Assessing Knowledge Graphs Accountability

Jennie Andersen(✉), Sylvie Cazalens, and Philippe Lamarre

Univ Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, 69621 Villeurbanne, France
{jennie.andersen,sylvie.cazalens,philippe.lamarre}@insa-lyon.fr

Abstract. Demand for accountability is increasing, driven by the growth of open data and e-governments. Accountability requires specific and fairly accurate information about people's responsibilities and actions. Studies on data quality or FAIRness do not have a specific focus on that aspect. Therefore, we describe our approach to evaluate the accountability of several knowledge graphs of the LOD cloud and the results obtained.

Keywords: Knowledge graphs · RDF graphs · Accountability

1 Introduction

Designing systems enabling individuals and institutions to be held accountable is increasingly important [6], especially as it enhances trust in the data. Dataset accountability means that "there is sufficient information to justify and explain the actions" on the dataset, "in addition to descriptive information and information on the people responsible for it" [4]. Concerning Knowledge Graphs (KGs), many works look for meta-information, either for evaluating some quality aspects [2,7] or conformance to recommendations, such as FAIRness [1,5]. None of them focuses specifically on accountability as a whole. They consider some elements of accountability but do not take into account all the elements required by it. For example, FAIRness requires meta-information such as creators but accountability goes further requiring affiliation and contact information of creators for instance.

Therefore, in this paper, we aim to conduct an evaluation of the accountability of KGs. For datasets in general, precise requirements are expressed as questions by the LiQuID metadata model [4]. It has been validated based on a real-world workload that relies on existing regulations and an expert survey. The use of this model and questions to evaluate KGs requires (i) adapting the hierarchical model and questions to KGs, (ii) expressing the questions into SPARQL queries, (iii) querying the KGs, (iv) computing accountability scores that can be detailed according to different levels of the hierarchy. These steps are illustrated by Fig. 1 and detailed in the next sections. Notice that we use the IndeGx framework [3]. This SPARQL-based test suite proposes several functionalities.

C. Pesquita et al. (Eds.): ESWC 2023, LNCS 13998, pp. 37–42, 2023.
https://doi.org/10.1007/978-3-031-43458-7_7

We only use it as an engine to submit multiple queries to many KGs, and to store the results in RDF, without querying its index. For further information, all our evaluation material and results are available on Github[1].

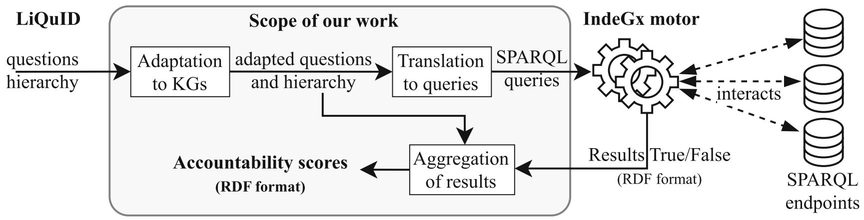

Fig. 1. Process to Define and Measure Knowledge Graph Accountability

2 Adaptation of LiQuID and Translation into Queries

LiQuID relies on a hierarchical structure: first, the steps of a dataset's life cycle, then the different question types (who, when, etc.) and finally the types of information (description, justifications, etc.). The authors provide an exhaustive and explicit list of questions to describe each leaf of this hierarchy, and so, what must be provided to be considered accountable.

Adaptation. Ideally, to assess the accountability of a KG, all this hierarchy and associated questions should be considered. However, not all of them can be adapted to KGs and translated into SPARQL queries, as shown by the comparison between LiQuID and the two metadata models Dublin Core and PROV [4]. Indeed, a lack of expressivity of the vocabularies prevents many questions from being translated into queries or often induces a loss of precision (e.g. inability to distinguish between data collection and data processing steps). Faced with these difficulties, we decide to consider only questions that are compatible with the current vocabularies of semantic web. For instance, we only keep the field "description" of the last level of LiQuID, removing too specific questions such as "Why is it ethical to create a dataset for this cause?". Therefore, we only keep a part of the LiQuID hierarchy, shown in Fig. 2, which leads us to consider a core set of 25 LiQuID questions (out of 207). These questions are adapted, as faithfully as possible, to the context of KGs, and made more precise. For instance, "Who has used/can use the published data set?" splits into "Who has the right to use the KG?" and "Who is intended to use the KG?". It results into 30 questions: 5 for Data Collection, 5 for Data Maintenance and 20 for Data Usage.

