Joined-up data for better decisions

Guiding Principles for Data Linkage

Version 1.0

Contents

Guiding Principles for Data Linkage

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 and Scotland will be recognised the world over as a hub of innovative and powerful statistical research, attracting investment and job creation.

The Data Linkage Framework for Statistics and Research aims 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.**

In order to support data custodians, researchers and other stakeholders in taking decisions about safe and effective linkage within this new culture, the foundation stone of the Data Linkage Framework is this set of Guiding Principles for data custodians, researchers, ethics or privacy committees, and others involved in data sharing and linkage in taking decisions about safe and effective linkage. Further work to collate good practice examples of the principles being put into practice will be conducted, and a Privacy Advisory Service will be established (see 'Joined-up data for better decisions: a strategy for improving data access and analysis', published at www.scotland.gov.uk/StrategyforImprovingDataAccessandAnalysis).

The Principles are heavily based on principles developed for The Scottish Health Informatics Programme (SHIP) by Professor Graeme Laurie and Nayha Sethi at the Edinburgh Law School, University of Edinburgh. The SHIP principles were adapted to be applicable to wider statistical and research data linkage activities and consulted on in *A Scotland-wide Data Linkage Framework for Statistics and Research: Consultation Paper on the Aims and Guiding Principles.* In addition, the approach and a summary of the principles were tested with members of the public through a deliberative research project. Details and results of the consultation and research are available through http://www.scotland.gov.uk/Topics/Statistics/datalinkageframework

We are grateful to Professor Graeme Laurie and Nayha Sethi, to all those who responded to the consultation and to the members of the public who gave up their time to participate in the deliberative research project.

FOREWORD

The Information Commissioner's Office (ICO) is pleased to have been involved in the establishment of a Data Linkage Strategy for Research and Statistics in Scotland. As the Assistant Commissioner for Scotland & Northern Ireland, it is my responsibility to engage with stakeholders in Scotland to assist in compliance with their responsibilities under the Data Protection Act 1998.

Data sharing between relevant organisations can result in better communication, better decision-making and more efficient and effective service provision, particularly in the public sector. However, in these days of scarce resources and restricted budgets, it is even more important to develop policy on an evidence-based approach to ensure best value and added benefit where possible. Data Linkage in a safe and secure environment under nationally agreed Guiding Principles will provide opportunities for researchers to acquire robust data for meaningful, evidence-based policy development.

I am pleased therefore, to commend these Guiding Principles to all involved in statistical data linkage activities.

Dr Ken Macdonald
Assistant Commissioner
Scotland & Northern Ireland
October 2012

SCOPE

The Data Linkage Framework that these principles underpin concerns linkages for research and statistical purposes only. 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.

INTRODUCTION

The principles presented in this document are designed to assist data controllers and other decision makers (e.g. ethics committees, privacy committees, data access panels) to adopt a common framework for decision-making and to take a proportionate approach to managing the risks inherent in any data linkage.

The principles are not rules and are not prescriptive. They are principles that we recommend are considered ahead of any data linkage activity and where they can guide deliberations on a given data linkage practice.

The principles are not a statement of legal rules. They flow from, but do not comprehensively restate or summarise:

- Human Rights Legislation
- The Data Protection Act
- Guidance issued by the Information Commissioner:
  - Guide to Data Protection
  - The Data Sharing Code of Practice
  - The Anonymisation Code of Practice
- The Scottish Government Identity Management and Privacy Principles

The added value of the principles lies in their guiding effect for decision-makers who must decide whether to approve data sharing or linkage. They provide a common framework for thinking about the kinds of issues in play and for justifying decisions about linkage or sharing. They operate most effectively when judgment must be exercised about whether linkage or sharing should take place. For example, a linkage might be perfectly lawful but there might still be reasons to ask on what basis it should take place, if at all. The principles assist these deliberative processes.

The principles are intended to promote the public interest in scientifically sound, ethically robust research while appropriately protecting privacy. They do not imply any changes to the legal requirements of data controllers under the Data Protection Act (summarised on pages 17-19) or any sector-specific legislation, and they do not alter the accountability of data controllers to existing regulatory bodies.

The only way to completely avoid risks to privacy from data linkage activity is to avoid data linkage activity. This would result in valuable research not being conducted. All or nothing approaches to risk management can be unhelpful. Rather, the principles presented here are intended to encourage a proportionate approach, whereby actions taken to reduce the risks to privacy are in proportion to those risks, factoring in the potential benefits of the research.

