

We would like to thank the reviewers for their careful reading and suggestions regarding this submission entitled “An Objective Assessment Framework & Tool for Linked Data Quality: Enriching Dataset Profiles with Quality Indicators” submitted to the IJSWIS Special Issue on Dataset Profiling and Federated Search for Linked Data.

Please, find below a detailed response to all comments explaining how we have addressed them in this revised version of the paper.

Addressing the comments made by Reviewer 1

1. I am still not satisfied with the state of the references, even though much has improved

We have revisited and cleaned the references one more time. We have fixed the names where appropriate and added extended information for journal papers where applicable (volume, number, pages).

2. I still do not like the notation in Formulas (1) and (2)

We have rewritten the formulas to make them clearer and improved the notations used.

3. Improvement of Figure 1 to make it look more like a pipeline

We have decided not to make this figure look like a pipeline because we wanted to emphasize the modularity of the framework as any module can be easily added or removed. For example, when integrating the tool directly into CKAN portal, one can easily remove the data portal identification module (step i).

4. Inclusion of Section 5 (now Section 4)

We decided to include this section as it shows that the current state of Linked Data quality checker tools lack in various aspects and to indicate that the Roomba quality extension fills in those gaps.

5. You exclude security-related aspects - that is OK. But the argument why you do this is not: group members should not restrict access, but they also should not use URNs, should not omit dereferencable URIs and should in general provide high quality data.

We have edited this part and indicated that the security aspects were excluded as checking them require checking special protocols (e.g., SSH) and is not the scope of the suggested tool.

6. The amount of notation introduced in 5.1 seems a bit much wrt. the concepts you want to describe. Also, at least I need an explanation as to what tags and groups are in this context

We have included a description of tags and groups right before as part of the CKAN data model.

Tags: Provide descriptive knowledge on the dataset content and structure. They are used mainly to facilitate search and reuse

Groups: A dataset can belong to one or more group that share common semantics. A group can be seen as a cluster or a curation of datasets based on shared categories or themes. The brackets denote an optional weight field.

Addressing the comments made by Reviewer 2

1. The new evaluation section (Section 5.2) is a great addition. It would be in my opinion however much more interesting to add some deeper discussion to the section: right now it reads as a very abrupt results dump. The authors should i) add more motivation/explanation concerning the experimental methodology in the context of this section and ii) discuss in more detail their results: are they surprising, good, or simply as expected and why? They should also probably change the title of the new section.

We have added few lines to describe the motivation of the evaluation and a description of the results as they were as expected because we have specifically tailored Roomba to completely cover all the mentioned quality indicators.

We have changed the section title to “Evaluation & Motivation”.

2. Regarding my second remark (according to the author's numbering scheme): thanks for the clarification, please discuss this in more detail in the paper as it can be in my opinion quite confusing for the reader.

We clarified that a bit more in the paper, especially in the quality score calculation section.

3. Finally, I still think that a brief discussion on the potential (/ experimental) scalability of the tool would be a very nice addition to the paper.

We have added brief descriptions in the conclusion and in the summary of the tool evaluation.

Addressing the comments made by Reviewer 4

1. Section 2 covers just some aspects of related work. It should be put in proper context – Section 2 just covers general initiatives (ODI and LDBC). The reader should be advised that Section 4 will survey quality assessment tools. Further remarks about related work

appear on page 16 under the “Summary” paragraph. These remarks should be moved to Section 2.

We mentioned in the last paragraph of the introduction that Section 2 cover the related work around data quality assessment methodologies where Section 4 covers the related work regarding the quality assessment tools. We have added in the end of Section 2 a note about that.

We have not moved the remarks in the “Summary” paragraph as they refer to some of the quality indicators mentioned afterwards. We feel that moving these remarks to Section 2 will not be suitable.

2. The authors should also clarify to what extent the version of Roomba differs from the previous one.

We have clarified that the tool proposed is not a new version of Roomba, but a new module that extends Roomba to produce a quality report.

3. Correctness. This should be called “Syntactical Correctness”

While correctness checks include Syntactical ones, they also include non syntactical ones like the size and MIME type correctness.

4. Quality Score Calculation. Why introduce weights just to dismiss them in the next sentence: “In Roomba, all weights are equal and set to 1”. This does not make much sense

We have fixed that. Roomba takes a configuration file with the quality indicators and their weights. We have specified that in our experiment we have set the weights into 1 as ranking the quality indicators is not in the scope of this paper.

5. Notation

We have fixed that (Please see also Note 2/ Reviewer 1)

We have also fixed the Structure of the paper as requested:

- Section 3 contains 3 pages before Section 3.1 The text on pages 3 (bottom) up to the end of page 5 should be put under a new subsection of Section 3.
- Section 5 suffers from the same problem. It contains 2 pages before Section 5.1. Again, the text on pages 12 and 13 should be put under a new subsection of Section 5.
- The first two paragraphs of Section 5 are misplaced and would be better of elsewhere. In fact, it would be interesting to briefly review CKAN to make the paper self-contained.
- Listing 1 should be placed before Figure 2, since you refer to Listing 1 before Figure 2.