Author’s response to reviews

Title: Examining database persistence of ISO/EN 13606 standardized Electronic Health Record extracts: relational vs. noSQL approaches

Authors:

Ricardo Sánchez-de-Madariaga (ricardo.sanchez@isciii.es)
Adolfo Muñoz (adolfo.munoz@isciii.es)
Raimundo Lozano-Rubí (rlozano@clinic.ub.es)
Pablo Serrano-Balazote (pserranob@salud.madrid.org)
Antonio Castro (alcastro@isciii.es)
Oscar Moreno (omoreno@isciii.es)
Mario Pascual (mario.pascual@isciii.es)

Version: 4 Date: 05 May 2017

Author’s response to reviews:

Madrid, 24 Apr. 17

Dear Editor,

This is a point-by-point response letter accompanying the revised manuscript after the minor revision.

We would like to thank again the reviewers and the Editor for their appropriate and useful comments on the manuscript after the major revision. We have tried to address every point suggested by them and we believe this has helped again to considerably improve the quality of the new version of the paper.

Editor comments.

Thank you for undertaking revisions to include an XML database system. As per reviewer comments, this was seen as an improvement in the overall methodology and interest in your manuscript. Given this, the reviewers have noted a few points for expansion and/or clarification
that we would like you to address. We believe this is necessary to further strengthen the manuscript.

We also believe the inclusion of the native XML database and the concurrence experiments has improved the overall methodology and interest of the manuscript. We address the rest of the points in the remainder of this letter.

The reviewer point regarding the use of the term NoSQL is not stating that you are using it improperly, but seeking clarification from you on how you define NoSQL within this context. As noted by the reviewer, the term NoSQL now is overloaded, and further defining early on what you exactly mean by NoSQL (for example, are you subscribing just to the concept of non-relational?) and ensuring consistent application of this definition within the paper would be helpful. Please also note appropriate capitalization with terms such as “NoSQL” (not “noSQL”) and “MongoDB” (not “mongoDB”) as you review.

In order to clarify the meaning and use of the term NoSQL database (and the meaning of the term document-based NoSQL database) we have added several lines (lines 98 through 102), with several references including the Wikipedia link suggested. Following that link, we have considered a NoSQL database one that is modelled in means other than the tabular relations used in relational databases. Lines 241 through 244 were also added in order to state the classification of native XML databases as a kind of document-based NoSQL database, i.e. a NoSQL database, following the criterion stated above.

With respect to the state of the art, please note some additional resources that Reviewer 1 has suggested which should be considered in your presentation of this as an expansion of the field with reference to existing works.

We agree that the state of the art in the previous version of the paper needed to be enhanced. We have carefully read and studied the seven resources provided and have used them to include an entire new section (1.4 NoSQL current state-of-art, lines 117 through 137) with 8 new references ([39] through [46]) and an expansion of the first part of the Background section (lines 60 through 67) with 4 new references ([47]-[50]). We believe this has considerably helped situate the reader in the current state-of-art of standardized EHR persistence and in the scope and meaning of EHR systems implementation in general.

In addition there are a few additional requests for clarification in other parts of the manuscript, as well as a need to ensure proper grammar. If any further questions or assistance is needed in these revisions, please let us know.
We have tried to clarify every point suggested by the reviewers. Regarding the English language correction, it has been proof read by an English native speaker.

Reviewer 1 comments.

1. The authors MUST refer to the wikipedia link: [https://en.wikipedia.org/wiki/NoSQL](https://en.wikipedia.org/wiki/NoSQL)

   (i) Authors are refering to non-relational and noSQL simultaneously in paper whereas they refer to noSQL as short form of non-relational in line 13. What exactly they mean by noSQL is still ambiguous throughout the paper.

   (ii) NoSQL basically denotes NOT ONLY SQL. In line 17, two noSQL databases (one document-based and one native XML database) is referred. XML is usually considered as semi-structured database. Can it be considered under the category of NOSQL DB? This needs to be examined.

   In order to clarify the meaning and use of the term NoSQL database (and the meaning of the term document-based NoSQL database) we have added several lines (lines 98 through 102), with several references including the Wikipedia link suggested. Following that link, we have considered as a NoSQL database one that is modelled in means other than the tabular relations used in relational databases. Lines 241 through 244 were also added in order to state the classification of native XML databases as a kind of document-based NoSQL database, i.e. a NoSQL database, following the criterion stated above.

2. Line 77 refers to EHR Extracts without defining what it is, nor any reference is given. Also later it is implied that the reader knows about it.

   A definition of EHR extract is given in lines 79 through 81. A new reference to the sources of this definition is also given. Later occurrences of the term in the paper may recur to this definition.

3. The line at 107 is not giving a clear meaning of what authors want to perform as experiment. It must be rewritten. Also at line number 113,"but also concurrently" is little confusing.

   Certainly the original paragraph did not give a clear explanation of what the authors wanted to perform as experiment. An entire new paragraph at the beginning of section 1.5 is provided (lines 140 through 147) giving a clearer explanation of these experiments. We believe this explanation is giving a clear idea of what is performed in our experiments.

We have checked the link originally provided in the reference and it does not work. We lament that the readers will not be able to access the XML schemas used by the ISO/EN 13606 standard; but we have opted to maintain the reference to the institution that developed them, suppressing the link.

5. The current state-of-art still needs to be enhanced. E.g.,

(i) Big Data Electronic Health Records Data Management and Analysis on Cloud with MongoDB: A NoSQL Database

(ii) Modeling Stack Framework for Accessing Electronic Health Records with Big Data Needs

(iii) Extraction of standardized archetyped data from Electronic Health Record Systems based on the EntityAttribute-Value Model

(iv) MongoDB Improves Big Data Analysis Performance on Electronic Health Record System

(v) Comparing the Performance of NoSQL Approaches for Managing Archetype-Based Electronic Health Record Data (Also References of this paper can be seen).

(vi) Evaluation of NoSQL databases for EHR systems

(vii) Modeling and Querying Data in NoSQL Databases

We agree that the state of the art in the previous version of the paper needed to be enhanced. We have carefully read and studied the seven resources provided and have used them to include an entire new section (1.4 NoSQL current state-of-art, lines 117 through 137) with 8 new references ([39] through [46]) and an expansion of the first part of the Background section (lines 60 through 67) with 4 new references ([47]-[50]). We believe this has considerably helped situate the reader in the current state-of-art of standardized EHR persistence and in the scope and meaning of EHR systems implementation in general.

7. English/ Grammar needs to be checked in paper. Tenses are present followed by past followed by present continuous. It MUST Be PROOF READ. E.g., Line 135 "This" should be replaced by "these".

The paper has been sent to be reviewed by an English language native speaker. We hope that the new version has corrected every error related to the correct use of the language.
Reviewer 2 comments.

Please take care of your text, especially the Abstract. For example, I cannot find any background information in section Background of your abstract.

We have tried to correct the text of the paper as much as possible. Regarding the background information in section Background of the abstract, we have added a few lines (lines 15 through 18) in that section of the abstract, in order to provide the reader with a perspective of the attention raised by NoSQL systems and the current coverage in the literature of the relational and NoSQL persistence of standardized EHR systems.