Reviewer's report

Title: The SAIL Databank: building a national architecture for e-health research and evaluation

Version: 1 Date: 7 April 2009

Reviewer: Mark McGilchrist

Reviewer's report:

(1) [Major] The paper describes the creation of a national healthcare databank that is compliant with UK legislation and recommended approaches to information governance. While the description is generally adequate I feel an opportunity has been missed to offer the rationale for the choices that have been made. Others wishing to pursue similar ends are seeking guidance. In particular, what additional principles have been employed to guide the solution? A central architecture has been chosen but no analysis is given of why this is the correct choice in this situation. That other architectures are possible is indicated in the section on future work. Architecture is a driver for the costs, institutional arrangements, types of risk that result (present and future), scalability and external discipline. I feel more discussion of this is merited.

(2) [Major] The mechanisms employed focus almost exclusively on a particular type of risk - the maintenance of patient anonymity - without an adequate discussion of what constitutes anonymous data and who should determine this. The paper seems to suggest that disclosure risks are mostly with the analyst community, but there are many other personnel within HIRU in contact with parts of the data, e.g. database administrators and other IT personnel. What is the rationale for placing trust in certain types of staff? Many other types of risk are possible relating to control, ownership and transparency, but these are not discussed in detail. From the receipt of data from DPOs through to publication HIRU has a high level of control and perhaps even veto. While I acknowledge the presence of the IGRP what are the consequences of this control for the external research community and the data providers given that HIRU will also use the data for analysis, and help vet the requests of others? Are there circumstances in which HIRU might reasonably argue in favour of its own analysts for certain types of requests? Does the review panel have the strength to deny requests it feels are not merited, even from HIRU? Does the panel draw on outside advice for this? Do the original DPOs have a say? Are the deliberations of the panel public knowledge? Some discussion on transparency and the extent of shared control of the overall process would be useful.

(3) [Major] A discussion of what motivated DPOs to participate would also be helpful. Does the mechanism actively assist DPOs to improve their function and the coverage, completeness and quality of their data, or are the beneficiaries elsewhere? Do DPOs have concerns about loss of ownership and control of their data?
The authors aspire to a much larger facility presumably involving non-NHS person-specific datasets. A more comprehensive discussion of barriers to scalability is therefore necessary. Beyond the danger of increasing workload on this central facility from the perspective of formats, coding, semantics, quality, and completeness of data as acknowledged in the text, is there a recognised limit as to the kind and size of organisation that would contribute data to HIRU? Would HSW be an acceptable linker to non public-sector bodies? It would be nice to know what boundaries are contemplated for this facility.

It would also be interesting to know the balance between data requests where HIRU plays an active part in the study/analysis and where it acts as a disinterested party in the provision of data views to external researchers. The issue of HIRU as an interested or disinterested party is important when considering the maintenance of discipline in its processes and future developments. While an independent audit has been performed, how often will audits like this take place? In considering future changes to the facility have any limits been placed on this now by other parties as a safeguard?

My reading of the paper is that no data from the databank (view) leaves HIRU, but I would like to be clearer about what actually can leave HIRU. Remote data access is mentioned as a future development but presumably at present external researchers must travel to HIRU to make use of the data. It would be useful to hear of feedback from external researchers with this approach.

In summary, I would like to have seen a clearer statement of why the databank and its procedures look the way they do with more space given to discussing control, ownership, transparency and conflicts of interest. However, the text is fairly clear about what SAIL is offering to those looking to do studies.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests