

Joined up data for better decisions: A strategy for improving data access and analysis











Joined up data for better decisions

Answering the important questions for Scotland through legal, secure, ethical and efficient data linkage

A strategy for improving data access and analysis

Our vision for the future is one where evidence of what works in delivering positive outcomes for all of Scotland is delivered quickly and efficiently with minimal burden on front-line services. By improving the ethical and legal governance arrangements, and the technical capacity to securely and efficiently link statistical data, we will enable the research needed to inform policy decisions. Scotland will be recognised the world over as a hub of innovative and powerful statistical research, attracting investment and job creation.

In <u>Scotland's Digital Future – Delivery of Public Services</u> Scottish Government committed to using information and communication technologies to achieve public services that are high-quality, continually improving, efficient and responsive to local needs (National Outcome 16). The Data Linkage Framework for Statistics and Research is a vital element of that commitment. Our ambitions are to:

- 1. build on existing successful programmes collaboratively to create a culture where legal, ethical, and secure data-linkage is accepted and expected;
- 2. minimise the risks to privacy and enhance transparency, by driving up standards in data sharing and linkage procedures;
- 3. encourage and facilitate full realisation of the benefits that can be achieved through data-linkage to maximise the value of administrative and survey data.

The publication of this document follows a consultation process, including a written stakeholder consultation exercise, deliberative research into public opinions, and many discussions with interested parties. The framework builds on Scotland's proven track record in effective and secure data linkage, and in particular on The Scottish Health Informatics Programme (SHIP)¹.

This document sets out how the strengths and successes of SHIP will be expanded beyond health. It is published alongside a set of guiding principles which we recommend as the basis for all decision making relating to data linkage.

¹ SHIP is funded by the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council.

Contents

3
6
6 7 7
9
10
11 12 12 12

SECTION 1: SCOPE AND EXPLANATION OF DATA LINKAGE

For the purposes of this Framework, data linkage is the joining of two or more administrative or survey datasets to greatly increase the power of analysis then possible with the data.

This framework is concerned exclusively with the linkage of data for research and statistical purposes where there is **no direct impact on an individual because of information about that individual being linked**. Examples can be seen as falling into three categories:

- Development and production of Official Statistics, including the production of aggregate statistical information.
- Production and dissemination of research resources, such as longitudinal statistical products like the Scottish Longitudinal Study.
- Ad-hoc research projects, or linkages conducted to answer specific research questions using statistical analyses, such as the West of Scotland Coronary Outcomes Prevention Study.

This framework concerns linkages for research and statistical purposes only. It does not cover the sharing of personal information about an individual between organisations in order to deliver a co-ordinated service to that person. Data linkage for that purpose raises a different set of legal, ethical, and logistical issues. The following examples are all **beyond the scope** of this framework:

- A Child Protection Officer sharing a particular family's case file with a School and the Police, in order that all three can work together to protect a child at risk.
- A Local Authority sharing information about named individuals claiming Housing Benefit with any other organisation for the purpose of combating fraud.
- A GP sharing information about an individual patient's symptoms or diagnosis with a hospital in order that the patient receives a co-ordinated service from all parts of the health service.

Two examples of data linkage for statistics are given below.

Example 1: The Scottish Health Survey

The Scottish Health Survey is a sample survey conducted through face-to-face interviews. Respondents are asked about a wide range of health issues including smoking, alcohol intake, diet, and levels of physical activity. Some biological measures are also taken, such as waist and hip circumference.

All aspects of the Scottish Health Survey, including data linkage, are approved by The National Research Ethics Service before being conducted.

All respondents are asked to consent to their name, address and date of birth being sent to the Information Services Division of NHS Scotland (ISD) so that their responses to the Health Survey can be linked with records holding data on medical diagnoses, in-patient and out-patient visits to hospital, and other information about cancer registration, GP registration and mortality.

