Reviewer's report

Title: A RESTful pseudonymization interface for use in modern web applications

Version: 2
Date: 3 November 2014
Reviewer: Wolfgang Hoffmann

Reviewer's report:

Manuscript
Submitted to BMC Medical Informatics and Decision Making.

Review
This is a description of a web-based solution for pseudonymization and record linkage. The technical concept is based on a standard suggested by the Technology and Methods Platform (TMF) in Germany in 2006. The description is considerably technical, and provides details on some fundamental processes of the interface. A reference implementation is provided under an open source policy.

The article reviews a few other available solutions for pseudonymisation of patient identifying data and discusses them in terms of their principal architecture, implementability in the context of multi-site research projects, firewall demands, and flexibility to accommodate additional tasks.

The article covers a topic of increasing relevance in both clinical and population research and addresses major points of the present discussion with a focus on the German speaking medical research community.

Major points
P 4, line 28-30 “… up to two errors can be detected …” “… the transposition of two adjacent characters can be corrected…” how is this achieved? how can be excluded that the respective pseudonyms with these kins of differences do not in fact refer to different patients?
P 4, l 37 what is meant with “… in an error tolerant way…” obviously, the degree of deviance, that can be handled by any algorithm is limited. How are these limits defined? is there empirical evidence to define equality vs. unequality of the identifiers in individual records?
P 4, l 45-46 I do not understand, why an additional “ability” renders a pseudonymization service less simple than one without this ability – do the authors rather mean “demand”?
P 5, l 31-33 How can this advisable feature be implemented in the typical use cases? Shouldn’t it be enforced?
The text reviews existing solutions from a single and current technical perspective. As different projects have different requirements on architectures and design, a more balanced review seems recommendable. E-PIX, for example, was designed a couple of years ago as part of a complex hospital-internal research platform, with no bandwidth or latency bottlenecks to consider, based on Java EE and ESB. Later platform and E-PIX extensions have added capabilities for the RESTful interface leaving the tool itself as-is.

Programming of a new tool inspired by the shortcomings of an existing one does not instantly make such new tool the “successor” of the existing one. Shortcomings of the existing one can easily inspire a multitude of new tools being designed and implemented. To (formally) become the successor of the existing one requires more than added features and running code. The statement made should consider strategic and technical aspects of the cooperative development of the PID generator 2.0 in the TMF community, including the option of not one tool being the 1:1 successor.

Minor points
P 3, l 43 please spell out UUID at first mention
P 10, l 34 Switzerland
P 11, l 2 what is meant with “…somewhat wide…” ? please clarify
P 1, l 4 (and further on) The article does not clearly distinguish between identity management with the assignment of some identifier as representation for the identity and pseudonymization referring to a domain-specific assignment of another identifier (the pseudonym) to protect the identity of a person. The discussion of the RESTful interface mixes aspects of both throughout the discussion. An explicit definition would improve the consistency of the arguments.

P 10, l 25 The German National Cohort operates an identity management that is, in fact, not based on the Mainzelliste but on E-PIX. However, the National Cohort’s system uses a RESTful interface which is compatible with the Mainzelliste RESTful interface to a large extent.

Level of interest: An article of outstanding merit and interest in its field

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