Author’s response to reviews

**Title:** An architecture and method of operation for improving the protection of privacy and confidentiality in disease registers.

**Authors:**

Dr Tim Churches (tchur@doh.health.nsw.gov.au)

**Version:** 4  **Date:** 3 Jan 2003

PDF covering letter
The author thanks both reviewers for their prompt, detailed and very helpful comments. Specific responses to the issues raised by each reviewer follow.

Reviewer 1 (Dr Jules Berman):

1. a) Reviewer comment: In the abstract, the author uses an abbreviation for Disease Register (DR), but abandons the abbreviation throughout the ms. He may as well drop it from the abstract.
   b) Author response: The abbreviation “DR” for Disease Register has been removed from the abstract.

2. a) Reviewer comment: He refers to his methodology as something to be used for population-based registers. It seems to me that the proposal is for registers that hold data extracted from patient records. There's nothing about the architecture that requires it to be population-based. The author even describes its use for tissue-registries and biological databases. I'm sure the author has a special interest in population-based studies and may have designed the methodology with this in mind, but there doesn't seem to be any reason to think of the protocol as something that is only suited to population-based registries. So I would suggest that the qualifier "population-based" be removed from the title.
   b) Author response: The qualifier “population-based” has been removed from the title, and the first paragraph of the “Background” section on page 3 has been expanded to explain that disease registers do not necessarily need to be “population-based”.

3. a) Reviewer comment: I was well into the paper before I realized that the architecture has never been implemented and that the supporting agencies don't actually exist (identifier translation agency and population register). I think the author should change the wording of the TITLE to indicate that the architecture is a proposal (A proposed architecture....) and not a done deal. The abstract should also clearly indicate that the proposal has not been implemented.
   b) Author response: The title has been altered to make it clear that the paper is proposing an architecture and method of operation, not describing its actual use. The abstract already makes this clear. Elsewhere in the paper, “proposed” has been substituted for “described” to further avoid confusion. As noted in the manuscript, the author is involved in a project to develop a freely available probabilistic record linkage engine, which is the most challenging part of the infrastructure from a technical perspective. It is anticipated that work will commence in the near future on freely available implementations of the other components, enabling individual organisations to adopt the proposed architecture for their “internal” databases.
Implementation on a society wide basis requires of course a great deal of necessarily lengthy community consultation. Two sentences have been added to the end of the manuscript (page 19) to reflect this reality.

4.  
a) Reviewer comment: In my mind, the chief problem with the proposal is that each party puts their trust in other parties. On page 8 (section 4), the author writes that the Population Register uses probabilistic record linkage or other methods to find the match. Otherwise it assigns a newly-created id. So you're trusting this agency to do the job right. When errors occur, patients end up with multiple id numbers, and the database becomes corrupted (unique identifiers that are neither unique nor identifiers). This is actually a real problem for hospital information systems and has been well-documented. At least in a hospital, the staff creates their own problems that they themselves need to somehow correct. But when someone else creates problems for you, you're really in trouble. So when you're proposing a multi-agency architecture, you need to have a way for the participants in the any agency to detect errors introduced by the other agencies and to remedy these problems when they occur. Extending this inter-dependence issue, what would happen, in this system, if there was a successful cyber-attack on an agency that resulted on a re-assignment of id numbers for the different patients, so that a subset of the patients had wrong id numbers? How would the system recover? The author seems to concentrate on discussing confidentiality breaches. But hackers may destroy a system like this without ever breaching anyone's confidentiality. The author should discuss how to deal with this issue.

b) Author response: A paragraph has been added on page 18 to the end of the section headed “Weaknesses and possible solutions” discussing the issue of the Population Register doing its matching job correctly, and proposing a possible solution. Two paragraphs discussing sabotage have been added on page 17 to the end of the section headed “Risk and hazard assessment”.

5.  
a) Reviewer comment: In the very last paragraph, the author raises an interesting issue. "To many readers, it might seem unlikely that the level of community consensus necessary for the commissioning of such a system could ever be achieved." Yes, but he answers the objection quite well indicating that the problem requires proposals that can be discussed by the community. That's the real strength of this paper. It serves as a focus for thought and discussion. This should be stressed in the abstract portion of the paper as well.

b) Author response: A sentence asserting the importance of widespread scrutiny of such proposals has been added to the abstract.
Reviewer 2 (Dr Greg E Simon):

1. a) Reviewer comment: The author might more clearly describe which of the various roles entities already exist, which roles might be served by existing organizations, and which would require creation of new organizations or entities. For example: Health care entities and disease-specific registries already exist while population registers and identifier translation agencies do not. Could these two roles be filled by existing entities? Why or why not?
   b) Author response: Three paragraphs discussing these issues have been added to the end of the section titled “Legal, legislative and governance considerations” on page 14.

2. a) Reviewer comment: The author might clarify how the proposed structure would relate to existing ethical review committees or institutional review boards. Would the proposed privacy and confidentiality protection committees replace or supplement existing ethical review processes?
   b) Author response: Three new sentences clarifying these relationships have been added to the 3rd paragraph of the section headed “Linking Data for Research Purposes” on page 9.

3. a) Reviewer comment: The author might discuss how patients might express preferences and provide consent at the time original data collection. For example: would patients be allowed the option of “opting out” of all research use?
   b) Author response: A paragraph on page 10 at the end of the section headed “Linking Data for Research Purposes” describes how patient preferences might be captured, analogous to the way in which some commercial organisations now capture privacy preferences in a standardised format (P3P). A reference to a just-published paper by Caulfield et al. which suggests an “authorisation model” as an alternative to one-time consent for DNA databanks has been added on page 10 (please see http://www.biomedcentral.com/1472-6939/4/1 for the paper by Caulfield et al.).

4. a) Reviewer comment: How would such an infrastructure be funded? Should research organizations be expected to contribute to the support and maintenance of such a system. Since initial fixed costs would be relatively high compared to ongoing maintenance or marginal costs, what funding mechanism would support initial development?
   b) Author response: two paragraphs have been added on page 14 to the end of the section headed “Legal, legislative and governance considerations” discussing some funding options. It is quite difficult to be specific due to the wide range of funding arrangements in place in different countries. The initial centralised funding described would be difficult, but certainly not impossible, to achieve in Australia.

Other changes to manuscript:
A quote from one of Ross Anderson's papers has been added on page 3 following the 4th paragraph of the “Background” section.

An extra sentence was added on page 13 to the 2nd paragraph in the section “Distinguishing features of the system”.

Added a paragraph on page 12 to the section headed “Related work” describing the resemblance of the proposed system to that proposed by the Liberty Alliance for e-commerce.

Modified the 2nd paragraph of the section headed “Weaknesses and possible solutions” on page 18 and added a reference to “fusion queries” described by Yernini et al.

Tim Churches
4 January 2003