medConfidential submission to the HES PIA consultation

medConfidential seeks to ensure that every flow of data into, across, and out of the NHS and care system is *consensual, safe and transparent*.

medConfidential was part of the consultation group panel involved when the PIA was originally written in 2014, before being buried for a couple of years. Comments below reflect the world as it is now. The PIA requires an update for 2016 in a post-Caldicott Review world.

While there are claims to reform, "Market Access" is the new euphemism justifying access to linked records on 1.5 billion hospital treatments, for that company's commercial customers - and their access to data was recently renewed.

This consultation takes place in the context of the public consultation into the Caldicott Review, and many of the issues around the PIA are tied up in the forward looking response to that consultation. A copy will be available from www.medconfidential.org/news when submitted, which may be shortly after this consultation deadline closes.

Response to consultation questions

4: The areas covered by the review are robust, but there are serious omissions.

5: The Review is extensive but there are omissions.

6: The actions taken by HSCIC are insufficient.

7 - Continues to end :

There is no reason why dissemination of HES data is not done in a manner which is consensual, safe, and transparent. As of July 2016, it is arguably none of those things, and the PIA is marginal in those cases, although it does reflect an improvement on the situation prior to the Partridge Review and the PIA being conducted.

It is of concern that the PIA believes dissemination of HES continues to be possible without a clear legal basis.[1] This does not necessarily accord with a plain reading of the Data Protection Act and so should be rectified. The use of an "implied" Direction is institutionally necessary otherwise every HES release would be in breach of Dissemination restrictions from the Health and Social Care Act 2012.[2]

[1] The current basis is described in the PIA as "an implied and unpublished pre-2013 Direction".
[2] s260-262 being relevant areas, primarily s261.

Billions of records

The PIA describes HES as where "Many millions of records are stored". The HSCIC website says it "contains around 1 billion records on patients".[3] The difference of 3 orders of magnitude should be corrected in the PIA, especially as commercial entities were touting the "1 billion" figure in 2014 when the PIA was written.

Disregarding of patient dissent

We will cover this in further detail in respect of the Caldicott Review, but while patients have expressed dissent from their data leaving the HSCIC for purposes beyond direct care, their data is still included within the releases.

We agree with the intent of the PIA that "However the public needs to understand what is done with information created from their hospital visits, and the potential privacy risks, so they can make an informed choice about whether they want to continue to support its use in indirect care"[4]

That the PIA is entirely silent on this topic, with the possible exception of the first bullet point in the preface, shows that the document has dated.

Lack of awareness of damage of impacts of such data

The PIA describes as an action taken since 2014 - "Separation of sensitive data items from other HES data", and assesses the risk that "An individual is re-identified from HES data used to produce an HSCIC publication" "is possible but unlikely to happen". While some sensitive data may be better protected than it was in 2014, event dates remain entirely unprotected - but are given in the glossary as an example of an "indirect identifier".

The first Caldicott Report from 1997, Annex 7, notes that event dates are identifiable, as the PIA later says:

> "A wide range of clinical and administrative data about the hospital care episode are included in the HES record, such as event dates, and procedure/diagnosis codes, and these can also sometimes be used to reveal the identity of a person. Different people may judge the sensitivity of health conditions differently, but hospital records can contain information that many people will feel is especially sensitive, such as a person's HIV status, sexually-transmitted disease, mental health treatment, or a termination of pregnancy. Records containing such especially sensitive data items are de-identified in the Secondary Uses Service"

[3] http://digital.nhs.uk/hesdid
[4] Introduction. PIA.

The footnote in the PIA notes that *"Sir Nick Partridge makes the point that clinical data in the HES record are codes and numbers and not "obviously personal descriptions of either patient or illness", implying that as a result they are less revealing."* does not necessarily seem to be grounded in fact. The fact that a condition may be coded as X71Do is only a quick google search away from translation.[5] It provides no protection for an individual at all. Either way, the implication should removed and replaced by the actual intent of the Partridge Review - and checked with Sir Nick for clarity.

HSCIC ignores opt outs

The PIA says:

> "The type 2 objection is the more significant objection in relation to HES. If a patient records a type 2 objection, the HSCIC is obliged to cease releasing HES (and other) data that identifies the patient which it holds about that person. It provides another layer of HES governance in addition to those set out in section 4.6.1, 4.6.2 and 4.6.3."

