Sent: 15 June 2012 15:20
To: 'datalinkageconsultation@scotland.gsi.gov.uk'
Subject: National Data Linkage Framework Consultation Response

Sent on behalf of Dr Corri Black

Please find attached a completed Respondent Information Form and Response to the Scottish Government Consultation:

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

This response was prepared by Dr Corri Black, Dr Sharon Gordon, Steph Hall and Katie Wilde and is submitted following consultation with and on behalf of the University of Aberdeen, Health and Data Linkage in North East Scotland (HEADLINES) research network, which includes:

Prof Siladitya Bhattacharya
Dr Sohinee Bhattacharya
Dr Corri Black
Prof Christine Bond
Dr Carl Counsell
Dr Leone Craig
Dr Mike Crilly
Prof Graham Devereux
Dr Sharon Gordon
Ms Stephanie Hall
Prof Phil Hannaford
Prof Peter Helms
Yvonne Hopf
Prof Amanda Lee
Prof Anne Ludbrook
Prof Gary Macfarlane
Dr Mary Macleod
Dr James Mclay
Prof Geraldine McNeill
Dr Robert Milne
Dr Peter Murchie
Dr Markus Steiner
Dr Janine Thoulass
Dr Melanie Turner
Dr Steve Turner
Dr Margaret Watson
Katie Wilde

Regards

Corri

Dr Corri Black
Senior Clinical Lecturer in Public Health Medicine
Division of Applied Health Sciences
University of Aberdeen
Polwarth Building
Foresterhill
Aberdeen

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)	**X**
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 **X** No, the benefits are described fully ☐

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

We believe that there are substantial benefits to linkage of high quality data for delivering a healthier, wealthier and smarter Scotland. Those that are not sufficiently described include: health service evaluation, audit, public health surveillance, public health research, epidemiology, health economics research and pharmacovigilance. These research and service areas should be highlighted as primary benefits of data linkage and as important as ensuring more cost effective research processes.

Scotland has particular strengths with long term, highly enriched cohort studies, prescribing data, primary care and education data to mention a few however, to maximise the research capacity of these rich resources, they need investment, good meta-data and the expertise to truly understand their strengths and limitations. Ensuring sustainable access to primary care data should be a priority for Scotland. The Primary Care Clinical Informatics Unit at the University of Aberdeen have been successfully collecting valuable primary care data from GP practices across Scotland since the 1990s and have demonstrated the willingness of GPs to participate, the technical feasibility and the high potential value for improving population health.

There is no mention of the usefulness of linked data in identifying patients eligible to take part in approved studies/trials. The Academy of Medical Sciences (AMS) recommended that, 'Accredited investigators and research team members should be considered part of a clinical care team to enable identifying patients eligible for approved studies.'[1]

The "statistical system" in Scotland should be recognised as extending to health boards with regional health intelligence services; health surveillance, audit and service evaluation making regular secondary use of data.

[1] A new pathway for the regulation and governance of health research. The Academy of Medical Sciences. Jan 2011.

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 **X** No, the challenges have been identified ☐

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

We agree that there is considerable uncertainty around the legalities and public acceptability of data sharing and linkage and this is a challenge.

In terms of the legalities, this uncertainty has resulted in the various bodies and individuals, who have a role in deciding access and approving linkage, operating both inconsistently and perhaps over-cautiously. The inconsistencies are both within bodies, and between bodies and across regions. This inconsistent and cautious approach can have significant impacts on research, with approvals taking more than 6 months in some cases, and with the substantial research time spent responding to various concerns this can significantly impact on the success of research projects, which are often funded for 2-3 years.

A major challenge is the lack of clarity about the level of risk studies pose and the proportionality of response to these risks. There is a need for clarity about what is "best practice" and proportionate governance for the potential risks relating to "non-interventional" data linkage studies. There is also a need that these processes are cognisant of the processes being adopted in the other nation states, particularly England with CPRD and other regional health data linkage projects. The move towards proportionate governance and Safe Havens needs to be considered in a UK context rather than just the Scottish context, with a view towards facilitating research across national boundaries rather than creating barriers where cross border differences exist.

