## HSCIC Pseudonymisation Review Steering Group

|  |  |  |  |
| --- | --- | --- | --- |
| **Date:** | **Thursday 16 July 2015** | **Meeting Nr:** | **13** |
| **Location:** | **Tavistock House, London and Conference Call**  |
| **Purpose:** | **For Ratification at September Steering Group** |

|  |  |  |
| --- | --- | --- |
| **Attendees:** | **Role** | **Organisation** |
| Antony Chuter  | Patient Representative |  |
| Paul Croft | Business Requirements Analyst | HSCIC |
| Harvey Goldstein | Academic expert on Data Linkage | UCL & University of Bristol |
| Wally Gowing | Pseudonymisation Advisor | Observer |
| Ian Herbert | Primary Health Care IT Specialist | BCS |
| Julia Hippisley-Cox  | Academic expert on Data Linkage | Nottingham University |
| Nicholas Oughtibridge | Lead – Code of Practice for Confidentiality | HSCIC |
| Chris Roebuck (Chair) | Benefits & Utilisation Director and Review Co-ordinator | HSCIC |
| Matt Spencer | Pseudo Review Project Manager | HSCIC |
| Martin Staples | Observer | NHSE |
| Tim Williams | Observer | Clinical Practice Research Data Link |
|  |  |  |
| **Apologies** |  |  |
| Kambiz Boomla | Observer | Confidentiality Advisory Group |
| Xanthe Hannah  | Observer | NHS England |
| 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 |
| Sean McPhail  |  | Public Health England |
| Dawn Monaghan | Observer | Information Commissioners’ Office |
| 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 |
| Marc Taylor | Observer | Confidentiality Advisory Group |
| James Wood | Head of Infrastructure Security | HSCIC |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **1.0** | **Welcome and Introductions** |
| 1.1 | The Chair welcomed everyone to Steering Group’s thirteenth meeting and thanked the group for its contributions over the past weeks.  |
| 1.2 | The Chair provided an update on HSCIC and his thoughts on the Review’s progress. |
|  | 1.2.1 | The Chair stated he is keen for the Review to complete its work rapidly, while enabling sufficient time for issues to be debated fully by the steering group to deliver the Review’s recommendations and final report. It would be helpful in today’s meeting to ensure the timeline required to complete the Review’s work is agreed. |
|  | 1.2.2 | The Chair updated members on Data Services for Commissioners (DSfC) programme. Data Services for Commissioners (DSfC) supports commissioners with their task of planning and managing services for patients by collecting information about patient activity efficiently and processing it in a way that protects patient confidentiality. NHS England and the HSCIC have created a joint programme team to transition the existing temporary arrangements to a sustainable future model.  |
|  | 1.2.3 | The Chair updated members on Patient Objections (PO), the very high importance this work is being assigned within the HSCIC and potentially large impact on users of data. The detailed policy and implementation plan is being developed by the HSCIC and DH in parallel with the Review. There are clearly implications for the Review around how Objections are applied, so current working assumptions will be used, as the detailed policy emerges.  |
|  | 1.2.4 | A member asked what is happening on Patient Objections in regards to Type 2 objections.  |
|  | 1.2.5 | A HSCIC participant stated for the Review the assumption should be, for any model being considered, that Patient Objections processes are in place. |
|  | 1.2.6 | The Chair stated PO are not just looking at pseudonymisation but also the context and purposes needed to be reflected in any of its processes. |
|  | 1.2.7 | Another steering group member stated it is important the Review proceeds based on current understanding of patient objections. |
|  | 1.2.8 | Another steering group member stated the issue is about risk and not about whether data should be identifiable or not. If people opt out and this becomes many then the data is degraded.  |
|  | 1.2.9 | The Chair stated that the HSCIC will consider in the future the high level analysis using anonymised data around the impact of objections on* users of the data for specific purposes
* on the demographics impacting the proportion of people objecting
 |
|  | 1.2.10 | Another steering group member stated it’s possible to obtain a practice level view of objections. The Chair responded that patient objections are expected to cover both primary and secondary care data. |
|  | 1.2.13 | The member commenting at 1.2.9 stated the difficulty is Researchers want the data and find it difficult to obtain and would need to understand specifically the impact of Objections on their study. |
| **Action No. 1** | The Chair asked the Review’s project manager to keep the steering group updated on the HSCIC’s plans for analysis of objections as they develop. |
|  | 1.2.14 | Another steering group member stated it is important that the HSCIC considers the impact of objections in light of the Mail newspaper headlines. The steering group member, who commented at 1.2.13, said the Review should look to understand the facts, which may not be represented by the headlines as it’s important to know what the issues are.  |
|  | 1.2.15 | The member who raised the original question on PO, at 1.2.4, agreed the Review shouldn’t be looking to define patient objections but should have an assumption that the PO processes need confirming. This needs completing as the issue is important for trust with some reports of GPs undertaking mass objections. A steering group member responded that such mass objections are not being done on empirical evidence. |
|  | 1.2.16 | The Chair in closing the debate on PO stated the implications of PO on the different pseudonymisation models need to be understood and should cover the following areas:* Operational issues
* Even if pseudo data were received that patient identifiable data would still possibly be required in order to fulfil any obligations to meet PO.
* Communication is needed around the benefits and safeguards of any process to manage POs and how this relates to other techniques to protect data, including pseudonymisation.

A steering group member, in agreeing with the above, offered the observation that any use of such data for the benefit of society has been lost in the heated debate. They stated that more should be done to communicate the benefits of not opting out. |
| **2.0** | **Review of Minutes & Actions** |
| 2.1 | **Minutes** |
|  | 2.1.1 | The Steering group minutes for May 21 were sent for review on 13th July. The Chair asked that in order to prioritise agenda item 4 (Assumptions & Key Findings papers) that they be reviewed, and for any comments, to be dealt with by email correspondence.Members agreed with the suggestion and accepted the prioritising of agenda items.  |
| 2.2 | **Action Log Review** |
|  | 2.2.1 | The Steering group Action Log was, similarly to the Review’s minutes, agreed to be reviewed and any comments dealt with by email correspondence. |
| 3.0 | **Pseudo @ Source sub-group update** |
|  | 3.1 | The sub-group Chair was unable to attend today’s meeting so HSCIC internal team members provided updates on the sub-groups individual deliverables. |
|  | 3.2 | **PS07 – Implications\_IG\_Transparency\_ Summary\_v0.5** This deliverable has had a number of reviews and has been sent to the Steering group for information only at this stage. There are a number of information governance (IG) areas involving patient identifiable data for SARS and S10 requests that need IG signoff before formal review by the sub-group and Steering group for ratification. A steering group member asked what different HSCIC sites and data they have.Another steering group responded that the HSCA caters for the reuse of existing data to facilitate SARS/S10 requests. The remit, in the act, is that effectiveness of data should be used for any such requests. The steering group Chair stated where data comes in under direction then HSCIC should be able to use identifiable data. Identifiable data can also be received under Section 251 approval, particularly where the HSCIC does not retain the data, for example the Trusted Data Linkage service that it offers to organisations such as CPRD.The steering group who commented about the HSCA further commented that Section 274 of the act has a list of HSCIC Functions for which the Secretary of State can give directions.The steering group member commenting above on other HSCIC sites stated as PS07 is not yet finalised then cannot agree to the current conclusions drawn in the paper.The HSCIC internal team member advised the group that the impact rating, in PS07, could be affected by any different understanding of the level of data available under a Pseudo @ Source model. |
|  | 3.3 | A wide ranging discussion was held on the status of deliverables, what work was required to be completed and how the group should reach agreement on the Review’s work including coming up with recommendations. A member expressed concern around proceeding to look at overall assumptions and findings in advance of all steering group papers being finalised, as well as the length of time taken to finalise some sub-group papers. Other views were expressed and the consensus was that the Steering Group would work on findings and recommendations in parallel with finalising some of the sub-group papers. However, there would be a strong focus on finalising sub-group papers and issuing these to members in a very timely way.The group discussed whether a meeting in August would be possible and practical for members. The consensus from the group was that final versions of deliverables should be completed during August and available for the group to review early September.It was also stated that there were a number of externally sourced papers referenced in the Review’s papers and that these should be circulated to the group for review.The Chair stated he was keen that the group should have as much opportunity as possible to focus on assumptions, key findings and on recommendations arising from the findings. The group agreed that a further, longer meeting is required to work through the document, checking for duplication. This longer meeting should also attempt to look at the cross cutting issues form the Review’s work. |