[1] https://github.com/Jendersen/KG_accountability.

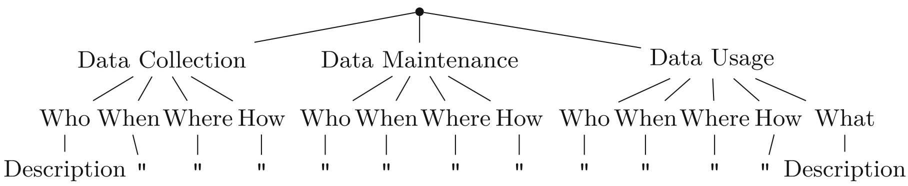

Fig. 2. Hierarchy Classifying Accountability Questions Adapted from LiQuID [4]

Translation. Finally, all the questions are translated into SPARQL queries. Ten vocabularies of reference are chosen regarding their pertinence, let them be specific to describe datasets (VoID, DCAT, SPARQL-SD, DataID and DQV), more general (the Dublin Core, FOAF and schema.org), or focusing on provenance (PROV-O and PAV). Each query uses all coherent properties and classes of these vocabularies to be as complete as possible. For instance, a query asking for a publisher accepts all publisher-like properties (using the Dublin Core, schema and PROV), see Listing 1.1. An important point about our approach is a strict interpretation and translation of questions into queries. This leads us to be very demanding as for the way KGs express metadata: we look for metadata explicitly linked to the IRI of the KG to ensure that the information is actually about the KG. To identify it, queries look for an entity of class Dataset which is linked to the URL of the endpoint interrogated. If the KG does not provide such an IRI, it will not answer any of our queries.

Listing 1.1. Query Associated with “Who publishes this dataset?”

```
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dce: <http://purl.org/dc/elements/1.1/>
PREFIX schema: <http://schema.org/>
PREFIX prov: <http://www.w3.org/ns/prov#>
ASK { # <kg> is the KG IRI obtained by a preliminary query.
      {<kg> dct:publisher ?publisher .}
      UNION {<kg> dce:publisher ?publisher .}
      UNION {<kg> schema:publisher ?publisher .}
      UNION {<kg> schema:sdPublisher ?publisher .}
      UNION {<kg> prov:wasGeneratedBy ?act .
             ?act a prov:Publish .
             ?act prov:wasAssociatedWith ?publisher .} }
```

3 Querying KGs and Aggregating Accountability Scores

To query numerous KGs, the framework IndeGx is used. To determine whether a KG contains the necessary information for accountability, we use ASK queries, ensuring that we get a True result if the information is present and a False otherwise. We embed the set of queries previously defined into the framework and configure the format of the results to be Data Quality Vocabulary (DQV).

The metric of accountability is defined as follows. If a query obtains the result True (meaning the information is in the knowledge graph), the queried KG gets a score of 1 for the associated question, and 0 otherwise. To determine the score on each aspect of the hierarchy, each question has a weight, showing its relative importance. We assume original LiQuID queries to be weightened to 1. When such a question is replaced by n (more precise) questions, their respective weights is set to $\frac{1}{n}$. Therefore, for a leaf of the hierarchy, the score of accountability on that aspect is the weighted average of the score obtained on each question. For the other elements of the hierarchy, we determine their score by recursively computing the (non-weighted) average of the score on the elements underneath. A fine analysis is possible as all the scores are available, including those obtained for each question and the intermediate ones.

4 Experiments

Each knowledge graph is queried at three different time points. Only the results of the last experiment for which it was available are kept. Among the 670 queried knowledge graphs, we keep those which have answered and provided their own IRI. We only get 29 KGs. This result is in line with [3]: only a few KGs provide their own IRI and thus a self-description.

