European Medical Informatics Framework (EMIF)

Simon Lovestone
KCL
Problems in the pipeline

- ‘Empty’ pipeline
- Toxic compounds
- Efficacy difficult to measure and to demonstrate
- Biomarkers
The difficulty in monitoring and in predicting progression in Alzheimer’s disease

- Controls
- Alzheimer’s disease
- Mild cognitive impairment

MMSE
NA-ADNI
n=800
~60M USD

WW-ADNI

J-ADNI
Planned n=600
4.7M USD / year

AddNeuroMed
n=700
~8.6M Euro

C-ADNI

AIBL
N=1111; 286 MRI
2.5M USD
Project Vision

To enable and conduct novel research into human health by utilising human health data at an unprecedented scale

‘Think Big’

• Access to information on 40 million patients
• AD research on 10-times more subjects than ADNI
• Metabolics research on > 20,000 obese & T2DM subjects
• Linkage of clinical and omics data
• Development of a secure (privacy, legal) modular platform

• Continue to build a network of data sources and relevant research
King’s Health Partners

Guy’s and St. Thomas’

King’s College London

South London and the Maudsley

King’s College Hospital
EMR – Patient Journey System

- Single integrated clinical record
- Covers all areas of specialist MH care – initial referral to full service discharge
- Total of 190,000 records with 35,000 active patients
- 5,000 unique users log-in per month
- 300,000 documents created per month

CRIS

- Complete clinical data – structured, semi-structured and unstructured
- Extracted, pseudonymised and deidentified, including free text
- Searchable repository for observational research, trial feasibility, identifying potential recruits, service evaluation and clinical audit
- Pseudonym attached to samples
- Linked to imaging and biological datasets
SLAM Firewall

BRC high performance cluster

Provision of analytics packages and pipelines
- Matlab
- IDBS InforSense
- Rich set open source tools and databases

Demilitarized Zone

Processing Capacity

CRIS Front End

CRIS SQL

FAST index

CRIS

SLAM Firewall

Clinical data sources

PJS

JAC

Research imaging

Research ‘OMICS

SLAM Bioresource

SLaM eMPOWER MENT

SLAM CDLS
CRIS Security Model

CRIS users

Require a trust contract or research passport

Source EHR

De-identification, including free text

CRIS

Audit log of all CRIS use

Record level Output

Findings

CRIS security model developed and managed by stakeholder / patient-led oversight committee

Project application

Project approval process

Trust firewall
Data processing

GATE text parsing software (General Architecture for Text Engineering) is used for ‘information extraction’, e.g. to generate structured data from free text, e.g.

- to extract MMSE score and date from text entries

Text: “MMSE done on Monday, score 24/30”

Run across all CRIS free text – produces 35000 validated structured scores/dates
Turning words into data

<table>
<thead>
<tr>
<th>MMSE ID</th>
<th>Table</th>
<th>Date Process</th>
<th>Brclid</th>
<th>CN_Doc_idx</th>
<th>Numerator</th>
<th>Denominator</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>21/09/2010</td>
<td>10007134</td>
<td>15229144</td>
<td>15</td>
<td>25</td>
<td>01/12/2009</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>21/09/2010</td>
<td>10007202</td>
<td>17749381</td>
<td>29</td>
<td>30</td>
<td>02/06/2008</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>21/09/2010</td>
<td>10007204</td>
<td>21280391</td>
<td>10</td>
<td>30</td>
<td>15/04/2006</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>21/09/2010</td>
<td>10007204</td>
<td>21280391</td>
<td>17</td>
<td>30</td>
<td>15/11/2005</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>21/09/2010</td>
<td>10007205</td>
<td>18799205</td>
<td>12</td>
<td>30</td>
<td>01/07/2008</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>21/09/2010</td>
<td>10007286</td>
<td>18528591</td>
<td>10</td>
<td>30</td>
<td>06/09/2005</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>21/09/2010</td>
<td>10007294</td>
<td>24880515</td>
<td>11</td>
<td>20</td>
<td>03/04/2011</td>
</tr>
</tbody>
</table>
Cholinesterase inhibitors and Alzheimer’s disease

Phase IV of AChEI

> 2500 patient years of therapy
> 8 fold dataset compared to Cochrane

Costs and effectiveness

precompetitive collaboration with pharma

Text mining derivation of service utilisation and costs

Predictors of response

Biomarkers and clinical
Trust Electronic Patient Record (ePJS)

Data Interchange with GP and other health care systems

Research Information System (CRIS)

Personal Health Record - HealthVault

Mobile Device Connection Center

Partner Devices

My SLaM Portal

...including Patient Reported Outcome Measures, trials information, engagement....

