

## HSCIC Pseudonymisation Review Steering Group

---

**Date:** Wednesday 11 February 2015 **Meeting Nr:** 9  
**Location:** Tavistock House, London & Conference Call  
**Purpose:** For Ratification at March Steering Group

<b>Attendees:</b>	<b>Role</b>	<b>Organisation</b>
Antony Chuter (by dial in)	Patient Representative	
Kambiz Boomla	Observer	Confidentiality Advisory Group
Wally Gowing	Pseudonymisation Advisor	Observer
Ian Herbert	Primary Health Care IT Specialist	BCS
Xanthe Hannah (by dial in)	Observer	NHS England
Julia Hippisley-Cox (by dial in)	Academic expert on Data Linkage	Nottingham University
Sean McPhail (by dial in)		Public Health England
John Parkinson	Observer	Clinical Practice Research Data Link
Jill Reeve (by dial in)	Project Support Officer	HSCIC
Chris Roebuck (Chair)	Benefits & Utilisation Director and Review Co-ordinator	HSCIC
Matt Spencer	Pseudo Review Project Manager	HSCIC
<b>Apologies</b>		
Paul Croft	Business Requirements Analyst	HSCIC
Paul Cundy	GP	General Practitioners Committee & BMA
Harvey Goldstein	Academic expert on Data Linkage	UCL & University of Bristol
Alan Hassey	GP	IIGOP
David Ibbotson	Programme Head, Care.data	HSCIC
Phil Koczan	GP	RCGP/Health Informatics Group
Geraint Lewis	Chief Data Officer	NHS England
John Madsen	Head of Productivity & Efficiency	HSCIC
Dawn Monaghan	Observer	Information Commissioners' Office
Nicholas Oughtibridge	Lead – Code of Practice for Confidentiality	HSCIC
Richard Pantlin	Social Care Representative	
John Parry	Medical Director	TechUK
Daniel Ray	Head of Chief Information Officer Network	University Hospital Birmingham
Hashim Reza	Consultant Psychiatrist	Oxleas NHS Foundation Trust
Eve Roodhouse	Director care.data	HSCIC
Dave Wilby	Business Analyst	HSCIC
Tim Williams	Observer	Clinical Practice Research Data Link
James Wood	Head of Infrastructure Security	HSCIC

## 1.0 Welcome and Introductions

1.1 The Chair welcomed everyone to Steering Group's ninth meeting and thanked the group for its contributions over the past weeks. The Chair further stated there was a lot of work still to be completed over the next few weeks to enable the Review to finish its work.

## 2.0 Review of Minutes/Actions

2.1 The minutes of the 14 January Steering Group were reviewed. There were no comments, on the minutes, from the group and the minutes were approved to be published on the Review's website.

### 2.2 Review of Actions

2.2.1 **Action 23** – There has been no update from Mark Elliott since early January. The Review's project manager is to chase Mark Elliott for dates of when the book is to be made available to the Review.

2.2.2 **Action 31** – This is in respect of Pseudo@Source sub-groups ToR and is covered in agenda item 3.

**Action No. 1:** Project Manager to publish January's minutes on the Review's website.

## 3.0 Pseudo@Source Sub-Group Update

3.1 The Pseudo@Source Chair could not attend today's meeting but had provided a progress report for review by the group. The project manager provided a walkthrough of the progress report's main points.

3.1.1 The Pseudo@Source ToR has been approved by the sub-group and was presented to the Steering group for approval.

3.1.2 The Chair requested that the date of December 2014 be removed from section 4, item 3 as the Review's initial findings have not yet been published.

3.1.3 It was noted in the updated ToR that the issue of Probabilistic v Deterministic linkage was not included together with the reasons why. The Steering Group members discussed this issue at length as there was concern the issue would not be dealt with in the Review's work. The Steering Group Chair stated that probabilistic matching approach is being covered by the Data Linkage & Data Quality sub-group and that the issue will receive consideration when that sub-group reports to the Steering Group. This will include on items such as availability of identifiers, example of where Probabilistic matching is needed e.g. where no NHS No. exists.

