

CONSULTATION QUESTIONS

Are you responding *primarily* 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)

Data Custodian

Data User (e.g.researcher)

Data Subject (e.g. member of the public or group representing citizens)

1. Are there any benefits of data linkage for statistical and research purposes that are not sufficiently described here?

Yes, there are further benefits No, the benefits are described fully

If you ticked 'yes', please describe the further benefits of data linkage for statistical and research purposes.

While Benefits 2 and 3 as described in the Consultation Paper are welcome and supported, there is no mention of the potential benefit following from data being available for a variety of geographical levels. One advantage of individual level data is that, provided the records retain some geographic referencing, the data can be aggregated to a variety of geographical areas. From a local authority point of view, small area data is vital since this allows us to further re-aggregate the data to larger areas of interest such as council wards or school catchments – data often currently built up from datazones. Care needs to be taken in aggregating data so that no disclosive data is produced and so that no individual's records can be identified by differencing, but the potential for useful geo-demographic or geo-epidemiological data for service planning and research to inform policy development and policy effectiveness is great.

There is no mention of the possibility of linking data at a household level, for example by using housing data from the One Scotland Gazetteer to link to individual records. A population spine is mentioned on page 22, but not the possibility of a housing/household spine. This seems to us to be an essential element of the ability mentioned in Benefit 2 in the consultation paper of producing demographic and census-type statistics. The One Scotland Gazetteer already exists and has the potential to be used for housing data which is currently sadly lacking (see for example the difficulties in generating a suitable housing domain for the Scottish Index of Multiple Deprivation). It could also be used to create a household database by linking to individual level data.

We would also like to reinforce the examples given in Benefit 1 where we have examples from criminal justice where we are being hard pressed to demonstrate that the expensive "inputs" have the desired effect in terms of improved well-being or the reduction of harm to the community. Data linkage is essential to provide the sorts of evidence required.

2. 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?

Yes, there are further challenges No, the challenges have been identified

If you ticked 'yes', please describe the challenges or barriers.

There is no mention in the document of data standards which are fundamental to data linking. One of the difficulties in actually doing data linking are the different standards which exist – perhaps most notoriously in addressing standards.

Once indices such as the CHI, NI number, Scottish Candidate Number etc are reconciled, linking data is reasonably straightforward. Getting these indices matched and reconciled is the difficult bit, and this can be made much easier if data standards are applied.

The Open Scotland Information Age Framework (OSIAF) (circa 2006) was an attempt at establishing good practice in this area, but its success was limited. The National Address Gazetteer and BS7666 address data standard have however, been a greater success in this respect, and like the BS7666 street gazetteer before it, now allow cross-referencing and linkage of address based data across a variety of organisations. To some extent this can be a tool for the national data linkage framework proposed, but a much more powerful tool would be wider application of a people data standard (i.e. BS8766) in public sector administrative systems, particularly those without standard indexes such as NI number.

The challenges under bullet point 4 relating to the limited capacity of public sector organisations to analyse and make use of linked data could be expanded. Mention is made of methodological know-how to conduct linkages. This is an issue which perhaps needs more emphasis. To this should be added the limitations of time to carry out such linkages (which is exacerbated by the lack of expertise meaning that linkages can take longer than would be possible if expertise were available).

Also, the benefits of data-linking can take time to materialise. There are always many calls on staff time and freeing up staff to undertake fairly time consuming data linking work can be difficult if the benefits are not immediately obvious – or may not flow directly to the organisation which holds the data.

One of the challenges of data linking is that 20% of the linking takes 80% of the time. Mention could be made of difficulties in dealing with the most difficult to match cases. Perhaps the standards could include advice on at what stage the laws of diminishing returns can be applied to cease matching while not damaging the data – since it is likely that certain types of cases will be more difficult to match which could result in bias in the data if data linking is stopped at too early a stage.

One barrier which is not mentioned is a general reluctance to release data (often dressed up as legal or data protection concerns) – “the data is mine and you’re not getting it” – or concerns about the time which would be involved in preparing data or doing the actual linkage. It is often not clear to the data holders what the advantages of data-linkage are as the beneficiaries are often not the data holders themselves but analysts, statisticians and researchers – and ultimately the general public in better services. This barrier needs to be overcome if more extensive data linking is to be achieved and some means of encouraging and facilitating data custodians to release data needs to be developed.

