



# Data Sharing and Data De-Identification

---

David Rodriguez Gonzalez  
(NeSC/SBIRC)

On behalf of the SINAPSE collaboration



The following Universities are charitable bodies, registered in Scotland, with registration numbers as below.





# Contents

---

- The SINAPSE project
- MIDAS meeting
- PrivacyGuard DICOM de-identification toolkit
- MRI QA centralised storage and analysis
- Normative Brain Imaging Bank prototype

# Contents

---

- **The SINAPSE project**
- MIDAS meeting
- PrivacyGuard DICOM de-identification toolkit
- MRI QA centralised storage and analysis
- Normative Brain Imaging Bank prototype

# The SINAPSE Project

---



- Stands for ***Scottish Imaging Network: a Platform for Scientific Excellence***.
- Pooling initiative of six Scottish universities: Aberdeen, Dundee, Edinburgh, Glasgow, St. Andrews and Stirling.
- Main objectives:
  - develop imaging expertise,
  - support multi-centre clinical research in conjunction with the Clinical Research Networks,
  - improve the ability of neuroscientists to collaborate on clinical trials,
  - have a direct impact on patient health.

# Centralised Architecture (pros & cons)

---

- Simpler Deployment
- Easier middleware release control
- Lesser impact in participant centres
- Easier to manage and use
- No default resilience
  - A second centre would be needed
  - But this is only necessary for critical services
  - With a good support a reasonable service can be provided using a single centre

# Deployment

---

- ECDF (Edinburgh Compute and Data Facilities)
  - <http://www.is.ed.ac.uk/ecdf/>
    - A singular facility along Scotland
- Disk space and CPU time can be rented
  - 1456 CPU cores
  - 275 TB of disk
- Also SINAPSE server hosted by ECDF:
  - For hosting SINAPSE specific services
  - ECDF provides basic hardware + software support

# Contents

---

- The SINAPSE project
- **MIDAS meeting**
- PrivacyGuard DICOM de-identification toolkit
- MRI QA centralised storage and analysis
- Normative Brain Imaging Bank prototype

# MIDAS meeting (18<sup>th</sup> March 2009)

---

- **M**edical **I**maging **D**ata **A**ccess and **S**haring
- Hosted in the e-Science Institute
- Brought together representatives from the NHS Scotland & the universities
- Successful meeting with useful discussion
  - Came out with a roadmap for improving the data sharing between both sides
  - Report circulated between attendees





# MIDAS contents

---

- SINAPSE
- National PACS
- Connectivity solution in Edinburgh
- Discussion about
  - Data Transfer
  - Data De-identification
  - Information Governance



# MIDAS: Action Points

---

- National minute of agreement
- Exemplar models: adapt to local constraints
- Research representative in the NHS PACS board
- Need for a National “champion”
- Summarise the barriers and potential solutions
- Exiting local NHS-HE image data transfer solutions
- Exemplar research questions
- Feed the key points from MIDAS into CSO

# MIDAS: local solutions report

---

- Very diverse situation in the SINAPSE centres along Scotland
- Aberdeen: recently entered the NHS PACS
- Dundee: anonymised CD-burning for exporting images
- MoU in place in Edinburgh. Scanners in no-man's land
- Glasgow: mechanical attachment/detachment

# MIDAS 2

---

- Next meeting to be hosted in Sept by Prof Edwin van Beek (SINAPSE Chair) and Iain Robertson (NHS Lothian)
- Meeting with e-Health group for NHS to further discuss the IT/PACS issues

# Contents

---

- The SINAPSE project
- MIDAS meeting
- **PrivacyGuard DICOM de-identification toolkit**
- MRI QA centralised storage and analysis
- Normative Brain Imaging Bank prototype

