

## 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)

**I feel my response reflects all of the above perspectives, so I have not ticked any.**

### Glossary

PAS	Privacy Advisory Service
PIA	Privacy Impact Assessment
NDLC	National Data Linkage Centre
SHIP	Scottish Health Informatics Programme
MOA	Memorandum of understanding
DC	Data controller

### **Consultation question 1**

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

**Answer:** The benefits identified in the document are:

- speed up cycles of improvement
- permit replacement of census
- more powerful statistics
- creation of economic value
- reduced cost
- increased capacity

Further benefits may accrue depending on the system design and infrastructure

- Randomised controlled trials of multi-agency interventions and robust cross-sectoral register based research becomes the norm
- Requirement to invest in an independent Proper statistical authority for health data – data linkage and associated R&D must be at arms length from Government.
- Build on local expertise (e.g. primary and secondary care data linkage in Lothian) to amplify national capacity
- Improvements in the quality of population health and health inequalities research by increasing the range of questions that can be addressed and the statistical methods employed

- More rapid development of research and development – reducing delays of up to a year caused by current governance requirements that are, of necessity, tailored to each study
- Improving the international competitiveness and capacity of Scottish researchers and service staff
- Provide greater economic value, which is more widely spread
- Improve access to the programme
- If matched by public and professional communication and training and feedback of findings, improve staff and public understanding of data as the essence of their/their patient or client’s story and personalise investment in data quality and security.
- Improve the responsiveness of e-health solutions
- Drive innovation and investment by all actors (sustainability)

## **Consultation question 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?

**Answer:** The challenges have been summarised in the consultation document as:

Public acceptability [1]

Legal restrictions [1]

Confidentiality [1]

The need for a **common identifier** (=Linkage Reference) [2]

Data **quality**, timeliness and completeness [2]

Technical **capacity** [3]

Analytical **capacity** [4]

Important challenges have been left out of this list (although they may have been hinted at in other parts of the document):

**separation of powers,**

monitoring behaviour

public representation/participation/control

**access** to the programme

actor roles

responsibility for PIA

ownership

semantic and syntactic **interoperability**.

*I would echo the following comments of my colleague Mark McGilchrist*

A comprehensive data linkage network requires the existence of a **common identifier** and mappings to operational unique identifiers within data sources. A common identifier can undoubtedly be constructed in Scotland and held by an *indexer*, but an important challenge is where the mappings are held. The unacceptability of the National ID card arose from the existence of the central database and its also holding other identifiers as mappings. This problem can be circumvented by placing the mappings with the data sources themselves, and the mapping operation can be designed to ensure the *indexer* does not know for sure the subject population of a data source. This allows the *indexer* to become a national ID database concerned with identification only. For example, the current SHIP indexer holds mappings centrally, which is problematic given that powerful inferences

can be drawn from this assembly of identifiers. In addition the SHIP *linker* does not actually perform a linkage function and is incorrectly named. There is a clear need for functional definitions of roles.

I would add:

National Records Scotland, in particular the research facility linked by NHS Central Registry, already has the competence to perform this function through its Safe Haven, linking CHI, Unique Citizen Reference Number etc. This function is vital because it is essential that a small number of individuals can change their CHI etc for reasons of personal safety etc *while their UCRN remains the same*. For Scotland to achieve its potential in this area, further investment in infrastructure and capacity building is essential.

### **Independence and governance**

It is essential that the public understands that this programme is set up on a secure statutory footing, overseen by an independent Chief Statistician.

**Separation of powers** can be achieved through autonomous institutions performing specific roles with no more than one role per institution.

Improved governance

The **monitoring** of behaviour needs to be comprehensive. The activities of data sources, Linkers, Recipients, should be monitored by an independent entity. Irregular behaviour would be prosecuted by a further independent entity. Neither the PAS nor NDLC should be used in this role.

A governance framework that builds on that developed by SHIP but with the standing of Good Research Practice and mirroring governance of clinical trials (as occurs in other countries) is essential. Currently much of the research in this area is not identified as such and this means that governance is local, reliant on enthusiasts, lacks the appropriate protection of the Data Protection Act or the access to investment in R&D infrastructure and capacity.

Access should be based on competence – currently capacity for training is limited.