Where the respondent gives consent for linkage the following process then occurs:

- First, respondents' name, address, date of birth and a unique serial number (which is different to that used on the publicly available survey dataset) are separated from the rest of the health survey dataset (all the responses to the health survey questions) and sent by the survey contractors to ISD.
- ISD then link respondents name, address, date of birth and a unique serial number, with the health records, and delete the respondents name, address, date of birth. This leaves a file of unique serial numbers and administrative health data. This file is then sent to a named analyst in Scottish Government.
- The Scottish Government analyst then merges that file with the data collected through the Health Survey, using the unique serial number. The unique serial number is then deleted and a new random one added.
- This dataset is then analysed, results are checked for risk of disclosure, and the aggregate results and conclusions and disseminated as widely as possible.

In this example, all data are sent between the three organisations by secure FTP (File Transfer Protocol) servers and can be accessed only by a small number of named people in each organisation. None of the three organisations – Scottish Government, the contractor, or ISD - has access to both survey and health records with personal identifiers attached at any time.

Example 2: The Scottish Health and Ethnicity Linkage Study

The Scottish Health and Ethnicity Linkage Study (SHELS) aims to explore the health experiences across different ethnic groups in Scotland. The project involves the linking of Census 2001, from which ethnic group can be obtained, to health data, which shows the number of times an individual has experienced a disease or health condition.

To do the linkage Census Division, within National Records of Scotland (NRS), sent a dataset to a safe haven, physically located within a NRS building. This dataset contained name, date of birth, address and a unique serial number for each person who completed a Census questionnaire in 2001. No other Census variables are included on this dataset.

Information Services Division of NHS Scotland (ISD) also sent a dataset, derived from health records, to this safe haven. This dataset also contained name, date of birth and address and also a unique serial number which is an encrypted version of the CHI number. No health related variables were on this dataset. A named ISD employee linked the two datasets in the safe haven and then removed name, date of birth and address to leave a file containing the encrypted CHI number and the unique Census number. This file, called the index, was sent to a named individual within Demography Division of NRS.

ISD then prepare a separate dataset containing the encrypted CHI number and health data but not name or address and send this to Demography Division. Census Division then send a file to Demography Division containing the Census unique serial number and some demographic information, but again not name or address.

Demography Division then put these two datasets together using the index, resulting in a linked dataset containing health information and ethnic group, but not name or address. This dataset is stored in the safe haven within NRS, where analysis takes place, and all results are checked for risk of disclosure before being released from the safe haven.

Neither ISD or NRS, nor any individual within these organisations, has access to both the Census and health data with personal identifiers attached. Explicit consent from data subjects was not sought for this study. All methods were approved by the Multicentre Research Ethics Committee (Scotland), the Community Health Index Advisory Group and the NSS-NRS Privacy Advisory Committee.

SECTION 2: BACKGROUND

In early 2012 we published the consultation document, *A Scotland-wide Data Linkage Framework for Statistics and Research*, which set out the aims, benefits and challenges to data linkage alongside a draft set of 'Guiding Principles' and suggested functions and objectives for a Data Sharing and Linking Service (previously named the National Data Linkage Centre) and a Privacy Advisory Committee.

In parallel to the written consultation process we had a number of meetings and discussions with stakeholders, sought advice from international experts, commissioned a series of public deliberative events and engaged with discussions of the Administrative Data Task Force.

Benefits of Data Linkage

In the consultation paper we asked for views on the benefits of data linkage and the barriers to more data linkage. Most of the responses concurred with the summary given in the consultation paper or added further detail or examples. Furthermore, all participants involved in the deliberative research events recognised the potential benefits of data linkage.

It is beyond the scope of this strategy document to restate or provide a comprehensive overview of the benefits of data linkage. Rather, we recognise that the benefits of data linkage are many and can, broadly speaking, be summarised as:

- Speeding up cycles of improvement through the delivery of a higher quality cross-sectoral evidence base to inform public policy and strategic planning, spending and delivery decisions.
- Maximising the value of existing data to develop efficient and reliable methods of producing statistics, including better statistics at sub-national level.
- Allowing relatively low cost longitudinal research to be conducted both retrospectively and prospectively.
- Increasing the capacity to robustly evaluate programmes, by providing the potential to answer far more sophisticated research questions than otherwise possible.
- Improving the quality and consistency of data, through general feed-back loops following linkage activities.

