ASH response -- A substantive proposal on Protecting Health and Care data based on a designed process incorporating safe settings.

The current processes to be replaced under this consultation do not exist by design. Rather they are the result of an accidental arrangement of the previous permissions granted to the past PCT model structure, forced to fit in emergency onto the new CCG structure, as a fix for commissioning purposes[1]. The patch, a blanket section 251 approval was issued in a bureaucratic panic on the 8th April 2013, after all existing permissions lapsed on 1st April 2013. This seems suboptimal. The level of complexity, conflation and confusion for this consultation is a partial result of the mixing of many multiple purposes and does not significantly improve on the status quo.

This consultation presents an opportunity to map structure, process and permissions appropriately to achieve the desired outcome: an effective model for a safe, consensual and transparent NHS data-management model, fit for the 21st Century. However, the choice of a tight time scale and unclear basis does not make this certain.

Multiple distinct areas of governance are required to acceptably operate a Safe Haven in any of the suggested areas is required. Carefully done, by deconflating stated purposes and methods, the outputs of that process may still be referred to collectively as an “accredited safe haven”, although the consultation provides no clear evidence to justify the assumptions and current silence on governance.

From the current consultation, we anticipate dysfunction and excessive data transfers from unclear requirements, and the sweeping of issues under the carpet for rediscovery via a future catastrophe with the public’s health records around commissioning (which is the extent of this consultation). This lack of clarity has caused confusion in an extremely tight time scale. As the starting point for a trustworthy system, this does not seem entirely wise.

We see no reason that data used for commissioning purposes, for invoice reconciliation, or for risk stratification, is exempt from consideration under the principles of balancing safe, consensual and transparent.

MedConfidential has previously introduced papers on the concept and implementation of a Health Research Remote Data Laboratory (HRRDL), to meet the needs of bona fide research to access individual level patient records in a safe, consensual and transparent manner.

A safe setting is a physical venue where (usually remote) data can be accessed under tightly controlled and audited conditions. Restrictions are placed on who and what enters the room,

[1] Risk stratification as indirect care, is not included as approved purposes for using identifiable data, and as such did not receive Section 251 approval from CAG for these purposes.

what they do when in there, and what they can take out. This allows for research to be conducted on individual level records which have minimal protections (which, for health data, has other problems). They were previously discussed for legitimate research, along existing models. This paper takes the proposal further.

While it is inappropriate for commissioners to have access to that Research Remote Data Laboratory, a similar construct could exist for each of the distinct functions of commissioning. The HRRDL proposed at HSCIC , is already the first of a class of facilities, with the Farr Institutes building their own safe settings to similar (or, likely, higher) standards than HSCIC. Those currently planned HRRDL-class facilities all hold the data requirement constant, but there are other uses of data than Research. The HRRDL itself, as proposed and publicly accepted by HSCIC, would be inappropriate for the purposes described in the consultation, but a HRRDL-style facility can serve other purposes and defined needs.

In all cases, staff remain with organisations, and clear governance responsibilities apply equivalent to the HRRDL. Clear lines of responsibility, governance, and understanding are vital.

In line with HRRDL, any facility that has sensitive data should be subject to similar audit and access revocation as researchers -- which includes single strike for egregious institutional/individual breaches. While researchers are capable of being held to a higher standard, those standards are appropriate to the data, and that must also be true for all other data access methods.

With the "research safe setting", and here two distinct commissioning safe settings, the need for confirmed and experienced leadership at HSCIC on this topic, along with an appropriately constituted broad high level advisory group (along the lines of other laboratories and data holding departments) would seem to now be necessary. If data laboratories, in which users can run their custom scripts and applications, become the norm for the nation’s individual level health records

---

What is the “Remote Data Laboratory” design for Invoice Reconciliation?

The key features of the HRRDL are the governance and confidence arrangements surrounding it. Those are extremely tight for HRRDL as a result of the detailed nature of the data. More specifically, the design of the HRRDL is wholly based on the needs of the research to be conducted, and the sensitivity of the data. But that leads to the first unanswered question of this consultation:

What data is required for Invoice Reconciliation?

In the same way that the HRRDL is a facility that recognises the needs of research, and as an improvement on sending 1bn health events to those who ask, a remote data facility for Invoice Reconciliation would improve on the blanket section 251 currently in place. HRRDL did not change the data available, but who could access it (after public outcry). The Invoice Reconciliation process needs to change what data is available.

We would expect that a custom, detailed, extract of aggregated data would be able to support the vast majority of routine invoices, which is generally 3 classes of questions:

1. Have we paid an invoice for this care before? (yes/no)
2. Should we pay this invoice? (ie is it care we commissioned?)
3. Did we pay this invoice now? (so it's paid only once, to answer (1))

There will likely be other questions, and we are not in a position to solely determine what the minimum dataset is required to answer those questions, but are confident that there is one. **We call on the relevant bodies to issue a Memorandum of Understanding of purposes and data required to meet those**. Due to the implementation of NHS England's care.data objection process (9Nu0/9Nu4), this process can not rely on individual level data. Aggregated data, extracted from the GP provider using GPES (and subject to IAG) does not engage 9Nu0.

With a list of questions that need answering, in what context, and the minimum dataset required, a set of security criteria can be applied under the HRRDL class facility headings. We would expect that any queries relating to invoices will be for well defined time periods, and while they will involve small counts, they do not need medical histories outside of particular time periods. As such, those questions are definable and automatable.

