The Scottish Federated Safe Haven Network

NRS Conference

1st Nov
The Scottish Federated Safe Haven Network

• One National Safe Haven (NSS (eDRIS)/EPCC)

• 4 Regional Safe Havens
  • A collaboration between an R&D Node of NHS Scotland and a University with a Medical School from within that region.
What is a Safe Haven?

• 3 Main Functions
  • **Analytic Platform**: provision of a virtual environment for researchers to securely analyse data without the data leaving the environment
  • **Data Repository**: secure handling and linking of data from multiple sources for research projects
  • **Research Support**: socio-technical environment that serves to bring together teams around health data science. The research coordinators provide support to researchers navigating the data requirements, permissions landscape and provide a mechanism to share the lessons from one project to the next.
High Level Overview

- **Data Repository**
  - Database of Patient Data (including identifiers)
  - Cohort Selection Process
  - Linking and Anonymisation Process
  - Anonymised Subset of Data for Research Project

- **Analytic Platform**
  - Security Permissions
  - Audit Logs
  - Virtual Desktops
  - Anonymised Subset of Data for Research Project
  - Secure Transfer

- **Cohort Selection Process**
  - Linking and Anonymisation Process
  - Anonymised Subset of Data for Research Project

- **Data can’t leave environment**
  - Connects via a web browser

**Researcher**

[Diagram showing researcher accessing data through a web browser]
Analytic Platform

• Researchers provided a remote log on to a Virtual Machine (VM)
• The VM generally has basic statistics packages installed
• Researchers can only see the data for their specific cohort to answer their specific research question
• Researchers can analyse the data in place and can’t copy the data out of the environment
• Aggregate level, analysis of the data can be copied out of the environment only once it has been checked by a Safe Haven team member for discloser control.
• Researchers can’t freely access the internet
• Researchers have to sign none discloser agreements and have to complete Information Governance Training
• Secure Access Controls:
  ✓ Firewall
  ✓ HTTPS Connection
  ✓ Password Protection
Data Repository Function

• Provision of linked anonymised datasets for research for a specific cohort to answer a specific research question
• The internals of how this works within each Safe Haven is slightly different
• Anyone Safe Haven staff member who has access to identifiable NHS patient data is either an employee of NHS or an Honorary employee
National Safe Haven

Data Repository / Data Processing Area

- Database of Patient Data (including identifiers)

Cohort Selection Process

Linking and Anonymisation Process

Anonymised Subset of Data for Research Project

Analytic Platform

- Security Permissions
- Audit Logs

Virtual Desktops

Anonymised Subset of Data for Research Project

Secure Transfer

Data can’t leave environment

Connects via a web browser

NSS Network

NSS / NRS / EPCC

University of Edinburgh Network - EPCC

Researcher

ISD

eDRIS / NRS Indexing / EPCC

NSS / NRS / EPCC

 NSS Network
Analytic Platform

Data Repository

Database of Patient Data (including identifiers)

Cohort Selection Process

Linking and Anonymisation Process

Anonymised Subset of Data for Research Project

Security Permissions Audit Logs

Virtual Desktops

Data can’t leave environment

Connects via a web browser

Cohort Selection Process

Anonymisation Process

Anonymised Subset of Data for Research Project

Secure Transfer

NHS Network

University of Dundee Network

Tayside and Fife

NHS Tayside and Fife Outsource this function to HIC

Researcher

Researcher

Researcher
Database of Patient Data (including identifiers)

Cohort Selection Process

Linking and Anonymisation Process

Anonymised Subset of Data for Research Project

Security Permissions Audit Logs

Virtual Desktops

Data can’t leave environment

Connects via a web browser

NHS Network

University of Aberdeen Network

Grampian

DaSH

Anonymised Subset of Data for Research Project

Secure Transfer

Cohort Selection Process

Linking and Anonymisation Process

Anonymised Subset of Data for Research Project

Analytic Platform

Researcher

Researcher

Researcher

Researcher

Researcher
Data Repository

Database of Patient Data (including identifiers)

