A Blueprint for Health Records Research in Scotland

July 2012

CONTENTS

FOREWORD

Scotland is fortunate to have some of the best health service data in the world. Few other countries have systems that combine high quality data, consistency, and national coverage with the ability to link data to allow patient-based analysis and follow up. Scotland has a long tradition of using linked health service data for research, and methods have recently been developed for combining health service datasets with other data sources such as the Census and population surveys. Data linkage is a highly efficient way to evaluate the capacity of interventions to deliver patient benefit. It allows us to measure long term outcomes in clinical trials, assess the safety of new medical interventions and to understand patterns of health and illness across the whole population. It is a key contributor to our aim, set out in the NHS Scotland Quality Strategy, of providing the highest quality of healthcare services to people in Scotland and, through this, to be counted among the best in the world.

Record linkage capability is being developed throughout the UK and elsewhere, opening up new opportunities for research using datasets from more than one country. Scotland must play a full part in these developments. The Scottish Health Informatics Programme Blueprint aims to create a secure and efficient infrastructure of linking health and relevant non-health records, which encourages research use while maintaining the highest standards of governance and data security. It supports a key strand of the Scottish Government's Research Strategy for health and healthcare, which aims to deliver patient benefit, improve population health and keep Scotland at the international forefront of clinical translational research.

I therefore welcome this Blueprint. It represents a collaboration of the Information Services Division of NHS Scotland, the Health Service as a whole, and academic institutions, all committed to public engagement as a critical component of future developments. Such teamwork is essential if Scotland is to create a world-leading system of data linkage for health research, which balances the need to respect patients' wishes for privacy with research for the public good .

I hope the Blueprint, will provide researchers, data controllers, Caldicott Guardians and others with a strong, clear and efficient model for linking health data, that will have wider applicability to other parts of the public sector. I encourage public bodies to review their practices of record linkage for research, and commend this common framework and infrastructure, which will ensure information is handled in a safe and secure way to protect patient confidentiality while meeting researchers' needs effectively.

SIR HARRY BURNS

Chief Medical Officer

Scottish Government

1. Using the SHIP Blueprint

We intend the *SHIP Blueprint* as a guide to primarily help everyone concerned with the use and governance of unconsented linked data for health-related research purposes. Its purpose is to promote an understanding of the common principles of good governance and the use of associated technical and administrative solutions, so that research capability can be enhanced and governance improved. We hope that the Blueprint and the infrastructure we are creating will be useful to data controllers and Caldicott Guardians, who are striving to do a difficult job better, to researchers, who are using an increasing number of health records datasets for research, and to individuals and groups, who have an interest in scrutinising the effectiveness of governance. The SHIP team has undertaken a comprehensive programme of public consultation and engagement to help shape these proposals and a separate less technical document and other communications describe the activities of SHIP to a more general audience.

1.1. Scope of the Blueprint

The SHIP Blueprint is intended for use by all organisations and partnerships that work with data or share data that are used for health related research or policy development. Most of these are public sector or academic organisations.

1.2. What does the Blueprint consist of?

The Blueprint and associated materials available on the SHIP website describe the following core issues:

- A Good Governance Framework;
- A national infrastructure within NHS Scotland, which promotes record linkage research in Scotland and adopts SHIP best practice for record linkage;
- A data linkage “tool kit”, which incorporates standards and document templates of governance, technology, audit and administration that we hope will provide practical help to others in developing safe and secure systems for data linkage research in the context of local safe havens

1.3. Relationship with other codes and guidance

The Blueprint potentially has a wide scope. We hope that those who develop and set codes for data sharing for research will refer to the Blueprint in updating and reviewing their own guidance and standards, and use it to harmonise the approach within and between different sectors. Where codes and guidance do not already exist, as in many formal and informal partnerships, we hope that the Blueprint will provide a shared understanding and a set of tools that constitute good practice of data sharing for health research.

1.4. Applying the Blueprint to different governance structures and sizes of organisation

The Blueprint principles form a standard of good governance, built upon a resilient infrastructure that will be applied to create a National Safe Haven for health research data in Scotland. We encourage all organisations as they develop local safe havens of infrastructure for data sharing, to show that they are putting the same principles into practice in a way that reflects their structure. We recognise that not all parts of the Blueprint will appear to be directly applicable to all types and size of organisation. The Blueprint comprises proposals for a "Governance Infrastructure" and a "Research Infrastructure" each with its supporting principles. By April 2012, SHIP is committed to deploying a national exemplar of this Infrastructure within NHS Scotland. In addition we will create a toolkit, illustrated with good practice examples, which will support the use of record linkage according to the principles of the SHIP Blueprint elsewhere.

BACKGROUND TO THE BLUEPRINT

2. Opportunities For Health Records Research In Scotland

2.1. Scotland's track record in linkage of health records for research

Scotland has some of the most comprehensive health service data in the world. Few other countries can lay claim to national indexed data of such high quality and consistency. The almost universal use of the Community Health Index (CHI) number means that Scotland's health data can now be easily linked to allow patient based analysis and follow up.

There are many examples of the benefits derived from analysis of anonymous health care data for research purposes in Scotland. This research has increased the knowledge base, helped to improve health outcomes, informed the effectiveness, efficiency and safety of the health services we offer and influenced international medical best practice. Scotland is also a strong contributor to the assessment of the effectiveness and safety of medicines we use, and more recently, understanding the genetic basis and biology of common complex diseases that remain a major challenge in Scotland and beyond (see examples in Appendix 1).

2.2. The complex governance landscape

Currently, health-related research is executed at local, regional and national level in both NHS and academic communities across Scotland. Despite the multiple examples of beneficial record linkage research studies, considerable uncertainty persists amongst the community about what are the legal requirements. This has sometimes led to confusion and frequent delay when applying for and using patient data in research. This was recently highlighted by the *Academy of Medical Sciences* review[1], 'A new pathway for the regulation and governance of health research' (2011), which concluded that "access to patient data for research is currently hampered by a fragmented legal framework, inconsistency in interpretation of the regulations, variable guidance and a lack of clarity among investigators, regulators, patients and the public."

Arguably the most difficult situations, highlighted by the Nuffield Trust Report "Learning from experience: privacy and the secondary use of data in health research"[2], are those in which consent is very difficult or impossible to obtain. Studies using databases in Scotland often need to analyse thousands, tens of thousands or even millions of records in order to gain coverage and statistical

[1] http://www.acmedsci.ac.uk/index.php?pid=99&puid=209

[2] http://www.nuffieldtrust.org.uk/publications/learning-experience-privacy-and-secondary-use-data-health-research

power. Consistent, clear approaches to support these types of analyses have been slow to develop and the landscape remains fragmented. There is therefore an opportunity to demystify the landscape and to clarify, strengthen and, where possible, simplify the legal and governance framework in Scotland.