| **Action No. 2:** | The Review’s project manager to circulate any external sources referenced, in the Review’s deliverables and other papers, for steering group review.  |
|  | 3.4 | **PS04R – Report on pseudonymisation products in the market**This deliverable approved by the sub-group earlier in July was presented for ratification by the steering group. No comments were raised so the deliverable was ratified with a recommendation that use of the conclusions in the deliverable be included in the Review’s final report. |
|  | 3.5 | **PS11 – Relative security benefits and risks of different pseudo models** A sub-group member outlined what this deliverable was seeking to say and offered an observation that there were not many findings for the Hybrid model in the Key Findings so personally felt that PS11 is relevant for consideration of Hybrid Pseudo model.  |
|  | 3.6 | A steering group member raised a concern that Pseudonymisation and Encryption words are used interchangeably in the deliverable. Pseudonymisation is not encryption, they are two different approaches. If encryption is in place then why bother with pseudonymisation. A HSCIC member stated one of the Review’s assumptions is that under a pseudo at source model that HSCIC couldn’t re-identify the data.The steering group asked whether the HSCIC and the Review is ruling out probabilistic matching as the above assumptions implies no identifiable data would be available.The Chair stated the Review is looking at the pros, cons and barriers for all models, including pseudo @ source, across a number of topic areas.Another steering group member commented that maybe pseudonymised NHS number could be used for probabilistic matching.  |
|  | 3.7 | Another steering group member commenting on PS11 deliverable raised a concern that the volume of positive/negative comments recorded around Hybrid model was inconsistent with the summary at the end. In addition there is the challenge of whether listing all comments could be misleading if some were absolute showstoppers. A HSCIC member suggested PS11 should be presented in a different way and to include what security arrangements exist within HSCIC and potential impact of using pseudonymised data on such arrangements.The Chair suggested and the Personal Demographics Service (PDS) offers potential for probabilistic matching either to be undertaken locally or centrally. The steering group member, commenting in 3.6, asked if the Review has a view on the percentages on identifiers to be available in datasets.Another steering group member stated that pseudonymisation should be demonstrated to improve patient trust in the use of its data.The Chair stated in view of the comments raised stated the PS11 deliverable is not ratified and should return to the sub-group to consider the comments made. |
| **Action No. 3:** | That Pseudo @ Source sub-group revises PS11 deliverable to address comments raised by the steering group members. The revised version should then be returned to steering group for review and ratification at September’s meeting. |
| 4.0 | **Review of Assumptions v10a paper**  |
| 4.1 | The Review’s Assumptions paper had been developed by the HSCIC internal team and presented to sub-group members at a workshop in June and followed further two weeks review by sub-group members including those that didn’t attend the June workshop event.The paper is therefore presented for review by the steering group.  |
|  | 4.1.1 | The Review’s Chair opened up discussions on the paper by stating that it is important to gain clarity and a common view around whether to use single or multiple pseudo keys under the different pseudonymisation models. 3 different scenarios were presented. Scenario 1) For disseminations from the HSCIC, each extract for a single purpose for a single customer is pseudonymised using a separate key.Scenario 2) For inbound flows to the HSCIC, if a pseudo at source model were adopted, where there is a bespoke extraction and flow of data to the HSCIC for each use of data (such as the GPES model), each extraction could be pseudonymised at source using a separate key. Scenario 3) For data that the HSCIC collects through time and builds a linked dataset for multiple purposes to benefit health and social care (such as HES), the Chair stated that he could only see how this could work with a single pseudonymisation key for the inbound flow..  |
