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.

I have answered the consultation questions in a particular order so that the answers build on each other.

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

Notes on programme actors for this response

The main actors for this programme are: (1) data controllers, (2) recipients of linked data, (3) the analyst community, (4) Linkers, and (5) Indexers.

For me, a Linker is an entity that creates or retrieves a unique number for a person-specific record based on a matching process of some kind. For example, a staff member at reception for a hospital clinic who identifies a patient against an hospital MPI using the patient details is operating as a linker. An algorithm matching records from two different record sets using patient details is acting as a linker. An pseudonymising entity that creates a new consistent set of random numbers for an existing set of person-specific records is acting as a linker. Any entity which simply uses existing numbers to join records together is not acting as a linker. Indeed this latter action is trivial to an analyst using any database or statistical package.

An Indexer is a linker matching against a reference standard for patient/citizen details, although they may have other capabilities/responsibilities assigned.

A recipient is a technical entity capable of receiving datasets from a variety of sources and securing these datasets for exclusive use by a designated set of analysts.

Consultation question 3

Are the guiding principles sufficient and appropriate?

Answer: The definition given for *principle* seems wrong. Principles are definitely not starting points for deliberation. In the context of the present discussion I have used this definition:

A principle is a law or rule that has to be ... followed ... such that a system is constructed. The principles of such a system are understood by its users as the essential characteristics of the system, or reflecting the system's designed purpose, and the effective operation or use of which would be impossible if any one of the principles was to be ignored [Alpa, Guido (1994) General Principles of Law].

We want our principles to help us reason toward a system design. But that design seems to have been decided already to a large degree with a centralised architecture based on the PAS and NDLC.

I am very uneasy with this and would ask why views on the architecture were not part of this consultation? We are already well down the road.

While the consultation document makes clear the purposes are research and statistics for the public sector, it would be naive to think that this segregation will persist. Non-public sector needs are growing so there are other forces at play which must be controlled now. Designs which rely on public trust in particular stakeholders may prove unforgiving as the nature of these institutions and their relationship to others changes.

Lest you think I am not 'on side' with the aims of the Scottish Data Linkage programme I would offer my definition of an information society to which I subscribe:

"A society which uses information technology and data flows to achieve greater self-understanding and adaptability of its politics, economy, society and culture"

The system should also reflect our values as a society. I posed the following question to an audience at a recent conference:

"Do the information infrastructures we build today reflect the values of our present society, or do they herald the society to come?"

We must take care what we build.

The stated principles must help us make design decisions. A few powerful principles will be more use than many 'smaller' principles even though the latter may be not unreasonable statements to most people.

I believe we should have a combination of general and specific principles that reflect our values as a society. While I acknowledge that the given principles may be perfectly reasonable, they do not, in my view, adequately encompass our aspirations as a society, nor impose the necessary constraints.

Below I set out 'alternative' principles that help us with the question of design and reflect our true aspirations, indicating (through numbering from the consultation document) how they encompass nearly all of the principles in the consultation document.

I would propose that the programme's principles should be those of democratic society (and its associated economic model), namely:

1. Respect for individual sovereignty [2,11,12,20,21,22,35,36]
2. Free association [11,12,29]
3. Equality and respect for others [11,12,29]
4. Ownership, free exchange, development [1,3,37,38]
5. Right to privacy [2,11,13,14,15,16,17,18,19,23]
6. Education/responsibility [30,37,38]
7. Executive legitimacy and accountability [7,11,12,31,32]

We should also embrace democracy's solutions: rule of law [34,37,38], competition, separation of powers [10,33], and the monitoring (FoI, mass media, regulation) [4,5,8,9] and influencing of official and other behaviour [39,40].

What do we notice from the comparison? The consultation document's principles are preoccupied with privacy. Privacy is very important (I am an advocate for privacy), but it should not be used to divert our attention from equally important matters; privacy is only part of what is at risk. **Few principles are directed at influencing official and other behaviour, separation of powers, and competition, presumably because the system design is already agreed?**

I introduce some more focussed principles in addition to the democratic ones above:

a) No central conduit as this would create a new, powerful sovereignty.
b) A multiplicity of all kinds of actors. For example, data sources entering into a linkage agreement with analysts should be able to select a Linker and Recipient for their data from a range of providers. **See my definitions at top for these actors!**
c) The public have explicit representation in the system. In particular they should be able to review (and preview?) linkage activity and, if they wish, block the use of their data. The coming EU regulation will have a 'right to be forgotten'. [Viviane Reding, Vice President, European Commission, The EU Data Protection Reform 2012: Making Europe the Standard Setter for Modern Data Protection Rules in the Digital Age 5 (Jan. 22, 2012)]. It is my personal view that this would not harm research and would be exercised by only a very small portion of the population, and it creates a new inclusive relationship with them.
d) Data sources perform all data handling, including pseudonymisation, before delivery to a Recipient. They should not leave privacy processing and assessment exclusively to others. Others may assist, but responsibility lies with the controller.
e) Access to the programme is a right on acquisition of a licence (i.e. actor is assessed to be compliant with standards/capabilities by relevant authorities). See Q4/5 below regarding PAS and NDLC.