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 presented should help balance these considerations, increase the public benefits from data usage and mitigate risks to privacy. A common framework of reference for decision-making 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. If this were so, they would be rules. 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 provided in coming months.

The Principles

1. Public Interest

Protection of privacy, efficient use of data, and scientifically sound and ethically robust research and statistics, are all in the public interest.

The public interest principles should be considered for all data linkage activity, regardless of the application of other principles.

1.1 The adequate protection of personal privacy should be a central consideration in all deliberations about the sharing and linkage of data.

1.2 The rights of individuals should be respected with adequate and appropriate privacy protection, recognising that data sharing and linkage is never risk-free. Acceptable risks are those that are relative to the benefits for all in the appropriate use of data for research and statistical purposes and this should be recognised.

1.3 The production and dissemination of statistics through data linkage should be in accordance with the Code of Practice for Official Statistics, The Pre-release access to Official Statistics Order (Scotland) 2008 and National Statistician's Guidance on Confidentiality of Official Statistics.

1.4 Benefits arising from linkage of personal data are public goods and should be shared as widely as possible.

1.5 Where linkages resulting in commercial gain are envisaged, this should be clearly and publicly articulated and widely communicated.

2. Governance and Public Transparency

Clear decision-making processes that are open and accountable to the public will help to ensure the appropriate balance of privacy protection, efficient use of data, and scientifically sound and ethically robust research and statistics.

The governance and public transparency principles should be considered for all data linkage activity, regardless of the application of other principles.

2.1 Data sharing and linkage should be carried out under transparent and proportionate controls and security processes, and the purposes and protection mechanisms should be communicated publicly and to oversight bodies/individuals with responsibility for data processing.

2.2 Information about all approved linkages; all privacy impact assessments; all data sharing agreements for linkage purposes and accessible summaries of plans for linked-data analysis should be made publicly available.

2.3 All practices, including all data linkages, shall be appropriately monitored and regulated by a relevant individual, organisation or governance body. It is possible that these activities will be monitored at an individual and organisational level simultaneously.

2.4 There should be a clear distinction in roles between those carrying out linkages, analyses and those policing governance and enforcing sanctions.

2.5 As far as possible, account should be taken of the full range of stakeholder positions in the development and implementation of governance arrangements.

2.6 The interests of one (or a few) stakeholder(s) should not dominate use/linkages or the conditions of the same, especially where this might be at the expense of other stakeholder interests.

3. Privacy

The law does not give absolute value to privacy, and a balance is needed between respect for privacy, through the proportionate mitigation of risk, and the potential benefits to all through the use of data for statistical and research purposes.

Methods for mitigating risks to privacy include anonymisation and security. Where data subjects consent to their personal data being shared or linked, privacy risk must still be considered.

3.1 Data Controllers should demonstrate their commitment to privacy protection through the development and implementation of appropriate and transparent policies and procedures and show how these operate relative to the public interest in promoting safe data sharing and linkage.

3.2 Every reasonable effort should be made to consider and minimise risks of identification (or re-identification) to data subjects and their families arising from all aspects of data handling.

3.3 Serious consideration should be given to carrying out privacy-impact risk assessments, following the most up-to-date ICO guidance. Where a PIA is not considered feasible or necessary, this should be clearly and publicly articulated. PIAs should be made publicly available (excluding sections as necessary for reasons of security), well ahead of linkage occurring so there is opportunity for data subjects to raise concerns.

3.4 Linked datasets should be kept for the minimal time necessary for the original purpose of the linkage to be met. The onus is on those wishing to hold datasets for longer to justify this, e.g. by demonstrating that adequate anonymisation takes the data outside the remit of the data protection regime. If a secondary purpose arises, a new Privacy Impact Assessment should be considered, and data-sharing agreements revised.

3a Consent

Consent of data subjects is an important consideration, although it is not a necessary requirement for data linkage under the Data Protection Act.

The consent principles should be departed from only where there is a strong justification and approval has been granted by an appropriate oversight body.

3a.i Where practicable, consent should be obtained from each data subject prior to the linkage of personal data for statistical and research purposes. Personal data are those from which an individual is identifiable or is likely to be identifiable.

3a.ii Where practicable, individuals or organisations collecting data should adequately inform data subjects of all material issues relating to the storage and use of their data. Material issues are those likely to affect a person in a non-trivial way.