Cohort Selection Process

Linking and Anonymisation Process

Anonymised Subset of Data for Research Project

Lothian

NHS Lothian

Virtual Desktops

Security Permissions Audit Logs

Data can’t leave environment

Connects via a web browser

EPCC

Anonymised Subset of Data for Research Project

Secure Transfer

Anonymised Subset of Data for Research Project

NHS Network

University of Edinburgh Network

NHS Network

EPCC
Data Repository Other Functions

• Safe Havens have many more tasks than provision of linked anonymised extracts

• As they often host identifiable patient records they can also:
  • Support recruitment to clinical trials
  • Running feasibility studies to see if relevant data is available for a specific research project
  • Providing audit data for NHS Boards (once it has been linked and cleaned)
  • Hosting and supporting permanently linked consented cohorts for research
  • Supporting Data Governance applications
  • Hosting disease registries
Historical Context
• Data Protection Act
• 7 Caldicott Principles

SHIP Guiding Principles and Best Practices (Oct 2010)

Guide for Data Controllers (and others) advising on best practices of linking data for research purposes across different organisations

Government adoption (Nov 2012)
Historical Context

• Scottish Government Safe Haven Charter – 2015
  • Set of standards for all Safe Havens

• Scottish Government Safe Haven Accreditation
  • Aligns with ISO27001 certification
  • Adds some additional controls on top of general ISO27001
  • Independent audit by eHealth (Scottish Government)

• All Safe Havens have signed up to the Safe Haven Charter.

• If a Safe Haven is Scottish Government Accredited they should work to exactly the same high standards of secure data management and analytic platform provision.
Researcher seeking access to health data for research can approach either a regional Safe Haven or the central eDRIS team. National dataset access is through eDRIS.

Within Scottish Informatics Linkage Collaboration

- Support from eDRIS:
  i) To provide researcher with advice and support to facilitate projects
  ii) To direct researcher to most appropriate safe haven(s)
  iii) Liaise with Data Controllers and Indexing Service on behalf of the researcher

- Advance Computing Facility UoE
  - Partitioned space for regional use
  - National Safe Haven Secure Analytic platform

- National Records of Scotland Indexing/linkage service

National SH = orange
Regional SH = green
Secure Environment = 

Once approvals are in place, approved researchers are granted access to de-identified research dataset on the secure analytic platform within one of the Safe Haven.
Data Governance

• All Regional Safe Havens have existing Data Governance Approvals processes working with their relevant Health Board(s) and Caldicott Guardians (there are small regional variations)

• Regional Safe Havens generally act as Data Processors rather than Data Controllers

• There are existing data sharing/user agreements and service agreements in place

• The NHS can withdraw support for a project at any time should they wish to do so
Data Flows

• We want to increase the projects which need data from multiple Safe Havens and improve the efficiency of doing this.

• Only a subset of the data held by a Safe Haven needs to be transferred i.e. the data for the specific research project
Cohort Selection Process

Linking and Anonymisation Process

Anonymised Subset of Data for Research Project

Secure Transfer of subset of data needed for the project

Data Repository

Database of Patient Data (including identifiers)

Glasgow

NHS Glasgow

Dundee Data Repository

University of Glasgow Network

Robertson Centre

Analytic Platform

Virtual Desktops

Anonymised Subset of Data for Research Project

Security Permissions Audit Logs

NHS Network

Data can't leave environment

Connects via a web browser

Glasgow

Example Data Flow
Costing Models

• The Safe Havens are currently meeting to try to standardise their costing model

• This is challenging due to the different way each is structured

• All Safe Havens try to recover their costs from research grant where appropriate.

• Feasibility studies and aggregates are generally done for free.
Thanks for listening!
Separation of Indexer and Linker functions

- Primarily used when linking data from different data controllers if the data is highly sensitive.
- Considered general good practice by Guiding Principles of Data Linkage.
- If there is agreement between Data Controllers to share data without this method then it reduces the coordination requirements of each project.
Source Data:
Different Data Controllers

- NHS Scotland
- Police Scotland
- Social Services

Identifiable Data
Indexing Service

Descriptive Data
Linking Service

Identity map

Data Governance Approval Required Before Data Flows

Data Extract
Source Data: Different Data Controllers

Identifiable Data:
Angus McFurgus
31/7/67
Lomond Street

Indexing code:
NHS: 12345
Police: 44444
SS: 55555

Identity map
44444=12345=55555

Descriptive Data:
NHS: 12345: Has great health
Police: 44444: Has criminal record
SS: 55555: Known to SS

Linking Service

Indexing Service

Data Extract

Safe Haven Researcher area to view project specific data

Secure Remote Access with disclosure control on output (only aggregates can be exported)