

The SAIL System

Challenges and opportunities in health data linkage



Dr Kerina Jones
Senior Research Fellow in Health Informatics

SAIL Databank
College of Medicine

Swansea University

Overview

- The potential and challenges of working with routine data
- Introduction to SAIL
 - How SAIL operates
 - How to work with SAIL
- Examples of studies using SAIL data

SAIL Databank
College of Medicine

Swansea University

The potential and the challenges

- Vast amounts of data collected routinely in the course of health and public service delivery
- How could these data be made available for research:
 - In accordance with Information Governance
 - To ensure data security, integrity and quality
 - To maintain data usefulness

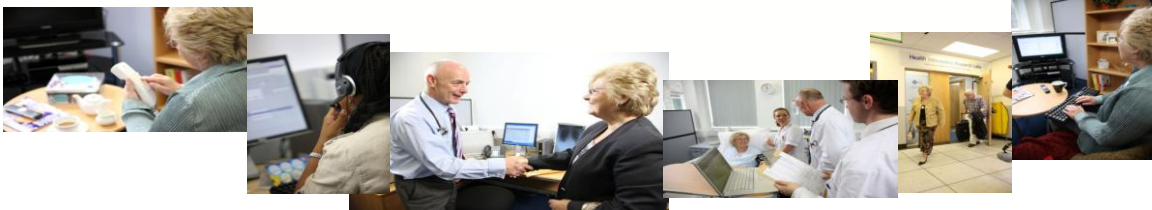


SAIL Databank
College of Medicine

Swansea University

Data linkage is the key

- Data linkage (at a person level) is essential to make the most of routine data
- To be able to link datasets from different origins
- Reliable data linkage needs consistent personal identifiers on which to link



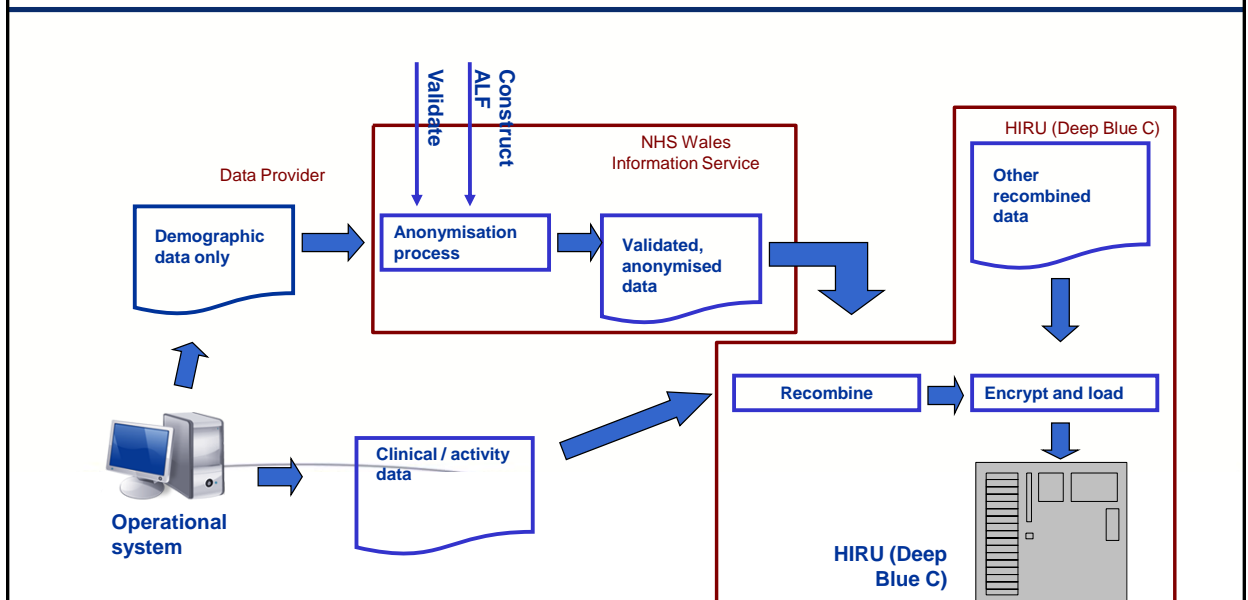
Governance frameworks in the UK

- Information governance – legislation, policies and structures to protect patient/personal information
 - How this is collected, transported, stored, used, shared and disposed of
- Research ethics and governance
 - In the patient’s best interest and not intrusive
 - Informed consent (where appropriate)

