

A Scotland-wide Data Linkage Framework for Statistics and Research: Consultation Paper on the Aims and Guiding Principles

Consultation Analysis

Contents

Executive Summary.....	3
Introduction	5
Consultation Question 1: Benefits of data linkage	7
Consultation Question 2: Challenges or barriers	10
Consultation Question 3: Guiding Principles	16
Consultation Question 4: Privacy Advisory Service	22
Consultation Question 5: National Data Linkage Centre	26
Next Steps	29
Annex A: Organisations and individuals invited to respond	30
Annex B: Respondents	32

Acknowledgements

The Office of the Chief Statistician and Performance would like to thank all those who participated in the consultation.

Executive Summary

The Scottish Government is working with a wide range of partners to establish a collaborative framework that will facilitate data linkages for research and statistical purposes to be conducted safely, securely, legally, ethically and efficiently. From 26 March to 15 June 2012 a written consultation exercise was conducted to seek views on the aims of the framework and a draft set of guiding principles. This report provides an analysis of responses received.

Overall Key Messages

- The majority of responses welcomed the framework and were very positive about the benefits that data linkage can bring to Scotland.
- Responses served to emphasise the challenges identified in the consultation paper with a significant minority feeling that privacy and legal barriers had not been properly considered. Levels of concern were highest in relation to consent and the potential for commercial gain.
- The draft guiding principles were favourably received. Some respondents suggested that the principles should allow more leeway for judgement and others suggested they needed to be presented as a firm set of rules.
- There was general support for the plans for a Privacy Advisory Service and a Data Linkage Centre. However, there was also uncertainty and concern over how the two would fit in with existing organisations and existing data linking arrangements.
- There were a number of suggestions on how both the overall approach could be improved and specific elements could be delivered most usefully.

Key messages on the benefits of data linkage

Respondents were very positive about the benefits of data linkage. A number of respondents used this question to emphasise their support for data linkage in general and their support for the benefits a strategic approach might bring.

A number of respondents highlighted specific areas of research which could be investigated to significant benefit using data linkage.

Key messages on the challenges or barriers to data linkage

The bulk of responses to this consultation question were detailed expansions or variations on the challenges identified in the consultation paper, i.e.: uncertainty about the legalities and public acceptability of data sharing and linkage; incomplete data, or data that cannot be linked; limited capacity for secure exchange and access to data and; limited capacity of public sector organisations to analyse and make use of linked data.

Some respondents felt that privacy and legal barriers had not been properly considered, and the lack of the necessary knowledge amongst staff who would be involved in data linkage was also repeatedly cited as a challenge.

Key messages on the guiding principles

Responses regarding the principles were broadly positive, nonetheless there were a number of detailed suggestions for improvements addressing a variety of the principles. A number of the respondents sought further information or clarity on the terminology used.

There were a large number of comments relating to the principles which covered consent with some differing views on the desirability of opt-in consent. There was also uncertainty and curiosity regarding the principle which referred to commercial gain.

Key messages on the objectives for a Privacy Advisory Service

The majority of respondents were broadly supportive of the Privacy Advisory Service (PAS) with over half of those who responded indicating that the objectives were the right ones. In particular the PAS was felt to be useful for situations where data custodians are unsure whether they can legally and appropriately make data available for linkages.

The general positive tone of responses was tempered with concern as to how the PAS would fit in with existing bodies and the proposed Data Linkage Centre, and it was suggested that there was scope for existing organisations to fulfil the functions of the PAS.

Other responses emphasised the primacy of data controllers in decision making and concerns were expressed over any plans for the PAS to make enforceable decisions.

It was emphasised in responses to the questions for both the PAS and the NDLC that any new arrangements should not merely present additional bureaucratic hurdles to data linkage and should not interfere with data linkage already being successfully conducted.

Key messages on the functions to be led by the National Data Linkage Centre

The weight of responses regarding the National Data Linkage Centre (NDLC) were supportive and many respondents suggested additional functions that an NDLC should take on, such as quality assuring the data and linkage processes and coordinating the maintenance of research-enabled national data ready for linkage.

To various degrees, respondents urged that the NDLC build on existing structures and best practice and several respondents commented on the importance of coordination, harmonisation and integration between all the bodies involved. It was highlighted by more than one respondent that there may be disadvantages to centralising data linkage functions through the establishment of the NDLC.

Introduction

The Scottish Government is working with a wide range of partners to establish a collaborative framework that will facilitate data linkages for research and statistical purposes to be conducted safely, securely, legally, ethically and efficiently.

The main purpose of this consultation exercise was to seek views on the aims of the Data Linkage Framework and a draft set of guiding principles.

This written consultation is just one part of a broader and on-going dialogue with a wide range of organisations and individuals. The findings outlined in the following sections are specific to this particular consultation exercise and do not necessarily reflect the weight or range of views within the population or sub populations as a whole. The respondents have not been representatively or purposively sampled and the majority of those who responded have a specialism or professional interest in the subjects covered by the consultation.

This report should be considered alongside the deliberative research conducted into the public acceptability of data linkage.
www.scotland.gov.uk/deliberativeresearchdatalinkage

Once a set of principles has been agreed, the details of the other parts of the framework (a National Data Linkage Centre and an Analytical Privacy Advisory Service) will be developed. Comments to help develop plans for both of these were also welcomed at this stage, but there will be further opportunity to comment in 2013. More information will be made available through the [Data Linkage Framework pages](#) on the Scottish Government website

Consultation process

The consultation was published on the Scottish Government website on the 26th March 2012 with a deadline for responses of the 15th June 2012.

[A Scotland-wide Data Linkage Framework for Statistics and Research: Consultation Paper on the Aims and Guiding Principles](#)

A number of organisations/individuals were invited to respond (Annex A). Where the Scottish Government received permission from the respondent, responses have been published on [the Scottish Government website](#).

