Response to the Scottish Government consultation:

"A Scotland-wide Data Linkage Framework for Statistics and Research".

Prepared by Dr. Geraint Bevan

scotland@no2id.net

June 15, 2012

Contents

1 Data linkage

NO2ID Scotland welcomes the opportunity to comment on the Scottish Government's proposals for a data linkage framework. We note that the main focus of the proposal is linkage of statistical data rather than linking records about individuals for case work. Overall, the principles laid out in the third chapter are generally good, but a worrying tone is set in the preceding chapter on current challenges. The data linkage framework offers potential to make a positive contribution to statistical research in Scotland while respecting individuals' rights to exert control over how their personal data are shared. It is important that the system is designed to enhance privacy rather than circumvent existing protections.

1.1 Scottish Health Survey.

The example of data linkage for the Scottish Health Survey on page 3 of the consultation document exemplifies much good practice for research involving personal data, specifically:

- requiring informed consent from individuals whose data are used;
- ensuring that data are transmitted and stored securely;
- stripping identifying information from records;
- ensuring that data are communicated between named handlers;
- ensuring that it is not a simple matter for any individual or organisation to have access to survey data, personal records and personal identifiers at the same time; and
- obtaining ethical approval from an appropriate expert body before undertaking the research.

That study is a good exemplar for the general approach that should be adopted when linking sets of personal data.

1.2 Official statistics available to all.

In contrast, it is not clear that the example given in "Benefit 3" on page 6 fully adheres to the same principles. For a start, the data relate to children

and it is not clear whether informed consent is (or can be) obtained to allow educational and social work records to be linked in this way. More detail about the mechanisms involved would be required before assessing whether this case exemplifies good or poor practice. Continuing with this example, it is noted that Skills Development Scotland is involved in the study. It has recently come to the attention of NO2ID that this organisation may have a poor approach to ensuring that informed consent is obtained from individuals for the personal data that it handles and that SDS insists on more personal data than should be necessary for the services it provides. We shall be investigating the allegations further, but suggest that the consultation paper would benefit from more detailed and explicit statements about privacy protection in this example and the others.

1.3 Consultation question 1.

Overall, the benefits of data linkage are adequately described, but the underlying principles for protecting personal data are not addressed sufficiently in each of the cases outlined. Each suggested benefit should include details of how those benefits can be obtained in practice without compromising individuals' control of their personal data. Without showing how the benefits can be achieved without violating privacy, they do not carry the weight that they otherwise might.

2 Current challenges.

The initial paragraph in the section on current challenges that impede effective and efficient data linkage is laudable; but the remainder of the section has a worrying tone, suggesting that these are barriers to be overcome in the future rather than constraints to be respected for good reason.

2.1 Balance of interests.

The suggestion of a "balance of interests" between personal privacy and benefits to the community is worrying and unwelcome. It is certainly reasonable to consider how benefits to the wider community may be obtained, but this must always be done with respect for the right to personal privacy and respect for informed consent; not in a manner that attempts to balance

away important rights. If data controllers are overly cautious in their approach because they misunderstand privacy principles or fail to recognise how privacy can be protected despite linkage, then it is entirely proper to address those issues. On the other hand it would be improper to seek to dismiss such concerns by appeal to a wider public interest.

2.2 Data that cannot be linked.

The section on different "unique identifiers" is also worrying, given its inclusion in the context of a section on *current* challenges. The recently published guidelines on Identity Management and Privacy Principles make very clear that persistent identifiers should not be shared. Whatever data linkage principles are finally developed must respect this constraint as it forms an essential element in affording individuals a degree of control over how and when their personal information can be shared.

2.3 Secure exchange.

The section on secure exchange of data is appropriate, but more information could usefully be provided for readers. For example, insecure exchange of data appears to often result from inadequate knowledge on the part of people using the data, such as a widely held but inaccurate belief that password protection provides any security whatsoever for unencrypted data. It is possible that some readers of this document will misunderstand the phrase "secure e-mail transfer" and not understand that it implies the necessity of encryption and good key management.

2.4 Analytical expertise.

The section on limited capacity discusses lack of resources for providing analytical training to staff. It would perhaps be better to focus on how analysis can be undertaken by people who do have sufficient training and oversight by ethical bodies, by involving universities or government specialists rather than encouraging analysis of sensitive datasets by people who do not have the requisite background to understand or mitigate the inherent risks of data linkage.

2.5 Consultation question 2.

The underlying premise of consultation question 2 is arguably poor. Rather than focusing on challenges and barriers, the focus should be on what mechanisms need to be put in place to ensure that good practice is adhered to. With a specification for the principles to which a good system must adhere, it would then be possible to analyse whether the proposed framework could meet those criteria.

2.6 Why a data linkage framework.

The section on why a data linkage framework is needed is generally good, but much depends on the interpretation of the word "inappropriately" in the first sentence. It would perhaps be better to rephrase the sentence along the lines of "... without impinging on personal privacy of data subjects except with their fully informed consent".

The concluding paragraph of the section discusses how the evidence base can be advanced, which is a good goal, but it should also focus on how privacy can be enhanced by adoption of good practice by researchers using an appropriate data linkage framework. It is essential that respect for personal privacy and consent is put at the very heart of any framework if it is to achieve its aims; privacy must not be seen as a barrier to be circumvented. Actively designing the framework to enhance privacy would pay dividends in terms of the research that it ultimately facilitates.

3 Principles.

Although supposedly flowing from appropriate sources, the principles described in the third chapter do not appear to embody respect for personal privacy to the extent that would be expected. It appears in places as if personal privacy can be dismissed when inconvenient. This may be a drafting issue rather than a fault in the intention; in which case more explicit statements might improve the section considerably.

