Reviewer's report

Title: A Generic Solution for Web-Based Management of Pseudonymized Data

Version: 3

Date: 25 August 2015

Reviewer: Axel Newe

Reviewer's report:

The manuscript has been greatly improved with this revision. However, some work remains to do.

Major Compulsory Revisions:
None.

Minor Essential Revisions:

* The requirements specifications (lines 153-161 and later) are much better and clearer now (actually, they are already now the best req. specs I have read in a biomedical informatics article so far), but they are still not matching ISO/IEC/IEEE 29148, especially as regards the use of keywords in a requirements construct (see 5.2.4 of ISO 29148). Since all requirements are probably mandatory requirements, the keyword "shall" should be used instead of "has to". Exception: R-P1 is a "should" requirement, because "shall" requires a definite reference and not a rather vague description of "well-established usability guidelines". Although this article is not a software engineering document, using appropriate keywords would improve it up to a quality level that could serve as excellent example for other work.

o In most cases the wording "designed in a way" is superfluous in req. specs. It draws away the attention from the actual req. (see proposals below).

o Lines 192-193: "The system has to be designed to support the hosting of different backends on different physical machines with different host names." can be shortened to "The system has to support the hosting of different backends on different physical machines with different host names." or even better "The system shall support the hosting of different backends on different physical machines with different host names." The same applies for the following lines: "must provide" should be replaced by "shall" (see description of problems as regards using the word "must" in req. specs in section 5.4.2 of IEEE 29148).

o Lines 213-215: better: "The reconstruction of the logical global dataset shall only be performed at the client-side to reduce the number of attack vectors."

o Lines 216-217: better: "Clients shall be unable to learn the pseudonymous identifiers used in the distributed databases." or "Clients shall be prevented from learning the pseudonymous identifiers used in the distributed databases."
Lines 229-230: better: "The system should allow for centralized installation and maintenance."

* Lines 204-206: The sentence "On the documentation level, and thus on the system level, re-identification requires reversing the separation between identifying data and payload data [20, 31, 35]." is more an explanation of the background of R-A1 than a requirement. Therefore, it should be separated from the specification text and moved to the descriptive text below.

* Fig. 3b, Step 3 reads "3. request(token MS (99->2))". Where does the "2" come from? Shouldn't it be "request(token MS (99->1))", since the describing text reads "The client forwards the token to the mapping service" (line 425)? Otherwise, please explain the origin of the "2".

* Lines 502-512 read inconsistent notations: K1 is written using lowered "1" and unlowered "1". Please make this consistent.

* Lines 591-592 read "Automated penetration-tests have been performed to detect weaknesses." -> This is important and the result might be obvious, but: What are the actual results?

* The descriptive text of Fig. 5 reads incorrect syllabification.

* Please rework the references again. There are typos in the following references: [3], [20], [29], [35]. The following references have missing or incomplete journal / conference names: [32], [44], [51]. Almost all journal references have missing issue numbers or other missing details.

Discretionary Revisions:

* Fig. 2: The reader could get the impression that different clients are involved in the loose coupling strategy, although it is only one client. Maybe it is better to replace the word "client" by "presentation layer" (or something similar) and to draw an additional box around named "client" around the presentation layer(s) (for each strategy). This would emphasize that there is only one client in each strategy.

* Access dates of web references could be updated.

* For the Discussion section: Legacy browser support is generally questionable... if outdated browser versions do not match security requirements, should they be supported then? However, this might be too far away from the article focus.

Level of interest: An article of importance 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.