# PrivacyGuard

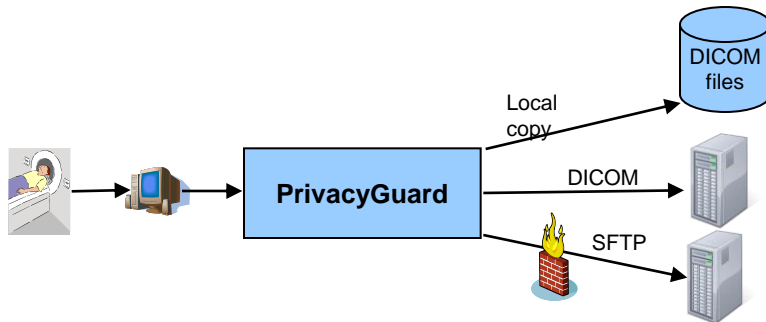
## A DICOM De-Identification Toolkit

David Rodríguez González<sup>1,2</sup>, Trevor Carpenter<sup>3</sup>, Jano van Hemert<sup>1,2</sup> and Joanna Wardlaw<sup>1,3</sup>

<sup>1</sup> SINAPSE <sup>2</sup> National e-Science Centre, School of Informatics, University of Edinburgh. <sup>3</sup> SFC Brain Imaging Research Centre, Division of Clinical Neuroscience, University of Edinburgh.

### Medical Imaging for Research and Teaching

Medical imaging acquired primarily for patient diagnosis is also of value for research and teaching, but the patient privacy must be respected. PrivacyGuard is a flexible DICOM de-identification toolkit that facilitates personal data protection for different scenarios.

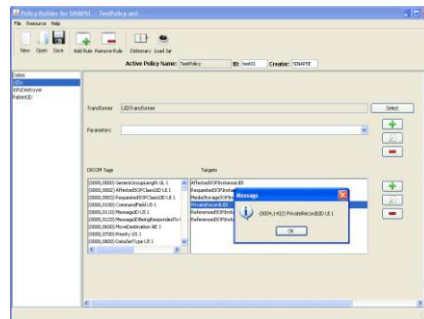


### Results

- Tested with DICOM objects from different modalities.
- Deployed in several centres in Scotland.
- Application and source code repository accessible from <http://research.nesc.ac.uk/privacyguard>
- Publication in *European Radiology 2010* (in press, already available online):  
"An open source toolkit for medical imaging de-identification"  
<http://dx.doi.org/10.1007/s00330-010-1745-3>

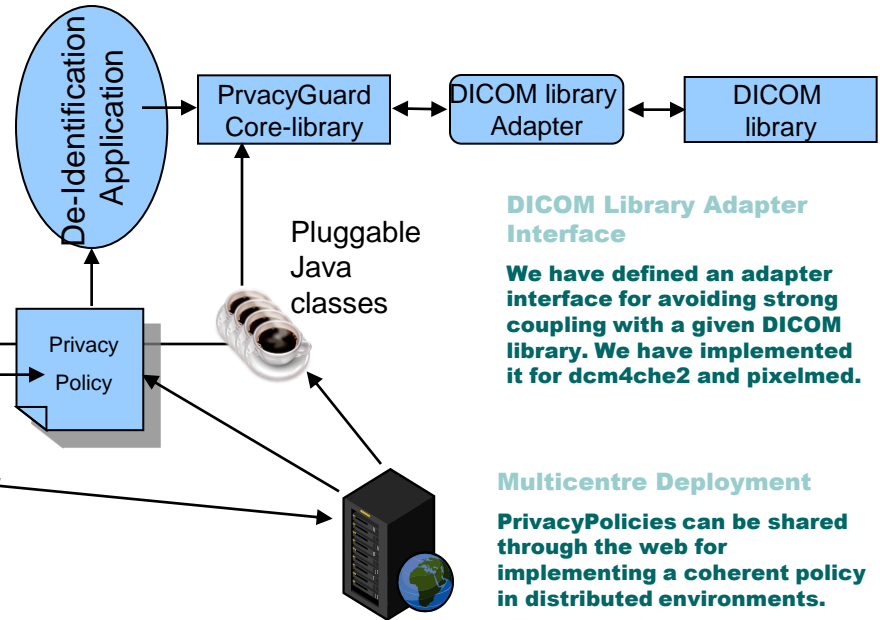
### PrivacyGuard Main Features

- Extensible: pluggable classes for anonymising DICOM headers and pixel data
- DICOM standard Attribute-level confidentiality mechanism implementation
- Adaptable processing pipeline allows for extra operations to be added by users
- The anonymous output can be routed to other equipment using either DICOM send and SFTP
- And also archived locally with a flexible path building mechanism.