HIRU and SAIL

- HIRU – the Health Information Research Unit
- SAIL – Secure Anonymous Information Linkage
- Main aim of HIRU is to realise the potential of electronically-held, routinely-collected, person-based data to conduct and support health-related studies
- The SAIL databank holds over 2 billion anonymised and encrypted individual-level records, from a range of sources relevant to health and well-being

How the SAIL system operates



National datasets - examples

- PEDW - in-patients & day cases and out-patients
- National Community Child Health Database
- NHS Direct Wales
- Cancer incidence registry for Wales
- National screening programmes
- Congenital abnormalities
- Ambulance service data
- National Pupil Database ... and more

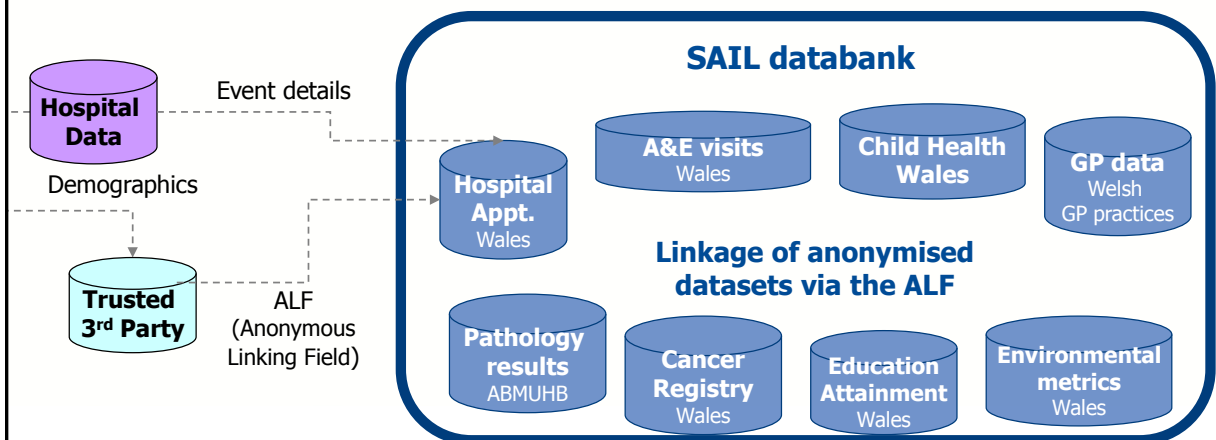
Local datasets - examples

- General Practice
- Pathology
- A&E departments
- Local social services
- Fire & Rescue services
- Local authority housing data
- And more....

SAIL Databank
College of Medicine

Swansea University

Opportunities for record linkage



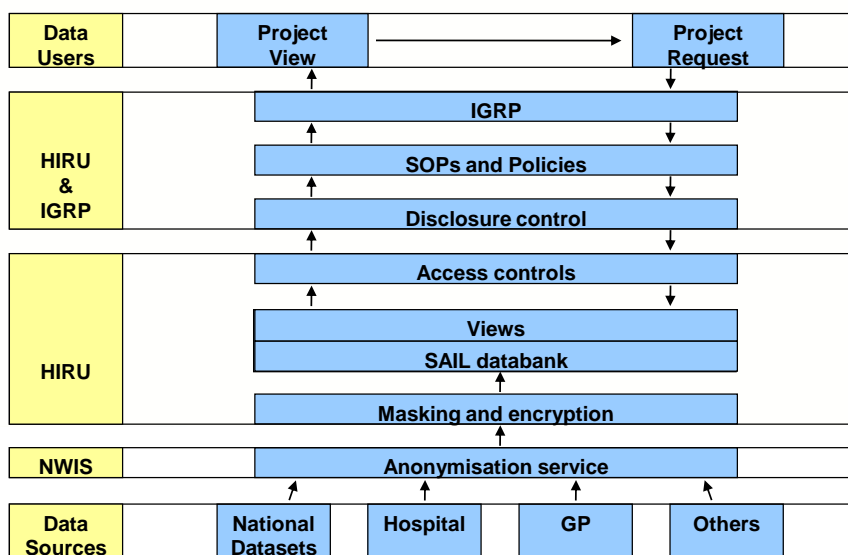
Working with SAIL

- Application and review process
- Data are accessed via a remote access system – the SAIL Gateway
- Subject to data access agreement, compliance with policies and approval by independent Information Governance Review Panel
- Views are created for the data user
- Results are scrutinised by data guardian before release

SAIL Databank
College of Medicine

Swansea University

The SAIL system



Study datasets

- Data collected as part of research studies where the aim is to use routine data as well – best of both worlds
- Permissions, consent and regulatory approvals
- Treat as dataset from DPO – study dataset anonymised and loaded into SAIL for linkage with SAIL data
- Project-by-project basis - study team and HIRU working together