The consultation responses have been analysed internally by the Scottish Government and this report represents a review of the responses rather than the Scottish Government's reaction to the responses.

Responses

In total there were 61 responses to the consultation, with 45 responses from organisations and 16 from individuals. The following table shows the breakdown of responses to the introductory question on the nature of respondents.

Are you responding <i>primarily</i> as a data custodian, data user or data subject? (We recognise all people are data subjects and many organisations act as data guardians and data users, but please tick only one box)	
Response selected	Number of respondents
Data Custodian	11
Data User (e.g. researcher)	30
Data Subject (e.g. member of the public or group representing citizens)	5
All boxes selected	2
Data Custodian and Data User selected	2
Data User and Data Subject selected	1
No answer given	10
Total	61

The table shows that almost half of respondents were responding primarily as data users. Five of the respondents chose to select multiple categories and ten respondents did not select any of the options. A list of those who responded is presented in Annex B.

Responses to yes/no questions

Care should be taken in interpreting the counts of “yes” and “no” responses to the consultation questions. Only responses where one of the boxes was actually ticked or an explicit yes/no was provided are included in the count.

In a number of cases respondents selected a yes/no option which was contradicted by the content of their written response. In these instances no attempt has been made to infer or alter the yes/no response and the responses have been recorded as entered by the respondent.

Consultation Question 1: Benefits of data linkage

The consultation paper set out a range of benefits that data linkage can bring and asked: *Are there any benefits of data linkage for statistical and research purposes that are not sufficiently described here?*

The table below shows that around half of those who provided a response to the yes/no question suggested that there were further benefits to those described in the consultation paper.

Are there any benefits of data linkage for statistical and research purposes that are not sufficiently described here?			
Type of respondent	Yes, there are further benefits	No, the benefits are described fully	No answer
Data custodian	4	7	0
Data user	18	11	1
Data subject	1	4	0
Multiple categories selected	3	1	3
No selection	0	1	7
Total count	26	24	11

General comments

In general, respondents were very positive about the benefits of data linkage. A number of respondents used this question to emphasise their support for data linkage in general and their support for the benefits a strategic approach might bring.

One such respondent noted that:

“Once the linkage infrastructure is in place, it will enable the creation of complex linked data and allow new and innovative research that can be used to enhance the delivery of public services. The benefits of such a programme would significantly improve research capabilities in Scotland and provide a robust platform to inform policy.” (Centre for Data Linkage (Curtin University))

It was noted by some, however, that the benefits were described too specifically to capture all the various advantages of data linkage. Even the phrase ‘data linkage for research and statistical purposes’ used in the consultation question was considered by some too narrow.

In the analysis of the comments, the type of respondent (data user, subject or custodian) was examined. The different types of respondents were found to

hold similar views on the benefits of data linkage, the main points of which are outlined in the sections below.

Additional benefits

It was suggested by a number of respondents that informing strategic and spending decision-making was a valid use of data linkage. As such, they argued that enabling activities such as tracking and predicting need for services, planning health and care provision, and monitoring the performance of public services was a key benefit of data linkage.

Data Linkage was also argued by more than one respondent to have benefits for local level decision making as it enables cost-effective data retrieval, analysis and comparison for lower geographical levels, and provides the potential to improve small area data. Additionally, linked data was argued to allow researchers to assess the representativeness of surveys of the populations from which they are drawn as well as potentially allowing for identifying patients eligible to take part in approved studies or trials.

There was a view expressed that the data linkage framework would lead to improved research skills and increased analytical competence with data linkage presenting a cost effective research tool. It was similarly argued that data linkage has the potential to drive innovation and investment by all actors and to improve understanding of available data.

One respondent suggested that there could be a benefit in terms of income for data gatherers through a charging mechanism such that carefully anonymised data can be shared with industry researchers. In a related point it was suggested that if secure access to linked datasets for commercial organisations was enabled this could enable further applications of the findings of linked datasets to be applied in an operational context without compromising privacy.

It was argued that the consultation document should have been more explicit regarding the benefits on outcomes or service usage relating to protected characteristics in the Equality Act. This was seen as an important means of assessing the extent to which people with protected characteristics are disadvantaged or experience poorer outcomes than others.

Specific research topics

A number of respondents highlighted specific areas of research using data linkage which could be of significant benefit. These were:

- the effects of social inequality and the impact of any policy changes designed to reduce inequalities
- long term or rare outcomes or side effects
- the accumulation of small exposures that may accumulate over a long period of time
- public health surveillance
- public health research
- health economics research
- pharmacovigilance

- health service evaluation
- health service audit
- the management of long-term conditions
- prognosis and prediction of major illness
- evaluation of health-care interventions
- epidemiological work to protect the health of the population
- evaluation of the responsiveness of e-health solutions
- areas which would benefit from more sophisticated analyses such as prescriptions, co-morbidity and serious adverse events.

It was also highlighted that new possibilities are likely to arise as work progresses and that the benefits realised would depend on the system of data linkage pursued.

Presentation of benefits

Additionally, there were a number of comments relating to how the benefits were described and presented in the consultation paper.

One respondent suggested it would be useful to rank the stated benefits in terms of their value to Scotland with another suggesting that each benefit should include details of how that benefit could be obtained without compromising individuals' control of their personal data.

Within the consultation paper examples of effective data linkage were presented alongside the mooted benefits and it was noted by respondents that it would have been helpful if the examples had included statements on how the solutions had benefited the population. It was also felt that the underlying principles for protecting personal data were not addressed sufficiently in each of the examples outlined.

It was further noted that the examples provided did not extend to areas such as counter-fraud or policing and would benefit from examples from beyond the health sector. Additionally, examples could have been included from Scandinavian countries where aggregate reports from linked information are readily available.

Consultation Question 2: Challenges or barriers

The consultation paper set out a range of challenges to data linkage and asked: Are there challenges or barriers preventing more effective and efficient data linkages for statistical and research purposes taking place that are not sufficiently described here?