In terms of the public acceptability, research from the Child Medical Records for Safer Medicines (CHIMES)[2] programme indicates that children/young people and parents/guardians have a limited knowledge of how routinely collected healthcare data is currently used but at the same time there is an assumption that the NHS use health data to improve and safeguard population health. Public awareness and engagement has been found to be critical to enable an understanding of the governance issues and solutions but that when this effort is made, the public are supportive of the process. Consent and assent are important to the young peoples' and parents'/guardians' support of data linkage with individual opt-out considered to be more practical[3], however, in many circumstances either approach would be extremely difficult, if not impossible, to implement effectively. Furthermore, opt out from previously linked datasets would be impossible to manage and incompatible with the aim to generate anonymised datasets. The time and costs to track individuals back through copies of datasets and all backups would be prohibitive. Concerns and anxieties of young people and parents/guardians increase as the number of linked data sources increases and when there are commercial interests[3].

Low profile public engagement and disproportionate negative media coverage, often around breeched data security, are a challenge and require investment to provide

[2] http://www.chimesprogramme.org/

[3] Scobie-Scott, *et al*. Is the use of linked routinely acquired NHS data for pharmacovigilance in children acceptable to parents and young people? Scottish School Primary Care Conference, April 2012, Glasgow.

the evidence to the general public of the many benefits of data linkage, including the valuable relationship between academia and the pharmaceutical industry in terms of protecting and improving health.

From the perspective of healthcare professionals work from CHIMES also identified some practical, ethical and legal issues with data linkage but again there is general support for data linkage to ensure patient safety, best clinical practice and to improve population health with comments such as: "*The benefits are so enormous and the risks are so minimal that denying this activity [data linkage] is unethical*." (A04) and "*As a professional I can see the huge advantages of linking the data […], I think it would be fantastic if it works.*" (FG05, Paediatric pharmacist)[4]

Data security is of course a priority; however, it is important to ensure data security without sacrificing research functionality. There are risks to research quality in building a system that requires multiple personnel to handle separate aspects of the data and by removing approved researchers, who often have an intimate knowledge of at least one of the data sources being linked, from the validation process. Some of those approved researchers are clinicians/health professionals who have access to the identifiable data as part of their clinical practice. While this does not give them the right to access it for research, it is important that their rich knowledge of their patients and data is not lost from the linkage process as it brings direct benefits to the quality of the research output and its impacts.

This is again consistent with the AMS recommendation that, 'Accredited investigators and research team members should be considered part of a clinical care team' and be allowed to access their patient identifiable data for research purposes and for data quality assessment, such that regulation and governance pathways do not undermine the reliability and security of data linkage[1].

<u>Challenge 2: Incomplete Data or data that cannot be linked.</u>
There is a discrepancy between the drive towards secondary use of data and a lack of systems to support this, including the awareness of health professionals with regards to data collection. High quality research outputs should be used to raise awareness of the importance of data collection and the quality of the data collected.

<u>Challenge 3: Limited capacity for secure exchange and access to data.</u>
There appears to be limited capacity within some of the national infrastructure to provide rapid access to data. Delays at NSS:ISD mean that researchers may wait 6 months plus for datasets even after approval. There is also a need to improve the communication of how linkages are done and for transparency around the validation of this linkage process, accuracy and the implications for interpreting the data.
The definition of "safe haven" in the consultation document is in fact that of a "dumb terminal" and does not reflect the range of access control mechanisms that could be offered – proportionate to the risk for individual studies. These include: restricted access via secure folders, remote access software, dumb terminals, training for approved researcher status and contractual agreements with approved researchers.

[4] Hopf, *et al*. (2012). "It's about recognising how many people you've poisoned by giving them medication": thoughts of healthcare professionals on pharmacovigilance**.** International Journal of Pharmacy Practice, 20 (Suppl. 1), pp. 5–15.

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 **X**

Please explain your answer fully and make suggestions for improvement.

In general terms there is no mention of the Caldicott Principles[5] within the consultation paper, in relation to the use of health data, and the list of 'Guiding Principles' seems extensive and better described as 'Guidelines' or as it is referred to in the document the 'Data Linkage Framework' to support e.g. working within an abridged version, similar to the Caldicott Principles. Care should be taken to make the distinction between the principles and the details of the guidelines on how to deliver against each of the principles.

