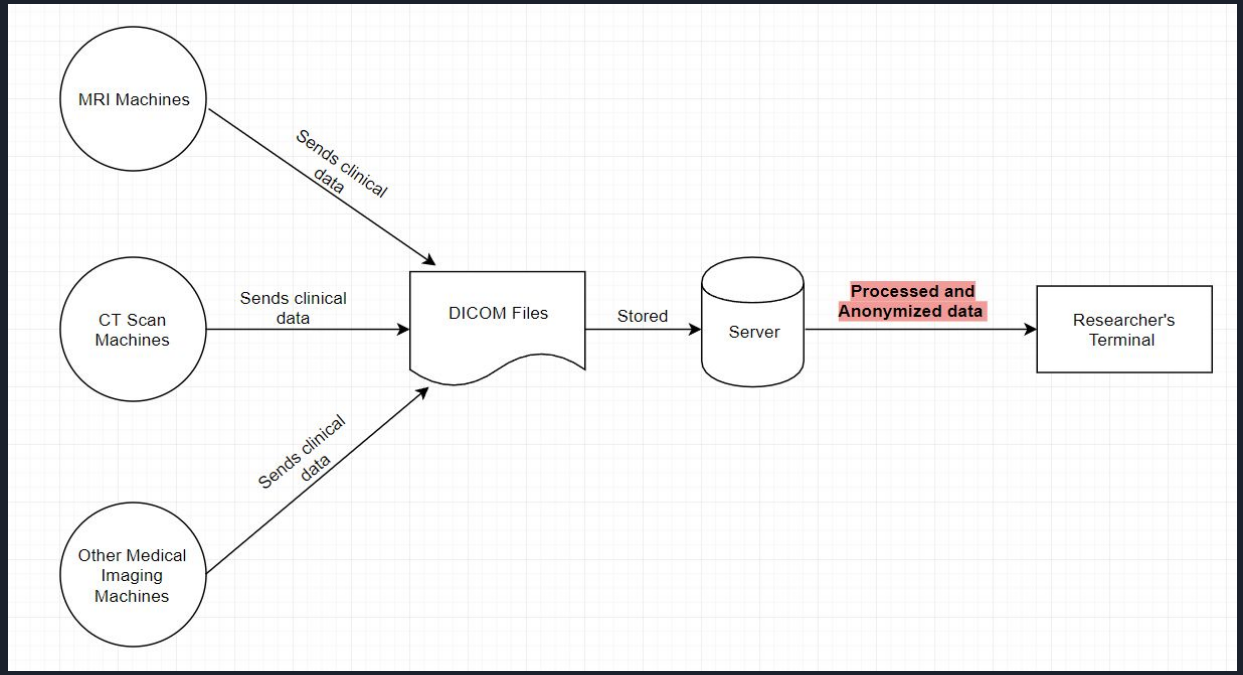
Medical Image De-Identification

Ben Wilson, Nigam Lad, Yidu Guo

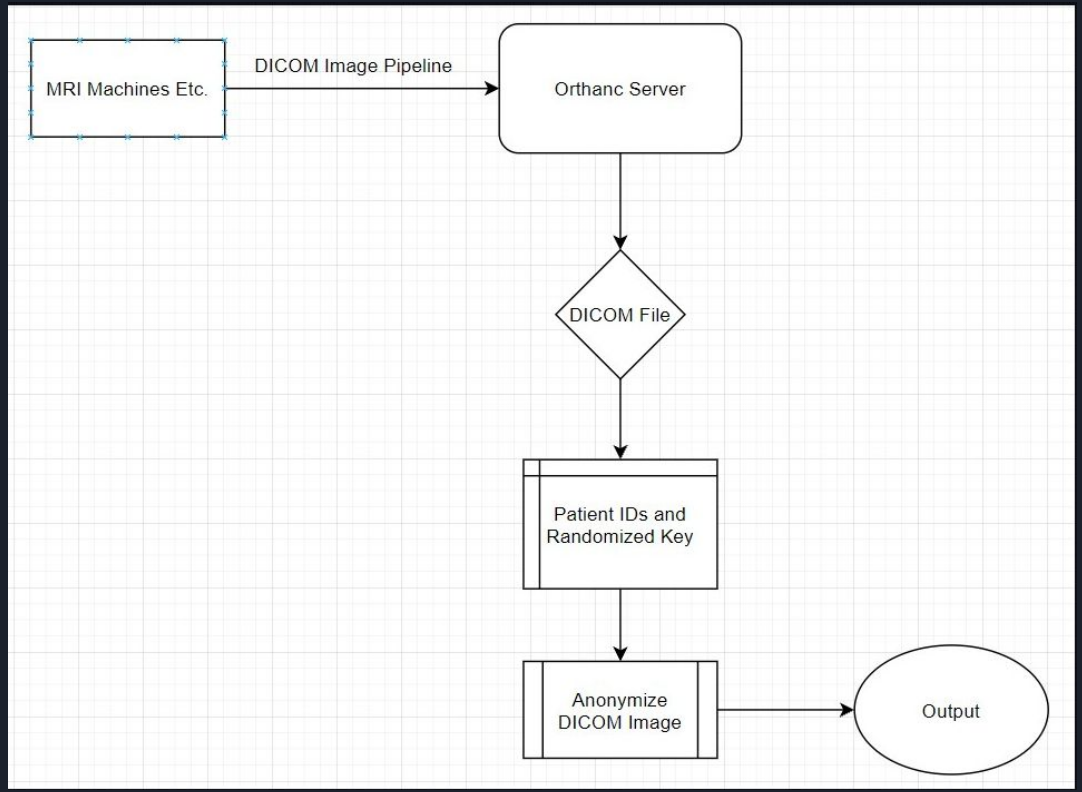
Project Description

- Medical research of the modern world is reliant on data
- AI has become an invaluable asset to the medical research field
- A major concern for researchers is lack of access to clinical data
- Introduces a pipeline that streamlines secure access to medical data
- Seek to anonymize real patient data from clinical studies

Level 0 Data Flow Diagram



Level 1 Data Flow Diagram



Functional Requirements Milestones

- Milestone 1
  - Complete Data Flow Diagram (DFD) level 0 and level 1
  - Complete Software Requirement Specification Document (SRS)
  - Set up weekly meeting with client
- Milestone 2
  - Set up Orthanc environment
  - Set up server
  - Determine language (Python or C)
- Milestone 3
  - Complete all functional requirements
  - Complete Peer testing # 1
- Milestone 4
  - Complete peer testing #2
  - Project deliverance

Non-Functional Requirements and Constraints

- Security
  - DICOM file before de-identification and anonymizing contain personal health information (PHI)
  - Access to anonymized data must be limited and monitored
- Reliability
  - Robust design to prevent breaches on PHI (Personal Health Information)
- Maintainability
  - Admin needs to able change the perimeter of which part of DICOM file is to be anonymized

- Lack of access to facility server for testing, as it contains PHI

Tech Stack

- Python

- Packages include:

  - Pandas

  - Numpy

  - Others

Testing

- Handling of mass amounts of data
- Feed DICOM file containing sample information through the plug in
- This method will reveal desired information on if the DICOM file is being anonymized

Conclusion

- Currently only one functional deliverable
- On completion of first deliverable, client will assign more functional deliverables.