The table below shows that around two thirds of those that responded to the yes/no question felt that there were further challenges to those described in the consultation paper.

Are there challenges or barriers preventing more effective and efficient data linkages for statistical and research purposes taking place that are not sufficiently described here?			
Type of respondent	Yes, there are further challenges	No, the challenges have been identified	No answer
Data custodian	7	4	0
Data user	20	9	1
Data subject	3	2	0
Multiple categories selected	3	0	4
No selection	1	0	7
Total count	34	15	12

General comments

One respondent took issue with the framing of this section of the consultation paper, feeling that challenges were presented as barriers to overcome rather than constraints to be respected.

There was a view expressed that having a centralised process as outlined in the consultation paper could itself act as a barrier to the development of local linkage work. In particular, this could affect the development of multi-agency linked data on service demand which is often required within relatively short timescales.

Additionally, respondents were keen to establish the impact of a limited number of data custodians agreeing to participate, raising the question of whether this would affect cost efficiency and the ability to produce meaningful results. It was pointed out that if just one data custodian is unwilling to release data for linkage then the linkage fails. A possible solution would be to create mechanisms not solely dependent on trust in particular people or bodies.

Comments relating to specific challenges

The bulk of responses to this consultation question were detailed expansions or variations of the challenges identified in the consultation paper. Comments are gathered below under the descriptions used in the consultation paper.

Challenge 1: Uncertainty about the legalities and public acceptability of data sharing and linkage

Legal/statutory issues

A number of respondents highlighted particular challenges around legal and statutory issues. More than one respondent made clear that, in particular, any data linkage must comply with the Data Protection Act.

It was noted that government departments or datasets often have distinct legal constraints which limit what is permissible for the data. One organisation (SCRA) noted concern over their “ability to share information except where an express or implied statutory power can be identified”. Such a power would need to be identified to enable participation in the framework.

One respondent felt that the consultation document was ‘cavalier’ when addressing the serious legal challenges of data sharing and linkage and did not pay sufficient heed to the legal barriers in place.

It was reported that uncertainty regarding what is legally permissible has led to various bodies and individuals operating “inconsistently and over-cautiously”. However, it was also suggested that some of the inconsistency in the level of willingness to share data may be down to cultural issues for different data holders or through a sense of ‘territoriality’ stemming from the level of resources already invested in developing approaches to data linkage.

Public Acceptability

With regards public acceptability it was felt that the current low levels of public engagement and disproportionate negative media coverage presented a challenge. A number of respondents identified that an important step would be to explain the purposes of data linkage to the public in order to both raise awareness and address the public’s concerns. To this end, one respondent proposed that high quality research outputs could be used to raise awareness of the importance of data collection and the quality of data collected.

Additional views on the public acceptability of data linkage were provided by one respondent who cited research from the Child Medical Records for Safer Medicines (CHIMES). The research indicated that children/young people and parents/guardians have a limited knowledge of how routinely collected healthcare data is currently used but assume that the NHS use health data to improve and safeguard population health. Consent and assent were seen as important in enabling support of data. The research also suggested that concerns increase as the number of linked data sources increases and when there are commercial interests involved.

Privacy and Public Interest

A number of respondents raised points relating to the issue of privacy, with one concerned over the suggestion of a "balance of interests" between personal privacy and public interest. They argued that it would be improper to seek to dismiss valid privacy concerns by appealing to a wider public interest.

It was suggested that the framework could be clearer or more forceful with regards anonymisation post linkage and outline what steps are taken to ensure that individuals cannot be identified from any outputs. The Information Commissioner's Office flagged up their forthcoming Anonymisation Code of Practice which would be relevant to this.

More than one respondent highlighted the role of data custodian as being a particularly challenging one with regards the management of privacy. It was felt that data custodians might require support in order to effectively balance the privacy of the individual citizen with the wider public interest. Several comments highlighted the confusion that exists about how to identify who a data controller is (particularly once data have been shared) and their responsibilities.

Challenge 2: Incomplete data, or data that cannot be linked

Data Consistency

Several respondents identified challenges around data consistency.

Inconsistency of data definition. i.e. the use of one term with several meanings or different terms with the same meanings could present challenges. Equally, inconsistency in the recording of time and place between different datasets was noted as potentially problematic if the datasets are to be linked.

Additionally, changes in coding systems between one period and another could result in discontinuity within the data to be linked, making comparisons over time problematic. Ensuring this comparability over time was argued to require specific attention as it could be compromised as data collections are 'improved'.

Unique Identifiers

One respondent suggested that once indices (for example the Community Health Index and National Insurance Number) are reconciled linking data is reasonably straightforward.

Another respondent argued that a comprehensive data linkage network requires the existence of a common identifier and mappings to operational unique identifiers within data sources. They felt that by placing the mappings with the data sources themselves and designing the mapping operation such that the indexer does not know for sure the subject population of a data source some of the public acceptability challenges could be circumvented.

References to unique identifiers led one respondent to caution that some demographic groups could withdraw cooperation from data gathering systems. It was suggested that some groups might feel threatened by the possibility of a single identifier being created and held centrally - potentially enabling a 'data profile' for each person in Scotland.

Linking Practicalities

A number of respondents commented on various challenges around the practicalities of data linkage. It was argued that there needs to be both robust assumptions on data linkages where information common to multiple individuals is linked, and robust systems for removing duplicates.

Mismatching was highlighted as a potential challenge, in particular the impact this would have on any inferences. This was predicted to become more problematic as more data sets are linked.

On a related point one respondent was keen to establish what the impact of the 'inevitably less than perfect' data linkage methods would be on the ability to produce accurate counts. This was a particular concern given that it is suggested that data linkage will provide Census type information and potentially replace the Census entirely. It was argued that there would be a continued requirement to invest in alternative methods to capture data on 'undocumented and socially excluded populations.'