At the start of the chapter it is suggested that resources allocated to applying the principles should be in proportion to the risk involved, but no guidance is given regarding assessment of risk, the factors that should be considered or what degree of consultation may be required, and with whom,

to ensure that the risks are not underestimated by researchers with a vested interest in proceeding with the research. Presumably the proposed Privacy Advisory Service would be involved here but it is difficult to comment on the reasonableness of this text as it stands.

3.1 Public interest.

The section on public interest must explicitly state that the public interest does not override individuals' right to personal privacy and the right to withhold consent. As written, item 2 appears to suggest that individuals' rights to privacy protection might be undermined by consideration of the research benefits, which is presumably (hopefully) not the intention. Similarly, the use of the phrase "necessary grounds" in item 3 might be interpreted to mean that privacy considerations might be overridden, which again is hopefully not the intent.

3.2 Governance and Public Transparency.

The section of governance principles is generally very good. The one problem is item 12 (page 16) which states that one or a few stakeholders should not dominate conditions at the expense of other stakeholders. If these "one or a few stakeholders" are the data subjects, then their rights should dominate. If this item is intended to describe how research should be set up to resolve conflicts between different users of the research, then that needs to be more explicit to make it clear that individuals' rights to control their personal data are not usurped.

3.3 Privacy.

The section on privacy is generally good, but could be enhanced with more detailed guidance. Element 14 suggests how every effort should be made to consider and minimise risk of identification; it is important that project planners have access to guidance on what measures are required. Without careful analysis, it is easy to believe that data are more securely anonymised than is actually the case.

3.4 Removal of names and direct identifiers.

Removal of names and identifiers is good practice, but it must be recognised that this alone is insufficient to ensure anonymity. A good example of the weakness of this as a privacy measure is the release by AOL [1] of supposedly anonymous search histories, from which many users of their service were identified quickly by the queries that they had submitted. The more data are linked, the easier it is to re-identify supposedly anonymous data subjects. This should be addressed explicitly in the principles by considering how the amount of data linked may affect the security of the data. There may be particular risks when supposedly anonymised records from a large collection of data sets are combined.

3.5 Consent.

Informed consent (opt-in, not opt-out) must be at the heart of any good privacy-respecting system. The principles appear to recognise this, but item 22 on pages 16 and 17 in the section on consent makes use of the words "practicable" and "reasonable efforts". It should be made clear that these are to be interpreted from the data subjects' perspective, not just the perspective of the researcher.

Consent must be obtained for specific purposes so that the individuals concerned can understand the implications of granting consent. It is important that recording of consent is managed to ensure that limitations on consent are respected and not violated by subsequent linkage of data. Furthermore, it must also be possible for individuals to withdraw consent for their personal data to be used. This is true for all individuals, who may change their mind at a later date, but is particularly important in the case of children where consent may be granted on their behalf by parents or guardians, but which the children may subsequently wish to revoke as they mature.

3.6 Clinical trials.

The need to allow re-identification in clinical trials is a reasonable one, assuming that the participants have been alerted to the possibility in

[1] Security Focus, "AOL search data identified individuals" 09 August 2006 http://www.securityfocus.com/brief/277

advance. There may be cases, however, where it is not appropriate for the researchers themselves to be made privy to re-identified data; instead, other agencies such as GPs may require that knowledge. For example, biomedical researchers have made it known privately to NO2ID that whilst they may support recruitment of people (including themselves) to clinical studies, they would be horrified for their colleagues to obtain access to their medical records; yet they would not feel comfortable expressing such concerns to managers in their organisations who are keen for research to be undertaken.

3.7 Sanctions.

The section on sanctions is appropriate, but it is also important to consider how data subjects might interact. What opportunities are there for them to express concerns and seek redress if they are uncomfortable with how their data are processed? For example, if it were deemed not to be practicable to contact data subjects and an extension to data linkage were authorised in accordance with item 23, but subjects subsequently became aware and concerned about the use to which their data was being put, should they seek to influence matters in the first instance through the ICO, through the Privacy Advisory Service or through the courts?

3.8 Consultation question 3.

The guiding principles are generally appropriate, but could be improved as described above.

4 Governance.

The composition of the data linkage steering group is generally appropriate, but consideration should be given to the inclusion of some lay members to provide a different perspective; and academics who are experts in computer security or digital forensics, able to identify weaknesses in proposed security measures.

5 Privacy Advisory Service.

The proposed privacy advisory service would appear to be a very welcome development. However, the notion of it helping to "strike the right balance between safeguarding individuals' right to privacy and the efficient use of data ..." should be rephrased: "promoting efficient use of data ... while always respecting individuals' right to privacy".

Consideration should also be given to how, or if, the Privacy Advisory Service would interact with members of the public. Would it take on some of the existing functions of the ICO or complement them? Would it seek representations from the public or civic society on particular issues? Could potential data subjects discuss their concerns or get reassurance from the service? If the service is to make recommendations on which data linkages are to be permitted, there must also be a transparent method for civic society to influence those considerations.

5.1 Consultation question 4.

Yes, NO2ID do wish to be consulted on firmer proposals for a Privacy Advisory Service.

6 National data linkage centre.

The creation of a national data linkage data centre to ensure secure data exchange is potentially very positive, but it must be implemented in such a way that it does not become a weak point in privacy protection. 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.

6.1 Consultation question 5.

Yes, NO2ID do wish to be consulted on firmer proposals for a national data linkage centre.