Barriers to Data Linkage

In terms of issues relating to incomplete data, or data that cannot be linked; capacity for secure exchange and access to data; and the capacity of public sector organisations to analyse and make use of linked data, consultation responses generally concurred with the summary given in the consultation paper or added further detail or examples of these challenges.

It is clear that in order to make better use of administrative data it will be necessary in many areas to first improve the quality and consistency of the administrative data systems that exist across the public sector in Scotland. Where this has been done, for example in relation to school and pupil data through the ScotXed Programme, substantial benefits to all are evident.

Likewise, building on and expanding from the existing high quality facilities for secure exchange and data access in Scotland, and further increasing the strong analytical skills and expertise to analyse and make use of linked data, are important components.

In the consultation paper we also suggested "Uncertainty about the legalities and public acceptability of data linkage". Responses provided examples and evidence of this, and it is clear that there is considerable variation in the interpretation of the legal and regulatory environment, and that data custodians are often unsure whether they can legally and appropriately make data available for linkages.

In addition though, the consultation responses, along side the broader range of views we have heard, have been helpful in beginning to unpack a range of challenges, ethical considerations and risks related to data linkage.

Public Interest and Public Acceptability

It is necessary that data linkage activity is conducted in the public interest, and that it is acceptable to the public. In order to deliver a data linkage framework that encourages and enables both of these things, we will engender an approach of transparency and public engagement in all aspects of the framework.

The consultation responses and deliberative events with the public showed that the scope for commercial gain arising from data linkage activity is a cause for concern amongst some.

This strategy is for the public, private and voluntary sectors in Scotland. The Guiding Principles are clear that public benefit should be at the forefront of data controllers' deliberations about whether or not to permit sharing and linkage of data.

In further developing the framework, to inform the deliberations and guidance of the Privacy Advisory Committee and the policies and procedures of the Data Linking Service, work with members of the public and other stakeholders to fully explore these issues will continue. We will facilitate a dialogue between members of the public and both public and private sectors about the appropriateness, the concerns, the benefits and the risks involved when the private sector is involved in the use and analysis of data collected and held by public bodies.

The Scottish Government does not intend to "sell" linked data or to encourage other public bodies to do so.

Privacy and Consent

Protecting identity and privacy in line with legal requirements (including but not limited to the Human Rights Act and the Data Protection Act) underpins this Framework and the Guiding Principles.

Whenever data are accessed, analysed, shared and linked there is some risk to privacy, however small that may be in most cases. The main risks are data loss (through accident or malice), or statistical disclosure (the potential to identify an individual within a dataset by their unique or rare combination of characteristics). It is therefore vital that in encouraging more use of data, and data linkage in particular, these risks need to be at the forefront of considerations and their mitigation must be a high priority.

The risks to privacy must be carefully considered and managed both where data subjects have given consent for a specific data usage and where alternative governance mechanisms (e.g. access granted by the data controller) are used. Consent of data

subjects for any specific usage is the 'gold standard' governance mechanism, and is a major concern of many members of the public, data controllers and researchers. Consent is not, however, a legal requirement for data sharing or data linkage to occur, and there are other ways to protect the rights of data subjects.

The European Commission is proposing new regulation to protect individuals relating to the processing of personal data and the free movement of such data. The ability to collect, analyse and transfer personal data is critical to data linkage for statistics and research. The EU Regulation is still under negotiation and the Scottish Government is liaising with the Ministry of Justice as the negotiations proceed. The regulations will be carefully examined and appropriately applied.

Archiving and Transparency

Service to amass data about individuals. Likewise concern was raised that any organisation, now or in the future, may hold excessive quantities of linked data. The principles relating to separation of functions will address this by encouraging separation of the indexing, linking and analysis functions, as well as the time-limited storage of linked data. Decisions on handling, storage and linkage of data will continue to sit within the established concepts of Public Interest, Governance and Public Transparency. This will place a clear responsibility on individuals and organisations engaged in this work to do so in an open manner which is consistent with the principles and responsive to evolving concerns of the public.

There is no intention to create a single archive, or store, of excessive quantities of linked data, or a single organisation that controls all data linkage in Scotland.