One of the less-populated local authorities noted that their statistical output often contains numbers which are too small to be published without risk of identifying individuals. National tables therefore often contain gaps for their data, making them of limited use. They felt that it was not clear whether this situation would be improved by data linkage, or whether the small numbers issue would still apply.

Challenge 3: Limited capacity for secure exchange and access to data

One respondent highlighted the challenges stemming from handling what may be large volumes of data, and the software needed for matching and for visualisation of said data. However, they also argued that close collaboration between the statistical and data mining communities may offer some solutions to these challenges.

More than one respondent acknowledged data security as a priority and it was highlighted that any increased concentration of data stemming from data linkage would increase the risk of mishandling or loss of sensitive personal data.

However, one respondent cautioned against sacrificing research functionality as a result of security being overly restrictive. A specific example was given around this where the application of rules can censor access to low but informative counts, even when the potential for deductive disclosure is remote and specific public health importance may attach to low counts.

It was noted that safe havens are not limited to stand-alone computers as suggested in the consultation paper. It would be possible to have access restricted to secure access points with no capacity of extraction of data.

Challenge 4: Limited capacity of public sector organisations to analyse and make use of linked data

A number of respondents reinforced the challenge identified in the consultation document regarding a lack of relevant expertise and knowledge amongst individuals and organisations to be involved in data linkage.

Firstly; respondents described a lack of knowledge regarding the datasets available and their characteristics. It was felt that even before practical challenges were considered it would be necessary to ensure that organisations and individuals know enough about what data could potentially be made available via linkage and how it could be used to answer research questions.

Secondly; there was concern that there may not currently be sufficient skills in data linkage methods distributed amongst relevant organisations and thirdly; there was a perceived lack of the skills required to analyse linked information.

Finally; it was argued that there is currently a lack of awareness of individuals' or organisations' responsibilities as data custodians or users. This can result in data being linked inappropriately or not being shared due to unwarranted risk aversion.

As one respondent noted:

“Speaking from experience of linking data across a number of organisations - there tends to be a general lack of awareness amongst researchers of their individual roles and responsibilities as data custodians/users if linked data are to be made publicly available” (Clare Baker)

With these potential shortcomings in mind, one respondent argued that it would be better to focus on how analysis can be undertaken by people who do have sufficient training by involving universities or government specialists rather than encouraging analysis of sensitive datasets by people who do not have the requisite skills and knowledge.

As well as staff issues, further resourcing issues were identified. It was argued that organisations will need improved systems for good quality, up to date, and safe data and it was suggested that the various financial and training costs borne by these organisations should be acknowledged.

As one respondent noted:

“Local Authorities may find it difficult to justify the cost of focusing on data systems when facing financial constraints. And yet it is the Local Authority data that will [be] needed for many research studies.” (Brigid Daniel)

This point was particularly relevant for smaller organisations and local authorities. Depending on the means of data linkage, the time, cost, and resources required could act as a barrier to data linkage and effectively act to exclude smaller authorities from the process.

One respondent felt that any resource gap could be removed or narrowed through effective partnership working with the private sector.

Commercial activity

More than one respondent raised the issue of commercial involvement in data linkage. One respondent argued that

“More consideration of the role of private enterprises and global corporations in the use of the Scottish population’s data would be worthwhile.” (Centre for Population Health Studies, Edinburgh University)

Another suggested that

“There needs to be a mechanism that allows a route to access to data for commercial researchers” (ABPI Scotland)

Differences between respondent categories

In the analysis of the comments relating to challenges, the type of respondent (data user, subject or custodian) was examined. The different types of respondents were found to hold broadly similar views.

However, two of the more substantial comments came from data subjects. These were: that concern over the creation of data profiles could lead some demographic groups to withdraw cooperation from data gathering systems; and that the consultation paper’s balance, as well as its ‘cavalier’ treatment of legal issues, were worrying.

Additionally, respondents who identified themselves as data users were most likely to highlight challenges around the role of data custodians and the inconsistency in organisations releasing data.

Consultation Question 3: Guiding Principles

The consultation paper presented a set of draft guiding principles for data linkage activity and asked: Are the guiding principles sufficient and appropriate?

The table below shows that the majority of data custodian respondents stated that they thought the guiding principles were sufficient and appropriate, whereas data subjects were more likely to say they were not.

Are the guiding principles sufficient and appropriate?			
Type of respondent	Yes	No	No answer
Data custodian	8	3	0
Data user	14	15	1
Data subject	1	4	0
Multiple categories selected	1	2	5
No selection	0	1	6
Total count	24	25	12

General Comments

In commenting on the principles respondents provided both general comments on the principles overall, and comments on individual principles. The most common comment in relation to the principles was to suggest that additional information and, in particular, definitions of specific terms are required. Some of the terms for which definitions were sought included: *every effort*; *proportionate*; *sound and robust*; *data controller*; *appropriate oversight body*.

There were a number of comments from respondents which were broadly supportive of the principles but went on to make suggestions for their improvement. One respondent stated :

“A general comment is that the principles are all reasonable, however in places they read more like a set of detailed requirements rather than a set of guiding principles. Ideally the principles should leave more latitude for proportionality rather than be too detailed and prescriptive.” (The Medical Research Council)

The counter-view was also expressed, with one respondent proposing a number of changes which sought to strengthen the wording with the effect of making a number of the principles into requirements.