**Action No. 2:** The Review Chair took an action to set out how individual providers could get their range of identifiers converted to NHS numbers through the Southport service. These could then be pseudonymised before submission

3.1.4 The Progress Report states that a final report for the sub-group is under development and an early draft will be provided to the March Steering group for review.

3.1.5 PS03 deliverable was circulated to Steering Group for comment. A member stated the list was helpful but did not include GP suppliers or Hospitals systems. After further discussion it was agreed that further columns should be added to capture additional information items – 1) To identify if product is Open Source and 2) Other proprietary products. The deliverable needs to illustrate a range of pseudo models in use.

3.1.6 Another member stated it would be useful to have a Data Controller view of the Market particularly as such products could be Cloud based.

3.1.7 It was noted that a clear definition of 'Pseudonymisation at Source' is needed. This was agreed by members and that this should be undertaken by the Standards & Terminology sub-groups Vocabulary work stream.

3.1.8 The Steering group Chair stated PS03 should be a live document and referenced by

the Review in its final report.

- 3.1.9 The Chair further stated the discussion on probabilistic matching was not a Pseudo question but more about how an organisation obtains correct identifiers to achieve linkages. Another member stated that as linkage is important to the HSCIC there is a need to have probabilistic matching. This was agreed by the Chair who referred to the earlier response that Data Linkage & Data Quality sub-group is looking into this.
- 3.1.20 The project manager in presenting the sub-groups progress report, explained how PS03 is linked to PS04, which is not available to present to Steering group, is to provide an assessment of the open market for pseudonymisation products.
- 3.1.21 The Groups members requested sight of PS04 as soon as it is available so it can review the assessment criteria proposed.

**Action No. 3:** The Standards & Terminology Chair to look into the definition of 'Pseudonymisation at Source' and send this round the group.

**Action No. 4:** The Review's project manager to ensure PS04 (Assessment Criteria) deliverable is made available to the Steering Group at the same time as it is reviewed by the P@S sub-group.

**Action No. 5:** Prior to approval of the P@S ToR the Steering Group Chair will liaise with P@S Chair on 'solution' to be included in sub-groups scope following supplier and provider events.

#### **4.0 Data Linkage & Data Quality Sub-group update**

- 4.1 The Chair of the sub-group updated the Steering Group on activity.
- 4.2 There are several papers in circulation. DLDQ03 looking at Data Quality has been reviewed by the sub-group and DLDQ04, a Theoretical view of impact of pseudonymisation on data linkage, is in development.
- 4.3 The original protocol to using real data from CPRD and HES datasets is still being progressed but approvals to obtain the data remains a challenge. The Review's project manager outlined the current position that CPRD have provided a position statement on use of CPRD GP data for the study is within their existing s251 cover. This is with the HSCIC DARS team to review alongside the application for HES data. The project manager is chasing daily for progress updates on the application. The objective is still to use live data for the study but may have to consider using test data based on scenarios outlined in DLDQ04 paper.
- 4.4 A member asked if matching statistics would be available in the papers. The Chair confirmed statistics on prevalence of identifiers and matching rates on linkages would be included in DLDQ03 and DLDQ04 papers respectively. The challenge is to provide meaningful statistics to provide the evidence base for any conclusions to be drawn from the papers.
- 4.5 Another member stated it is important to look at the statistics in deterministic and probabilistic matching approaches. Another member mentioned Post Codes as being an identifier that should be looked at in linkage rates. The Chair agreed and said the challenge of Post Codes as an identifier needs to be considered in the sub-groups papers.

