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. The references are quite a mess

We have fixed and cleaned the references all over the paper which are now written very consistently.

2. A major claim is that previous frameworks are not applicable because LOD is more difficult that relational data. I disagree. You mention as sources of complexity ontologies, vocabularies, links and query endpoints. RDBMS can have all of those, too. Be more specific and also explain how this adds to the complexity of measuring quality

We have toned down these claims and highlighted that ensuring Linked Data quality is complex as it included multiple dimensions (ontologies, vocabularies, links and query endpoints). This adds to the overall complexity of ensuring quality only at the data instance level.

3. I really like your four main categories and how they correspond to the linked data principles. What I do not understand though,. is why you dismiss the first category and place it under the third. E.g., one could ask about the completeness of a specific entity, i.e., how many properties it uses. I can easily imagine such a question for all dimensions at entity level

We did not mean to dismiss the first category. However, we thought of those two categories as deeply connected thus merged the indicators to check for them into one. The example you gave perfectly resembles that idea as you need the model (properties) to ensure an entity completeness. Moreover, the merge is also due to the fact that we noticed a variety of tools that can check for entities and models qualities, in contrast with other indicators as shown in the quality tools survey.

4. The title mentions objectivity, and I was expecting the abstract to at least give an intuition of how this is achieved

We changed the abstract a little bit to reflect that an objective quality measure should not depend on the underlying use case and can be automatically checked by tools.

5. "ODI is a great initiative": How much/often is it used?"

We added some information about the current coverage of the ODI certificates issues at the time of writing this paper

6. Your discussion of LDBC does not mention quality at all. Is it part of the benchmark?

We added some more description about the LDBC and how they aim at promoting graph and RDF data management systems to be an accepted industrial solution. LDBC is not focused around measuring or assessing quality. However, it focuses on creating benchmarks to measure progress in scalability, storage, indexing and query optimization techniques to become the *de facto* standard for publishing performance results.

7. You explain that completeness highly depends on scenario and use-case. Can you name any dimension for which this is not the case?
8. Availability, as you define it, seems to be a Boolean measure. Is that what you have in mind?

Completeness is generally a subjective measure depending highly on the scenario and use-case in hand, opposite to other measures like availability where one can measure if a dataset is available or not independently of the underlying use case.

9. First you say that completeness is a measure. Later you say that completeness has some measures. Something is off here.

It was an error, we fixed that by saying that completeness has some quality indicators.

10. For completeness you list on page 6 a whole bunch of aspects. However, they do not correspond at all to those mentioned in Table 1.

The indicators in the table are one-to-one mapped to those in the paragraph

11. Underneath Figure 1 I learn that there is apparently a CKAN standard model to assess quality issues!?

CKAN does not have a model to assess quality. However, by examining the issues reported in Table 2 we found out that the quality score can be calculated by examining the dataset metadata/profile. Since Roomba is primarily developed for CKAN-based portals, we check the quality for datasets presented by measuring their conformance to the CKAN standard model.

12. The amount of notation introduced in 4.1 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 added descriptions for what tags and groups mean in a CKAN model.

13. I do not understand the notation of formula (1). What are the pointy brackets doing there?

The brackets denote an optional weight field.

14. At the bottom of page 10 you suddenly talk about errors (for Figure 2). Is this inverse quality?

Yes, and we have made that clearer in the paper

Addressing the comments made by Reviewer 2

1. In the introduction, the authors say that ensuring data quality in LOD is "much more complex" than for Web documents; I personally don't agree with that statement (I would say it is actually the other way around since Web documents are unstructured), please discuss in more detail.

Please, refer to the answer provided for the reviewer 1/remark 2

2. In general, the elements from that list make sense, although the various properties and metrics are in my opinion not sufficiently discussed (e.g. "contains different data access points" or "uses datasets description vocabularies like DCAT or VOID": how do the authors rate this? Are the lists given in Table 2 final? Are the resulting scores binary values? I yes, is this sufficient as a metric? If no, how do you come up with the score?).

