



# Brain Imaging: data sharing & privacy protection

---

David Rodríguez González  
Data Intensive Research  
Brain Research Imaging Centre  
University of Edinburgh

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

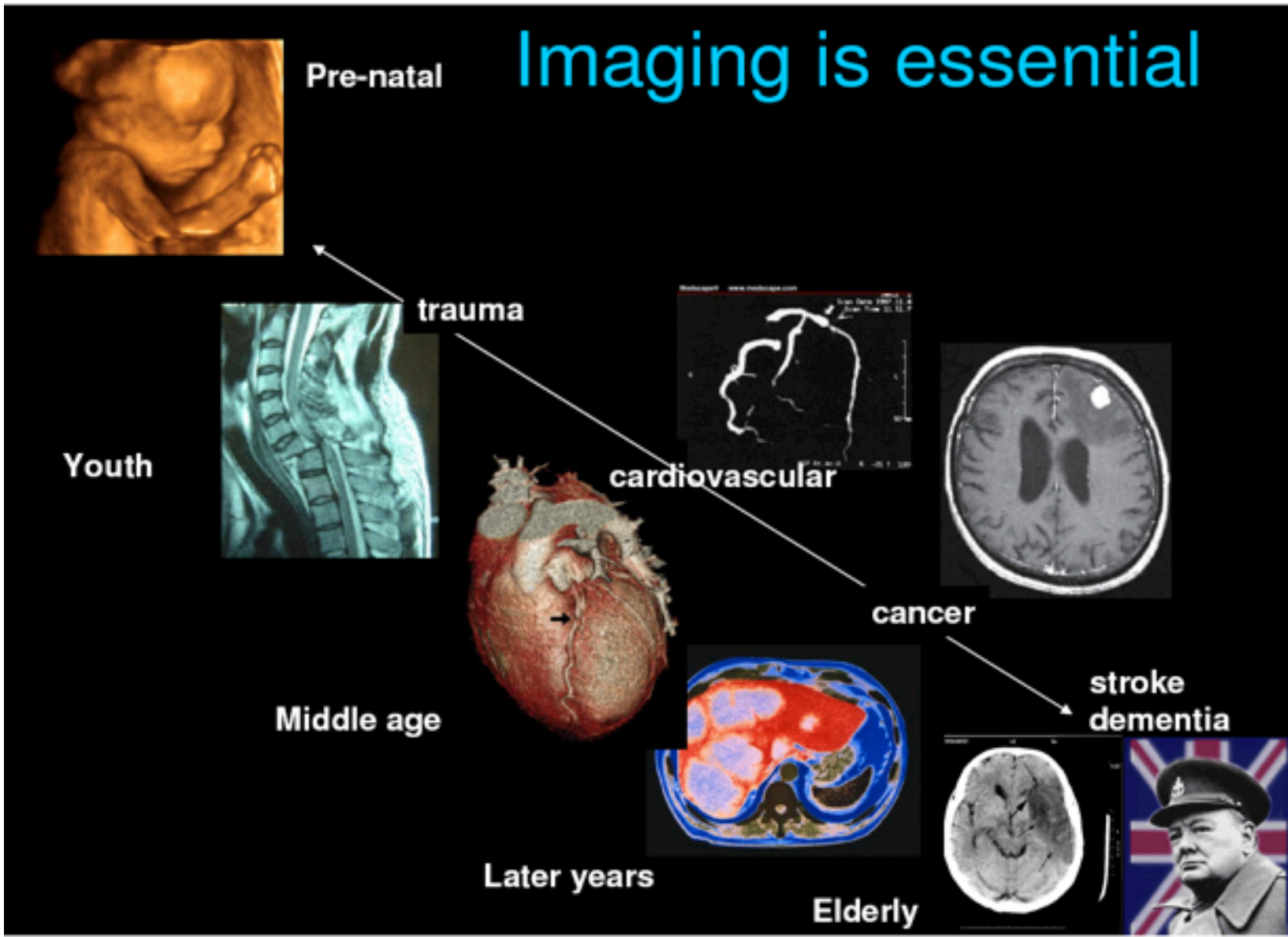


# Contents

- Introduction
- SINAPSE
  - DICOM
- BRAINS
- Information governance & data protection
  - Data Protection Act
  - DICOM de-identification
  - PPDP

