

Below are our responses to the specific points raised by the editor and the reviewers.

> Editor's comments

> Please address the remaining comments raised by reviewers 1 and 2

Done. See below.

> Reviewer 1 (Daniel Garijo)

> What I would like to recommend the authors is to include the link to the  
> repository with the data from the evaluations  
> (<https://bitbucket.org/tkuhn/trustypublishing-study/src/>). That way anyone  
> interested in the data would be able to access it.

That is a very good idea. The links to the two repositories are now included in the beginning of the evaluation section.

> Reviewer 2 (Anonymous)

> C1. The authors did not address my first comment in the previous round. "The  
> authors stated two main requirements: Having a reliable mechanism for hosting  
> and referencing datasets, but also the ability to reference and retrieve  
> datasets at different granularity levels. With respect to the second  
> requirements, the authors did not discuss the process by which a datasets is  
> transformed into a set of nanopublications. The model proposed by the authors  
> support nano-publication and nano-publication indexes, which references other  
> nano-publication. The question the reader may ask is how nan-publication and  
> their container nano-publications indexes are obtained given a dataset".

We now understand your point about data being represented in RDF but not in nanopublication format. We previously misunderstood it as being covered by our response to C2. To address it, we added a few sentences to the first paragraph of the approach section, sketching the problem and a starting point for its solution:

"We furthermore exploit the fact that datasets in RDF can be split into small pieces without any effects on their semantics. After Skolemization of blank nodes, RDF triples are independent and can be separated and joined without restrictions. Best practices of how to define meaningful small groups of such triples still have to emerge, but an obvious and simple starting point is grouping them by the resource in subject position. We focus here on the technical questions and leave these practical issues for future research."

With respect to the generation of nanopublication indexes, we now explain this explicitly in the end of the section Nanopublication Indexes, and refer to the subsequent section, which explains how the nanopublication library can be used for this.

> C2. In their response to C2 in the previous round, the authors suggest that  
> heterogeneity of the data model used in datasets (e.g., CSV and relational)  
> can be resolved by using existing state of the art technique to translate data  
> in those models into RDF. This approach may be expensive, I was wondering if

> lightweight approaches which do not attempt to translate the original data,  
> but instead create metadata that describe them using nanopublication would be  
> more realistic and cost effective as a solution. Of course the granularity of  
> retrieval in this case would be a whole dataset, but there are scenarios where  
> this solution would be acceptable. I think that a discussion in these lines  
> that clarify the options to the reader would be helpful, specially that linked  
> data form only a small proportion of available scientific data.

This is a very good point. Thank you for raising it. We have in fact been thinking about such light-weight dataset announcements, in particular by using the HCLS Community Profile for dataset descriptions. We now include the following paragraph at the end of the section on nanopublication indexes:

"As a side note, dataset metadata can be captured and announced as nanopublications even for datasets that are not (yet) themselves available in the nanopublication format. The HCLS Community Profile of dataset descriptions (Gray et al., 2015) provides a good guideline of which of the existing RDF vocabularies to use for such metadata descriptions.