Many of the problems associated with the use of health records data for research were again described in the Walport Thomas Data Sharing Report (2008)[3], which concluded that:

- There is a lack of transparency and accountability in the way organisations deal with personal information.
- There is confusion surrounding the Data Protection Act, particularly the way it interacts with other strands of law.
- Greater use could be made of the ability to share personal data safely, particularly in the field of research and statistical analysis.

The Walport/Thomas Review recommended specific actions including:

- The need to improve leadership, accountability and training within organisations.
- To ensure all organisations are as transparent and open as possible about how and with whom data are shared, with what authority, for what purposes and with what protections and safeguards.
- To clarify and simplify the legal framework governing data sharing, including provisions to guarantee better and more authoritative guidance for practitioners.
- To develop mechanisms that will enable population-based research and statistical analysis for public benefit, whilst safeguarding the privacy of individuals.
- To help safeguard and protect personal information held in publicly available sources.

The Report recommended the establishment of Safe Havens, which would act as a focus to ensure the proportionate and safe sharing of data for research. This SHIP Blueprint directly addresses the recommendations of the Walport/Thomas report, and describes how we propose to develop an infrastructure in Scotland that will support a network of national, regional and local safe havens for the execution of safe, secure health records research.

[3] http://www.justice.gov.uk/reviews/docs/data-sharing-review-report.pdf

3. The Scottish Health Informatics Programme (SHIP)

SHIP is a Scotland-wide research collaboration that was established in 2008 to enhance the safe and secure collation, management, dissemination and analysis of Electronic Patient Records for research. The programme brings together the Universities of Dundee, Edinburgh, Glasgow and St Andrews with the Information Services Division (ISD) of NHS Scotland. SHIP was funded by the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council. A £3.6M grant was awarded to the SHIP collaborators to support its work from 2009-2013. A full list of members of the collaboration is given in Appendix 2.

The SHIP team believes that a step-change in the quality, quantity and governance of research using electronic patient records can now be achieved with a more joined-up Scottish-wide strategy. Instead of the *ad hoc* linkages used to date, the SHIP programme aims to provide a platform for Scottish record linkage that will drive health related research throughout the UK and abroad.

3.1. The SHIP Vision

For Scotland to set an international standard for the safe and secure use of electronic patient records and other population-based datasets for research purposes in order to create new knowledge that informs the delivery of quality health care and the development, implementation and delivery of policy at a Scottish, UK and International level.

Our four values are: ***collaboration, transparency, innovation, and excellence.*** These are underpinned by a commitment to perform research, which is relevant to Scotland and the careful stewardship of all information that we use.

3.2. SHIP deliverables

The initial SHIP deliverables are to address directly the challenges set out by the Walport/Thomas review by:

- Providing access to a new **national research facility**, firmly embedded within and supported by NHS Scotland, providing the basis for numerous future studies using electronic patient records.
- Creating a **research portal** for electronic patient records already held by NHS Scotland that will **provide rapid, secure, access** to the type of data that clinical scientists require.

- Developing and evaluating systems which work across institutional boundaries to allow linkage between large, federated, **third party** research datasets and the NHS research portal.

In addition, SHIP is:

- Increasing the research capability and know how in Scotland by running a programme of training seminars and workshops, and a biennial St Andrews conference on "*Exploiting Existing Data for Health Research"* in 2009, 2011 and 2013.
- Engaging with the public, building on considerable experience in the field of the public's attitudes to genetic studies, to define a transparent and publicly acceptable approach to the governance of electronic patient record research.
- Producing novel research using electronic patient records and major longitudinal cohort databases, specifically in the areas of clinical trials, safety of medicines, diabetes epidemiology, and research resulting from the linkage of health data to socioeconomic and environmental data.
- Exploring the feasibility of taking major genetic studies in Scotland back through time by linking historical vital events data for the >20,000 study members and their families.

Full details of the collaboration, its work programmes and membership can be found on the SHIP website.[4]

The immediate priority of SHIP is the development of an infrastructure for the provision of national datasets to researchers for health related research. However, we anticipate that the infrastructure will be effective in delivering a range of data linkages, across a range of health and non-health datasets and potentially involving a range of data controllers, in a streamlined, safe, secure and efficient manner.

The anticipated benefits of this approach are:

- High standards of governance and technical processes
- Clarity around governance and processes for researchers
- Reduction in regulatory burden in securing access
- Proportionate approach to assessing benefits and risks
- Better data security and less data travel

[4] www.scot-ship.ac.uk

- Dedicated research co-ordinators to assist researchers with projects
- Register of approved researchers and bespoke online training
- National one-stop-shop for guidance and advice on applications
- Statistical disclosure control to protect patient confidentiality
- Excellent research that has a beneficial impact on society and the economy

SHIP aims to protect the rights of individuals and to serve their interests whilst facilitating high quality research. Public engagement is also a key component of SHIP, and the focus of a related work programme, but out of scope of this document.

4. The Creation of the SHIP Blueprint

4.1. Review of Existing Practice

In 2010, in the light of changing legal and professional requirements and of the need to ensure full partnership with the public, we undertook a wide-ranging review of international best practice in the use of electronic patient record information for research, and the systems, processes and safeguards necessary to maintain public trust and privacy. These findings are published on the SHIP website in our scoping report. We also hosted seminars and workshops at which international colleagues (from Australia, Wales, England, Canada and Scandinavia) presented existing models. We took account of proposals for protecting and using patient identifying information being developed by the National Information Governance Board of the Department of Health in England[5], the Report on the Regulation of Health Related Research by the Academy of Medical Sciences[1] the House of Lords Science and Technology Committee Report on Genomic Medicine (2009)[6] commissioned by the UK Government, and the Nuffield Trust *Learning from Experience: Privacy and the Secondary Use of Data in Health Research*[2].

4.2. Expert Working Groups

Following the initial stage of information gathering we convened three short life groups in June 2010 to produce draft recommendations on the governance, administrative, and IT standards and procedures required to achieve our objective of placing Scotland at the forefront of international best practice. The membership of these groups is listed in Appendix 5. Public engagement is addressed by a separate core workstream within SHIP (see www.scot-ship.ac.uk).

The three groups concluded their deliberations in October 2010. Their reports were then used as a basis for this Blueprint.

4.3. Consultation with Key Stakeholders

The expertise and experience of many colleagues contributed greatly to our work (see Appendix 2). We are also indebted to members of our International Advisory Board (Appendix 3) who guided our considerations of the issues, particularly those of privacy, consent and anonymisation. A list of all groups who have been consulted is given in Appendix 4.

---

[5] http://www.nigb.nhs.uk

[6] http://www.publications.parliament.uk/pa/ld200809/ldselect/ldsctech/ldsctech.htm

5. The value of SHIP to Scotland

Many of the outputs of SHIP may also have wider reach and applicability to an emerging strategic framework for data linkage. Examples include *the Quality Strategy*[7] of NHS Scotland. The aim is to ensure that the most appropriate treatments, interventions, support and services will be provided at the right time to everyone who will benefit, and wasteful or harmful variation will be eradicated. Central to this ambition is the need to apply measurement data to drive reliable and better care across NHSScotland. Such a system will require data linkage across parts of the health system, and the learning from SHIP may help with the co-ordination of planned activity across Scotland.