#### **5.0 Standards and Terminology sub-group update**

- 5.1 The sub-group Chair updated the steering group on the sub-groups activities. The Context paper, distributed to the Steering Group, has had a number of revisions and is presented now for review and approval.
  - 5.1.1 The sub-group Chair explained the background to the paper in that it was initiated following the comments made in August 2014 Steering Group that the Review needs to be clear on the uses of and instances of pseudonymisation.
  - 5.1.2 A member queried the statement in Section 5 second bullet. What was the sub-groups view on data items for pseudonymisation? The sub-group Chair said this item

and other items are to be addressed in the sub-groups Standards paper which is not yet available for review by the Steering Group.

5.1.3 The Steering Group Chair stated a definition of Anonymisation and Pseudonymisation is required for the purposes of the Review. .

5.1.4 The sub-group Chair advised that coherent definitions for De-Identification, Anonymisation and Pseudonymisation are needed in order for the range of outputs of the sub-group and the Review to be consistent.

5.1.4 A member asked if the CAG definition on identifiable data should be used. As it would be useful to reference definitions already in use. Another member stated lots of definitions have been looked at by the sub-group, including the ISO25237, and that it was important that the sub-group came up with terms that were practical to use.

5.1.5 A member queried the statements in Section 4.4 on dissemination of pseudonymised data to other countries. The controls of such data, in other countries, were outside the UK governments, and by extension HSCIC, control so the section needs to recognise this.

The Steering Group Chair stated the Context paper's central tenet should be that Pseudonymised Data needs controls to be in place. The Chair further stated that section 4.4 needs to be finalised before Steering Group approval.

**Action No. 6:** Steering Group members were encouraged to review the current version of the Context paper and the S&T sub-group Chair is to propose a guillotine date for comments to be sent to the sub-group.

5.2 The S&T Chair outlined the position on the sub-groups Standards paper. This has had several revisions and is currently not yet available for Steering Group review. The expectation is a version will be made available in the next week for review by the sub-group and subsequently by the Steering Group before the March's steering group meeting.

5.2.1 The sub-group Chair did however update the Steering Group on one recommendation from the Standards paper. This is that a standard needs to be developed for supporting pseudonymisation and de-identification, and that the Information Governance Alliance (IGA) should look at developing the standard on behalf of the group.

Discussions with the IGA have centred on the new standard should be on De-Identification not just pseudonymisation as the process of de-identifying data goes beyond pseudonymisation and involves technical, legal and organisational measures as well, which need to be clearly stated as part of the new standard.

5.2.1 The Steering Group Chair asked the group for approval for IGA to develop the standard. The group agreed this was something that should happen and that the IGA should be asked to undertake it.

A member asked if the IGA standard is about Pseudo and associated controls and not anonymisation. The S&T sub-group Chair confirmed this was correct.

5.2.2 The sub-group Chair stated that the Caldicott Review recommended that a standard on release of de-identified data should be developed; one standard, ISB 1523, has been developed about data for publication and that the proposed IGA standard would cover the 'the supply of de-identified person level data that is not for publication'. It was also stated, by the member, that the process of controlling identifiable data needs to be looked at.

5.2.3 A member stated that the environment dealing with non-identifiable or identifiable data needs to be documented. Whether the data is identifiable or not is based on the context of where it is produced.

**Action No. 7:** The S&T sub-group Chair to confirm, to the IGA, that the Steering Group approves the recommendation for a new standard, called 'De-Identification', is developed on

behalf of the Review.

**6.0 Pseudonymisation Review – Work Plan**

6.1 The Chair commented that there is still much to do and March will involve a significant number of deliverables to be reviewed or approved. The project manager was asked to monitor progress and to make deliverables available as early as possible before March's meeting to facilitate progress.

**7.0 AOB**

7.1 The Chair asked for members view's on the date for March's meeting as there will be a lot of papers to review. Opinions were sought on bring the meeting forward or putting it back to later in March to allow for review periods.

A member suggested short meetings before the March meeting. The Chair suggested that more meetings would not be conducive to progress. No firm decisions were arrived at so the project manager is to look at the work plan with sub-group Chairs and consider options.

**8.0 Next Meeting**

Tavistock House, London - Tuesday 17<sup>th</sup> March 13:30 – 16:30

APPROVED