By establishing facilities whereby cross-sectoral linkage can be achieved but personal identifiers (e.g. names, address, or persistent identifiers such as Community Health Index number) are transferred and held separately from any 'characteristic' data, the Data Linkage Service will increase the capacity within Scotland for optimal methods of data linkage.

By establishing a set of principles in which the importance of transparency is central, and a Privacy Advisory Committee with representation from members of the public, we aim to develop a system of checks and balances that prevents any one organisation from gaining too much personal data.

SECTION 3: AMBITIONS AND AIMS

A strategic approach to improvements in data linkage processes, from Privacy Impact Assessments and Data Sharing Agreements, through the logistics and technical aspects of linkage to the dissemination of results, will ensure Scotland remains a world leader in data linkage.

We have an ambition to build on existing successful programmes collaboratively to create a culture where legal, ethical, and secure data-linkage is accepted and expected

- By fostering collaboration between existing data linkage programmes and initiatives we will co-ordinate what is currently a fragmented landscape of activities to achieve immediate benefits through the sharing of ideas, solutions, best practice and methods.
- By encouraging collaboration in the use and procurement of data linkage ICT across the public sector, we will avoid unnecessary duplication and so reduce purchasing and running costs.
- By increasing the value of datasets, increasing the wealth of good practice and experience in data linkage research, and demonstrating that Scotland is a worldleader in this field, we will continue to encourage research investment into Scotland.

We have an ambition to minimise the risks to privacy and enhance transparency, by driving up standards in data sharing and linkage procedures

- By recommending a set of guiding principles for all data linkage activity we will both raise standards and create a clear and consistent approach to data linkage across Scotland.
- By working with the Information Commissioner's Office to increase the understanding of the Data Protection Act and other legislation across all those involved in linkage activities we will encourage respect for privacy and proportionate and effective approaches to mitigating privacy risks.
- By encouraging transparency, openness and public involvement in decision making we will increase public understanding about how and why personal data are used for statistical and research purposes, and ensure the public value of research involving linkage methods.
- By co-ordinating and harmonising data access and approval processes across sectors, without adding layers of bureaucracy, we will streamline the establishment and management of data linkage projects.

We have an ambition to fully realise the benefits that can be achieved through datalinkage to maximise the value of administrative and survey data

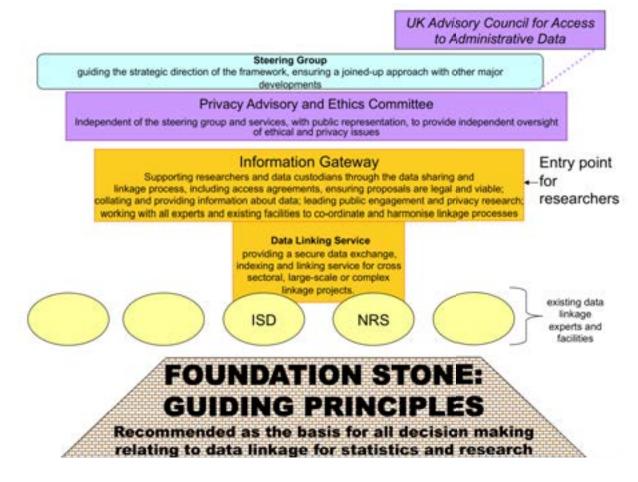
By enhancing the data standards and statistical capacity we will improve the quality of data that exists and make advances on the evidence base, particularly in terms of a joined-up understanding of how outcomes are achieved, allowing for more informed spending on public services and early interventions that save money in the long run.

SECTION 4: DELIVERY

The Data Linkage Framework is composed of four blocks, which will be delivered and supported through a set of work streams. The four blocks are:

- The Guiding Principles to assist data controllers and decision makers (e.g. ethics committees, privacy committees, data access panels) take a proportionate approach to managing the risks inherent in any data linkage;
- A Privacy Advisory and Ethics Committee to advise on data sharing and linkage and associated ethical, legal and social issues;
- Information Gateway and Data Linkage Service to support researchers and data custodians develop legal, ethical and practical linkage projects, provide separated indexing and linking functions for cross-sectoral data, and secure access for analysis;
- The Steering Group to oversee the strategic direction of the Linkage Centre and Advisory Committee, ensuring a joined-up approach with other major developments in Scotland.