|  | 4.1.2 | A steering group member disputed the assertion made in 7.1.1 as instances of multiple keys have been used in a pseudo @ source model. The Chair in responding stated that the different scenarios needed to be viewed separately – this may work for scenario 2) or local data flows. However, in the final scenario, he could not see how the HSCIC receiving multiple keys would decrease the risk as the HSCIC would link the dataset together using the same key or else lose the ability to analyse patient outcomes that are reliant on tracking the same patient through time. |
|  | 4.1.3 | A steering group member stated that each provider could submit a very large number of fields containing data pseudonymised in different keys to the HSCIC, which would still enable the HSCIC to link the data, but decrease the risks associated with using a single key.An HSCIC participant gave his view that this would in fact increase the risk as there would be a number of keys that could be broken instead of just one.The view of the majority of the members of the group was that multiple keys would not work under scenario 3) in a pseudonymisation at source model. This would be provisionally reflected in the assumptions. However, one steering group member did not agree with this and would arrange a call with the chair of the Pseudo @ Source sub-group specifically to look at the proposal for whether each provider submitting a large number of fields containing data submitted to different keys could increase security. If merit were found in this proposal, the assumption would be reconsidered.The Review’s Chair agreed with this suggestion. |
| **Action No. 4:** | That the Pseudo @ Source sub-group Chair convenes a conference call meeting to consider whether each provider submitting multiple fields of the same data pseudonymised to different keys would enable linkage to occur and enable any security benefits compared with a same key model. |
|  | 4.1.4 | Assumption 4 A steering group member stated as patient objections policy was unclear then this assumption is made out of scope for the review. The Review’s Chair suggested that patient objections is included as a reference as the Review needs to be aware of objections policy as it emerges as it has strong implications for the pseudonymisation approach.  |
|  | 4.1.5  | Assumption 3This is to be reworded for clarity and to include the word ‘re-identification’.  |
|  | 4.1.6 | Assumptions 5 – 8 Data ControllersThese should be removed as duplication of points made and merged into Assumption 9 – 10.  |
|  | 4.1.7 | Assumption 20This to be amended to make it clearer what is being said about pseudo @ source. A steering group member suggested that definitions, used by the review, should be included in Assumptions paper.  |
|  | 4.1.8 | Assumption 22A steering member suggested this should be amended to say ‘point of pseudo @ source means that you cannot re-identify the data’.Also the member stated that encryption is reversible but pseudo @ source implies no reversibility. The steering group who commented on encryption stated that irreversibility would make it difficult to get the data. If the data was reversible then everyone can get the data. Another steering group member commented that if reversible then everyone would have the same key, same algorithm etc.The Review’s Chair suggested that a clear assumption on whether pseudonymisation under the different models would be reversible is needed. Those that cite the fact that pseudonymisation should be reversible often cite actions following risk stratification. Where Risk Stratification has occurred using pseudonymised data, the patients with health risks would need to be identified to enable their clinicians to take appropriate action. After some discussion, it was agreed that the pseudonymisation algorithms applied under a pseudonymisation at source model should be irreversible, otherwise any security benefits from this approach would not be realised. The organisations performing the pseudonymisation could retain a look up of identifiable to pseudonymised data. |
|  | 4.1.9 | A steering group member suggested there needed to be three other assumptions in the paper. These are:* Any recommendations from the Review’s report will need to be costed.
* Public and professionals expect HSCIC to be aware of and make best use of privacy enhancing technology.
* Review’s report will need to be reviewed on a regular basis and take account of changing requirements and advances in technology.