Challenges - data readiness for use

- Dependent on information from data providers
- Variations in coding systems and recording standards
- Data quality, completeness, permissible values, etc.
- Data preparation and cleaning - analytical capacity and expertise
- Subject-specific expertise with data skills
- Information Governance / security

Examples of studies using SAIL data

- Welsh Electronic Cohort for Children (WECC) - to show the breadth of datasets that can be used in a study
- Biomarker Discovery and Validation – to show the linkage of cancer-related laboratory data to routine data
- UK Multiple Sclerosis Register – to show how clinical data, routine data and patient-reported outcome measures are being used in the Register data model

SAIL Databank
College of Medicine

Swansea University

Welsh Electronic Cohort for Children

- WECC is designed as the first complete population e-cohort of children in the UK
- It brings together and anonymously links many longitudinal datasets, including: GP, PEDW (hospital), National Community Child Health Data, ONS births and deaths, All Wales Perinatal Survey, Congenital abnormalities, National Pupil Database, Residential housing data, and others
- Many opportunities for child health-related research

SAIL Databank
College of Medicine

Swansea University

Welsh Electronic Cohort for Children

- Over 800,000 children
- 74% born in Wales, 26% moved into Wales
- What factors at birth are associated with increased risk of respiratory admissions?

Vulnerable baby group	Hazard Ratio	95% Confidence Interval
Very low birth weight (1.0 to 1.5kg)	2.35	1.90 to 2.91
Very preterm (29 to 33 weeks)	2.09	1.59 to 2.24
Extremely preterm (24 to 28 weeks)	3.49	2.10 to 4.77

Biomarker discovery and validation

- Development of a platform and user interface to support biomarker discovery and validation studies
- Prostate cancer, lung cancer, endometrial disease and Barrett's oesophagus
- Standard method – NHS employee searches the clinical system to construct a list of potentially suitable specimens to be passed to archivist and researcher

Biomarker discovery and validation

- Query Pathology, PEDW and cancer registry (WCISU) data in SAIL and identify cases of interest
- To request tissue samples from the archive via the Health Board
- To upload the results of laboratory data to be linked to SAIL data
- To augment the study of possible biomarkers with the routine datasets in SAIL



UK MS Register

- Estimated 85,000 – 100,000 people with MS in the UK.
- But a lack of high quality information
- To develop a prototype Register that can be scaled to national deployment
- To 'make sense of MS' and support research and service delivery
- An innovative data model

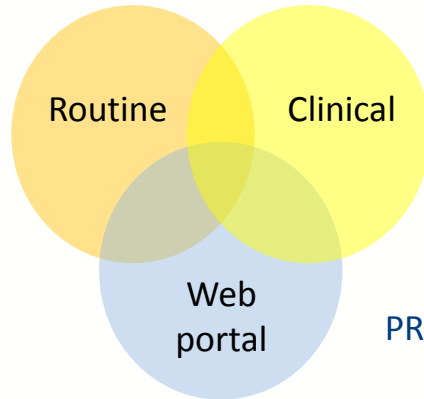
MS*
REGISTER

SAIL Databank
College of Medicine

Swansea University

UK MS Register

GP, hospital, ONS
births & deaths




iMed
OpenEMR

PROMs


SAIL Databank
College of Medicine

Swansea University

Email: Password: [Forgot Details?](#)



MS
REGISTER



***Help Make Sense of MS**

The MS Register is a ground-breaking study designed to increase our understanding of living with MS. [Join the study](#) by completing a simple questionnaire every three months to help us influence policy and improve healthcare.

Email*

Join the Study

If you are over the age of 18 and living in the UK, with a confirmed diagnosis of Multiple Sclerosis made by a consultant neurologist, you are eligible to take part in this ground breaking study.

Once you've completed the registration form, you'll find a simple multiple choice questionnaire to complete; this should only take around 10 minutes. [Join The Study](#)

Password*

Confirm Password*

*Required

***8300 people have joined the study so far.**

Summary and future work

- Many opportunities for working with multiple routine datasets
- Rapidly growing field
- More challenges to be addressed
- 'To infinity and beyond...'



SAIL Databank
College of Medicine

Swansea University

Acknowledgements

- David Ford, Ronan Lyons, Caroline Brooks, Martin Heaven, Simon Thompson, Cynthia McNerney, Rod Middleton, Paul Lewis, and many more....
- Thank you for listening
- Any questions?

SAIL Databank
College of Medicine

Swansea University