The principles should also reflect the duty to improve and protect public health not just the duty to protect the identity of the individual and place greater emphasis on the need for proportional governance to the risks posed.

There is an urgent need to be explicit and for agreement on the precise definitions of the terms and language used within the principles. For instance: What is the definition of 'sound and robust' research in Principle 1? What are the 'legal and ethical standards' referred to in Principle 3? What are the 'benefits' that should be shared widely in Principle 6? What is meant by 'publicly available' in Principle 8? Principle 9 suggests all data linkages should be 'appropriately' monitored by a 'relevant' individual or organisation. In Principle 14 what determines 'every effort' to minimise the risk of identification or re-identification? What is the 'relevant expertise' required for those judging the risk of re-identification of data subjects (Principle 19)? The ambiguity in the language used is unhelpful and it is imperative that a National Framework is explicit and provides clarity and the signposting required for the data linkage process.

Moving on from language to the data linkage process: the implications of the Official Statistics Orders need to be clarified in relation to the wider use of linked data (Principle 4). Relating to "commercial gain" in Principle 5 – this needs a clear definition - if a pharmaceutical company invests in post marketing surveillance – is this commercial gain or good health protection? Does sharing of the 'benefits' of linkage in Principle 6 include the linked data itself? Is that feasible/fair for all research?

Principles 8 and 15 suggest all privacy impact risk assessments and data sharing agreements should be available to anyone – these documents are study specific and can be fairly complex - would it be more appropriate for a subset of key information to be made available and in an accessible format?

Principle 10 asks for a clear distinction of roles for those undertaking the data linkage, analysis and those policing the governance which in itself is fine however, for many institutions to define "linkers" as someone who does nothing else is not practical. It may be workable on a project by project basis but it is not useful as a blanket rule – particularly where there is "low risk" of identification. There are also research quality benefits from having a single data management person seeing a project through from linkage to final dataset support.

[5] http://www.scotland.gov.uk/Resource/Doc/340362/0112733.pdf

Principle 16 refers to the minimal storage of linked data – it seems sensible to include a review process before linked datasets are destroyed to evaluate the potential for further approved secondary use, as part of standard data management plans.

There is a need for clear guidance on what is considered ‘appropriately trained’ and ‘proportionate’. This must also be done in the context of potential international users of the data and commercial uses to ensure that researchers are not disadvantaged and therefore that Scotland is not disadvantaged compared to other countries.

It is also important to recognise that it is not just clinical trials that may need re-identification.

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

Yes, the objectives are right **X** No, they are not ☐

Please explain your answer fully and make suggestions for improvement.

A central advisory service to provide consistent and clear legal and governance advice for Scotland seems a sensible approach but 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.

It would be valuable to have clear, published advice about the risk levels and what proportionate governance and disclosure risk management responses are required.

We are supportive of objectives 1 and 2.

For objective 3, care should be taken not to add an additional tier of permissions on top of ethics and Caldicott approvals. Any function for approval to start a project must also be timely.

For objective 4, it would be appropriate for the privacy advisory service to advise on security, linkage methodology where it protects privacy and data management to improve security. Advice on study design, methodology and analysis would sit better with a Scottish Health Informatics Research Centre or individual academic centres.

Objective 5 could be delivered through clear website based guidance that makes explicit the different levels of risk and the appropriate proportionate responses.

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

Yes **X** 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 **X** No, they are not ☐

Please explain your answer fully and make suggestions for improvement.

Academic institutions across Scotland hold many research datasets and have extensive expertise in data linkage research built up over many years of experience in this research area. 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. 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.

An important function is missing from the remit of a National Data Linage Centre

around coordinating the maintenance of research-enabled national data ready for linkage – this is something separate to the technical function of linkage.

A trusted data exchange service would need to be, at least, sufficient to meet the security requirements for the transmission of any of the individual data sources.

It will be critical to the success of any National Centre that it embrace and works collaboratively with the NHS National Safe Haven at ISD and the regional Safe Havens, where there has been considerable investment already, as part of the regional network. The regional Safe Havens have a range of focus and strengths that provide a rich resource for linkage skills nationally, deep knowledge of local datasets and are viewed as easily accessible and trusted partners within research teams.

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

Yes **X** No ☐