### **Approach to investment in e-health**

Traditionally, investment in e-health has been in large systems that limit access to raw data. A system based on secure data linkage requires investment in inter-operability and adapters so that data based on different database structures can be linked without losing functionality and relational ability. These data remain with their original data controller with specific variables forming part of a specific, de-identified, linked individual level dataset.

The ability to codify free text to enhance traditional data sets is developing rapidly and this functionality is essential.

### **Continued requirement to invest in alternative methods to capture data on undocumented and socially excluded populations**

If the census is abandoned without investment in alternative tools of equivalent quality being placed on a statutory footing, there is a serious risk that additional gaps in our knowledge of the needs, potential and circumstances of undocumented and socially excluded populations will develop.

### **Consultation question 3**

Are the guiding principles sufficient and appropriate?

The guiding principles are what underpins the rationale for the system and its design. These are routed in a democratic society with solidarity of peoples, equality and respect for individuals, legitimacy and accountability of parliament and those in power (public or private sector) and the separation of powers.

Privacy is only part of what is at stake

To public interest, I would add that our current, limited ability to link data results in an incomplete picture of our society and its people, the distribution of affluence, deprivation and discrimination. The experiences of those whose data are not captured are invisible and their voice unheard.

A statement of public goods does not capture fully the requirement to consider how the people of Scotland will retain the intellectual property rights to their data, a say in how these data are used and a share in the economic benefit that we anticipate will be created. Some statutory instrument is likely to be necessary to ensure that data are held in common for the benefit of the population of Scotland with individual linked datasets being licensed for a time limited period.

A commitment to registration of datasets, with the register being open to the public, and a requirement for publication in open, peer-reviewed journals would reinforce this development as a public good.

With such a framework in place, and supported by statute, individual level consent is unnecessary and inappropriate. Any request for opt-out of inclusion in a data set should be handled by a local Caldicott Guardian on a case by case basis.

Several of the principles in the privacy section already apply to health data under the governance of the Caldicott Guardian. This role and competence should be replicated across the public and private sector by the appointment of an appropriate independent professional. This already exists through the network of data sharing partnerships, for example, the Lothian and Borders Data Sharing Partnership.

Aligning the governance of register based research with that of trials would incorporate several of the governance mechanisms included in this section and minimise the risks of inappropriate variation between organisations. Some register based research using linked datasets *may* require (and some studies already use) a similar process of code breaking to that which already exists for trials. The requirement for this process would arise if study results suggested harm.

Currently, universities are not classed as public bodies and most are charitable institutions. Some advice is necessary to ensure that the appropriate legal and governance relationships exist between the NHS, universities and their strategic partners, particularly those in the public and third sector.

There should be an expectation of public and participant engagement in research and unbiased communication of findings, action to be taken and next steps unless people have opted out.

### **Consultation question 4a**

Are the objectives for a PAS set out in section 3c the right ones?

**Answer:** The 5 objectives make clear that the PAS is there to help (PIA, ethical, legal, regulatory), advise (reputation), and make recommendations to other actors. A Privacy Advisory function is essential but should be a formalisation of the existing expert groupings. Those associated with this function should be able to provide expert methodological advice. There is some lack of distinction between governance which is appropriate but powers are already vested in the Information Commissioner of Scotland. The National Data Linkage Centre should provide technical advice. It is important that the role of the centre and those within it is not confused.

### **Consultation question 4b**

Do you wish to be consulted on firmer proposals for a PAS as and when they are developed?

**Answer:** Yes

### **Consultation question 5a**

Are the functions that will be led by the NDLC set out in section 3d the right ones?

**Answer:** Qualified yes

See 4a – As set out this confuses the development of a centre with that of an independent Proper Statistical Authority. This should not take responsibility away from individual data controllers but be a function under the auspices and governance of a national statistics service, led by experts. Some linkages would need to be undertaken directly but others could be licensed to competent bodies acting as ‘outreach’ e.g. research safe havens. There is a clear need for a partner to the PAS that provides support, advice, education and development on technical matters. This could include the provision of software tools and services to various actors.

### **Consultation question 5b**

Do you wish to be consulted on firmer proposals for the NDLC as and when they are developed?

**Answer:** yes