Another respondent stated that there was overlap between the principles and there would be scope for reducing them in number. In contrast, some of the respondents suggested that additional principles could be added to those presented in the consultation document. These additional principles related to the following areas:

- Commitment to encourage the use of a single identifier on all databases.
- Management of the risks of indirect as well as direct identification when data are disseminated as there is a potential increase in identifiability when datasets are combined.
- Specification of the process to agree the nature of valid uses that can be made of the linked datasets and the approval mechanism to be applied to applications using the datasets, as well as any control mechanisms to be applied to such use.
- Removal of personal identifiers as soon as they are no longer required - where they need to be kept they should be kept separate from the integrated dataset.
- Assertion that the type of matching used should be the minimum needed and range of attributes used to establish a common identity should be the minimum necessary for the linking operation to succeed.
- Development of minimum standards for secure management of information.
- Management of data access (and risk of privacy and/or confidentiality breaches) for research projects ensuring: Confidentiality, data protection, information security, record management, data access agreements, international information security standard.
- Acknowledgement of the valuable contribution that research populations could make to research design. They can offer significant insights to research teams and help more effective dissemination.

A summary of the main comments on specific principles are provided below under the headings given to the principles in the consultation document. In the analysis of the comments on the principles, the type of respondent (data user, subject or custodian) was examined, however, in the majority of cases there was no relationship between the type of respondent and the principle commented on. Indeed, many of the comments were made by respondents that did not select any of the type of respondent options.

Public Interest

Respondents made comments on the balance between individuals' rights to privacy protection and the public benefits from linking data. One respondent stated that principles 2 and 3 must explicitly state that the public interest does not override the individuals' right to privacy and the right to withhold consent. A further respondent requested clarification on the mechanism that would be used to achieve the balance.

There were a number of comments in relation to principle 5 (Where linkages resulting in commercial gain are envisaged, this should be clearly and publicly articulated and widely communicated). These included requests for clarification and further debate on the reference to "commercial gain".

It was also suggested by two respondents that principles for private sector data linkage might be different to those for public sector linkage.

Finally, one of the respondents suggested that findings from research should be widely publicly disseminated in a way which is accessible to the widest audience possible.

Governance and Transparency

Two respondents enquired how information about linkages would be made available to the public (principle 8). One further respondent suggested that, rather than making complete Privacy Impact Assessments and data sharing agreements available to the public, a subset of key information could be made available as the full documents are study specific and can be complicated.

In relation to monitoring and regulating practices (principle 9) one respondent suggested that there might be a risk of conflicting requirements, delays and duplications as a result of multiple overseers. There was also a query about how the costs of the governance body monitoring data linkages will be met. Another respondent suggested that there should be appropriate public representation for any organisation or governance body to ensure that the general public and patients have confidence in the use of their data for research.

Privacy

For principle 14, one respondent stated that it is important that project planners have access to guidance on what measures are required to minimise risks of identification. One respondent expressed concern that applying all measures to achieve data privacy could be an overreaction and would make the data harder to use. In contrast, another respondent stated that the wording needs to be stronger to protect public privacy and that data linkage

must never be allowed where there is a possibility of identification or re-identification.

A number of respondents stated that more information should be provided in relation to Privacy Impact Assessments (PIA). Specific comments and queries on PIAs included:

- PIAs should be a requirement – rather than only suggesting that “serious consideration” being given to their completion (principle 15).
- The robustness of PIAs should be part of the evaluation process.
- Where are the PIAs submitted to? Would it be the PAS?
- Would the ICO’s version of the PIA be used?
- What are the implications and procedures if a PIA is not carried out?
- Limited resources will have an impact on the consideration of completing a PIA – it could become common practice to not carry out a PIA if there is no sanction for not doing so.

One particular respondent stated:

“This, or relevant supporting information, should contain a clear statement about the extent and nature of Privacy Impact Assessments (PIAs) required; e.g. as part of a current cross sectoral data linkage work with which NSS is involved, a PIA of over 100 pages has been produced. By contrast, this is not the sort of PIA that has previously been required of academic researchers prior to granting access to de-identified linked data.” (NHS National Services Scotland)

Four respondents commented on principle 16 (Linked datasets should be kept for the minimal time necessary...) as they suggested that there might be requirements to hold the data for longer for research purposes.

Removal of names and direct identifiers

Several respondents raised issues associated with indirectly identifying variables in relation to principles 17 to 19. It was suggested that the principles should include an explicit reference to indirect identifiers as the removal of direct identifiers alone is not sufficient to guarantee anonymity. Furthermore, one respondent stated that the more datasets are linked the easier it is to re-identify anonymous data subjects. One respondent suggested that a direct identifier definition would help as de-identifying a dataset is complex.

Two respondents sought clarification on the role of “Data Controllers”. For the assessment of the risk of re-identification, two respondents requested information on what would constitute a “suitable body”.

Consent

The highest number of comments on the principles were in relation to the consent principles, with 19 of the 61 respondents providing comments specifically relating to principles 20 to 23 in the consultation document. All types of respondents (data users, custodians and subjects) provided comments in relation to the consent principles.

Several respondents stated that additional guidance and definition of terms would be beneficial for the principles on consent. In particular, respondents requested further information on what and how it could be deemed to be “practicable” in relation to explicit consent and what would constitute an “appropriate oversight body” (principle 23). It was suggested by one respondent that an example could be included to demonstrate the application of the principles on consent.

Opposing views were expressed regarding informed consent. For example, one respondent argued:

“Informed consent (opt-in, not opt-out) must be at the heart of any good privacy-respecting system.” (No2ID)

With another observing:

“‘Opt-in’ processes on a study by study basis would result in very low uptake. This usually invalidates and removes the reason for the linkage. Thus explicit opt-in consent should only be sought when there is an overwhelming reason to seek this in the interests of privacy.”(Anonymous)

There were also comments on the practicalities associated with explicit consent in terms of how questions seeking explicit consent are worded and how individuals are informed about uses of data. Two of the respondents suggested that the principles in the consent section might conflict with the Data Protection Act in relation to use of data without consent. A number of respondents stated that it would not be possible to obtain explicit consent for the data they already hold.

Security

One respondent requested further information on the role of the National Data Linkage Centre in relation to principles 24 to 28, in terms of access, data management, retention, standards for storage and transfer, and information on how a security breach would be managed.