While this is what *should* happen, and what the public were told would happen in 2014, it has been confirmed by HSCIC that this is not what happens for dissemination of HES.

The PIA says what should happen, but the opt out has been reinterpreted in the interim.[6] This is not noted within the PIA.

This issue is currently the subject of a complaint to the ICO by medConfidential.

Are safe settings covered?

The discussion of safe settings is perfunctory, incomplete, and insufficient.

It is currently unclear why commercial entities keep cutting corners with patient data.

While it can not be done publicly, HSCIC may wish to compare the classes of organisations which opposed the data laboratories, and the HSCIC Data Sharing Audit results for that class of organisations.

The arguments given against data laboratories are self-serving for those few respondents who argue in favour of data users being able to continue to use data in an unsupervised and unauditable manner: Turkeys don't vote for Christmas.

[5] In this case, "shot with crossbow"

[6] See letters between medConfidential and Prof Martin Severs in June/July 2016.

No forward view.

The PIA does not look forward, for reasons at the time. While this may have been appropriate in 2014, it is now dated, and there should be a roadmap for an updated PIA based on events since it was written.

We support the decision to publish the 2014 PIA now, to avoid the additional delays from additional work required to update it. The decision not to publish in 2014 was taken by HSCIC, when the length of the delay and future required revisions by "HSCIC Executives"[7] were not forseeable.

Anonymisation

Whether HES data are anonymous is covered by the UK Anonymisation Network, which says quite clearly on page 16 (emphasis added):

> *"Anonymisation – refers to a process of ensuring that the risk of somebody being identified in the data is negligible.* ***This invariably involves doing more than simply de-identifying the data, and often requires that data be further altered or masked in some way in order to prevent statistical linkage.***
>
> *We can highlight further* ***the difference between anonymisation and de-identification (including pseudonymisation)*** *by considering how re-identification might occur:*
>
> 1. *Directly from those data.*
> 2. *Indirectly from those data and other information which is in the possession, or is likely to come into the possession, of someone who has access to the data.28*
>
> ***The process of de-identification addresses no more than the first, i.e. the risk of identification arising directly from data. The process of anonymisation, on the other hand, should address both 1 and 2.*** *Thus the purpose of anonymisation is to make re-identification difficult both directly and indirectly. In de-identification – because one is only removing direct identifiers – the process is unlikely to affect the risk of indirect re-identification from data in combination with other data."*

As such, claims in the PIA that the ongoing release of data is compliant with the ICO's Anonymisation Code are deeply flawed. To quote the PIA description of the risk to patients of re-identification from HES: "This may happen in future" (risk 7). As such, HES can not be considered anonymised, and according to the PIA, is not - it is only de-identified.

Defining "anonymisation", the glossary in the PIA says "Also commonly referred to as "de-identification"". This definition should be replaced with the definition from the UK

[7] Document Management Revision History, for version 1.2

Anonymisation Network guidance,[8] and a check conducted of related terms where the HSCIC definition is not identical to the UKAN definition, and all such definitions replaced with UKAN or Office of National Statistics definitions.

Other minor comments

- The PIA says HSCIC "May *require* health and social care organisations to provide data;" - it should be noted that HSCIC may "*request or require*" organisations to provide data.
- It is a serious omission that the (then) Mental Health Minimum Data Set (MHMDS) was excluded from the PIA, given the structure of the data and linkages. Given that it was excluded, it should be mentioned in some fashion
- Given risk 7 is categorised as "this may happen in future', does the HSCIC (and parent department) have a contingency plan for responding to privacy impact of the unintentional unrestricted publication of a HES dataset? Independent of any PIA assessment, given the number that are disseminated, it is highly optimistic to believe that there will never be an error which leaks 1.5 billion linked health events.
- Regarding SUS, the PIA says "Since *2014*, the BT contract was terminated and this processing is now done "in house" by the HSCIC." Should that date be *2015*?
- Information on risk 13 is contradictory - the "past issue" field says "no", whereas the likelihood field says "this has happened". The past issue field should be amended in line with past events.

medConfidential

August 2016

[8] http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf