Reviewer's report

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

Version: 2
Date: 30 October 2014

Reviewer: Adolfo Muñoz Carrero

Reviewer's report:

• Major Compulsory Revisions

1. We started to read the paper with great interest but got a bit disappointed by the content because it doesn’t offer what we expected from the title. The treated subject, pseudonymization of clinical information, is of great importance for the transference of information which is one of the pillars for the continuity of care and also for secondary use (research and statistics). However, the paper is only focused on the technological aspects of one possible solution, avoiding to give any opinion about a lot of aspects of this problem (some of them, by the way, might not be answerable with the technical solution proposed, like the possibility to group the patients by age or by their residence place). In order to fix this problem and avoid the potential audience the same confusion, we propose the two following changes:

2. Change the title to clearly reflect the content of the paper. Something like that could do the trick: A RESTful interface to access pseudonymization services for modern web applications.

3. Clarify the objective in the background section. In page 2, line 2 it is said that one of the problems to fix in order to be able to use the information for research is the pseudonymization of the records. This is not solved by simply changing the ID of the patient: The ID or the name can be found in several parts in the record, like in the multimedia information, in free-text sections or in information regarding one of the patient’s relatives. The identity can even be established combining some other data (quasi-identifiers) like birthdate, sex, and home place. So, in order to keep privacy, the record must be filtered and modified before it can be safely disclosed. Probably, the objective of this paper is not to solve this question; however, it should clarify that to avoid misunderstandings in the audience.

4. Transference of sensible information could be dangerous because of the risk of interception. In background section it could be necessary to specify if there is some safeguard implemented to protect the communication against that risk.

5. After reading the manuscript it seems that all demographic data are in the IDAT while all medical data reside in the MDAT, which is the only information accessible for secondary use. This could not be enough for a majority of research usages where knowledge about age, sex or residence of patients is necessary in order to extract valid conclusions. On the other hand, it is true that
these data can lead to the re-identification of the patient. This risk is treated through several models (k-anonymity, l-diversity, t-closeness) and there are techniques (generalization, removal of unusual values ...) to minimize it. We think that the paper should deal with this subject or the authors should justify the elimination of all the demographic data.

**• Minor Essential Revisions**

6. Figure 2 shows the components of the Mainzelliste reference implementation. The relationships depicted inside the Patient List components lack of clarity and might cause ambiguity. We think that this figure should be improved to show in a clearer way how these components interact.

**• Discretionary Revisions**

7. There is an ISO standard, the Technical Specification ISO/TS 25237 that defines the concepts for identification and pseudonymization (along many other things). It could be interesting to check it and to state in the paper that whenever the authors use the pseudonymization related concepts in the text their meanings are those defined by this standard.

**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:**

We declare that we have no competing interests