Access and Personnel

One respondent expressed concern at the statement in principle 33 that linkers should be separate from data custodians as there are currently organisations which perform both functions, namely ISD and education services. Conversely, one respondent stated that principle 33 is ineffectual as “a clear distinction” is open to any convenient interpretation and it is suggested that where data from multiple controllers is linked the functions must be physically, technically, financially and organisationally separate. There was also a request for more information in relation to “robust governance mechanisms” referenced in principle 32.

Clinical Trials

In commenting on the principles associated with clinical trials two respondents suggested that principles 34 to 36 should not just be applicable for health data and clinical trials as the need for re-contact might arise in other scenarios. There were also requests from respondents for clarity on who has responsibility in relation to re-contact (principle 35). It was suggested by one respondent, in relation to principle 36, that improved data linkage is likely to increase the need for re-contact of individuals who have participated in a clinical trial.

Consultation Question 4: Privacy Advisory Service

The consultation paper set out a draft set of objectives for a privacy advisory service and asked: Are the objectives set out for a Privacy Advisory Service in Section 3c the right ones?

The table below shows that over two thirds of those that responded to the yes/no question felt that the objectives for the Privacy Advisory Service as set out in the consultation paper were the right ones. Of the five respondents who identified themselves as data subjects, three felt that the objectives were not the right ones.

Are the objectives set out for a Privacy Advisory Service in Section 3c the right ones?			
Type of respondent	Yes, the objectives are right	No, they are not	No answer
Data custodian	7	4	0
Data user	23	5	2
Data subject	2	3	0
Multiple categories selected	2	1	4
No selection	1	0	7
Total count	35	13	13

General comments

The majority of respondents were broadly supportive of the Privacy Advisory Service (PAS) as set out in the consultation paper. One such respondent remarked that

“We strongly support the concept of a ‘privacy advisory service’ that would provide a recognised source of expert advice to researchers, and offer assistance to them in managing associated risks.” (Wellcome Trust)

In particular the PAS was felt to be useful for situations where data custodians are unsure whether they can legally and appropriately make data available for linkages.

However, one respondent felt it unlikely that PAS would address requirements of the law in accessing data and that such a service would only be valid if it was set up by legal instrument.

In the analysis of the comments on the PAS, the type of respondent (data user, subject or custodian) was examined. The different types of respondents

were found to hold similar views on the PAS, the main points of which are outlined in the sections below.

Relationship with other bodies

A number of respondents commented on the relationships between the PAS and other bodies. In particular, organisations that performed related functions to the PAS. One respondent stated:

“It is not clear how this body would fit in with existing organisations that offer such facilities.” (Medicines Monitoring Unit, University of Dundee)

The relationship between the National Data Linkage Centre (NDLC) and the PAS was also questioned. An overlap in roles was noted and it was suggested that the separation of technical and ethical skills could lead to a loss of knowledge on both sides.

The question was raised as to whether the PAS and the NDLC should in fact be separate. As one respondent noted:

“We are supportive of the objectives for a Privacy Advisory Service but see no reason to separate the it from the National Data Linkage Centre (NDLC). Issues of privacy should sit at the centre of the Framework's operation and rather than hiving it off to a separate body, the NDLC seems the best place for it to be sited” (SCRA)

Some respondents went further and questioned the need for a PAS at all, particularly in the current financial climate.

“We are unclear why this consultation is not asking us whether we need to create a Privacy Advisory Service in the first place, rather than about its objectives. There are already data protection officers in every relevant organisation, the NHS has Caldicott Guardians and all organisations in Scotland have recourse to the Information Commissioner for Scotland. Rather than create a new body it would make more sense for the Information Commissioner's Office to take on this task.” (Orkney Islands Council)

The above respondent was not alone in questioning the need for such a service whilst pointing out perceived duplication of roles with either the Information Commissioner's Office or the Scottish Information Commissioner.

The Information Commissioner's Office itself was positive in their response:

“The ICO particularly welcomes the proposed establishment of a Privacy Advisory Service (PAS) and believes that this will be of fundamental importance in realising the benefits of privacy assurance and control mechanisms for the research community.”

An alternative to the PAS was outlined whereby data controllers are educated, supported and empowered to work within existing data-protection arrangements, rather than creating a body that will make their decisions for

them. This was argued to provide data controllers with “freedom to innovate and invest in directions they think are viable.”

Whilst not going as far as the above scenario other respondents did emphasise the primacy of Data Custodians with regards decisions for their data. The view was expressed that the PAS should be there to help, advise and make recommendations rather than to make decisions.

On this point, but taking a different view, one respondent suggested that there could be greater value from the PAS if data controllers authorised the service to make decisions on their behalf.

Responsibility and liability

The question was raised as to where the responsibility for decisions to link data would lie in circumstances where the Privacy Advisory Service had offered advice.

As one respondent noted:

“how best to avoid organisations abdicating accountability and responsibility for...decisions they make as a result of receiving advice and support from the Service merits careful attention, as does the question of liability” (Jane Dargie)

Additional Bureaucracy

A number of respondents made the point that the Privacy Advisory Service should not lead to an increase in regulatory burden and that:

“There needs to be an emphasis on making things easier and quicker for researchers and evaluators” (Anonymous)

However one respondent noted that in their experience:

“making a procedure ethical actually does have a procedural cost. Ethical procedures are not to be sidelined as soon as they become costly.” (Scottish Council on Human Bioethics)

Resourcing and staffing issues

More than one respondent identified that a variety of skills would be needed from members of Privacy Access Service. Their remit would require them to advise on improved methodology for linkage and improved analyses of linked datasets whilst also assessing proposals realistically for possible public effect or benefit.