The Chief Scientist Office Strategy (2010-2015) *Investing in Research/Improving Health*[8] is committed to work with NHS Information Services to create a world-leading research information linkage and data exchange programme within NHS Scotland, supported by an infrastructure that is adaptable and developed in response to the research needs of the NHS, academia and industry.

The *Life Science Advisory Board* for Scotland published a refreshed strategy (2011)[9] to optimise the impact of the sector generating new opportunities and more jobs in Scotland. Its aim is to identify where existing academic and commercial Scottish strengths are aligned with near-term, global market opportunities to deliver greater benefits for the Scottish economy. SHIP has the potential to deliver greater alignment between life sciences R&D support from Scottish Enterprise, the Scottish Government and NHS procurement.

Data linkage in Scotland is also vital if our social services, justice, education and other services are to function well, for the benefit of citizens. SHIP is committed to support ongoing discussions, which are exploring opportunities for integration and advances in data linkage, ultimately to offer better information for statistical and research purposes at no extra burden on data providers. There are clearly opportunities for data linkage across policy and topic areas. The current somewhat piecemeal developments are occurring uniquely within specific policy areas (eg health), rather than across them and without a more strategic approach potential for linking data across portfolio areas may take longer and be harder to achieve.

---

[7] http://www.scotland.gov.uk/Topics/Health/NHS-Scotland/NHSQuality

[8] http://www.cso.scot.nhs.uk/Publications/research.pdf

[9] http://www.lifesciencesscotland.com/about/lisab.aspx

SHIP IN PRACTICE

This section of the Blueprint explains the SHIP infrastructure that we propose to develop by March 2012 for research that uses unconsented healthcare data. The two key elements are the Research Infrastructure and the Governance Infrastructure. Together they generate linked research datasets and provide secure access for researchers within a framework of proportionate governance.

SHIP has developed a unique governance system embodying high standards for data access and linkage and which adopts a proportionate approach to decision-making. This helps to ensure that the right governance pathway is found for each application. The system has three main features (see further section 7):

• The Principles and Best Practices template sets clear and appropriate standards for research governance and these serve as a benchmark for data controllers and researchers.

• A national Privacy Advisory Committee will provide a one-stop-shop for consistent and expert advice to data controllers and researchers on proposed used of data.

• Research coordinators will assist users of SHIP to navigate the system, providing guidance on each step, and acting as gatekeeper to training and accreditation services.

Section 6 describes the services, administrative structure and audit trails that will support a researcher throughout the life of a project. This includes recognition of approved researcher status through the application for access to data, the analysis and, finally, to the statistical disclosure control of outputs.

This Blueprint identifies the parties that are legally responsible for administering the key functions, and defines their specific roles and responsibilities. SHIP's two research promises are to safeguard the privacy of individuals (the Governance Infrastructure) and to promote research (The Research Infrastructure).

6. The SHIP Research Infrastructure

The SHIP Research Infrastructure is designed for the provision of linked datasets. It is anticipated that researchers will apply directly to data controllers for access to non-linked data, i.e. extracts of single source datasets. However data controllers may request that researchers use a safe haven facility to analyse their extract. In this case the data controller may request safe haven facilities from SHIP or may consider setting up a safe haven within their own organisation and SHIP can provide advice about how to do this. The detail of how the data is handled will depend on whether use falls within any consent associated with it and on the privacy impact assessment conducted as part of the proportionate governance process (see section 7). Although anonymisation will be the default mode of working, where a duty to warn or health protection issue arises the matter will be referred to PAC and the directors of public health as appropriate. It should also be noted that there may be occasions when data controllers will be happy to release innocuous data to the researcher directly, outwith a safe haven environment.

6.1. Key Components of the Research Infrastructure

The SHIP Research Infrastructure has three key components which together will provide a timely and consistently high performance research service:

i. **A SHIP indexing service** will maintain a population index based on a unique patient identifier (UPI; eg the Community health Index (CHI) in Scotland). The indexing service will add anonymised identifiers (referenced to UPI) to individual records for the purposes of linking these records across two or more datasets. The indexing service will be separate from the linkage agent.

ii. **A SHIP linkage agent** will use anonymised identifiers to perform the matching of records belonging to individuals from two or more datasets to form a single linked dataset. The identifiers for the linkage will be provided by the indexing service.

iii. **SHIP safe havens.** These have three key characteristics, as defined by the Thomas/Walport Data Sharing Report:

- The safe haven will provide a secure environment for the linkage, storage and analysis of personal data.
- Access to data within safe havens may be from a dumb terminal within the safe haven or remotely via secure thin client technology dependent on risk assessment.
- Only ‘approved researchers’ will be permitted to access the data and there will be penalties for anyone who abuses personal data. Researchers will be bound by a strict code, which prohibits disclosure

of any personal identifying information. Safe havens will carry out statistical disclosure control on outputs to prevent accidental disclosure. The extent and level of disclosure control checks for a given project will be agreed with data controllers.