# Contents

- **Introduction**
- SINAPSE
  - DICOM
- BRAINS
- Information governance & data protection
  - Data Protection Act
  - DICOM de-identification
  - PPDP



## Massive expansion in research imaging

---

- All branches of medicine – particularly brain
- Not just medicine – psychology, linguistics, engineering, parapsychology, etc.
- In Scotland too!!!
  - 8% UK population
  - 12.5% of all highest rated departments.
  - Highest concentration of biotech in Europe
- Neuroscience – much larger than NIH
  - But in 2006 there were machines, pockets of excellence, but little cohesion

# Contents

- Introduction
- **SINAPSE**
  - DICOM
- BRAINS
- Information governance & data protection
  - Data Protection Act
  - DICOM de-identification
  - PPDP

# 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.

# DIR involvement

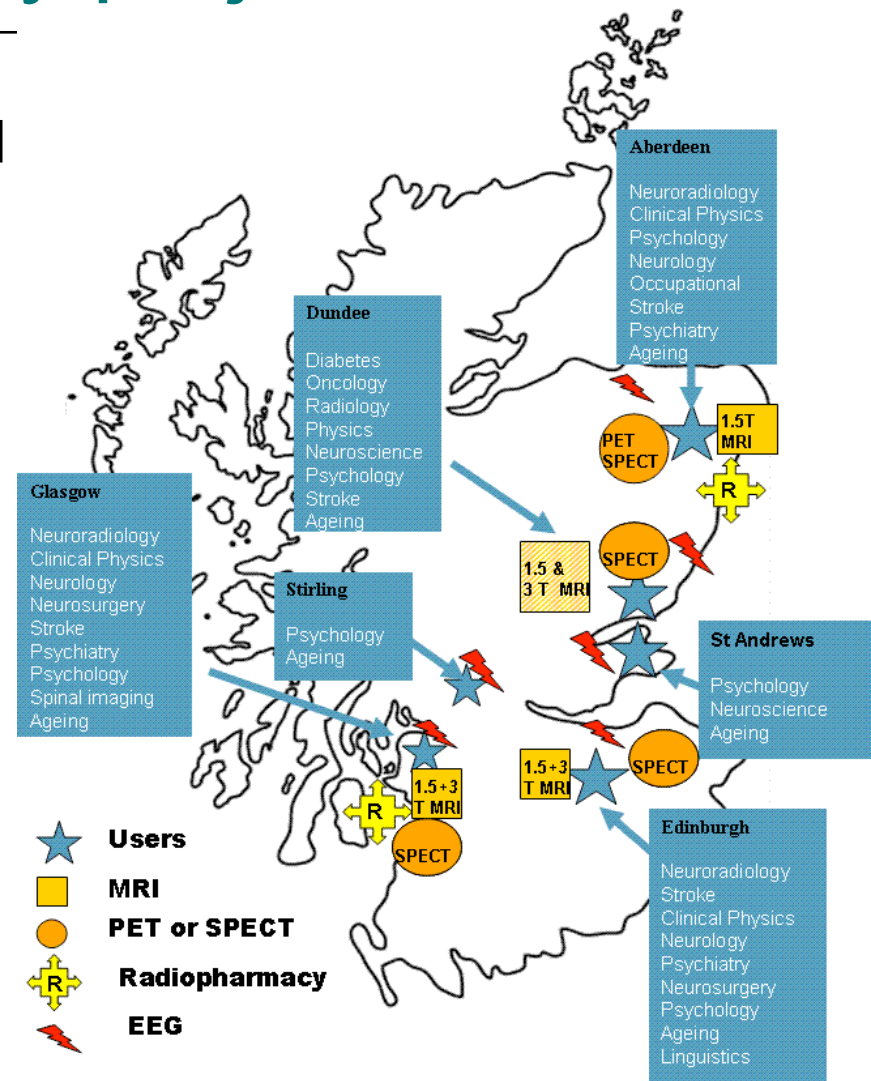
---

- Information governance & data de-identification
  - Networking
  - Development of de-identification tool
- Data sharing infrastructure
  - Facilitating multi-centre studies
- Portal for brain imaging
  - Improving usability
- Image Analysis methods (Fan's talk)



# SINAPSE priority projects

- Stroke, the brain and the blood-brain interface
- Ageing brain to dementia
- Novel molecular imaging markers for major psychiatric disorders
- Innovative radiotracers for CNS inflammation



# Data in SINAPSE

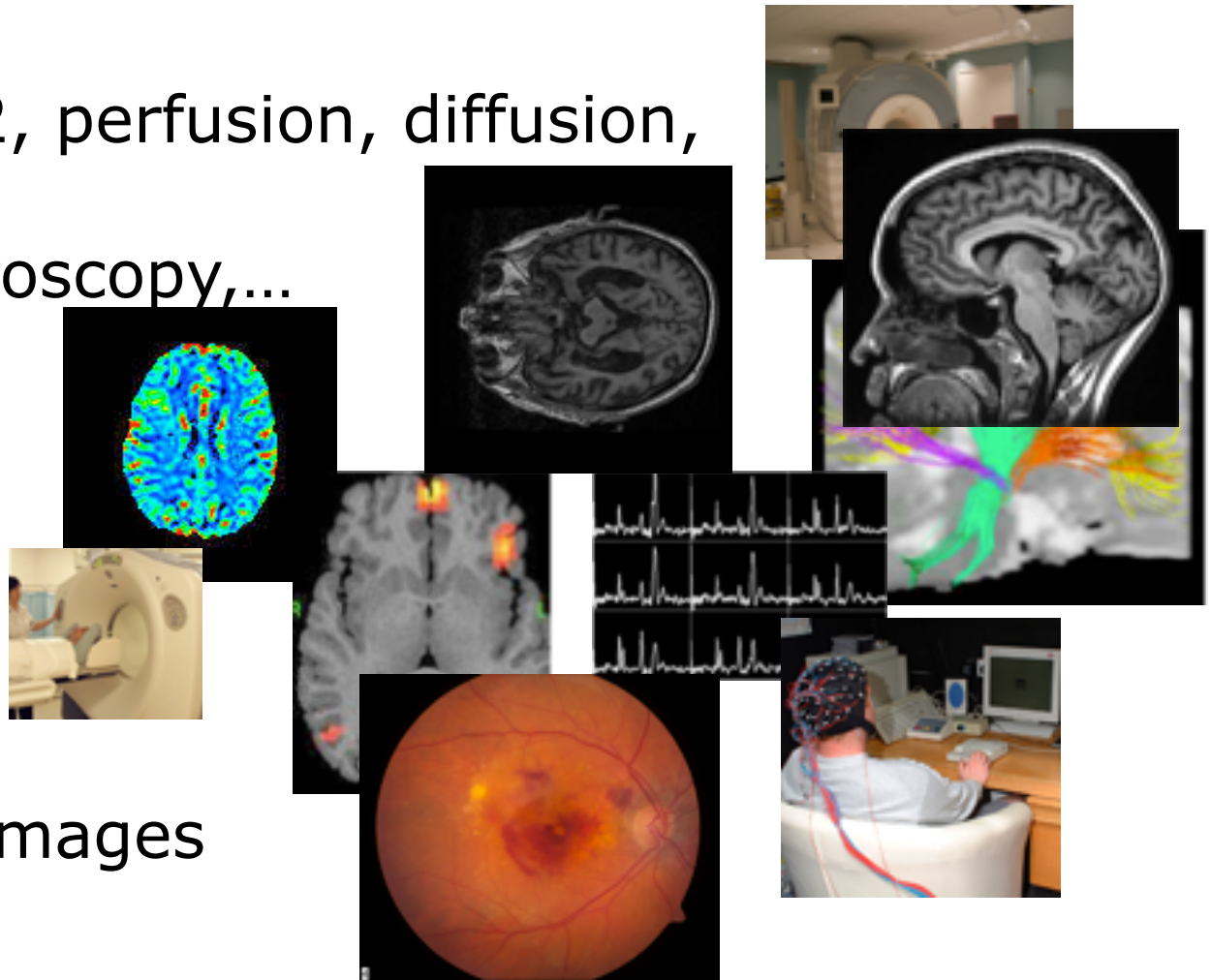
---

- MRI:
  - T1, T2, perfusion, diffusion,
  - fMRI,
  - spectroscopy,...
- CT
- PET
- EEG
- Retinal Images

Images from SINAPSE, NIH and D. Clunie

# Data in SINAPSE

- MRI:
  - T1, T2, perfusion, diffusion,
  - fMRI,
  - spectroscopy,...
- CT
- PET
- EEG
- Retinal Images



Images from SINAPSE, NIH and D. Clunie



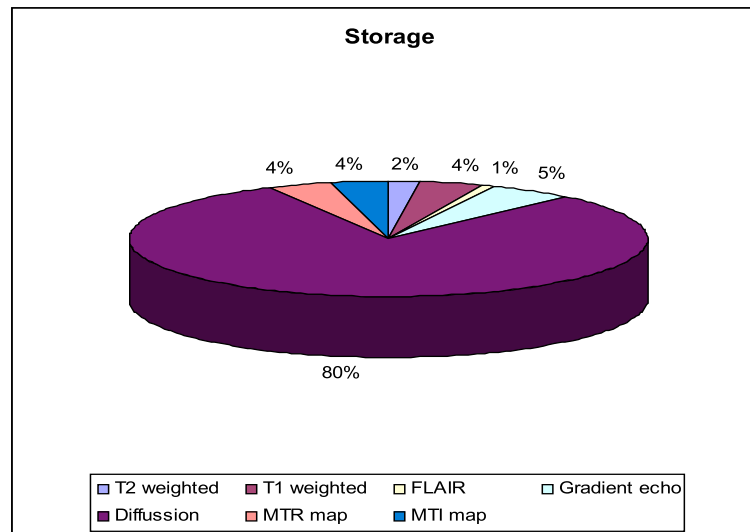
# Data Formats

---

- Raw data
  - Imaging data usually in standard data format (DICOM)
  - EEG & spectroscopy format is manufacturer dependent
- Processed data
  - Varies from project to project

# Storage sizes example (100 subjects)

<i>Sequence</i>	<i>Raw (Gb)</i>	<i>Process.(Gb)</i>	<i>Total (Gb)</i>
T2 weighted	1.09	1.0241	2.1141
T1 weighted	2.14	1.9969	4.1369
FLAIR	0.546	0.51	1.056
Gradient echo <sup>(1)</sup>	4.37	1.024	5.394
Diffusion	21.9	65	86.9
MTR map	1.96	2	3.96
MTI map	1.96	2	3.96
<b>Total</b>	<b>33.966</b>	<b>73.555</b>	<b>107.521</b>



# Digital Imaging and Communications in Medicine (DICOM)



- Standard for handling, storing, printing and transmitting medical imaging information
  - Supports CT, MRI, PET, Nuclear medicine, ultrasound,...
  - Several types of objects: Images, Presentation States, Structured Reports, Encapsulated Objects
- Data format:
  - "Header": includes metadata
  - Pixel data
- Also defines communications, confidentiality profiles, ...



