

Draft Meeting Minutes

Pseudonymisation Steering Group

Meeting 3, Wednesday 13th August 2014

Skipton House, London, SE1 6LH

(Ratified at 7th October Steering Group meeting)

Attendees:	Role	Organisation
Kambiz Boomla	Observer	Confidentiality Advisory Group
Antony Chuter (dial in)	Patient Representative	
Garry Coleman	Head of Data Management Services	HSCIC
Paul Cooper	Member of techUK Health & Social Care Council	techUK
Paul Cundy(dial in)	General Practitioners Committee and BMA	
Harvey Goldstein	Academic Expert on Data Linkage	University College London & University of Bristol
Wally Gowing	Pseudonymisation Advisor	
Alan Hassey	Member of Independent Information Governance Oversight Panel	Retired GP
Ian Herbert	Primary Health Care IT Specialist Group and GP Extraction Service Independent Advisory Group (GPES IAG) Member	British Computer Society
Julia Hippisley-Cox (dial in)	Academic Expert on Data Linkage and EMIS National User Group	Nottingham University
David Ibbotson (dial in)	Programme Head care.data	HSCIC
Max Jones	Director of Data & Information Services and Pseudo Review Chair	HSCIC
Phil Koczan	Representative of the Royal College of GPs and member of the Health Informatics Group	GP, RCGP
Sean McPhail	Senior Information Analyst	Public Health England
John Parkinson	Consultant	Clinical Practice Research Datalink
Chris Roebuck	Director of Benefits and Utilisation and Review Co-ordinator	HSCIC



Matt Spencer Pseudo Review Project Manager HSCIC

James Wood
(dial in) HSCIC Head of Infrastructure Security HSCIC

Apologies

Xanthe Hannah	Observer	NHS England
Geraint Lewis	Chief Data Officer (Observer)	NHS England
John Madsen	Head of Productivity & Efficiency	HSCIC
Dawn Monaghan	Observer	Information Commissioners' Office
Nicholas Oughtibridge	Leading on Code of Practice for Confidentiality	HSCIC
Daniel Ray	Head of NHS Chief Information Officer Network	University Hospital Birmingham
Hashim Reza	Consultant Psychiatrist and Mental Health Information expert	Oxleas Trust
Eve Roodhouse	Director care.data	HSCIC
Ralph Sullivan	GP	RCGP

APPROVED

1. Welcome and Introductions

- 1.1. Colleagues were welcomed to the third meeting of the HSCIC's steering group on pseudonymisation. The Chair thanked the group for its contributions over the past weeks
- 1.2. Introductions were completed including the representative from techUK (Paul Cooper) attending the steering group for the first time following a request from the Review group.

2. Review of minutes/actions

- 2.1. The minutes of the 3rd July Steering group were reviewed and no comments were raised. The minutes were therefore ratified by the steering group and will be published on the Review website within 7 working days of today's meeting.
- 2.2. Action log was reviewed and a number of actions were closed and updates applied to actions which were to be carried forward. Please refer to action log for details.
- 2.3. It was noted that sub-groups terms of references should be produced and presented to the steering group for approval. All sub-groups are to submit ToRs for next steering group meeting.

3. HSCIC Update

- 3.1. An update was provided to the steering group on two areas, approach to Data Labs and Data Access Request service.

3.2. Data Labs

3.2.1. The approach to Data Labs has undergone requirements gathering and as part of this a stakeholder event was held on 21st July with a wide range of stakeholders from patient groups, GP practices and other groups. One point raised at the event was that Data Labs should not become a technical solution and should not inhibit innovation. Notes from the stakeholder event have been published and further work will take place with stakeholders to develop the requirements for the service. A number of approaches to access Data Lab are currently proposed as follows:

- 3.2.1.1. Data Lab 'reading room' concept . This is a service where you can view and read HSCIC data. The data is presented for online reading. There are no print or download options. Access and viewing is supported by Data Sharing agreements being in place and no information will come in or out of the reading room.
- 3.2.1.2. Statistical analysis – A tool providing a guided way to explore analysis of statistics. This will be a controlled set of data and available to the general public and would only produce aggregated, anonymised output. Virtual Environment Data Lab – Would be used where data sharing agreements are in place. It would provide registered users with remote access to a wide range of analytical

functions, including the ability to access, manipulate, interrogate and then report on various data sets. An example for access would be ALBs access HES data within the data lab. Trusted Organisations – Where access by trusted organisations to the data lab would be able to link HSCIC data to their own data. Local data stays local, for rapid access. The less frequently used remote data held by the HSCIC is still accessible, as needed, and queries are optimised by the federated server to ensure that they are retrieved as efficiently as possible. Traceability of access and data accessed would be required to demonstrate control of access. An example of a trusted organisation would be those that would be potential users of NTS.

3.2.2. Progress on Data labs is being made and further engagement with stakeholders will be undertaken. The HSCIC feels it is a good opportunity to pursue in order to meet the Partridge Report recommendations of implementing adequate controls and ensuring transparency of access and use.

3.3. Data Access Request Service

3.3.1. The Data Access Request Service was launched on 31st July 2014. This service, delivered via a dedicated web page, provides HSCIC customers with a single point of access for all new data applications, operating a revised process which has been simplified into Application, Approval and Access. The simplified process has greatly improved the efficiency with which requests can be managed and progressed by the HSCIC, and provides far greater clarity for customers in understanding the status of their requests. This will be further supported by the introduction of Service Level Agreements for each stage in November 2014. The target service level for simple requests will be 14 days; the service level for the most complex requests will be 60 days. One area being looked at is the correct terminology to be used, a subject which is relevant to the Standards & Terminology sub-group.

