Health Information Research Unit (HIRU)

Secure Anonymised Information Linkage (SAIL)
Primary Care Practice Sign Up Pack
INTRODUCTION TO SAIL

The Health Information Research Unit (HIRU) has set up the Secure Anonymised Information Linkage (SAIL) system to link together the widest possible range of robustly anonymised, person-based data, which is building into a Wales-wide research resource that is used to address important research questions, focused on improving patients’ health and the services they are provided.

Over 170 practices in Wales are providing anonymous data to the SAIL system and we are very grateful for this participation in enabling the creation of such a rich resource for research.

HIRU would like to invite your practice to participate in the SAIL system. Your practice’s data, once anonymised (at both a patient and practice level), will make an extremely valuable contribution to this national effort and will ensure that any analyses conducted on the SAIL data reflects your patients and the local circumstances, whilst preserving anonymity.

This pack provides answers to frequently asked questions about the SAIL system, in addition to a Data Collection / Confidentiality Agreement form that needs to be completed and returned, should your practice wish to take part.

The extraction of data from your practice’s system will be via the Audit+ system. Contributing anonymous data to SAIL requires very little to set up, and is fully automated thereafter. There are no costs to the practice to take part and the practice can cancel the arrangement and opt out of SAIL (retrospectively if desired) at any time.

We hope that your practice will be able to join with the very many practices in Wales already signed up to take part in SAIL.

If you have any queries regarding SAIL please do not hesitate to contact:

Cynthia McNerney, HIRU Information Governance Coordinator,
Telephone: 01792 606661, Email: c.l.mcnerney@swansea.ac.uk
FREQUENTLY ASKED QUESTIONS

What is HIRU?

The Health Information Research Unit (HIRU) has a programme of funding from the National Institute of Social Care and Health Research (NISCHR). The main aim of HIRU is to realise the potential of electronically held, person based, routinely collected information for the purposes of conducting and supporting health related research. HIRU works collaboratively with researchers and service professionals across Wales to conduct and support research and to improve service delivery.

What is the SAIL system?

HIRU has established the Secure Anonymised Information Linkage (SAIL) system to link together the widest possible range of anonymised, person-based data and make this available to researchers whilst safeguarding privacy. Initial work was conducted to devise, test and refine the processes of data anonymisation, acquisition, linkage, data quality management and analysis. This work also provided opportunities to establish refine and evaluate data flows and transport mechanisms. Since then, HIRU has been expanding the databank, both in types of dataset and in geographical coverage, and many additional organisations have since provided, or agreed to provide, their datasets. Through the robust processes that have been developed and implemented, this growing databank represents a valuable resource for health-related research and service development, whilst complying with the requirements of data protection legislation and confidentiality guidelines.

Who has already agreed to provide datasets?

Many organisations have agreed to provide data to the SAIL system. Some key examples are:-

- NHS Wales Informatics Service (formerly Health Solutions Wales)
- Office of National Statistics: Birth and deaths registrations
- The Department for Children, Education, Lifelong Learning and Skills (DCELLS)
- National Community Child Health Database
- Screening Services: Breast Test Wales & Cervical Screening Wales
- Welsh Cancer Intelligence and Surveillance Unit
- Many Health Boards/Trusts, GP Practices, and Social Services departments
- Local Government Data Unit ~ Wales
All data received and held by HIRU are securely anonymised and encrypted so that individuals cannot be identified. The commonly recognised individual identifiers in the dataset are removed and replaced with a unique, non-identifiable data item, so that data linkage can take place at the individual record level without HIRU having access to any identifiable information. This is accomplished through a reliable pseudonymisation service at NHS Wales Informatics Service (NWIS), which then provides anonymised datasets to a secure anonymised data repository at HIRU via robust data transfer mechanisms. The data processes that have been implemented retain the facility to link the data, from the various data-providing organisations (DPOs), at the anonymised individual record level so that they are useful for research studies and service development initiatives, without compromising individual privacy.

**What processes are involved?**

**Data Acquisition**

Many organisations have electronic data stored at an individual person level; this could be a register of individuals receiving a care package or a more complex database collecting for example, individual pathology results. With the support of the SAIL team, the Data Providing Organisation (DPO) identifies from their dataset a system-linking field that can be applied to each individual within their database to act as a secret, unique identifier. The DPO then sorts their data into two distinct categories: identifiable demographic data (such as name, address, date of birth, gender, etc.) and confidential clinical data (relating to events, diagnoses, interventions and outcomes, etc.) When the two data entities are defined, the DPO generates output data files assigning the unique system-linking field to each individual in the demographic dataset and to each corresponding row in the clinical dataset. The resulting output is two distinct data files each carrying the common key to enable them to be rejoined. These are referred to as File_1 (demographic data) and File_2 (clinical data). Using secure HTTPS upload facilities, File_1 is transferred to the trusted third party, NWIS for pseudonymisation, and File_2 is securely transferred directly to HIRU, thus ensuring that no one party other than the DPO itself ever has access to both demographic and clinical data files.