These additional principles provide strong constraints on system architecture and design and do not realise a centralised system. Data sources, recipients and analysts are free to associate, but Linkers and Indexers and those providing the necessary workflow coordination are part of the mechanism only.

Finally, although benefits arising from linkage are stated to be public goods for this consultation [6], there is no reason why this should be so in a future that includes non-public sector actors where companies may perform inter-company linkages for the good of their customers. What matters is that a linkage is not against any clear public interest. We must build with an eye to the future.

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 will comment on a few of these challenges.

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 partly arose from the existence of the central database and it's 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.)

If data sources, recipients and analysts are free to associate we may take a 'market' view of data sharing, and by permitting revenue generation for all actors from secondary uses of data we have a mechanism to improve access, quality, technical and analytical capacity, and data interoperability. (For the latter, data sources and analysts must be in direct contact to assist with the exchange of meaningful data. A centralised entity trying to take on the role of semantic mediator is problematic.)

Separation of powers can be achieved through autonomous institutions performing specific roles with no more than one role per institution. The use of 'Chinese walls' where a single institution performs more than one role is problematic and in other domains, such as banking, has been shown

to fail. Given the wealth as well as health agenda, serious consideration should be given to this issue and the ultimate design of the system.

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

It is vitally important that there is open access to the programme to spur investment and innovation. In a distributed system a licensing regime would probably suffice with licence acquisition open to all. This leaves the issue of the licensing agencies (PAS, NLDC?).

Regarding the PIA, I would suggest that the actors for this are the recipient and analysts since they are best placed to know what possible additional uses could be made of the data when in their hands. They should provide this assessment to the PAS and data controllers for their consideration and insertion into the MOA that is suggested. The PAS and data controllers can request further information until they are satisfied with the disclosure. Data controllers should be permitted to reject a linkage at any time and the PIA would be a valid reason for the authority monitoring behaviour of controllers (see above).

I will not discuss ownership & control as these are too complex to discuss here. However, there is a definite need to resolve these issues, and I believe we must give patients/citizens a more direct role to play in the system, far beyond representation and 'participation', but some kind of control.

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 employed. A system which permits revenue generation and autonomy for data sources, recipients, analysts and other actors will, for example:

- Provide greater economic value, which is more widely spread
- Improve access to the programme (for consumers such as analysts)
- Improve availability and quality (from producers such as data sources)
- Drive innovation and investment by all actors (sustainability)

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. This is as it should be; since the data controllers must be the ultimate authority on whether to link data (satisfying the democratic principles presented in question 3 above.) We must educate and support data controllers, recipients, analysts and other actors so that they can drive a neutral mechanism for linking data and have the freedom to innovate and invest in directions they think are viable. No centrally planned information economy!

At present the PAS has no 'powers'. It is important NOT to give this entity real powers to approve/block linkages as this would interfere with the sovereignty of the data sources/controllers (see question 3). Giving it statutory powers would open it up to accusations of possible interference and bias for particular linkages, and who would hold it to account? What would its membership be? Who decides that? It is not clear why the PAS should offer technical advice as this sits more comfortably with the NDLC acting in an advisory role (function 1 in section 3d).

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: No.

As things stand the NDLC would represent an alternative sovereignty to the data controllers, countering the democratic principles suggested in Q3. All the functions suggested vest considerable power in this one entity: (3) linker (your definition), (4) data-exchange service, (5) population spine. There is moral hazard in such an entity. It centralises the systemic risk and creates an entity that is too important to fail! If it does not behave as expected there is no alternative. What sanctions would be brought on such an entity? To avoid such an entity is one purpose of the 'no central conduit' principle and 'multiple actors' principle suggested in Q3 above. With these principles actors can be punished and removed from the system.

There is no requirement for special pleading with the census. In a distributed system, the census is a completely normal linkage, but with a specifically chosen recipient and set of analysts. In systems where the data controllers do all data manipulation and where the final data moves directly from data source to recipient the only entities seeing the 'census' will be the designated recipient and analysts.

The NDLC function 1 is certainly valid. 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