3a.iii The minimum amount of personal data should be used to achieve the stated objective; the reasons and justification for its use should be adequate and clearly explained; and reasonable efforts should be made to inform data subjects of the purposes of the use.

3a.iv Where obtaining consent is not practicable, then (a) removal of direct identifiers should occur as soon as is reasonably practicable and/or (b) approval from an appropriate oversight body should be obtained which can confirm that the public interest in data linkage is met and appropriate safeguards are in place.

3b Anonymisation

There are degrees of data anonymisation and it may not be possible to completely remove the risk of reidentification. Nevertheless, data can be anonymised sufficiently for data controllers to make a reasonable risk-based judgment that data can be shared.

The anonymisation principles may have less importance if consent for linkage of non-anonymised data has been given or if linkage has been approved by an appropriate oversight body.

3b.i Procedures to link data should involve the separation of identifiers (e.g. name, or unique reference number) from the rest of the data, and consideration should be given to separating the indexing, linking and analysis functions and personnel.

3b.ii The linkage method used should be that which requires the minimum necessary identifiable data.

3a.iii The default position should be that data users have access only to data from which names and direct identifiers have been removed, and data users should be subject to an obligation not to attempt to re-identify individual data subjects. Any requirement for researchers to have access to data containing identifiers should be fully justified and risk assessed.

3b.iv Data controllers should determine and agree upon the appropriate extent of anonymisation to be applied to any given dataset or linkage exercise. Particular consideration should be given to indirect identifiers (e.g. individual reference numbers), combinations of data (e.g. gender, date of birth and qualifications) and geo-references (e.g. postcode). The balance to be struck is between the level of risk to privacy relative to the likely benefits from linkage.

3b.v The risk of re-identification of data subjects must be assessed by a body/individual with the relevant expertise to make such judgments, including risks arising from indirect means such as statistical disclosure.

3c Security

Security of data transfer, storage and use is vital for the protection of privacy, especially where there is any risk of reidentification.

3c.i Appropriate and proportionate physical and technical security measures should be applied to ensure the confidentiality, integrity and availability of information and should reflect the assessed risk level of information assets.

3c.ii All personnel involved in data linkage activities should be properly trained on the data security policies and procedures, and should undertake periodic refresher training.

3c.iii The importance of data security should be reflected in the business objectives of all organisations involved in data linkage.

3c.iv Information about data security policies and procedures should be highly visible within organisations conducting indexing or linking or sharing of personal data.

4. Access and Personnel

These principles are important for publicly demonstrating security and respect for privacy and in avoiding any one person or organisation having access to large quantities of personal data.

4.1 Roles and responsibilities of parties with regards to data linkages should be identified from the outset, and all personnel involved in data linkages should be fully aware of their roles and responsibilities.

4.2 Terms and conditions for data sharing should be set out in the form of a data sharing agreement. Where researchers wish to deviate from or modify the terms of the data use/sharing agreement, new terms must be agreed by all parties.

4.3 All data recipients should be appropriately vetted to ensure they have adequate training. Vetting procedures should be robust and transparent and proportionate to the requests made and the sensitivity of the data requested.

4.4 Whether a single data controller or otherwise, a clear distinction should be maintained between each of the functions of linker, indexer, and recipient. Linkers should be responsible only for linking data.

5. Clinical Trials

Data linkage as a method to support or enhance clinical trials presents specific requirements.

5.1 Mechanisms for linkages involving clinical trials must permit re-identification by the principal data source, this is particularly important for pharmacovigilance purposes.

5.2 The specific circumstances and conditions governing whether or not patients involved in clinical trials can be contacted and by whom, should be clearly set in place in transparent policies.

5.3 Researchers should only seek to contact participants directly with respect to information arising from a clinical trial in which they took part where prior consent to be contacted for specific purposes has been obtained. Any dilemmas in this regard should be referred to the ethics committee that approved the original protocol.

6. Sanctions

Where organisations or individuals break the law then legal sanctions apply. Other sanctions should be considered where the Guiding Principles are breached.

6.1 Sanctions for failure to respect terms and conditions should be clearly stipulated in the data use/sharing agreement and should be proportionate to the sensitivity and quantity of data in question.