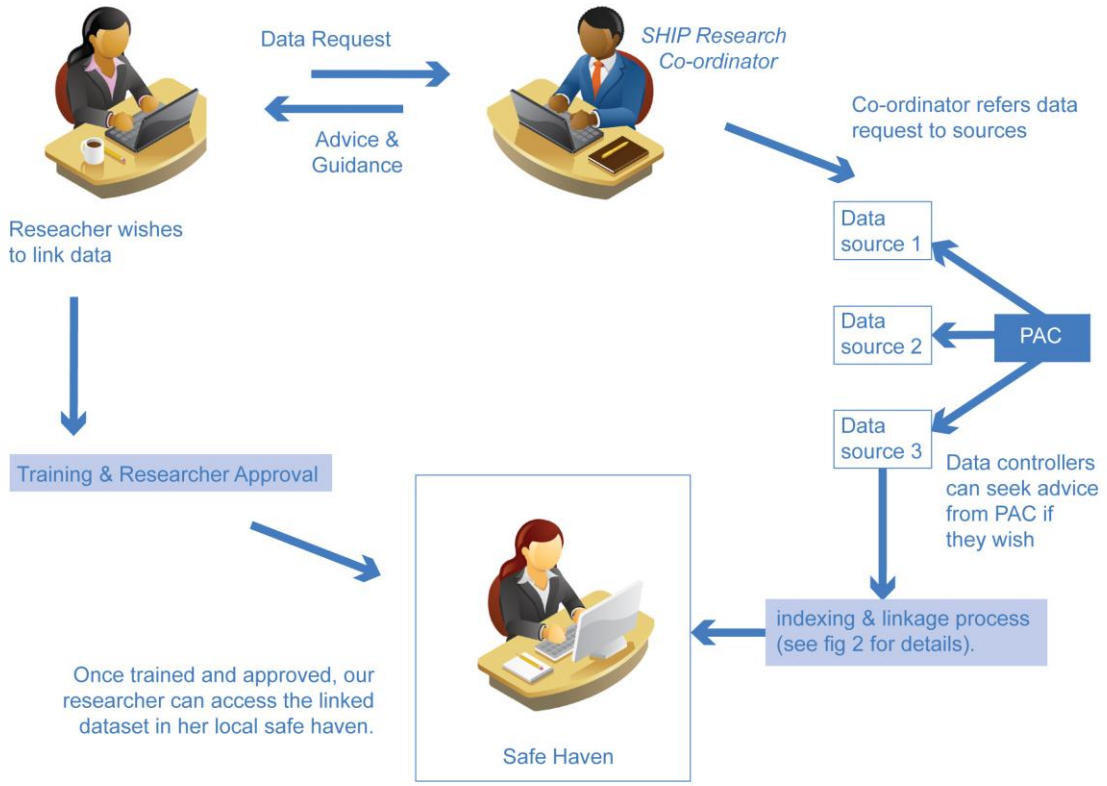


Figure 1: Overview of the SHIP infrastructure showing the interactions between organisations/individuals. Note that in some cases data controllers may deem the linked dataset sufficiently innocuous to be analysed out with a safe haven environment. Once all approvals are in place demographic data is sent to indexer for coding and de-identified data then goes to the linkage agent in the safe haven. The researcher can then analyse the linked dataset in the safe haven. See figure 2 for details of the linkage process.

To ensure high levels of information security and the protection of subject confidentiality the storage of contributory datasets, indexing, linkage of data, and storage of the final dataset will be carried out separately. In practice this means that no individual should be directly involved in any more than one of these processes, but a single organisation could host more than one activity with appropriate segregation of roles and IT facilities. The Indexing Service will be 'stand alone', because this is the only function for which patient identifiers are required. The Safe Haven will be responsible for the remainder of the processes, which use anonymised data: linkage of data, provision of analytical software, the separate storage of the source and linked datasets and the analytical outputs. The Safe Haven will also be responsible for the implementation of other key functions, including an inventory of datasets with associated metadata, statistical

disclosure control, accreditation of researchers (as part of a central register of approved researchers) and adherence to the good governance framework.

The Blueprint proposes to establish a **National Safe Haven** within NHS National Services Scotland (NSS), which will be the central point of the SHIP infrastructure. It is intended that the national safe haven be used for nationwide research, or studies that require linkage of datasets from multiple regions or that have multiple data controllers. This is within the scope of current funding.

It is also proposed that local safe havens will be established at the Scottish Academic Health Science Collaboration (SAHSC) nodes in NHS Grampian, NHS Greater Glasgow & Clyde, NHS Lothian, and NHS Tayside. Other local safe havens may follow.

SHIP has defined a framework of principles and good practices that will support the development of local safe havens. All safe havens participating in SHIP will be encouraged to operate to the same high standards of governance, privacy, confidentiality and audit, as outlined in the Good Governance Framework (GGF).

SHIP is committed to supporting innovation at the local level, and to encourage harmonisation of governance and technical standards within this framework.

SHIP will also work with Clinical Trials Units and epidemiology units that hold consented datasets to ensure the SHIP infrastructure add value to their activities, and does not add unnecessary barriers to the use and linkage of consented data.

6.2. Indexing

SHIP has established a separate **National Indexing Service** to facilitate deterministic linkage of datasets. An indexing service receives only a project code, local identifiers and subject identifiers from the data sources for each of the datasets that are to be linked and no other data. The indexing service creates a study specific anonymised identifier for each subject (called a study number) and returns this with the associated project code, the local identifier and a score describing the likelihood that the linkage is correct. The study number for any individual subject will be different for each submitted dataset in which (s)he appears, to minimise the risk to subject confidentiality. The indexer will also supply a linkage key to the linker in the Safe Haven so that the datasets can subsequently be joined together by matching up the study numbers. The linker receives no other information from the indexer. This process is shown pictorially in Figure 2 below.

Figure 2: Details of the data flow in the linkage process: the indexing service uses demographics to assign project specific identifiers (study numbers) which are returned to the data sources in encrypted form. The key to decrypt the study numbers is sent to the linkage agent within the safe haven. Then the data sources send anonymised data to the linkage agent identified by encrypted study numbers only. The linkage agent decrypts and links the data on the study numbers. See figure 1 for an overview of the system.

6.3. Linkage

The linkage agent and secure access facility will lie within NHS net. The indexing service must be able to receive and transmit information across the NHS net. Data may be supplied by other secure modes of transmission if these comply with SHIP governance principles. All information must be encrypted before transmission between data controllers, safe havens, indexing and linkage services.

The **National Linkage Agent** in the National Safe Haven will carry out linkages involving either data from multiple Health Boards or data held at national level. Linkage of data from a single Health Board or from two or more Boards that have appropriate data sharing agreements in place may be undertaken locally in local safe havens.

The Linkage Agent receives the study code and the study numbers together with the information the researcher needs ("payload data") from the data sources, as approved by the data controller(s). The Indexing Service supplies the linkage key so that study numbers can be matched across datasets, as described above. The Linkage Agent does not receive identifiable information (e.g. names, addresses or CHI numbers). The payload will generally consist of a subset of fields from any given dataset, being those that the researcher requires and for which permission has been obtained.

The Linkage Agent uses the linkage key received from the Indexing Service to join datasets for the study and deposits the linked dataset in a separate area of the Safe Haven. The linkage score will be included for the researcher. The Safe Haven is the Data Controller for both the received datasets and the newly created linked dataset. See Figure 2 for a schematic representation of the data flow between data sources, indexing and linkage services.

6.4. Storing and accessing linked data files in Safe Havens

The safe haven holds the linked datasets and ensures that only approved researchers can gain access.

Researchers will access the data held within the Safe Haven either remotely or via a dumb terminal in a secure access facility dependent on risk assessment. The secure access facility may be situated either within the same safe haven that holds the data or in another safe haven (with appropriate permissions).

Analytical software will be available within the safe haven for use by researchers. The dumb terminals will be configured so that the researcher cannot download or remove any of the data or outputs held at the Safe Haven. A dedicated file space will be provided for the researcher to store their outputs pending release by the safe haven.

De-identified data will be held separately from any data that carry identifiers eg consented datasets. Every Safe Haven must keep a suitable record of the use of its facilities for security and audit purposes. Software is available to provide a full log of terminal access.

6.5. Statistical Disclosure Control and Archiving

The Safe Haven is responsible for undertaking Statistical Disclosure Control prior to release of analytical outputs to researchers. This will be done by appropriately trained employees of the safe haven. Once the output is deemed safe it will be sent to the researcher electronically.

The level of disclosure control required will vary between studies. It is the responsibility of the data controllers for the contributory datasets and the Caldicott Guardians to decide upon the appropriate level of disclosure control at the beginning of the project before the datasets are linked and access is provided to the researcher.