In a climate where working in partnership with other organisations is essential, we find that some partners are more forthcoming than others about the process of data sharing. We would hope that a clear set of principles and sanctions and the support of an advisory service would clear up doubts around privacy and confidentiality so that any remaining barriers to effective research and forward planning can be removed.

3. Are the guiding principles sufficient and appropriate? Please explain your answer fully and make suggestions for improvement.

Yes, they are sufficient and appropriate No, they are not

Please explain your answer fully and make suggestions for improvement.

Very thorough and comprehensive. However, we have a number of queries about the detail of some of the principles:

- Data Protection principle 2 states that 'personal data shall be obtained only for one or more specified and lawful purpose, and shall not be further processed in any manner incompatible with that purpose or these purposes' – is there any potential conflict with this principle if data subjects had not been made aware of the purpose of 'data use for statistics and research' at the time of providing the data (p16 of the consultation document refers to explicit consent being obtained from data subjects 'where possible and practicable' – is this adequate? P17 refers to authorisation from 'an appropriate oversight body' where this possible and practicable rule cannot be met – this needs clarification on who/how/ etc)*
- Related to that, there is a lack of clarity over who acts as 'Data Controller' (or Data Controllers) where data is shared and linked for the purpose of statistics or research – presumably this would be formalised via information sharing protocols (ISP) (or memorandums of understanding), but it is not clear where accountability for ensuring the robustness of these ISPs sits (p15 refers to 'a relevant individual, organisation, or governance body').*
- Also on ISPs and accountability, it is not clear where responsibility lies in terms of the 'provider' of a data set ensuring that the security arrangements of the 'recipient' of that data set are robust – i.e. where does responsibility for, and ownership of, a shared or linked data set start and end? And who is responsible for making a judgement as to the legality of use of that data set once 'ownership' has passed to a third party?*
- The consultation document talks about limited capacity for the secure exchange and access to data – has the emergence of cloud technology and of data being held in third party (or internal) cloud data centres been given the appropriate consideration as a potential opportunity/risk?*
- Our Social Work Services have expressed concern about data privacy and state that Social Work records require the same level of protection as Health records.*
- The following specific comments are offered on this section:
Public Interest section, note 5 refers to potential commercial gain from data linkage, but there are moral and political implications of allowing access to extensive public sector datasets to those whose interests are not only commercial, but who may also be motivated for other reasons, such as religious, or political organisations and whose motives may be either explicit or merely implicit. Even the apparently straight forward commercial opportunities have economic and therefore also political implications that need to be considered, subjected to public debate and addressed explicitly within the proposal in order to safeguard the public interest.*

These matters must be addressed prior to any data access linkages being provided, as there is the potential for this public data - even when it is anonymised – to be misused in ways that would not necessarily be in the public interest. For example if such data were to be provided to private companies such as insurance companies, would we be happy for this to be used in their financial and risk calculations? Or to commercial advertisers to market products that many people would see as socially undesirable, or even harmful? For example adverts could be targeted to specific populations, such as young people in specific localities in potentially undesirable ways.

Privacy

Note 14

The wording in this section is too weak as needs to be worded in a robust way to protect public privacy. Data Linkage must never be allowed where there would be the possibility of identification or re-identification using other publicly available datasets, such as Electoral Rolls, or Council tax data to enable identification of individuals. Data Linkage

must never endanger public privacy, so it is suggested that the paragraph should be worded:

"Data linkage will not be allowed if the dataset, or any subset of it, could allow identification of data subjects arising from any aspect of data handling."

Note 15:

In our view, Privacy Impact Assessments (PIA) should be mandatory prior to any and all data linkage applications being approved and the robustness of the PIA should be part of the evaluation process carried out by the agency holding the data prior to granting access to public data.

Consent

Note 20

The reference to 'where possible and practicable' is too weak to protect individual privacy and we do not think it would meet our Data Protection Act responsibilities. Personal data which might allow identification of an individual data subject should never be released via data linkage, as we would be in breach of the Data Protection Act. So, if the data has been anonymised, this paragraph seems unnecessary.

Note 21

Our Data Protection Act registration records the potential use of our service user data for statistical or research purposes. However, this use would always be used on anonymised records; never records which could allow any data subject to be identified.

Note 22.

It is not clear from this paragraph to whom the reasons and justifications for its use should be 'adequately and clearly' explained. Is it to the data controller, or the data subject? Removal of identifiers must be a pre-requisite to the release of personal, confidential social work data prior to data linkage in order for us to comply with the Data Protection Act.