### Policy Editor

A GUI to edit the Privacy Policies that incorporates a DICOM dictionary to ease the users work. Can inspect jar files loaded at runtime in search of Transformer classes.



### DICOM Library Adapter Interface

We have defined an adapter interface for avoiding strong coupling with a given DICOM library. We have implemented it for dcm4che2 and pixelmed.

### Multicentre Deployment

PrivacyPolicies can be shared through the web for implementing a coherent policy in distributed environments.

# PrivacyGuard

---

- A DICOM de-identification toolkit
  - Implemented in Java
  - Highly configurable
  - Configurable pipeline for chaining different operations
- Privacy Policies expressed in XML documents
  - PolicyEditor: a graphical policies creation tool
- Transformation classes distributed in signed jar files
- DICOM read/write through an interface that allows using different libraries
  - dcm4che2
  - pixelmed

# Privacy Policies

---

- XML documents containing the rules for anonymising the data
- A rule specifies:
  - The target fields
  - The class used for the transformation including:
    - Version
    - Digest
    - Resource (jar file)
  - Parameters

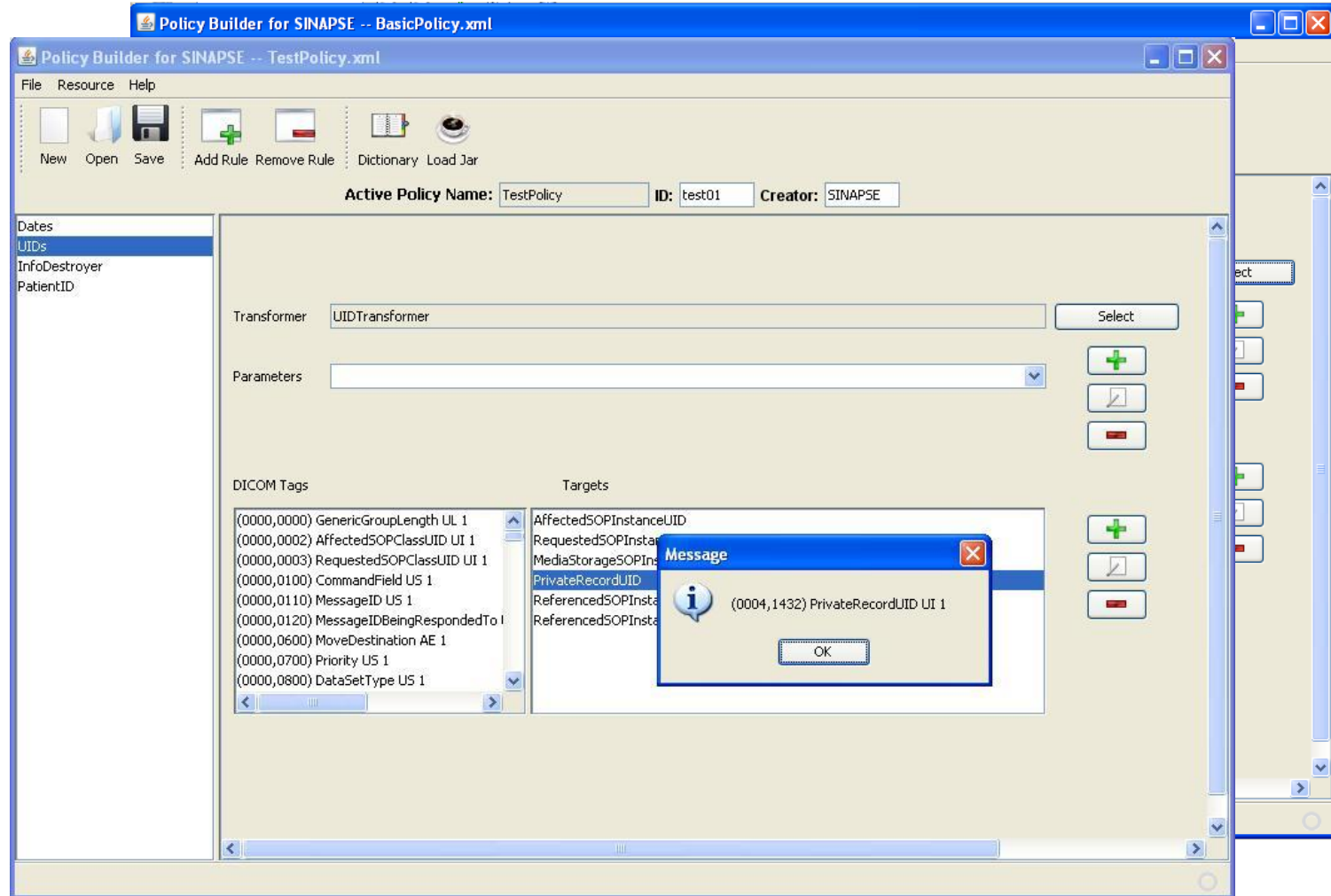


# Policy Editor

---

- A policy creation graphical tool
- Includes a DICOM dictionary
- Can look for “Anonymisation” classes in jar files
- Can sign and check the signatures of policies

# Policy Editor



# Contents

---

- The SINAPSE project
- MIDAS meeting
- PrivacyGuard DICOM de-identification toolkit
- **MRI QA centralised storage and analysis**
- Normative Brain Imaging Bank prototype

# MRI QA in SINAPSE

---

- QA is used to monitor the performance of MRI scanners
  - particularly important in multicentre imaging studies
- Previous work in SINAPSE towards establishing a common QA protocol
  - 7 participant MR scanners in 4 centres
  - Framework for monitoring the quality of the data
  - It will facilitate the combination of data between centres