All elements of the framework will both facilitate and rely on collaborative working, sharing of best practice and harmonisation of activity across the **network of experts and facilities in Scotland**.



The Guiding Principles

In order to support data custodians, researchers and other stakeholders in taking decisions about safe and effective linkage, the foundation stone of the Framework is a set of Guiding Principles, published at www.scotland.gov.uk/GuidingPrinciplesforDataLinkage.

The principles are a guide for thinking about the kinds of issues in play and for justifying decisions about linkage or sharing. They are not rules, they are not prescriptive and they do not restate the law.

There are three central considerations that the principles aim to assist:

- do the potential public benefits from the research justify the risks to privacy?
- what can be done to mitigate the risks to privacy?
- what can be done to increase the public benefits of data linkage and sharing?

Consideration and proportionate application of the principles should help balance these considerations, increase the public benefits from data usage and mitigate risks to privacy. They should help to promote consistency of decision-making and also to foster a degree of trust in the high levels of protection and transparency that the system delivers.

It is the very nature of principles that they do not specify exactly how they can be met. Rather, principles must be considered and applied in the context of a particular project, with its particular objectives and particular risks. Examples of good practice in specific instances and suggestions for implementation will be collated and shared by the Information Gateway Service (see below).

Network of experts and facilities

Scotland has long been a pioneer in the use of linked health service data for statistical research, and there is well established capacity to use data linkage in evaluating health interventions for the benefit of the population's health. Many excellent data linkage facilities and safe havens for data access operate within Scotland and the Data Linkage Framework aims to encourage and facilitate collaborative working, sharing of best practice and joined-up approaches to resource investment across them all.

The Scottish Longitudinal Study (SLS) has proved successful in linking administrative data with a sample from the 1991, 2001 and soon the 2011 censuses. Having attracted substantial Research Council funding the SLS team is now developing a major intergenerational life course study using civil events data (data on births, marriages and deaths) back to 1855, creating an extremely powerful statistical resource which will put Scotland at the forefront of intergenerational research worldwide. This new dataset will be used in conjunction with existing studies but also as a resource on its own.

Since 2009 The Scottish Health Informatics Programme has developed a vast body of expertise in all aspects of data linkage, from public attitudes through good governance and linkage methodology to researcher training and safe havens. The next phase of SHIP will go further in pursuing cutting-edge research, delivering innovative linkage and analysis of health-related data sets, building research capacity and expertise and providing an interface with industry, practice and policy. This will not only contribute to a Scotland that is Healthier. Because of the SHIP commitment to share expertise, experiences and

opportunities widely, it will contribute to all of Scotland's strategic objectives to be Wealthier and Fairer, Safer and Stronger, Smarter and Greener as well.

Privacy Advisory and Ethics Committee

The Privacy Advisory and Ethics Committee will be made-up of people from a range of organisations and sectors with experience and expertise of legal, technical and public acceptability issues from different organisations. It will include lay-representation and will be informed by information and data on the views of the broader public.

An important element of the committee will be strong channels of communication with related bodies across the UK and, in time, internationally. The Committee will have at least one member who also sits on the UK Advisory Council for Access to Administrative Data.

The primary functions of the committee will be to oversee and guide the direction and operations of the Data Sharing and Linking Centre, and to offer advice to data custodians on cross-sectoral linkage applications. The development of this capacity will be taken forward in close discussion with public bodies across Scotland to avoid overlap, additional bureaucracy and to ensure added-value.

There will be no legal or regulatory requirement for committee advice for any linkage to occur, but we aim in the longer term to bring together organisations and to create a cross sectoral committee that has the mandate from the organisations to make recommendations as to whether or not data linkages are conducted, giving advice that is proportionate to the risks involved, such as privacy and reputational risks, and in light of the public benefits that would accrue if the linkage took place. The prototype will be the N-PAC (National Privacy Advisory Committee) being developed for SHIP, itself an extension of the current Privacy Advisory Committee that serves NRS and NSS.

To deliver a Privacy Advisory Committee we will work across the public sector, in particular with the Information Commissioner's Office, to develop more detailed proposals for the operation and management of the Committee. There is no intention to replace or undermine the role of The Information Commissioner's Office or any other regulatory body.