The Safe Haven will provide an archiving service for all linked datasets so that researchers can return to the dataset for an agreed specified period of time following the initial analysis. While an extension to the time may be easily arranged, the analysis must still relate to the research question in the original application. If not then another application must be submitted. It is recognised that a well-developed research dataset is a valuable resource that may be useful over a long lifespan. Automatic destruction may not always be appropriate.

6.6. Support services for Researchers

6.6.1. Training

A system for approving researchers will be established in order to ensure that people who access data are adequately trained in such key areas as patient confidentiality, statistical disclosure, and the Data Protection Act (1998). SHIP is developing a dedicated online training package for this purpose. To attain "approved" status a researcher must meet three criteria:

- complete the prescribed training course;
- be sponsored by an approved institution (eg University, NHS Board board);
- either demonstrate a suitable track record in research or be sponsored by an established researcher.

The approved researcher (or a suitable person within their institution) will then be required to sign a data access agreement for each project requiring use of the SHIP infrastructure. This agreement clearly states the researcher's obligations in respect of their access to the data and the sanctions that may be imposed for any for breach of the agreement.

The National Safe Haven within NSS will maintain a register of approved researchers. Approved status will last for a fixed period of time and refresher training will be required if the researcher wishes to remain on the register after this time. Sanctions may be imposed against both the researcher and their employing institution if the rules for using the Safe Haven are breached. These sanctions will include suspension of access to SHIP facilities for a period of time.

6.6.2. SHIP Research Co-ordinators

Research coordinators will support approved researchers throughout the life of a project from advice on data access through to the statistical disclosure control and release of outputs. Figure 3 shows all the processes involved from the initial

contact with SHIP through to the release of outputs following completion of the data analysis.

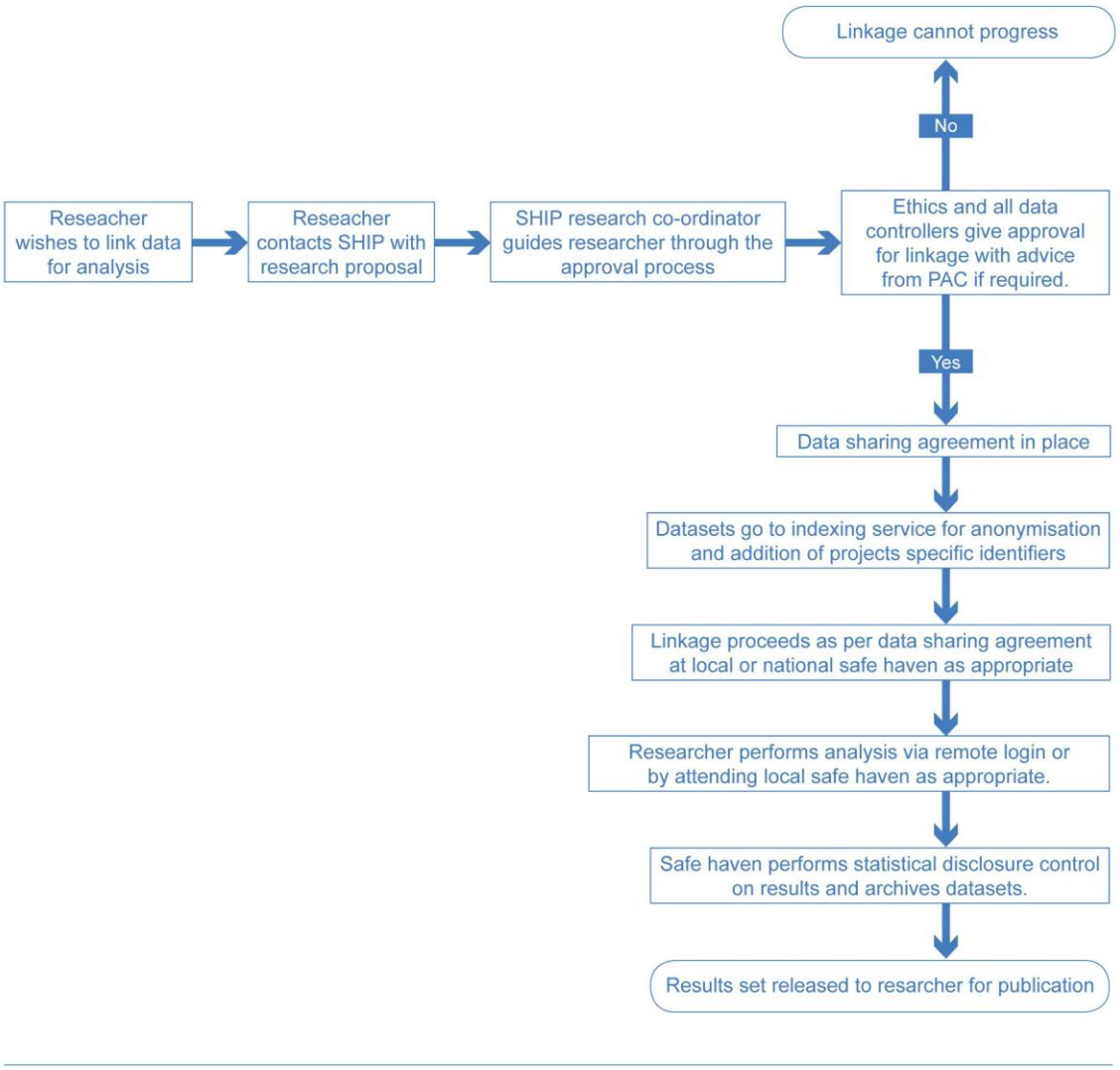


Figure 3: the sequence of events in the data linkage pathway. See figure 1 for the interactions between organisations/individuals.

The necessary permissions from data controllers and ethics committees must be obtained before any project commences. The responsibility for gaining ethics approval for a research project lies with the researcher, but SHIP research coordinators will assist approved researchers in this process.

Once data controllers have agreed to release the data and the researcher has signed a data access agreement then the data will be indexed and linked, and stored securely in a safe haven.

6.6.3. Inventory of Datasets

Safe havens will provide an inventory of datasets to which they have access. This will increase awareness amongst researchers of record linkage capability with associated metadata to encourage research activity. Safe havens will facilitate dialogue with data custodians about data quality and interpretation which will benefit both researchers and custodians.

7. The SHIP Governance Infrastructure

7.1. The Good Governance Framework

To underpin these proposals, SHIP has developed a Good Governance Framework which is presented in detail in Appendix 7 and on the SHIP website. The structure of this framework is based on **principles** and **best practice**, following the OECD Guidelines on Human Biobanks and Genetic Research Databases.[10]

The Good Governance Framework is intended as a guide to inform the public on how data are used, for researchers who wish to use the SHIP infrastructure and for others involved in data sharing and information governance both within and beyond the health sectors. It is not intended to cover exhaustively all aspects of governance, nor is it a statement of legal rules. It is assumed that all parties involved in data sharing and working with the SHIP Infrastructure are aware of their legal responsibilities and comply with them. It serves to define a high quality framework for Scotland according to which the SHIP Infrastructure will be governed and against which users will be held accountable. It is an expression of commitment to promote the public interest in scientifically sound, ethically robust research while appropriately protecting the privacy and other interests of the people whose data are used in such research.