# DICOM Files

---

- Enhanced Multi-frame DICOM CT & MR objects support storing a whole series in a single file
- Unfortunately this is still not widely adopted/supported
  - Thousands of very small files (even of 20KB)
  - Performance problems

# Contents

- Introduction
- SINAPSE
  - DICOM
- **BRAINS**
- Information governance & data protection
  - Data Protection Act
  - DICOM de-identification
  - PPDP



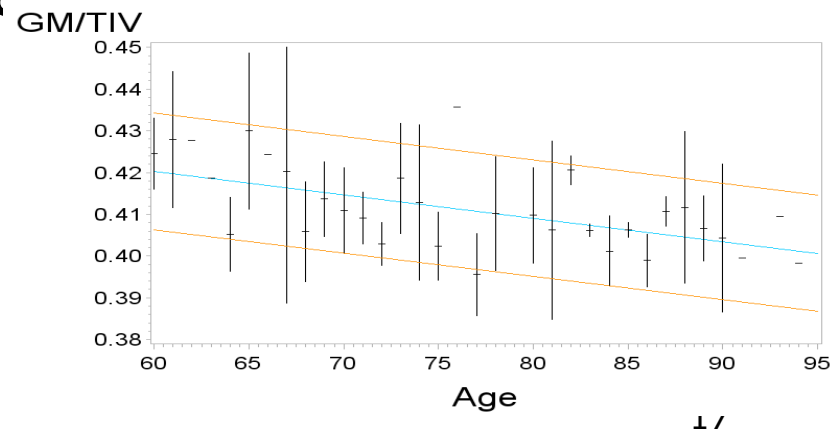
# BRAINS

---

- Pilot project for a Normative Brain Imaging Bank
- Server for databases
  - ECDF storage space
  - Portal
- Includes clinical, demographic and cognitive data along with imaging data
  - Cleaning process required before importing it to the centralised system
  - Looking for collaboration with other databanks

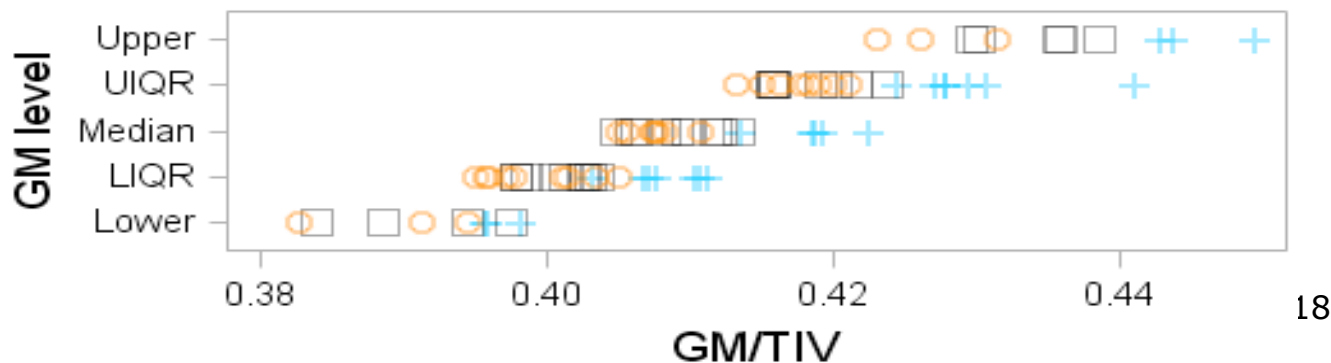
# Statistical Models

- Definition of normal ageing brain structure will improve diagnoses of pathological changes in brain structure associated with ageing
- statistical summaries (like GLM) make assumptions often not met in brain imaging



# Alternative: databank

- A “data model”, based on the whole distribution rather than summary of brain image data,
  - to distinguish normal and pathological ageing brain structure.
  - Preliminary results show that brain tissue loss may actually slow in normal



# Existing Data Banks

- ADNI, AIBL, fMRIC, IXI, OASIS, ICBM, X  
NAT central
- From 9 databanks about 900 (>60  
years) normal subjects
  - Several repeated in different banks
  - Many not representative of normal  
brain ageing
    - 25 different individual, representative  
normal subjects aged 60-69, 201 aged  
70-79, and 101 aged  $\geq 80$  years but in all  
only 98 subjects were openly accessible