BioResource

Critical Path Institute

National Institute for Health Research

efpia
EMIF – platform for modular extension

EMIF governance

Research Topics

EMIF - Metabolic
  - Patient generated data
  - Risk stratification

EMIF - AD
  - Risk factor analysis
  - Prevention algorithms

CNS
  - Call 5

TBD
  - Predictive screening

Data Privacy
Analytical tools
Semantic Integration
Information standards
Data access / mgmt

IMI Structure and Network
Key objectives – EMIF-Platform

• Access to harmonised data
  – Access to harmonised patient medical information from different data sources across Europe
  – Comprehensive health data comprising clinical, biomarker and other detailed health information on a number of populations and specific cohorts (pediatrics, adults, including vulnerable groups).

• Governance
  – Procedures and SOPs that govern access and utilisation of patient level data
  – Robust measures to enable linkage and sharing whilst preserving privacy

• Tools
  – Solutions in the areas of data privacy and ethics, standards and semantic interoperability
  – Patient health data linkage and access to a combined patient health information base

• Business Model
  – That governs the use of the project output as well as the support for future research projects
EMIF-biomarkers

Biomarkers
- Clinical
- Genomic
- Proteomic
- Lipidomic
- Metabolomic
- Lipid turnover

Sample Sources
- EHR/Biobanks
- Large Prospective studies with Endpoints
- Medium Size Cohort Studies
- Small biomarker-rich Cohort or intervention Studies and cell/animal-based studies

Number of Samples

Number of Biomarkers

Time
Key objectives – EMIF-AD

1. Collection of data required for the development and validation of new biomarkers for AD
2. Characterisation of study population and definition of extreme phenotypes
3. Discovery of new biomarkers for the diagnosis and prognosis of predementia AD
4. Validation of new biomarkers and development of strategies for selection of subjects in AD prevention trials
EMIF: one project – three topics

1. EMIF-Platform: Develop a framework for evaluating, enhancing and providing access to human health data across Europe, to support the two specific topics below as well as research using human health data in general

2. EMIF-Metabolic: Identify predictors of metabolic complications in obesity, with the support of EMIF-Platform

3. EMIF-AD: Identify predictors of Alzheimer’s Disease (AD) in the preclinical and prodromal phase, with the support of EMIF-Platform
Thinking – and doing - Big

- 58 partners (3 consortia + Efpiia)
- >200 individuals involved
- 14 European countries represented
  (14 – UK, 8 – Germany, 6 – Belgium, 6 – Italy, 5 – France, 5 – Sweden, 4 – Netherlands, 3 – Denmark, 3 – Finland, 2 – Spain, 2 – Switzerland, 1 – Portugal, 1 – Estonia, 1 – Luxembourg) [# partners – country]
- ~56m € worth of resources (in-kind / in-cash)
- “3 projects in one”
• **EMIF general**
  – Bart Vannieuwenhuyse (bvannieu@its.jnj.com)
  – Simon Lovestone (simon.lovestone@kcl.ac.uk)
  – Johan van der Lei (j.vanderlei@erasmusmc.nl)

• **EMIF-Platform**
  – Johan van der Lei (j.vanderlei@erasmusmc.nl)
  – Patrick Genyn (pgenyn1@its.jnj.com)

• **EMIF-Metabolics**
  – Ulf Smith (ulf.smith@medic.gu.se)
  – Dawn Waterworth (Dawn.M.Waterworth@gsk.com)

• **EMIF-AD**
  – Pieter Jelle Visser (pj.visser@maastrichtuniversity.nl)
  – Mike Krams (mkrams@its.jnj.com)

• **www.emif.eu** (coming soon)