7.2. Format of the Good Governance Framework

'Principles' are fundamental starting-points to guide deliberation and action. They reflect the values that underpin the SHIP Infrastructure and its commitment both to promote the public interest and to protect individual interests. Principles are not rules; indeed they sometimes conflict. This is why they are starting points for deliberation or action. However, because of their fundamental importance, it is expected that they are followed where they are relevant to a given data use, storage, sharing or linkage practice. Any departure must be fully and appropriately justified.

[10] http://www.oecd.org/dataoecd/41/47/44054609.pdf

"Best Practices' are examples of principles in action. These are instances of optimal governance and in that sense they are aspirational. As with principles, where instances of best practice are not or cannot be followed, clear justification should be offered. Together, these principles and best practices are an indication of the standards expected within and upheld by SHIP, and to be adopted by the SHIP infrastructure across Scotland.

In considering its two key principles at stake: (1) promotion of the general public interest and (2) protection of the privacy and other interests of individual citizens, the Good Governance Framework has defined 15 key areas of responsibility and accountability. These are:

- Public Interest
- Privacy
- Consent
- Anonymisation
- Authorising/advisory bodies
- Governance
- Access
- Trusted Third Parties
- Data Controllers and Data Processors
- Clinical Trials
- Cross Sector Sharing
- Data Sharing Agreements
- Public and Stakeholder Engagement
- Sanctions
- Benefit Sharing

It is hoped that these proposals for change will provide for a shared responsibility between the public, healthcare professionals and researchers over the use of health data to the benefit of public health and the future care of the population at large.

7.3. Safe, effective and proportionate governance.

The proposed SHIP model is founded on a system of effective, safe and proportionate governance, and the concept of **data stewardship**, which has, as

its main objective, the management of data assets in order to improve their reusability, accessibility and quality, whilst ensuring good governance.

The system we aspire to create is safe and also proportionate. It reduces unnecessary regulatory burden while maintaining the highest governance standards for data linkage and research approvals. The Good Governance Framework provides decision-makers within SHIP with a common framework of reference for deciding which linkages should be approved and which checks and balances should be in place.

7.4. Who is responsible?

Data Controllers are persons who (either alone or jointly or in common with other persons) determine the purposes for which and the manner in which any personal data are, or are to be, processed. They are primarily responsible in law for overseeing data protection and their responsibilities are discussed in Appendix 6.

Caldicott Guardians are senior persons in each of the health boards responsible for protecting the confidentiality of patient and service user information and enabling appropriate information sharing.

Privacy Advisory Committees advise on the correct balance between protecting personal data and making data available for research, audit and other important uses and ensure that any information releases are carefully controlled. For example, permission to use data held by NHS National Services Scotland, NHS Central Records and National Records of Scotland (NRS, formerly GROS and NAS) is obtained via the NSS Privacy Advisory Committee (PAC).

These individuals/organisations, are charged with ensuring that those processing data under their authority comply with both the spirit and the letter of the SHIP Blueprint. Decisions taken by authorising and advisory bodies in this context should be publicly available and fully justified.

7.5. SHIP Authorising/Advisory Bodies – an enhanced role for the Privacy Advisory Committee.

Data Controllers and Caldicott Guardians can authorise the use and sharing of data under their custodianship. Advice can also be sought from bodies such as PAC or local research ethics committees on the appropriateness of specific requests to use or share data. Thus individuals and/or independent bodies can act in an authorising or advisory capacity with respect to data use and linkage.

We propose to strengthen the role of PAC to undertake approval of use of national datasets. As described in Sections 7.3 and 7.4 above, the use of proportionate governance will ensure that applications are ‘triaged’ and the appropriate regulatory burden applied depending on the risks involved; some applications being fast tracked through the system, while others requiring a full PAC review.

In addition, over time it is envisaged that a new PAC would be established with a role beyond NSS/NSR, which would allow for advice pertinent to non- health datasets.

The new PAC will clearly articulate and make readily available the criteria and procedures by which they decide whether or not to sanction data use. We propose that the new PAC will be re-constituted as a group and include members from diverse backgrounds, who possess the necessary expertise to make appropriate and justifiable decisions on use/access. Consideration will be given to how it reports to NSS and other public bodies.

7.6. The role of proportionate governance in reducing bureaucracy

The PAC will take a proportionate approach to governance by considering four key issues:

- **Identifying what is at stake?** The Good Governance Framework is a living instrument that reflects the core purposes of SHIP, sets high standards for the workings of SHIP, and provides a template to identify the core issues that should be taken into account when deciding on appropriate data linkages and associated downstream terms and conditions;

- **Who is involved and who is responsible?** SHIP provides detailed guidance to data controllers and data stewards (such as Caldicott Guardians) on their primary responsibilities and also offers a streamlined and supported mechanism to receive one-stop advice on requests for data use and linkage;

- **What are the beliefs, burdens and risk involved with each application?** Our decision-making systems ensure that core considerations are taken into account, together with robust privacy impact assessment, which helps to identify the particular governance issues related to each application;

- **What is an appropriate research pathway for each application?** This is about engaging the right people and principles and directing applications down the right governance pathway. At present there is only one pathway – full review. SHIP proposes mechanisms to conduct bespoke risk assessment that is linked to five possible options for disposal (rejection and Categories 0-3 as shown in Figure 4). The degree of regulatory burden is proportionate to the risks involved. Less risk means less burden. Privacy concerns remain at the heart of our governance approach.

Examples of applications that might follow Low Impact research pathway – no need for full review, approval and standard terms and conditions (Fast Track)

- No concerns raised at stages one or two – application is for linkage which is non-disclosive and non-sensitive and safe haven system will be used

- Application is for a particular linkage in keeping with broad purposes already approved between SHIP and trusted researchers for long-term project, eg the Scottish Longitudinal Study
- Applications is for a non-contentious extension of a previously approved linkage

Examples of applications that might follow Medium Impact research pathway – triaged but with option for full review

- Moderate risks or concerns arising from the privacy impact assessment at stage two
- Repeat requests from multiple sector/international/researchers who are able to demonstrate a trusted track record with respect to SHIP
- Application is for a non-sensitive and non-disclosive linkage but safe haven system will not be used

Examples of applications that would follow the High Impact research pathway – requiring full review

- Failure to satisfy any one of the criteria for assessment at stage one (eg – questions over the public interest in the research, safe people, safe systems or safe environments, or wider risks such as reputation of the data controller)
- Concerns arising from the privacy impact assessment at stage two (eg – very sensitive data; serious risks of disclosiveness)
- Multiple sector or international linkage being requested for the first time.

In all cases, appropriate terms and conditions for sharing and linkage will reflect the nature of the governance pathway followed by any given application.