Information Gateway and Data Linkage Service

In the consultation paper we sought views on the functions for a 'National Data Linkage Centre'. Overall there was a high level of support for the proposed functions, but some concern that a "Centre" may serve as a data warehouse.

In light of these concerns we will not develop a "Centre" but we will develop a Data Linkage Service and associated information gateway, or concierge support service.

The Data Linkage Service will provide the following functions:

- Lead development of data linkage IT and expertise, generating capacity for more and better data linkage for research and statistics across Scotland.
- Develop and maintain methods for read-through between different individual referencing systems, and support the development and maintenance of a 'population spine'.

- Provide a linkage service: conducting approved within and cross-sector data linkages where necessary and efficient, delivering improvements on the existing range of services available to potential users of linked data.
- Provide a trusted data-exchange service.
- Provide support and encourage co-ordination across the network of data linkage facilities and safe havens that already exist, with a focus on collaboration in procurement and use of ICT and sharing of developments, good practice and methods for linkage.
- Provide support and guidance on the development of linkable local and national sources in order to enhance the quality of strategically important data resources being shared and linked for statistical research purposes.

The SHIP infrastructure, which includes an indexing service and a 'state of the art' high performance database server and remote safe haven access, will form a crucial element of the technical capacity.

To increase the understanding of the Data Protection Act and other legislation across all those involved in linkage activities, to encourage dialogue between researchers, data custodians and the public about how and why data are used, and to co-ordinate and harmonise data access and approval processes across sectors, we will develop a Information Gateway in tandem with the Linkage service. Again, the Information Gateway will build on existing capacity in Scotland. Its functions will include:

- Provide secretariat and support to the Privacy Advisory Committee.
- Assist researchers through the process of identifying appropriate data sources, establishing data sharing agreements, providing technical, logistical or methodological suggestions for mitigating privacy risks and increasing public value from for statistical and research purposes.
- Lead work to further explore the ethical, legal and social issues related to data linkage, in particular the most appropriate forms of benefit sharing.
- Lead work to further explore public attitudes and raise awareness amongst the public as to how data are and could be used, the risks and the benefits of that.
- Collate and publish examples of good practice examples in the application of the Guiding Principles.
- Work with public bodies across Scotland to co-ordinate and harmonise data access and approval processes.
- Work with the Information Commissioner's Office to provide training and guidance in relation to the Data Protection Act and good practice in data linkage activity.
- Develop and disseminate training materials to raise levels of understanding across the public sector in the legal frameworks governing data access and sharing for research and statistical purposes.

Steering Group

A Steering Group was established in 2011 to guide the development of the Data Linkage Framework and it will continue to meet biannually to guide delivery, including the strategic direction of the Information Gateway and Data Linkage Service and the Privacy and Ethics Committee.

Chaired by Director General: Governance and Communities (Scottish Government), the Steering Group consists of representatives from a range of sectors and agencies many of whom are also on other groups and committees, helping to ensure that this framework is developed harmoniously with other initiatives and policies across Scotland.

The remit of the group is to provide high-level strategic direction for framework development and delivery. It does not consider operational issues of data linkage in detail. We currently have in place a cross-sectoral multi-agency operational working-level group that agrees co-ordination issues and takes forward delivery of the strategy. As the framework is developed working arrangements will evolve.

Membership of the Steering Group and minutes of all meetings are available at http://www.scotland.gov.uk/Topics/Statistics/datalinkageframework/steeringgroup

This framework strategy sets out a long term vision for data linkage in Scotland which will take many years to realise fully. The approach will be iterative, engaging with stakeholders and the public as plans are developed, and joining-up existing linkage programmes and projects to make better use of existing resources.



© Crown copyright 2012

You may re-use this information (excluding logos and images) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/ or e-mail: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

ISBN: 978-1-78256-215-3 (web only)

The Scottish Government St Andrew's House Edinburgh EH1 3DG

Produced for the Scottish Government by APS Group Scotland DPPAS13570 (11/12)

Published by the Scottish Government, November 2012

www.scotland.gov.uk