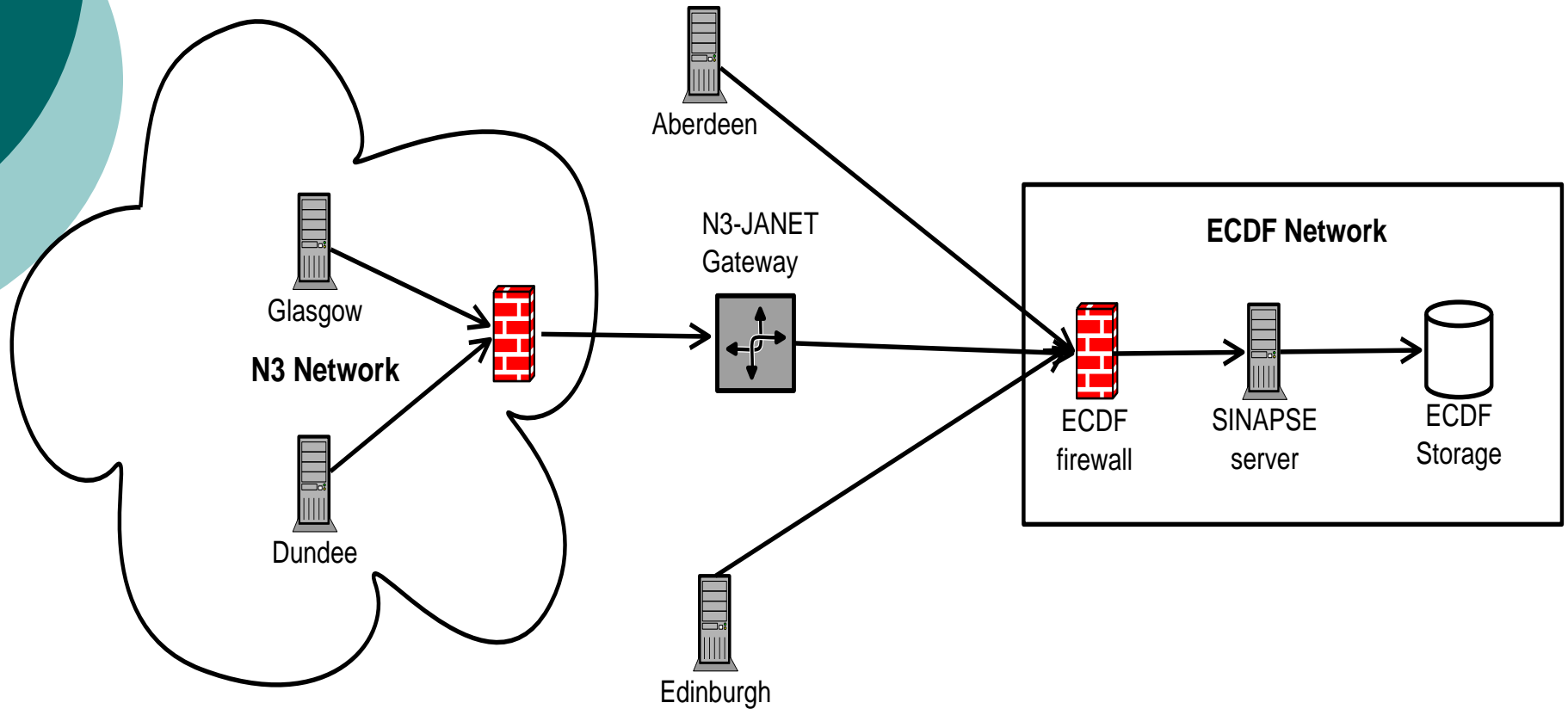


# Motivation for an automatic system

---

- Remove the burden of some manual tasks currently being done in the centres
- Allow checking the correctness of the sequence parameters used
- Ensure the consistency of the software used for the analysis and
- Facilitate the reanalysis of the data
- Enforce (pseudo-)anonymisation policies across collaborations