One respondent noted:

“the very nature of cross-sector linkages and the objectives outlined in the consultation would require an extensive advisory network to cover all areas of expertise with a very broad remit to fulfil this service” (HEADLINES, University of Aberdeen)

Additional Functions

Respondents suggested the following further functions for the PAS:

- Overseeing statistical disclosure
- Facilitating and encouraging the release of data where data custodians may be reluctant
- Extolling the benefits of data linkage to the wider public and allaying fears regarding the safety of the use of linked data for research

Questions

Additionally, there were a number of more specific questions which respondents were keen to clarify:

- How is the Analytical Privacy Advisory Service to be established and maintained?
- Who would be eligible to join the PAS?
- Could the service be used for local linkage projects?
- Would there be any formal requirement for researchers to gain approval of this body and if so, what impact would this have on the regulatory body?
- Whether the centre will cover only those data linkage projects for 'research and statistical purposes' as set out in Section 1 of the consultation paper or whether this will have a wider remit and advisory capacity?
- In practice how would organisations set up data linkage agreements? Would requests for data come directly to organisations or would the Privacy Advisory Service work as an intermediary or would the National Data Linkage Service provide this role?
- Would ISD dictate the release of NHS data for linkage purposes?
- How would the PAS interact with members of the public? Would it take on functions of the ICO or complement them? Would it seek representations from the public or civic society on particular issues?

Consultation Question 5: National Data Linkage Centre

The consultation paper set out a range of functions for a National Data Linkage Centre (NDLC) to provide and asked: Are the functions that will be led by the National Data Linkage Centre set out in section 3d the right ones?

Around two thirds of those that provided a response to the yes/no question felt that the functions of the National Data Linkage Centre as described in the consultation paper were the right ones.

Are the functions that will be led by the National Data Linkage Centre set out in section 3d the right ones?			
Type of respondent	Yes, they are the right functions	No, they are not	No answer
Data custodian	7	1	3
Data user	20	8	2
Data subject	2	2	1
Multiple categories selected	2	2	8
No selection	0	0	3
Total count	31	13	17

General comments

The weight of responses regarding the NDLC were broadly supportive. One such response noted:

“The benefits of having a centre to capture the collective knowledge, skills and memory and to deliver efficiency savings but at the same time ensuring research excellence, is an exciting prospect.”(HEADLINES, University of Aberdeen)

However, a number of respondents suggested that whilst broadly reasonable the functions were expressed too generally. It was felt that a comprehensive and detailed list would allow an idea of how the centre would operate in practice and allow for a more informed response.

In the analysis of the comments on the NDLC, the type of respondent (data user, subject or custodian) was examined. The different types of respondents were found to hold similar views on the NDLC, the main points of which are outlined in the sections below.

Proposed additional functions for the NDLC

A number of functions additional to those in the consultation paper were suggested for the NDLC:

- promoting a better understanding of the constituent data sets
- quality assuring data and the linkage process and establishing and monitoring standards
- performing a screening role concerning the validity of requests for data linkage
- performing an oversight role concerning the uses of the linked data
- coordinating the maintenance of research-enabled national data ready for linkage – as separate from the technical function of linkage
- providing expert commentary on the linked datasets
- facilitating and encouraging the release of data for linkage purposes
- providing information and communication technology support
- acquiring and maintaining data linkage research equipment
- leading the standardisation of components of datasets that facilitate linkage
- providing data linkage and analytical workforce training
- providing a “clearing house” role as a single portal for researchers to approach (this would be a positive differentiator from other jurisdictions when competing for research investment).

Staffing and resource issues

A number of respondents emphasised that members of NDLC staff would require specific technical skills with one respondent noting that the appropriate level of knowledge would have to be acquired rapidly. It was suggested that the NDLC would require sufficient resources to draw on the best available Scottish talent, some of which may be located in private data mining companies.

As well as ensuring adequate staffing, more than one respondent suggested that designing and building a sufficiently high performance infrastructure with comprehensive security features would be a challenge.

One respondent suggested that thought be given to how processes and procedures can be modernised, given that the model on which the NDLC is based is over 15 years old another called for ‘innovative and forward thinking approaches’.

Additional bureaucracy

As with the Privacy Advisory Service a number of respondents were concerned that the NDLC might present further bureaucratic hurdles and lead to a lengthy process.

In particular, a number of respondents made clear that whilst they could see the merit of the NDLC, they did not want it to impact adversely on pre-existing, successful data linkage:

“this centre with these functions may help to facilitate future linkage. My concern is that where linkage is currently happening appropriately, securely and successfully...any requirement to go through this centre may result in long delays and may actually hinder future linkage.” (Anonymous)

Relationship with PAS and other bodies

Related to this, there were a significant number of responses discussing the relationship with the NDLC and other bodies

“It is not clear how this Centre would interact with existing bodies that currently deal with such data” (Medicines Monitoring Unit, University of Dundee)

In this regard, respondents urged that the NDLC to build on existing structures and best practice and several respondents commented on the importance of coordination, harmonisation and integration between all the bodies involved.

“We strongly urge that the National Data Linkage Centre is taken forward as an inclusive, collaborative network involving the regional safe havens and other experienced data management teams to build on the national strengths.”(HEADLINES, University of Aberdeen)

Respondents suggested variously that relevant stakeholder groups, the research community, and the Scottish Longitudinal Studies Centre should be involved with the NDLC.

As was the case for the PAS, the need for a NDLC at all was questioned.

“we question whether enough thought has been given to the need to create a new body when there may be existing organisations which could provide this service.”(Orkney Islands Council)

The relationship between the NDLC and PAS was again raised. Some responses suggested combining the two bodies. However, the counter-argument was also made:

“An important point to recognise is that privacy advisory service must be wholly independent and publicly recognised as such, from the data linkage centre”(Glasgow Housing Association)

Centralisation

A number of concerns were expressed relating to the centralisation implied by the formation of the NDLC and it was highlighted that the stated functions vested considerable power in one entity and this presented moral hazard and created an entity that is ‘too important to fail.’