**Validation and Anonymisation**

The File_1 that is sent to NWIS undergoes a process of validation and anonymisation. The person identifiable information is matched and validated to an NHS number on the Welsh Demographic Service (WDS); and it is then replaced by a unique encrypted code to act as an Anonymous Linking Field (ALF), with NWIS securely keeping the only key. Other data items, such as postcode, date of
birth and gender are generalised into categories to provide further safeguards. The final stage is to produce an output file containing just the DPO generated system linking field, the categorised data and the Anonymous Linking Field (ALF). This is known as File_3 and it is securely transferred directly to HIRU.

Building the SAIL Databank

All data are loaded into a data warehouse on Blue-C for linkage, data management and analysis. The FILES 2 & 3 can be linked together to produce a file that contains clinical/service data against the ALF. HIRU then encrypts the ALF again, producing an Encrypted Anonymous Linking Field (ALF_E) thus creating the Anonymous Analytical Dataset called SAIL Databank. This double encryption provides an added layer of security and ensures that neither HIRU nor NWIS can independently decrypt the patient identifier. These processes are illustrated in Figure 1.

**Figure 1: Example of data extract and process flow from a data provider’s system to the SAIL Databank**

Has the SAIL system been independently Quality Assured?

Yes - HIRU has developed a Data Anonymisation Policy & Process (DAPP) and a suite of Standard Operating Procedures (SOPs) to set out the information principles that we abide by and the procedures for the anonymisation and encryption of all data in the SAIL databank. The work of HIRU has been developed in consultation with the Welsh Government, Informing Healthcare (now

Primary Care Practice Sign Up Pack – November 2012
reorganised as part of NHS Wales Informatics Service) and the Office of the Information Commissioner. HIRU also undergoes a programme of voluntary independent audits of our information systems for compliance with Information Governance. These reports have been notably favourable and have confirmed that the design of the system is sufficiently robust to provide assurance that the activities and procedures in place will achieve the objectives for the system.

**Is privacy protected?**

Yes – as explained above, at no point will person identifiable information and clinical/service information be held together by any organisation other than the original data provider. Because we are only dealing with anonymous data, we are not required to obtain informed consent from individuals whose records come to SAIL. However, although HIRU receives only anonymised data, due care is taken to ensure compliance with Information Governance, including all relevant regulatory and legislative procedures at all stages in the development of the databank, from data acquisition to data utilisation.

HIRU takes an integrated approach to Information Governance by combining: robust technological and data management processes; strict permissions and data access arrangements; and independent review of proposals for data utilisation via the Collaboration Review System (CRS). The CRS comprises reviews by the HIRU management team - to consider the feasibility, practicality and resource implications of each proposal - and by an independent Information Governance Review Panel (IGRP) - to ensure the proposed work conforms to the principles of Information Governance. Through the IGRP, representatives from the British Medical Association, Public Health Wales, National Research Ethics Service, NHS Wales Informatics Service (NWIS) and the Involving People network are all formally involved in assuring the work of HIRU.

The SAIL system conforms to the HIRU Data Anonymisation Policy and Process (DAPP); which has been endorsed by Informing Healthcare (now reorganised as part of NHS Wales Informatics Service) as well as a number of Caldicott guardians and information governance officers in the NHS and Local Government. The DAPP document covers the areas of:-

- Information governance and legislation requirements
- Research Ethics
- HIRU Information Policy
- Anonymisation procedures
- Security procedures
The authorised analysts interrogating the SAIL databank are required to conform to the HIRU SAIL Databank Access Agreement, which ensures there is no risk of an individual being identifiable through the disclosure of small numbers.

**How will the data be used?**

Once the data have been securely anonymised and loaded into the SAIL databank the data will be used to help answer important research questions. The SAIL dataset has a huge range of potential applications. HIRU is committed to working with data-supplying organisations and research groups to identify analysis that can make a difference, leading to enhanced patient care and improvements in health and well-being. HIRU is not involved in performance management and does not carry out analyses for such purposes unless requested by the individual data provider.

In particular cases involving the linkage of data from research studies where:

- informed consent has been obtained from individuals in the study population
- this included consent for the participants health records to be accessed
- the research proposal was given a favourable opinion by a research ethics committee;

special conditions may be applied whereby results can be passed back to the data custodian at an individual record level if required. In these cases, study data are anonymised and processed into SAIL via NWIS in the usual manner. But in addition, the study data are assigned a project-specific linking key which can also be supplied to the data custodian to enable them to associate participants’ data to their study data. This project-specific linking key has no currency beyond the study and so cannot be used to link wider datasets together.