3.4. A question was raised as to whether the level of anonymisation, in Data Labs, is being looked at. HSCIC Data Lab team will reference the ASH consultation, which closed on 8th August, and the team will consider the outcome of the consultation on this topic. The steering group member stated the question is to do with the quality of the anonymised data as the process of anonymisation could degrade the quality of the data. The level of anonymisation will potentially affect researchers.

3.5. Steering group members discussed the levels of risk that needs to be considered. From CPRD's viewpoint there had been no issues with any research data released in the last 25 years. However it was noted that the landscape around public data and its dissemination had changed considerably in the last 2 years.

3.6. The different approaches to data labs offers opportunities to get the release of data to other users, beyond the HSCIC, right but it was recognised by the HSCIC that there will be challenges on areas of exploiting the data for various purposes.

3.7. A steering group member commented that the use pseudonymisation on the different data lab approaches could impact the level of transparency.

3.8. The HSCC will, in publishing the Data Access and Data Lab approaches, ensure that transparency, and levels of anonymity, are considered.

3.9. The Independent Information Governance Oversight Panel (IILGOP) will also be reviewing the data lab approaches.

3.10. One steering group member raised a concern that Data Labs could make GP data available over and above that agreed for care.data and without taking patients' views into account. The Chair advised that this would not be the case and that Data Labs would only make data available under any terms and conditions previously recommended by the GPES IAG.

3.11. The recommendations, from IAG, have stated that data must only be available in anonymised or pseudonymised form.

Post Meeting Note: HSCIC are currently preparing to request at the next IAG meeting, 11th September, to remove the restriction around only commissioning purposes only.

3.12. A question was raised as to whether there has been any discussion on the impact of Data Labs to other areas of data release. The HSCIC stated that all scenarios are being looked at and evaluated. The question, on the evaluation of any rollout, was raised as to whether there was an assumption of starting with a pilot rollout across different datasets. It was agreed that the HSCIC would not be expecting to go to a big bang rollout.

4. Probabilistic Record Linkage – Presentation

A presentation was delivered by Harvey Goldstein.

4.1. A steering group member suggested to the sub-group looking at probabilistic record linkage needs to be mindful that linking NHS data with NHS data just replicates the quality issues even if 99.1% linkage was achieved.

4.2. A general update from sub-group followed this agenda item and the chair commented that sub-groups should produce tangible results quickly from their deliberations but it was accepted that topics needed to be explored.

5. National Back Office (NBO) and Medical Research Information Service (MRIS)

A presentation on NBO and MRIS was delivered by Garry Coleman. The presentation has been emailed to steering group members and is attached for reference.



Garry Coleman
Presentation for 2014-

5.1. A number of comments were made by members as follows:

5.1.1. One steering group member queried whether there was any demand to pseudonymise the data underpinning the National Back Office (NBO), comparing it to the telephone directory. The reply was that there was no demand for doing this, but it was important for the steering group to be aware of all uses of identifiable data at the HSCIC.

5.1.2. A question was raised around how MRIS ensures the quality and validity of NHS number needed to perform its linkage and more how this can be assured more widely across the system. A member replied that it is the responsibility of submitting organisations to check this. A question on the numbers reported by GC on MRIS data. GC to look at and confirm the make-up of the numbers used in the presentation.

6. **Reversibility/Irreversibility of Pseudonymisation** – an understanding of current practices

The representative from CPRD outlined some of the current issues around reversibility/irreversibility with following statements:

6.1. Difference between Anonymisation and Pseudonymisation is not always clear.

6.1.1. Anonymisation is not Reversible and Pseudonymisation is Reversible.

6.1.2. The difference between anonymisation and pseudonymisation has become blurred in recent discussions leading to different approaches being implemented.

6.1.3. Members contributed to a general discussion with following comments specifically outlined:

6.1.3.1. Reversibility of pseudonymisation could be important for direct clinical care, whereby Risk Stratification identifies some patients with likely clinical needs, who then need to be identified by the clinician for intervention.

6.1.3.2. The Chair suggested that the issue of Reversibility/Irreversibility should be referred to the Standards & Terminology sub-group as part of its work programme to look at standards involved in this topic.

Action Point: Standards & Terminology sub-group to add Reversibility/Irreversibility to their work programme.

6.1.3.3. The Chair asked members to consider the subject of Pseudonymisation and Patient data and how to make pseudonymisation more effective.

Post Agenda Item Note 1: A presentation on this agenda item was provided by John Parkinson – CPRD and has subsequently been sent to steering groups members.

Post Agenda Item Note 2: In addition a number of definitions were provided by John Parkinson – CPRD and these have been forwarded to the Standards & Terminology sub-group for them to consider in their work programme.

John Parkinson presented that there can never be a technical solution that removes all risk of patient reidentification, whilst maintaining the usefulness of the data, so the most important element is around controls around who has access to data and appropriate contractual safeguards. He presented his specific concerns that an approach to provide each customer with a specific study specific pseudonym does not add much extra security, as it would be technically possible for data to be relinked on characteristics not usually pseudonymised, such as event dates. He presented his view that steps could be taken to blur these, but this would impact on utility of the data for certain purposes.

7. **Current Legislature points**

An update on current legislature covering the number of applicable national and international standards was provided. These will be explored further by the Standards & Terminology sub-group which has just met for the first time today, 13th August.

7.1. A member suggested it would be useful for the steering group to provide a synopsis of this area along with examples of how various known standards work.

7.2. The Chair finished the agenda item by encouraging the sub-group to consider any proposals for the wider system. However the Steering Group will have to approve any proposals produced by the sub-group as the scope of the sub-group will have to be careful not go too far beyond the remit of the Review.

8. **AOB**

No items were noted.

9. **Next meeting:** Tuesday 9th September 1.00pm to 3.00pm at Skipton House, London.

APPROVED