 |
|  | 4.1.10 | The Review’s Chair stated that HSCIC business as usual functions will be responsible for managing any recommendations resulting from the Review. |
|  | 4.1.11 | Assumptions 24/25 A steering group member had an issue with these two assumptions for postcodes, stating that there are two scenarios involved; where post code has changed and where it hasn’t changed. A HSCIC member responded that individual items would be pseudonymised and not partial items. This would allow these items to be used in linkages.The Review’s Chair stated that a new assumption should be added that *individual data items would be pseudonymised’*. The steering group agreed with this suggestion.The steering group member from CPRD left the meeting at this point saying he would provide further comments by email correspondence.  |
| 5.0 | **Review of Key Findings v10 paper**  |
|  | 5.1 | A steering group member stated a general point that there were not many findings for the Hybrid model.There was also further discussion on evidence used to support findings need to be checked to ensure they are correct and traceable. |
|  | 5.2 | Key Finding 1An HSCIC participant advised the context of this finding should be amended to state ‘majority of national data flows that the HSCIC processes do not include patient name or address (with the exception of postcode). Where name and address is collected this is often to support direct care purposes such as to enable patients to be invited for national screening’.  |
|  | 5.3 | Key Finding 11A steering group member thought this was the wrong way round and should be reworded. |
|  | 5.4 | Key Finding 12A steering group member stated that S251 has to be used when disseminating data, except where a legal basis exists, so the Finding needs rewording to reflect this.For inbound data there has to be a direction to allow the data to be collected. |
|  | 5.5 | A steering group member commented that the documentation, for example Finding 12, implies Chinese walls are implied. Is this really appropriate.Another steering group member asked if the Review has any view on what the Review’s recommendations should be. The Review’s Chair stated that recommendations should reflect the views of steering group members, aiming to reach consensus where possible.A further steering group member started recommendations from the Review should be clear but state where different views exist and what the issues are. |
|  | 5.6 | The steering group in reviewing the Key Findings paper discussed potential recommendations to be considered at the group’s September meetings. A steering group member outlined the following perspective:* Pseudo @ Source model is too difficult to adopt for a number of existing HSCIC datasets
* Central model – thus needs to apply to a number of existing HSCIC datasets
* The HSCIC should look at the Hybrid model specifically for new data flows and particularly for any flows involving GP data.
 |
|  | 5.7 | A HSCIC participant asked if the hybrid model, whereby the HSCIC held and used the same pseudonymisation key as is held and used locally and applied to data submitted under pseudo at source would provide any greater security compared with the HSCIC pseudonymising centrally. The steering group member, commenting at 5.6, responded that the HSCIC in adopting Hybrid model should be split into two separate entities to ensure that risks of re-identification are reduced as much as possible. An HSCIC participant advised that this would not be practical, so there was discussion of establishing a clear separation between staff who would come into contact with patient identifiers and those who wouldn’t.The Review’s Chair stated that of re-identification is to be available then the level of risk needs to be clear and perhaps the presumption on data flowing should start at Anonymisation should be the starting point of any assessment. A steering group member commented that technology has moved on – Anonymisation should be the model’s starting point. The member advised he was writing a paper on this and its application for linkages and would be happy to make the paper available to the Review.The Chair stated this would be beneficial for the Review and HSCIC to be aware of. |
|  | 5.8 | A steering group member commenting in general, and in response to above comments at 5.6, felt that different people and views should be made clear in the review’s work. The public has a high level of trust with GPs but a low level of trust with government organisations.Whatever the process is the issue is how to make the public comfortable about how their data is being used. The member felt that the Hybrid model could be the best model but any recommendations on the use of this model would need to be specific to address the GP and public concerns.A steering group member summarised the discussion that for datasets currently collected by the HSCIC in identifiable form and pseudonymised centrally, this should continue. New data collections should be collected in the form with the lowest risk of patient identification that enables the full benefits to health and care to be realised and this would start with exploring completely anonymised data, followed by data pseudonymised at source, followed by data pseudonymised centrally and that this should be considered on a dataset-specific basis as part of a Privacy Impact Assessment. There appeared to be some agreement from other members around this, although one member expressed the view that the presumption should start with the HSCIC collecting identifiable data and only if there are specific issues around this, look to collect the data in less identifiable form. |
| **6.0** | **AOB**No items were raised in AOB and the Review’s Chair thanked everyone for working through the substantive issues outlined in the Assumptions and Key Findings papers. The Chair further stated he looked forward to the group coming up with a set of recommendation at its September meetings to allow the Review’s report to be produced. |
| **7.0** | **Next Meeting – August and September dates to be advised.** |