Figure 2 shows one example of the benefits of linked-anonymised data in gaining an integrated view of a patient’s journey through the care system and which could assist service providers in the development of individual care bundles.
Figure 2: Illustrative Synthetic Example

Figure 2 shows health and social care events for an anonymous individual over a twelve month period. Underlying each point on the graph is a cluster of data items relating to their condition, procedures, tests and results.

SAIL analysis benefits from the use of the Anonymising Linking Field (ALF) to link across existing datasets. The original datasets, left in silos, would be able to show that this individual had two A&E events in this period. However, linking the data sets together an analyst can see that after the second A&E event the individual was admitted to hospital; required a social care intervention of having a lifeline installed; and is on regular medication from their GP.

Providing benefits to the data provider

Examples of benefits are:-

- Providing incidence and prevalence data for chronic diseases to support planning
- Profiling health and social events for care pathways
- Providing deprivation, access and rurality profiles for participating primary care practices
- Providing hospital utilisation rates adjusted for age, sex, deprivation and access metrics
- Responding to data provider requests for analyses
Supporting research questions

Examples of research questions are:

- Is disease x increasing or decreasing?
- If care is redesigned in a particular way what will be the likely impact on primary/secondary care and different populations?
- What is the impact of poverty/deprivation on the need and demand for health services?
- What additional workload is needed to meet national service frameworks (NSFs)/changes in contract etc?
- How many patients would benefit from a new treatment supported by NICE and how much would this cost?
- Are there enough patients suitable for a new clinical trial in an area?
- What are the long term outcomes of trials, policy developments, and public health interventions?
- What chronic health conditions in childhood influence educational attainment?
- Are children who do well at school less likely to develop health conditions?
- Does birth weight affect a child’s learning at school?
- Are there any correlations between early childhood medications and educational achievement?

What’s in it for our practice and patients?

The SAIL system is building into a Wales-wide research resource that is being used to address important research questions, focused on improving patients’ health and the services that are provided. Your practice’s data, once anonymised (at both a patient and practice level), will make an extremely valuable contribution to this national effort and will ensure that any analyses conducted on the SAIL data reflects the needs of your patients and the local circumstances. In addition, we are working with NHS organisations and Social Services, to put in place ways of ensuring locally-focused analyses are supported, and to establish ways in which your practice will be able to suggest ideas for research or analysis using the SAIL data. We hope this will bring benefits to your patients and to those in surrounding areas.
How much work is involved for the practice?

Very little, we have worked hard to make it as easy as possible for practices to contribute. First, the practice has to agree to participate in the SAIL system and complete the “Data Collection / Confidentiality Agreement” form. Once we have received this, the HIRU team will explain how to start the automated SAIL Data Extract process by changing the settings on the Audit+ system. Subject to your approval to contribute data to SAIL, the software will send us weekly extracts from your system via our anonymising process without any intervention from the practice. The suppliers of the Audit+ system have been instructed to ensure that the extraction and transmission process is scheduled so that there is no interruption to the operation of your clinical system (usually scheduled at night and/or weekend). There are no costs to the practice to take part, and the practice can cancel the arrangement and opt out of SAIL (retrospectively if desired) at any time.

Contact the HIRU Team

Professor David Ford (Co-Director)  d.v.ford@swansea.ac.uk
Professor Ronan Lyons (Co-Director)  r.a.lyons@swansea.ac.uk
Dr Kerina Jones (Senior Research Fellow)  k.h.jones@swansea.ac.uk
Cynthia McNerney (Information Governance Co-ordinator)  c.l.mcnerney@swansea.ac.uk

Health Information Research Unit (HIRU)
College of Medicine
Institute of Life Science 2 (ILS2)
Swansea University
Singleton Park
Swansea SA2 8PP

Tel: 01792 606661

http://hiru.swansea.ac.uk
PRACTICE AGREEMENT

<table>
<thead>
<tr>
<th>Practice Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice Address:</td>
</tr>
<tr>
<td>Contact Name:</td>
</tr>
<tr>
<td>Contact Telephone No:</td>
</tr>
<tr>
<td>Contact E-mail:</td>
</tr>
</tbody>
</table>

This is an agreement between the above-mentioned practice to participate in the SAIL System. This practice agrees to provide data through HIRU anonymisation process as described in its Data Anonymisation Policy and Process (DAPP). HIRU will approach the named contact at the practice to discuss the methods of data acquisition.

Signed on behalf of

Senior GP or Practice Manager

Date

Signed on behalf of HIRU

Contact Name

Job Title

Date

Please return the signed agreement to:

Cynthia McNerney
Information Governance Coordinator
Health Information Research Unit (HIRU)
College of Medicine
Institute of Life Science 2 (ILS2)
Swansea University
Singleton Park
Swansea SA2 8PP

01792 606661  c.l.mcnerney@swansea.ac.uk