In the paper, we have two main contributions. The first is the list of objective measures and their quality indicators. For that, we specify, for example, the "contains different data access points" indicator. However, checking this indicator highly depends on the dataset model being used. As a use-case, we show the quality calculation process on CKAN-based portals as CKAN is the model examined by Roomba (the tool of choice). As a result, we end up with a finite set of values permitted by the underlying model that we can check against.

As for the score calculation, some indicators such as "dataset availability" have a binary value (available or not) but others have a score based on the set of values the tool check against. For example, having a complete set of access points for CKAN is measured against 6 values. So if the dataset have 3 of them then the completeness score for that quality indicator is 50%.

3. Contrary to what the authors say, they do not directly evaluate the proposed framework. The run their tools on a series of datasets and report the resulting values but

do not evaluate the performance of their tool. This is from my perspective a severe limitation.

We have added a dedicated section (section 5.2) for evaluating the tool according to two scenarios. The correctness and completeness for quality scores calculation.

Addressing the comments made by Reviewer 3

1. The paper is apparently based on [3], but [3] is not discussed in the related work section. It is mentioned in section 3, but not in detail. I miss a description of what is done in [3], why [3] was insufficient, and why there is a need to extend the work in [3]

We argue that the measures discussed in [3] are mixed between the subjective and the objective ones. We refine those by focusing only on the objective measures. We also refine the quality indicators in details by surveying the landscape of data quality tools and frameworks. We propose tooling for automatically assessing those objective measures.

2. You refer to LDBC. Can you compare their work to yours in terms of content? Currently, the only comment is that the final LDBC rapport is not in yet. But LDBC is a 4 year project (afaik) so there must have been other publications.

Please, refer to the answer for reviewer 1/remark 6

3. For most criteria in the list you cite a paper. How did you select these? Are all criteria mentioned in these paper also in your list? For some criteria you do not cite anything. Where do these come from?

The objective criterias are all mentioned in this list. The indicators have been gathered by:
- Transforming the objective quality indicators presented as a set of questions in [4] into more concrete quality indicator metrics;
- Surveying the landscape of data quality tools and frameworks;
- Examining the properties of the most prominent linked data models from the survey done in [3].

4. Not all criteria are obviously good ones (at least, not obvious to me). For example, is it always true that 2 entities in a dataset should not have the same preferred label? In a dataset of people there are often two people with the same name.

We have explained in the paper that the quality indicators are weighted. These weights give the flexibility to define multiple degrees of importance. For example, a dataset containing people

can have more than one person with the same name thus it is not always true that two entities in a dataset should not have the same preferred label. As a result, the weight for that quality indicator will be set to zero and will not affect the overall quality score for the consistency measure.

5. In section 4, an introduction is needed on what is Roomba, and why Roomba was chosen as a starting point

 An introduction to Roomba has been added.

6. Here is says: "we have presented 30 objective quality indicators" Before, in table 1, there were 64. What is this subset?

In our framework, we have presented 30 objective quality indicators related to dataset and links quality. The remainder 34 indicators are related to the entities and models quality and cannot be checked through the attached metadata. Excluding security related quality indicators as LOD cloud group members should not restrict access to their datasets, the Roomba quality extension is able to assess and score 23 of them (82%).

Misc. Fixes

The following fixes have been applied throughout the paper:
- Place Survey of assessment tools section before the section about Roomba **(Review 3)**
- Listing 1 is never referenced **(Review 1)**
- In formula (2) you have a sum over all i, but i does not appear **(Review 1)**
- Figure 1 looks more like an architecture that a process. Maybe you can change around a little, for instance by adding arrows. Also, you got the roman numeral (iv) wrong **(Review 1)**
- Check for commas around "e.g." **(Review 1)**
- For the third paragraph on page 2, a second more domain-specific example besides DBpedia would be good **(Review 1)**
- I would call section 2 simply "Related Work" **(Review 3)**