Clinical Trials

Note 34

Great care is required here with data linkage relating to clinical trials and this paragraph is not sufficiently clear or explicit about the circumstances or the arrangements required to avoid data being re-identified by other parties potentially by linking one set of data with other public datasets.

In a climate where working in partnership with other organisations is essential, we find that some partners are more forthcoming than others about the process of data sharing. We would hope that a clear set of principles and sanctions and the support of an advisory service would clear up doubts around privacy and confidentiality so that any remaining barriers to effective research and forward planning can be removed.

4a. Are the objectives set out for a Privacy Advisory Service in Section 3c the right ones?

Yes, the objectives are right

No, they are not

Please explain your answer fully and make suggestions for improvement.

We do not disagree with the objectives as set out, but would like to add to them.

We believe that one of the barriers to data linking is the general reluctance of organisations to release data – for which they often cite reasons such as data protection (see comments under item 2 above). Therefore, another objective of the proposed Privacy Advisory Service should be to facilitate and encourage the release of data for data linkage purposes in conjunction with the National Data Linkage Centre.

While legislation would probably be required in order to make consulting the Privacy Advisory Service (PAS) mandatory before data could be linked, there should probably be some form of obligation for those wishing to link data to consult with the on data linkage projects – particularly larger and higher risk projects. Assuming that consulting the PAS would be voluntary (as a mandatory scheme would require legislation which might be difficult to enforce), it would be useful if the PAS could provide an accreditation scheme so that approved projects which met the standards set out in the guiding principles could show that they met the required standards.

It would be unfortunate if this new body proved ineffective because potential customers did not use it because they were not obliged to do so.

There would need to be publicity about the existence of the PAS in order that those doing data linkage would know to consult them and data users would know about the accreditation scheme and be assured that the data they were using met the required standards.

In a climate where working in partnership with other organisations is essential, we find that some partners are more forthcoming than others about the process of data sharing. We would hope that a clear set of principles and sanctions and the support of an advisory service would clear up doubts around privacy and confidentiality so that any remaining barriers to effective research and forward planning can be removed.

We would like to suggest that a charge could be applied for access to data by non-public sector organisations. This would allow some offset of costs of providing both the Privacy Advisory Service and the Data Linkage Centre.

4b. Do you wish to be consulted on firmer proposals for a Privacy Advisory service as and when they are developed?

Yes No

5a. Are the functions that will be led by the National Data Linkage Centre set out in section 3d the right ones?

Yes, they are the right functions

No, they are not

Please explain your answer fully and make suggestions for improvement.

As in our comments on 4a above, we do not disagree with the objectives as set out, but would like to add to them.

We believe that one of the barriers to data linking is the general reluctance of organisations to release data – for which they often cite reasons such as data protection (see comments under item 2 above). Therefore, another objective of the proposed National Data Linkage Centre should be to facilitate and encourage the release of data for data linkage purposes in conjunction with the Privacy Advisory Service.

We would like to suggest that a charge could be applied for access to data by non-public sector organisations. This would allow some offset of costs of providing both the Privacy Advisory Service and the Data Linkage Centre.

5b. Do you wish to be consulted on firmer proposals for a National Data Linkage Centre as and when they are developed?

Yes No

A Scotland –Wide Data Linkage Framework for Statistics and Research

Response from Falkirk Council

In addition to the comments included on the response form, Falkirk Council wishes to make the following additional comments.

Falkirk Council welcomes this consultation document and the possibilities which it opens up for the better use of linked data on individuals for statistics and research. We believe that this is a way forward to provide better statistics and research. This is a timely and worthwhile document and we look forward to having access to better data and research produced as a result and to swift progress being made to implement the proposals in the document.

We are generally comfortable with the proposals as outlined. We are already involved in data linking for statistical purposes in our Education and Social Work Services, for example to provide statistics on the educational outcomes for Looked After Children.

The consultation response form invites respondents to select their role as either a data custodian or a data user and we have selected Data User. We could potentially make a great deal of use of statistical data for our service planning, performance management and as an input to statutory documents such as our Single Outcome Agreement, Community Plan, Local Development Plan and Housing Strategy.

However, we are clear that as a local authority we are also custodians of a great deal of personal, and not so personal, information about our citizens including information about our school pupils, social work clients, housing tenants, council tax payers, housing and council tax benefit claimants as well as more ad hoc information on service users acquired as part of our statutory duties and service provision.