# Network Configuration

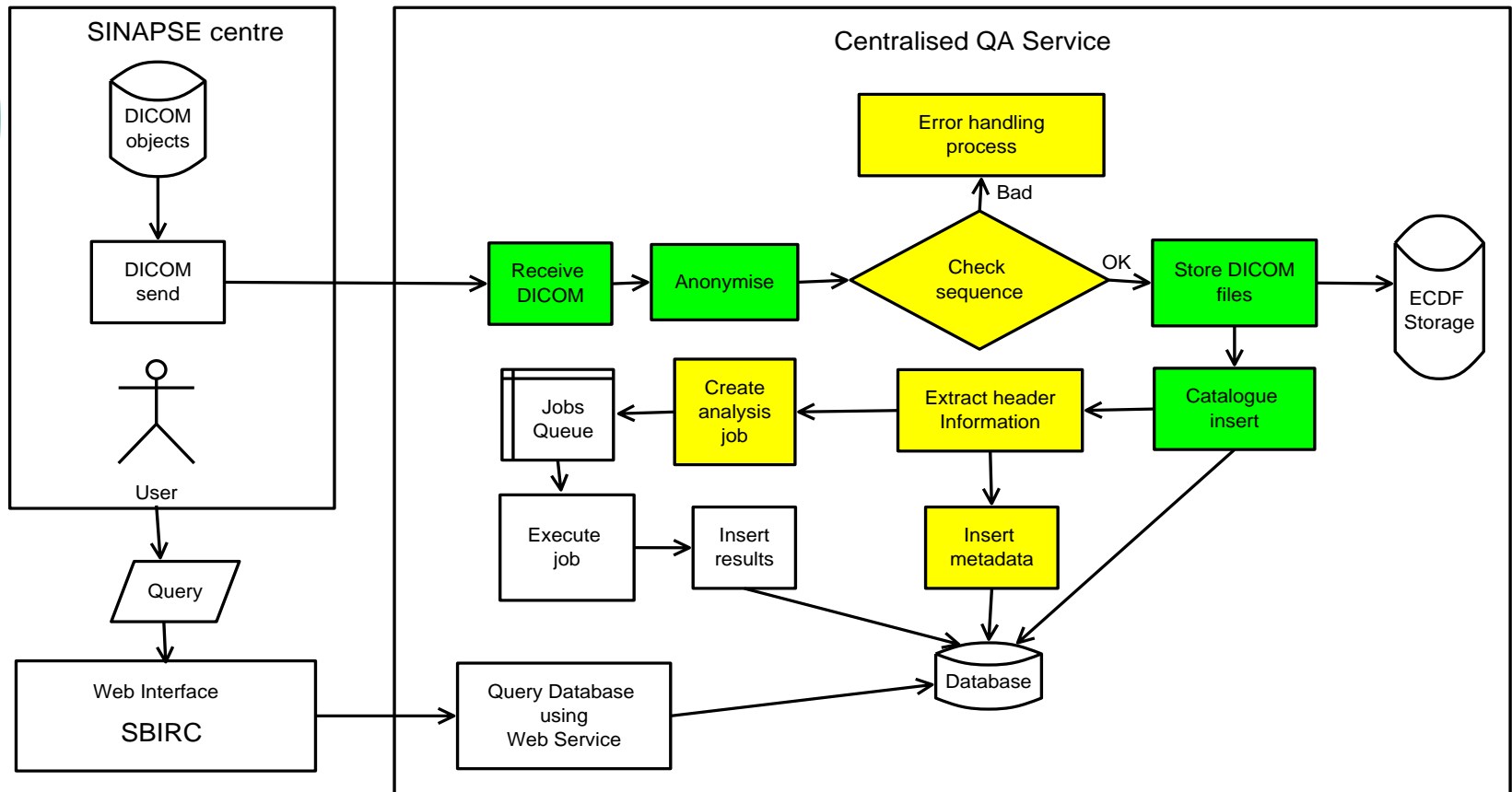


# Networking Issues

---

- MRI scanners are connected to N3 (NHS network) in some SINAPSE centres
  - Glasgow & Dundee
  - Application included in the N3-JANET gateway
  - Gateway already configured
- And port open in ECDF firewall
  - For the gateway?
  - Limited number of machines from the Universities
  - But still some connectivity issues

# MRI QA flowchart & PrivacyGuard



- PrivacyGuard
- PrivacyGuard extensions



# Data storage and Analysis

---

- After checking the sequence
  - DICOM data is stored in ECDF
  - An entry is inserted in SINAPSE catalogue
- Predefined information is extracted from the header and inserted in the QA database
- An analysis job is created
  - Executed asynchronously
  - The results are also inserted in the QA database
- A web application is used to monitor the QA parameters evolution
  - Accessible from all SINAPSE centres
  - Uses the QA database as backend

# Extending PrivacyGuard

---

- PrivacyGuard provides a mechanism for adding new functionalities
- We are using it to:
  - Check the sequence correctness
  - Extract metadata from the DICOM header and insert it into the QA database
  - Create the analysis jobs

# Contents

---

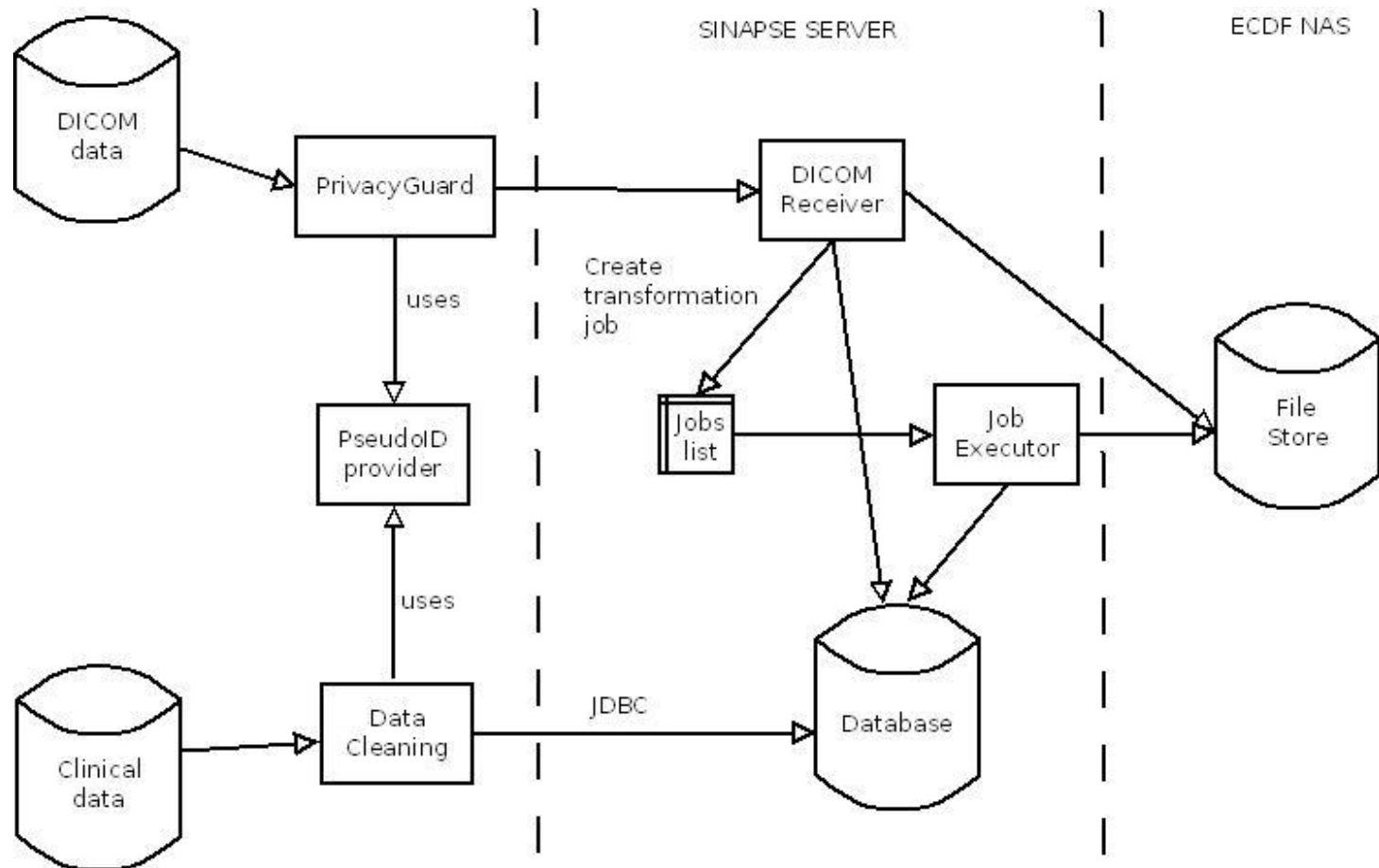
- The SINAPSE project
- MIDAS meeting
- PrivacyGuard DICOM de-identification toolkit
- MRI QA centralised storage and analysis
- **Normative Brain Imaging Bank prototype**

# Image Bank

---

- Pilot project for a Normative Brain Imaging Bank using SINAPSE infrastructure
  - Server for databases
  - ECDF storage space
  - Portal
- Includes clinical and cognitive data along with imaging data
  - Cleaning process required before importing it to the centralised system

# Data cleaning and import process



# Other eScience activities in SINAPSE

---

- Portals:
  - Using Rapid a portlets building technology developed at NeSC
  - Proof-of-concept prototype last summer
  - A production portal to be deployed at ECDF is being built now
    - ECDF and EPCC collaborating
    - A general solution for portal single sign-on authentication to the cluster in place
- GPGPUs
  - Implementation of deconvolution algorithms for brain perfusion imaging (Fan Zhu, SINAPSE PhD student)

# Summary

---

- SINAPSE project wishes to impulse data sharing
  - MIDAS meeting
- PrivacyGuard DICOM anonymiser developed by SBIRC/NeSC
- Exemplar projects:
  - Centralised MR QA data storage & analysis
  - Brain Image bank prototype

# Questions

---