6.2 Sanctions should relate to both the individual and the organisation, and should be both proportionate and specific in relation to financial penalty; period of time that the organisation will be refused further data access; and reports of improper use of data to senior management and/or funding bodies.

Functions, Roles and Responsibilities of Data Controllers

The following is taken from the SHIP Blueprint Appendix 6 08/12/2011 http://www.scot-ship.ac.uk/sites/default/files/Reports/Appendix_6.pdf.
See also ICO guidance on Identifying 'data controllers' and 'data processors' Data Protection Act 1998.

All UK individuals and organisations must ensure that their use and disclosure of personal data complies with the requirements of the Data Protection Act (DPA).

Identifying the Data Controller

The DPA confers the responsibility and liability for compliance with the requirements of the DPA on the Data Controller. Identifying the Data Controller(s) in relation to a set of personal data and its processing operations is therefore key to ensuring that data protection obligations are known and adhered to. It is sometimes challenging to identify the Data Controller where a number of actors and processing operations are involved.

The opinion of the Article 29 Data Protection Working Party[1] published in 2010[2] recognised the challenge in this area. The Working Party made some unambiguous observations:

- In identifying a Data Controller, identifying who sets the purposes of the processing is the paramount consideration;

1 The Article 29 Working Party was set up under Article 29 of Directive 95/46/EC. It is an independent European advisory body on data protection and privacy.

2 Opinion 1/2010 on the concepts of 'controller' and 'processor', 00264/10/EN, WP 169, adopted 16 February 2010.

- The actors involved must have the legal and factual capacity to fulfil their role, i.e. a Data Controller is not a Data Controller unless in facts and law they have the capacity to set the purposes for the processing of the personal data;
- A pluralistic situation, with a number of Data Controllers, including with different degrees of responsibility and liability, is both possible and acceptable.

Key messages:

- It is essential to be clear as to who is acting as a Data Controller with respect to any given data set which involves the processing of personal data.
- It is possible that one or more parties can act in the capacity as a Data Controller and will accordingly be held jointly liable.
- It is possible to agree between parties who will act as a data Controller with respect to a given dataset and/or to agree difference levels of responsibility and liability.

Data Controllers and Data Processors

The Data Controller is defined as the person or persons who determines the 'purposes for which and the manner in which personal data are to be processed'.

The Data Processor is defined as any person '. . . other than an employee of the Data Controller who processes data on behalf of the Data Controller'.

Data Controllers and Data Processors are typically organisations, authorities or businesses, e.g. the Data Controller of the personal data used across NHS hospitals in the Lothians area is Lothian NHS Board. There are also cases where a Data Controller is an individual, for example General Practitioners are Data Controllers for patient information provided to them.

An important feature of the Data Controller/Data Processor relationship is that the Data Controller retains liability under the DPA for all processing of personal data undertaken by the Data Processor on their behalf. There is a legal requirement that a written contract between the Data Controller and Data Processor governs processing undertaken by a Data Processor on behalf of a Data Controller.

Data Controllers may only disclose personal data in accordance with their Register entry in the Information Commissioner's Register of Data Controllers, and the Data Protection Principles set out in Schedule 1 of the DPA. **Whilst the Data Controller is legally required to ensure that all disclosures of personal data meet these requirements, they do not retain these obligations after the data are disclosed. These obligations essentially flow to their recipient, who then becomes the Data Controller and liable for their use and disclosure in accordance with DPA.**

Key messages:

- Data Controllers retain legal liability with respect to processing of data and the activities of Data Processors who work on their behalf until such time as data are disclosed.
- It is imperative to be clear with respective parties as to the capacity in which they are entering a relationship and also the point at which the responsibilities of Data Controller(s) will pass (if at all).

Glossary

Consent

Freely given and informed agreement by the Data Subject for his or her personal data being processed for a specific purpose.

Data Controller

An individual, organisation or body that determines the purposes for which and the manner in which any personal data are, or are to be, processed.

Data Sharing Agreement

Agreement between Data Controller and data recipient clarifying: the purpose or purposes of the sharing; Who will have access; What will be shared; How the data will be transferred: Quality issues (including accuracy, relevance and usability); Data security; Retention and deletion; Review of effectiveness of sharing.

Data Subject

Individual who is the subject of personal data. A Data Subject may be identifiable, directly or indirectly, through reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, social or cultural identity.

Linker

Individual (or body) who receives datasets from data controllers and links them together using a key created by the indexer.