# Contents

- Introduction
- SINAPSE
  - DICOM
- BRAINS
- **Information governance & data protection**
  - Data Protection Act
  - DICOM de-identification
  - PPDP



# Regulatory Framework

---

- Data Protection Act:
  - Personal data should not be held longer than necessary
- R&D Ethics Committee:
  - Time frame should be placed upon retention period
- MRHA:
  - Data should be retained

# MIDAS meetings (2009-2011)

---

- **M**edical **I**maging **D**ata **A**ccess and **S**haring
- Motivation: data sharing depends on the NHS regional boards
  - Different rules in different parts of Scotland
- Brought together representatives from the NHS Scotland & the universities
  - Roadmap for improving the data sharing between both sides
  - Difficult to implement



# Data Protection Act

---

- UK's Data Protection Act (1998). Implements the European Community Data Protection Directive 1995.
- Establish individuals' rights on data held about them and obligations for organisations or people processing personal data.
- Personal data must be processed in a fair and lawful manner.
  - 8 DPA principles.
- Other legislation pieces apply to medical data.
  - Common law: duty of confidentiality.
  - Human Rights Act 1998 (article 8).



# DPA in research

---

- The DPA does not define the term “research purposes” apart from clarifying that it includes statistical or historical purposes
- Data processing for research should be “compatible” with the purpose for which the data were originally obtained
- The data subjects should be aware that their personal information will be used for research purposes

# Anonymous Data

---

- Coded (pseudonymised or linked anonymised) data:
  - the identifiable information has been substituted by alphanumerical sequences with no plain meaning
  - The data is anonymous to the research team
  - The key to reverse the transformation shall be held securely by a third party to avoid falling into the DPA
- (Fully) Anonymised data:
  - all personal identifiers or codes have been irreversibly removed

# Personal Data in DICOM

---

- As they are used in clinical workflows DICOM objects include many attributes with personal information
  - Some times personal data is found also “burned in” the pixel data
  - There is a potential risk for face recognition in 3D reconstructions (MRI)
- Considerable number of de-identification tools, but
  - Some do not do the job
  - Lack of flexibility
  - Bad performance
  - Linked to specific suites or frameworks