Equally it was argued that the NDLC must be implemented in such a way that it does not become a weak point in privacy protection. As one respondent noted:

“If one organisation ends up with access to a large collection of data sets, much of the protection that results from ensuring that data are not centralised would be lost. The design and implementation of such a centre requires great care.” (No2ID)

Additionally, there was a caution against increased centralisation on the grounds of 'efficiency':

"It would be much more efficient if the police and legal system were one body but there are good reasons to keep them separate and these reasons apply equally to data linkage systems." (Violet Warwick)

Questions

Respondents also raised specific questions relating to the functions of the NDLC

- Would the NDLC hold linked datasets, if so for how long? Would it be an archive in any sense?
- What is meant by "satellite linkage units"?
- Would ISD be the source of all health data? Would Health Boards be consulted about access to datasets collected in their areas for their residents?
- What would the role of the NDLC (and PAS) be would in relation to data sharing by private organisations? Would the NDLC restrict the activities of commercial organisations regarding data linking within Scotland?

Next Steps

The Scottish Government and the Data Linkage Steering Group will now consider this report in detail alongside that from the public deliberative events, and issues raised at meetings, workshops and conferences on this subject over recent months, including the UK wide "Administrative Data Task Force" which is considering data linkage issues¹. Taken together, the range of issues, suggestions, concerns and ideas will be used as the basis for:

- revisions to the principles, which we aim to publish over the coming months
- a strategy for the development and implementation of the framework
- further work to scope and plan a National Data Linkage Centre (likely to be renamed the Data Sharing and Linking Service) and a Privacy Advisory service for consultation in 2013.

All relevant materials will be made available through the Scottish Government website at <http://www.scotland.gov.uk/Topics/Statistics/datalinkageframework>

¹<http://www.esrc.ac.uk/funding-and-guidance/collaboration/collaborative-initiatives/Administrative-Data-Taskforce.aspx>.

Annex A: Organisations and individuals invited to respond

Age Scotland
Association of Chief Police Officers Scotland
Association of Directors of Social Work
BCS: The Chartered Institute for IT
Big Brother Watch
Black and Ethnic Minority Infrastructure Scotland (BEMIS)
BMA
Brigid Daniel
British Security Industry Association
Caldicott Guardians
Capability Scotland
Care Inspectorate
Chi Advisory Committee
Children First
Community Planning Partnerships
Consumer Focus Scotland
Convention of Scottish Local Authorities
Council of Ethnic Minority Voluntary Sector Organisations
Department for Work and Pensions
Economic and Social Research Council
Engender
Equalities and Human Rights Commission
Equality Network
Experian
Glasgow Women's Library
Graeme Laurie
HM Revenues and Customs
Iain Atherton
Improvement Service
Inclusion Scotland
Independent Living in Scotland
Information Commissioners Office
Janice McGhee
LGBT Youth Scotland
Liberty (The National Council for Civil Liberties)
Local Authority Chief Executives
Local Authority Research and Information Association
Lothian & Borders DSP
Medical Research Council
MRC|CSO Social & Public Health Sciences Unit
National Institute for Social Care and Health Research
NHS Chief Executives
No2ID
Northern Ireland Statistics and Research Agency
Office for National Statistics

Privacy International
Public Health Directors
Registers of Scotland
Royal College of GPs
Royal College of Nursing
Royal College of Nursing
Royal Statistical Society
Sarah Cunningham-Burley
Sarah Lowe
Scottish Centre for Social Research
Scottish Children's Reporters Administration
Scottish Collaboration for Public Health Research and Policy
Scottish Council of Voluntary Organisations
Scottish Disability Equality Forum
Scottish Human Rights Commission
Scottish Inter Faith Council
Scottish Longitudinal Studies Centre
Scottish Members of European Parliament
Scottish Privacy Forum
Scottish Refugee Council
Scottish Transgender Alliance
Scottish Women's Convention
SHIP Management Board and International Advisory Board
Steve Platts
Stonewall Scotland
Susan McVie
Wellcome Trust
Welsh Assembly Government

Annex B: Respondents

Individuals

Anonymous x5
Clare Baker
Sheila Bird
Allison Craig
Brigid Daniel
Jane Dargie
Mark McGilchrist
Professor Emeritus Charles Raab
Colin Simpson
Cameron Stark
Violet Warwick

Organisations

ABPI Scotland
Argyll and Bute Council
Australian Institute of Health and Welfare
BEMIS
BMA Scotland
Centre for Data Linkage, Curtin University, Australia
Dr Alison McCallum on behalf of CHI Advisory Group
Dumfries and Galloway Community Planning Partnership
East Ayrshire Council
East Dunbartonshire Council
Evaluation Support Scotland (ESS)
Experian
Falkirk Council
Glasgow City Council
Glasgow Housing Association
Glaxosmithkline research and development ltd
Greater Glasgow and Clyde NHS Board
Health and Data Linkage in North East Scotland (HEADLINES), University of Aberdeen
The Highland Council
Information Commissioner's Office
The Market Research Society
The Medical Research Council
Medicines Monitoring Unit (MEMO), University of Dundee
MRC/CSO Social and Public Health Sciences Unit
NHS Ayrshire & Arran
NHS Tayside and other
NHS 24
NHS Education for Scotland
NHS Health Scotland
NHS Lothian x 2
NHS NSS
NO2ID Scotland

Orkney Islands Council
Royal Statistical Society
ScotCen Social Research
Scottish Borders Council
Scottish Children's Reporter Administration
Scottish Consortium for Learning Disability
Scottish Council on Human Bioethics
SCOTTISH SCIENCE ADVISORY COUNCIL (SSAC)
SCOTTISH WATER
Scottish Ambulance Service
sportscotland
The University of Edinburgh, Centre for Population Health Sciences
Wellcome Trust



© 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: 978-1-78256-044-9 (web only)

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

Produced for the Scottish Government by APS Group Scotland
DPPAS13320 (08/12)

Published by the Scottish Government, August 2012

w w w . s c o t l a n d . g o v . u k