Indexer

Individual (or body) who receives personal data from one or more Data Controllers and determines which records in each dataset relate to the same individual (or entity). The indexer then creates a unique reference for each individual (or entity) and a corresponding key to allow the data from the different sources to be joined.

Privacy Impact Assessment (PIA)

A process which helps assess privacy risks to individuals in the collection, use and disclosure of information. PIAs help identify privacy risks, foresee problems and bring forward solutions.

Individual Reference/Identifier

Frequently a sequence of characters and/or numbers that is used and/or assigned by an organisation to a person to identify uniquely the person for the purposes of the organisation's systems and operations.

Further Reading

Anonymisation Code of Practice
(ICO forthcoming publication, November 2012)
http://www.ico.gov.uk/for_organisations/data_protection.aspx

Consultation Report: A Scotland-wide Data Linkage Framework for Statistics and Research
(Scottish Government, 31 March 2012)
http://www.scotland.gov.uk/Publications/2012/08/3287/0

Consultation Responses: A Scotland-wide Data Linkage Framework for Statistics and Research
(Scottish Government, 13 June 2012)
http://www.scotland.gov.uk/Publications/2012/07/7705

Data sharing code of practice (ICO, 2011)
http://www.ico.gov.uk/for_organisations/data_protection/topic_guides/~/media/documents/library/Data_Protection/Detailed_specialist_guides/data_sharing_code_of_practice.ashx

Data Sharing: Legal Guidance for the Public Sector
(Scottish Government, 2004)
http://www.scotland.gov.uk/Publications/2004/10/20158/45784

Identifying 'data controllers' and 'data processors' Data Protection Act 1998 (ICO, 2012)
http://www.ico.gov.uk/for_organisations/guidance_index/~/media/documents/library/Data_Protection/Detailed_specialist_guides/data_controllers_and_data_processors.ashx

Identity Management and Privacy Principles
(Scottish Government, 2011)
http://www.scotland.gov.uk/Topics/Government/PublicServiceReform/efficientgovernment/privacyprinciples

Information Governance Of Use Of Health-Related Data In Medical Research In Scotland: Current Practices And Future Scenarios
(G Laurie and N Sethi, SHIP Core Programme No.2, Working Paper No.1, 2011)
http://www.scot-ship.ac.uk/sites/default/files/Reports/Working_Paper_1.pdf

Information Governance of Use of Health-Related Data in Medical Research in Scotland: Towards a Good Governance Framework
(G Laurie and N Sethi, SHIP Core Programme No.2, Working Paper No.2, 2012)
http://www.scot-ship.ac.uk/sites/default/files/Reports/Working_Paper_2.pdf

Government Security Policy Framework (Cabinet Office, 2011)
http://www.cabinetoffice.gov.uk/sites/default/files/resources/hmg-security-policy_0_0.pdf

National Statistician's Guidance: Confidentiality of Official Statistics
(Government Statistical Service, 2009)
http://www.statisticsauthority.gov.uk/national-statistician/ns-reports--reviews-and-guidance/national-statistician-s-guidance/confidentiality-of-official-statistics.pdf

Privacy Impact Assessment Handbook (ICO, June 2009)
http://www.ico.gov.uk/upload/documents/pia_handbook_html_v2/index.html

Public acceptability of cross-sectoral data linkage: Deliberative research findings (31 August 2012)
(Ipsos MORI with Sarah Cunningham-Burley and Claudia Pagliari on behalf of Scottish Government, 31 August 2012)
http://www.scotland.gov.uk/Publications/2012/08/9455/0

Scotland-wide Data Linkage Framework for Statistics and Research: Consultation Paper on the Aims and Guiding Principles
(Scottish Government, 26 March 2012)
http://www.scotland.gov.uk/Publications/2012/03/3260/0

SHIP Guiding Principles and Best Practices
(Scottish Health Informatics Programme, 2010)
http://www.scot-ship.ac.uk/sites/default/files/Reports/Guiding_Principles_and_Best_Practices_221010.pdf

Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office.
(Cabinet Office 2011)
http://www.cabinetoffice.gov.uk/sites/default/files/resources/transparency-and-privacy-review-annex-b.pdf

The Scottish Government
Riaghaltas na h-Alba

ISBN: 978-1-78256-204-7

This document is also available on the Scottish Government website:
www.scotland.gov.uk

APS Group Scotland
DPPAS13559 (10/12)