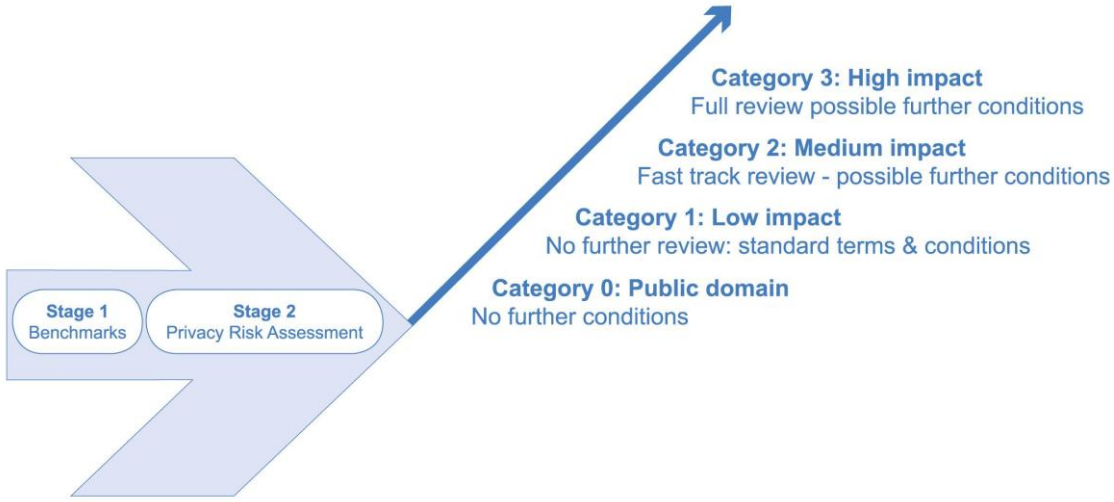


Figure 4: the proposed stratification to be developed by PAC. Benchmarks are public interest, safe people, safe systems, safe environment, relative risks. The privacy risk assessment is based on criteria such as disclosiveness, sensitivity etc.

7.7. Governance Infrastructure in Practice

7.7.1. SHIP Administrative and Managerial Structures

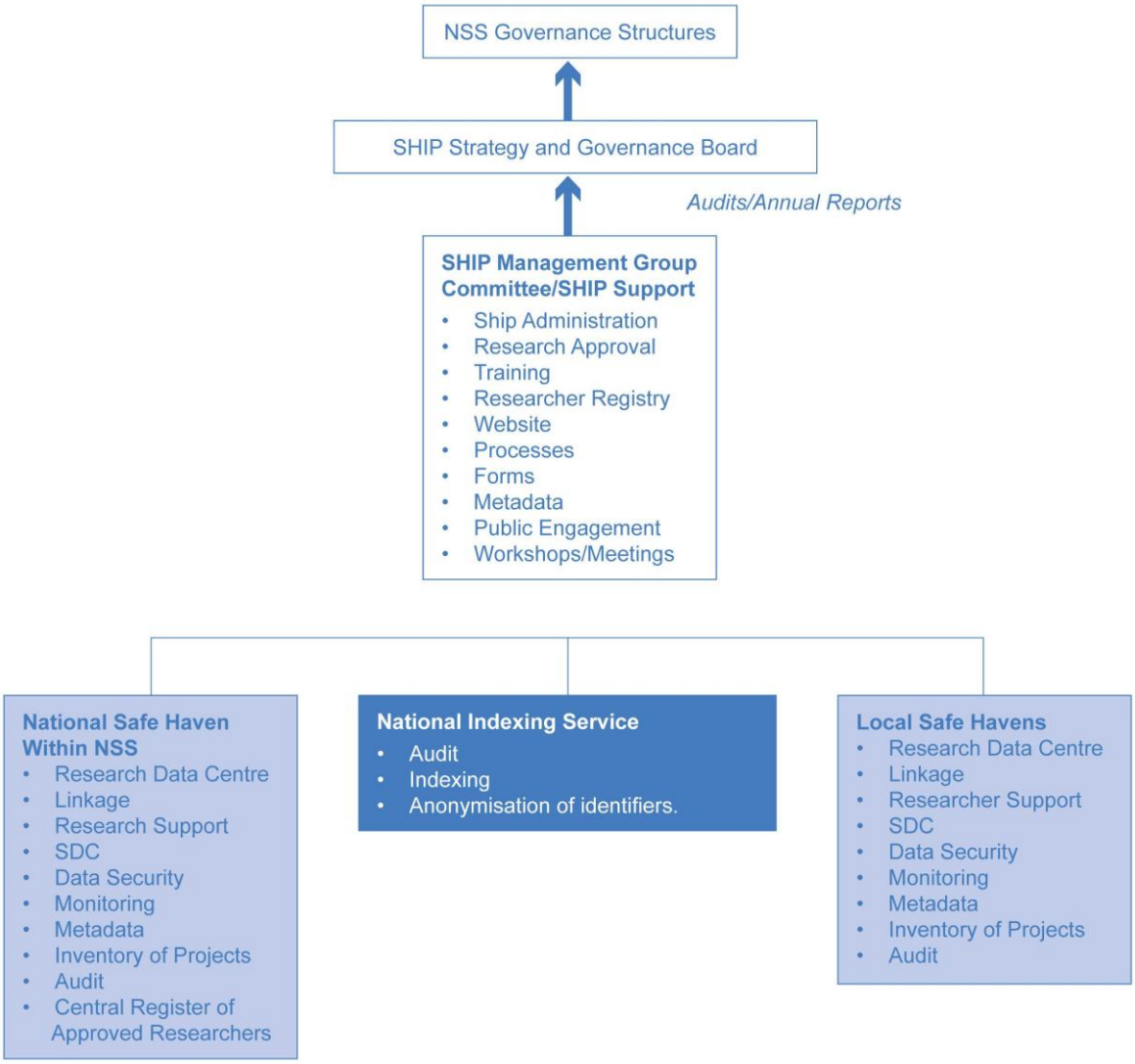
A schematic representation of administrative and managerial functions is presented in figure 5, below. National activities, which form the backbone of SHIP, are shown in detail. Local safe havens which interact with the National SHIP Infrastructure will be responsible for adopting the same standards of governance, security and audit for local structure and also be represented on the SHIP Management Committee.

We propose that a SHIP Strategy and Governance Board (SGB) will play the central governance role and will be established by the Scottish Government Health Department. We propose that the SGB Board will assume responsibility of the oversight and strategic direction for the SHIP Infrastructure and monitor progress of the SHIP Programme. This group will have representation from the major stakeholder organisations, including the national and territorial Health Boards, Universities, lay representation, Chief Scientist's Office and other agencies collaborating with the SHIP Infrastructure. The Board is also responsible for resolving risks and issues escalated by the SHIP Management Group. .

Its remit is to:

- Oversee progress of the SHIP Programme;
- Provide the programme mandate and investment decision;
- Create an environment in which SHIP can thrive;
- Provide continued commitment and endorsement of the SHIP programme;
- Monitor progress of SHIP against its strategic objectives;
- Provide visible leadership and commitment to SHIP; and
- Co-ordinate applications for further funding and receiving reports from independent audits of processes.