With N3, data may remain in a high security data facility at HSCIC, and custom queries be run on it from relevant locations over a secure network. Those queries being predefined, logged, and run on a custom data extract to provide exactly the information needed for the purpose.

Properly designed, the queries of the system should confirm or refute information on the invoice, and provide no additional data. This will affect the level of physical security required.

With a clear data model, and deconflated purposes, such decisions should be comparatively simple.

What are the "Remote Data Laboratory" designs for Risk Stratification?

Risk stratification has multiple aspects, conducted in roughly two categories. The first is the creation of models, treatment effects - the design of strata; the second is the use of models, which is far more prevalent than their creation. Neither of are direct care, and attempting to use identifiable data with a direct care justification is questionable for the reasons NHS England explain[2].

Once again, deconflating purposes to separate out the various data needs for the different users, and the differing purposes, and hence the required different data products, allows for much clearer data flows, audit and consent. This can be safe, consensual and transparent.

Generating Risk stratification models

The generation of models, the testing of hypotheses to form them, the development of publications justifying them, is research. The process for a model developed by MRC is almost indistinguishable from a model for the Mayor of London's office. The models should be subject to publication, peer-review etc. As such, that development should take place within the main research HRRDL, on the rich, detailed patient data required to distinguish (or at least, potentially investigate) between correlation and causality.

As part of creating a risk stratification model within HRRDL, part of the requirement can be to design the data specification required for the model to be run. Those outputs should be aggregated, and area (CCG) specific, and can then be made approved by HSCIC before being copied to a different facility for access by users of a risk stratification model.

Where a model needs to include data from other sources, we would elect that HSCIC's safe setting, when operational, would be welcomed into the club of equivalent government safe settings (ONS, MoJ, HMRC, DWP) with cross-departmental acceptance of standards, such that data of lower sensitivity could be available in the HSCIC HRRDL. This is established process, and while it can not be confirmed before such a laboratory is operational, it would be a cause for public concern if other departments were not satisfied by HSCIC's laboratory.

Using Risk stratification models

While larger organisations will generate their own models, smaller organisations will primarily apply models used elsewhere to their own areas. An organisation may even draw these

[2] http://www.england.nhs.uk/wp-content/uploads/2014/03/130614-s251conditions-0314.pdf

distinctions internally, with a research and a commissioning function which interact but are separate.

Models developed for stratification can include the specification of the data that is used to drive them, is will be aggregated statistical outputs, with up to some number of dimensions of data tabulated. Those tables will include small numbers, and so can not be used in unprotected environments. However, they are significantly less disclosive than individual patient records.

As a result, there will still need to be a controlled environment for an organisation to look at the risk stratification data that they are interested in, consider different iterations and boundaries, based on the specification of the model using constraints designed by the model's designer. No model allows an unlimited amount of variation, which is why different models are required for different purposes.

However, models can be used within a HRRDL type facility, but which is appropriate to the data input to the model (ie the custom data product designed by the model creator), which will be of much lower sensitivity than raw patient data. The primary output of the application of a model to an area is not a list of patients, but a set of types and thresholds of treatments and counts, to enter into other documents.

Where appropriate, after agreement, those types and thresholds can then be run by GPs against their full patient dataset, which will include individuals who have opted out of secondary uses for data sharing. They are not included in the generation of the model, but by careful and correct data design.

Unanswered questions needing clarity for implementation of the above

1) What is the definition of 'commissioning purposes'? We would expect that these definitions remain in line with Calidcott2, however, this is unclear. When agreeing what data is required, a prerequisite is a shared coherent understanding of what will be done and for what purposes. This understanding has not been conveyed, and does not appear to be consistent across involved organisations.

Do DH believe "commissioning purposes" covers any activity undertaken by a commissioner, or are there some limits? Is there a statutory definition? Is there a comprehensive list of purposes, or at least activities from which purposes may be inferred? Can DH or NHS England provide specific examples of activities and/or purposes which are 'boundary cases'?

If there is not a defined list of commissioning purposes, then there can not be a defined list of data and data products needed. As such, one should be produced (and iterated as necessary).

2) How will the creation of an ASH affect other data flows? Will CPRD and the other research databases, e.g. QResearch, Thin, ResearchOne, be required to meet such standards? (para 11) How will CPRD continue into the future? How will the need be met for additional detail in CRPD (e.g. free text) that has been excluded from care.data via public statements from NHS England?

3) Is para 19 referring to opt out on a statutory basis? If so, how will this apply if Patient Objections Management is only being done via Directions to HSCIC? Opt out should be done via Regulations to make this completely clear. The nature of NHS England's implementation of 9Nu0/9Nu4 has made various of these decisions preordained. The assurances given to patients may not be weakened.

4) Is the list of "broad purposes" in para 26 complete? Will ASHs be the only bodies able to provide, e.g. risk stratification? Conversely, will (commercial) providers performing risk stratification need to utilise an ASH? How does this interact with Caldicott2? Bullet point 4, para 28 says "subcontracting of the processing work would not be permitted" so ASHs clearly won't be able to subcontract the processing to commercial providers.

5) Will an ASH be a legal entity in its own right, able to form contracts and agreements, etc. or are they akin to DMICs (para 21)? To whom will ASHs be accountable? This question, together with 4, suggests the relationship and legal status is not entirely clear. The above provides high levels of clarity, using the terminology and specification of HRRDL-class facility. If this is not adopted, there will be a range of questions which need answers.

Sam Smith
medConfidential
July 2014