# DICOM Confidential

---

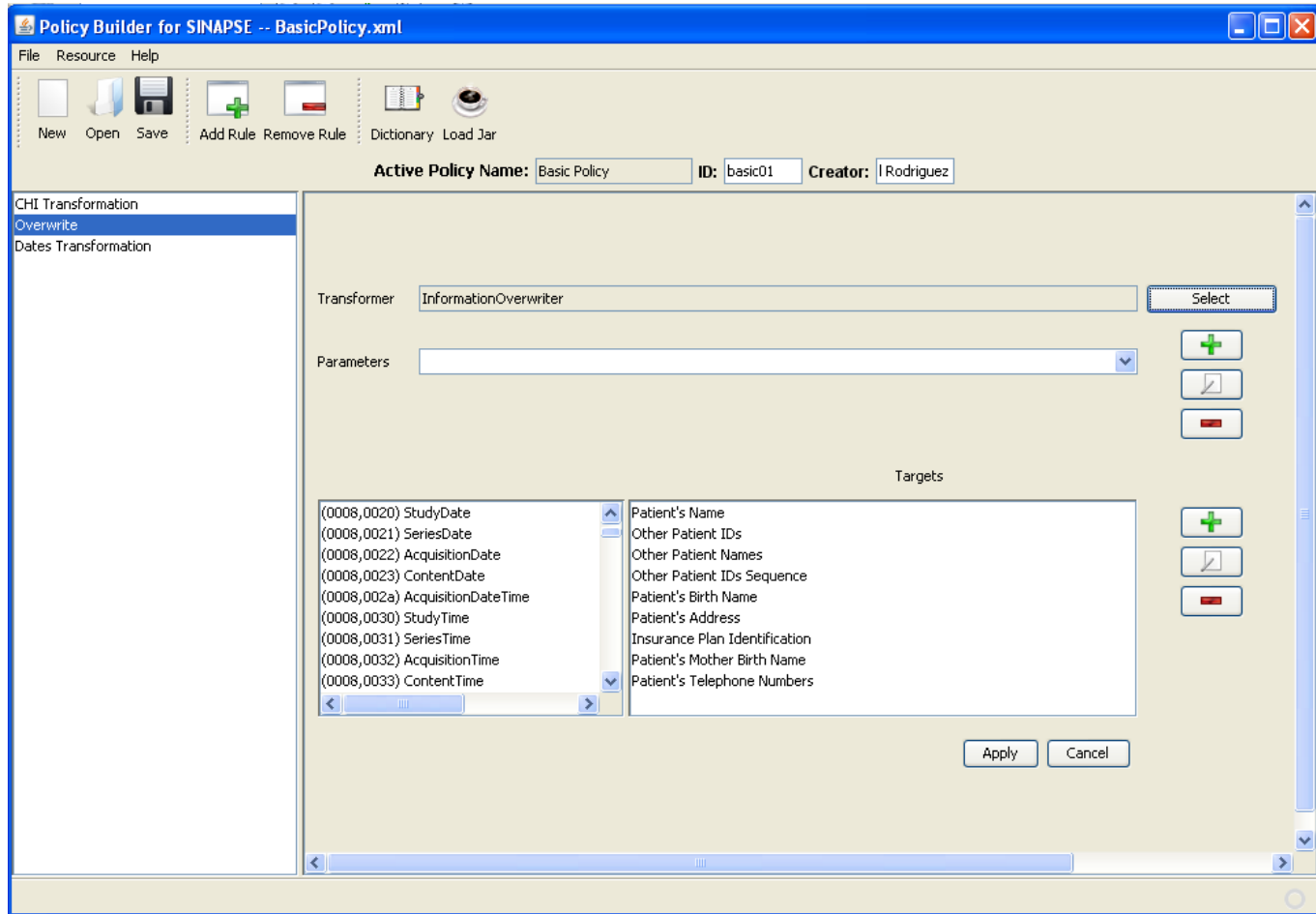
- A DICOM de-identification toolkit
  - Implemented in Java
  - Highly configurable
- Privacy Policies expressed in XML documents
  - PolicyEditor: a graphical policies creation tool
- Transformation classes distributed in signed jar files
- DICOM read/write through an interface that allows using different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



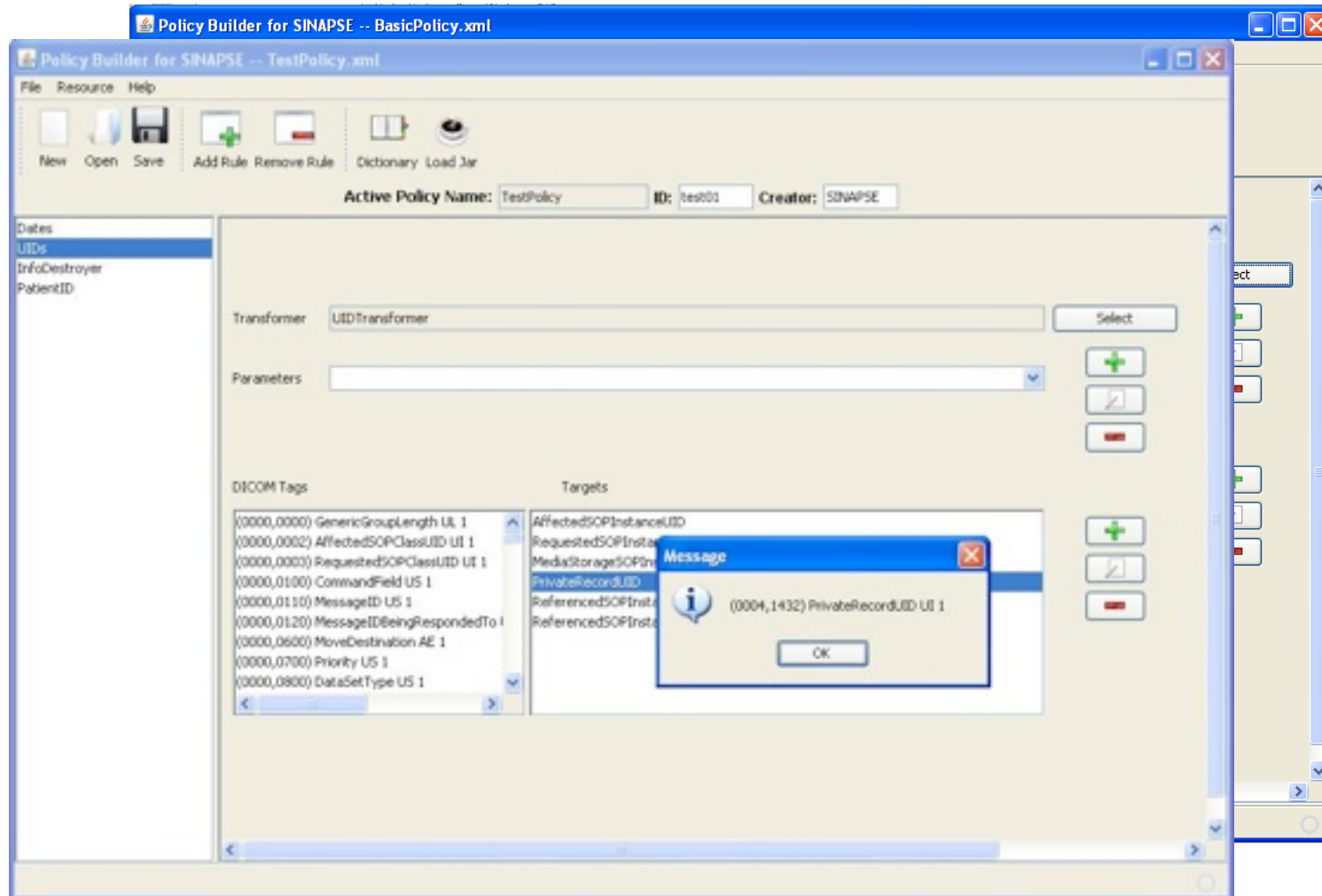
# Policy Editor

---

# Policy Editor



# Policy Editor





# Pending: Face de-identification

---

- This is done by independent applications after volume reconstruction (Analyze or Nifti format)
  - So it is done once the data has already been passed to the researchers
- Recently there have been a publication doing it directly in DICOM



# Future problems: Data Publishing

---

- Data Publication increasing due to:
  - Regulators
  - Funding agencies
  - Collaboration spirit
- This also applies to the imaging data
- Even after the removal of personal data items there is risk of disclosure through linkage with other published data
  - The unknown background knowledge about the victim complicates the problem

# Privacy-Preserving Data Publishing

---

- See survey by Fung, Wang, Chen and Yu:
  - *Privacy-Preserving Data Publishing: A Survey of Recent Developments* (ACM Computing surveys, vol 42, 4, June 2010)
- Demand for publication of microdata
  - Publish data not the data-mining result
- Current practices lead to excessive distortion or insufficient protection
  - Truthfulness at the record level required in some scenarios
  - Unknown background knowledge



# Privacy Protection

---

- When publishing data the attacker should not learn anything about the target (victim) compared with not publishing the data
  - Practical approach: assume limited background knowledge
- Two types of attacks:
  - Linkage (record, attribute or table)
  - Probabilistic
- Minimality attack on Anonymous data

# Anonymisation Operations

---

- **Generalisation:** replace a value with more general one
- **Suppression:** remove a value or record
- **Anatomization:** deassociates the relationship between the quasi-identifier and the sensitive attribute by grouping
- **Permutation:** the same by partitioning and shuffling the sensitive values
- **Perturbation:** preserve statistical information by replacing the original data with synthetic

# Questions

---