The national SHIP Infrastructure functions will be directly accountable to the SHIP Management Committee for operational purposes, but accountable to established NSS Structures (eg the Clinical Governance Committee) for Governance purposes. Sub-groups of the SHIP management group will take responsibility for specific activities (e.g. website).



Local Safe Havens and Indexing Services will have their own management and governance structures, but will be expected to mirror the National SHIP's Governance standards.

A SHIP Management Committee will be established. It will be responsible for operational matters and will report to the SHIP Strategy and Governance Board. The management committee will ensure that essential functions such as maintaining the processes for researcher approval and applications to link and access data are completed efficiently and effectively. It will be responsible for the following activities:

- Research training.

- Research approval.
- Register of approved researchers.
- Website maintenance.
- Source of information on SHIP processes, approval and forms.
- Metadata definitions and archiving.
- Public engagement.
- Workshops/Meetings.

The SHIP Management committee will be supported by the national safe haven manager, co-ordinators, a small amount of technical resource and a website manager.

7.7.2. Audit and Accountability

All SHIP processes will be subject to regular independent audit to ensure adherence to legal requirements and to SHIP's data security and confidentiality principles. This audit will be carried out annually and the results published on the SHIP website. Local Safe Havens wishing to connect with SHIP will be expected to have equivalent audit procedures. Safe Havens will be accountable to governing bodies. These bodies will have representation from data controllers, researchers and the public.

7.7.3. Safe Havens: Relationships with Data Controllers

The Safe Haven will be responsible for obtaining permission from the data controller(s) before any dataset is used for research or linked to other datasets within the Safe Haven. Once permission has been given, data controllers cannot subsequently influence the publication of the work resulting from their dataset.

The role and function of data controllers are laid down in the Data Protection Act and are explained in appendix 6. It is important to note that assessing privacy risk is an integral part of the data controller's role. It is also the responsibility of data controllers to ensure the development of transparent policies that demonstrate their understanding of public interest and the basis upon which they will disclose data. These policies will be the responsibility of the Safe Haven in its capacity as data controller for the received data and linked data sets that it holds.

7.7.4. Safe Havens: Security Principles

While being autonomous, local safe havens will be encouraged to adhere to security principles common to SHIP. The manner of implementation however can be locally determined.

- A data controller must be defined at each stage of the process and that controller must be aware of their responsibility.
- An indexing service should never receive any information about the patient/client/research subject other than the required identifiers.
- All information must be encrypted before transmission between data controllers, safe havens, indexing and linkage services.
- All data in the safe haven are held on secure servers located on the NHS network.
- Data should be de-identified unless otherwise agreed.
- Where the safe haven must hold data with identifiers it must be held on separate servers from de-identified data.
- All data processes are carried out within secure offices, on the NHS network, prior to secure releases to external data users
- Safe haven to check that all approvals are in place before receiving data for processing.
- A record of all projects, approvals and data releases is to be kept on a Project Management System
- All processes are audited annually by external auditors and actions taken in response to issues raisedwith periodic reports to relevant bodies/persons.
- All procedures are reviewed annually.

7.7.5. Safe Havens: Functionality Principles

Effective functionality is of vital importance lest the benefits of research be lost to the community. Therefore safe havens should address the following issues:

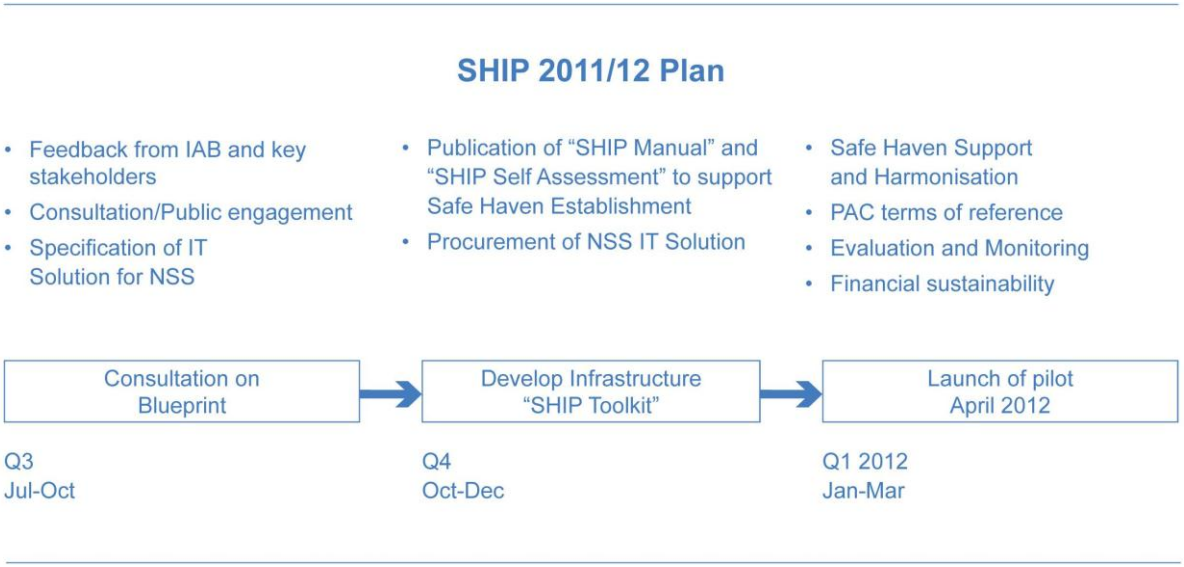
- understand research functionality as well as security
- be physically and virtually accessible
- be physically and virtually sustainable
- allow SAS, SPSS, Stata and specialised programmes

- have capacity - terabytes per project
- allow creation, storage and backup of intermediate files
- allow sufficient views of the data to allow proper debugging
- design suppression of low copy numbers in secure views in consultation with experienced analysts to preserve functionality
- make intermediate output available in short order - about an hour or by arrangement
- allow analysts to store and remove their own programmes via the safe haven manager
- provide a reasonable workplace environment

8. SHIP milestones and project plan

The SHIP team are committed to a comprehensive programme of public and professional consultation. A full list of groups and bodies who have already commented on the Blueprint, or will be in the near future is given in Appendix 4. Allied to consultation is the development of the support tools (standards, policies, terms of reference, administrative structures) necessary to implement the research and governance infrastructures.

A high level overview of the next 12 months is given in figure 6 (below). More information is available on the SHIP website.



9. Resources and funding

SHIP has funding until May 2013. It is proposed that local safe havens will be self-supporting, however a funding model will need to be developed for the national facilities. Discussions are on-going with Scottish Government, including the Chief Scientist's Office, to gauge support for such a development, which will be coupled with a cost recovery model for all applications that source data from such a facility. We believe this joint model will be sustainable in the long term and a business plan is being drafted accordingly.