Considering these 29 KGs, Fig. 3 shows their overall accountability with the scores detailed on the three lifecycle steps. Our evaluation of accountability allows to discriminate between them, with values distributed between 2.2% and 44%, with a mean and median of 22%. On average, KGs are more accountable on Data Usage than on Data Collection, than on Data Maintenance. In addition, the results enable to compare KGs in more detail using a radar chart[2].

Regarding our evaluation, one must keep in mind that some meta-information may be provided by KG producers outside of the KG itself, for instance on its web page. It can also be inside the KG, but in such a way we did not identify the IRI of the KG: if it is not related to the URL of the endpoint nor to an entity of type Dataset. In any case, the meta-information is not detected and therefore not considered. While this may penalize some KGs, it points out the fact that they are less accountable because information is less accessible.

Concerning the measure itself, it is important to notice that some queries never succeed, 9 out of 30. A part of them is due to some strict choices we made that resulted in over-constrained queries. The rest is due to a lack of information provided by the KGs.

[2] https://github.com/Jendersen/KG_accountability/tree/main/results.

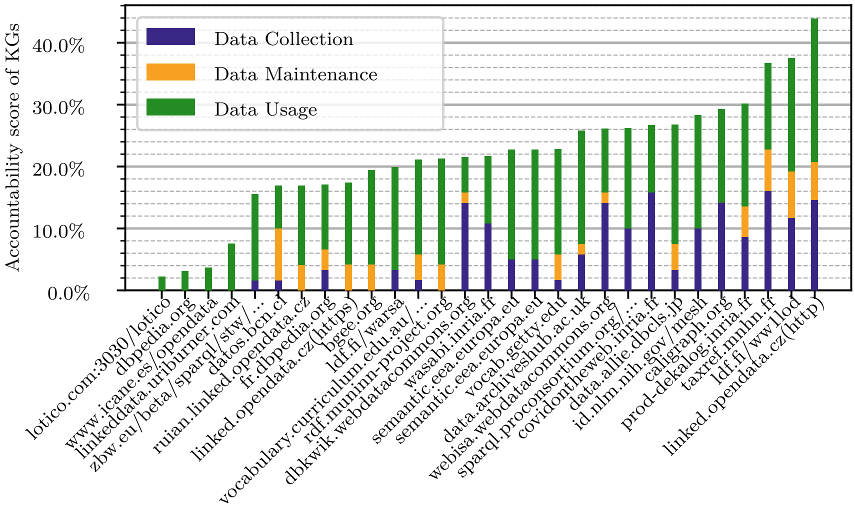

Fig. 3. Accountability of Knowledge Graphs

5 Conclusion

Our assessment of knowledge graphs accountability highlights the weaknesses of many KGs regarding accountability, even if the requests for certain information is very common. The obtained low or no scores show that there is room for improvement, and the presence of good ones shows that linked data is very suitable for accountability. Our measurement is detailed enough to help any KG producer to precisely identify missing information and therefore to improve on these aspects. Having started with a strict interpretation of the questions, we plan to relax some queries and introduce gradations in the definition of accountability requirements.

Acknowledgments. This work is supported by the ANR DeKaloG (Decentralized Knowledge Graphs) project, ANR-19-CE23-0014, CE23 - Intelligence artificielle.

References

1. Amdouni, E., Bouazzouni, S., Jonquet, C.: O'FAIRe: Ontology FAIRness evaluator in the agroportal semantic resource repository. In: Groth, P., et al. (eds.) ESWC 2022–19th Extended Semantic Web Conference, Poster and Demonstration. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-11609-4_17
2. Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of DBpedia, Freebase, OpenCYC, Wikidata, and YAGO. Semant. Web **9**(1), 77–129 (2018)
3. Maillot, P., Corby, O., Faron, C., Gandon, F., Michel, F.: IndeGx: a model and a framework for indexing RDF knowledge graphs with SPARQL-based test suits. J. Web Semant. **76**, 100775 (2023)

4. Oppold, S., Herschel, M.: Accountable data analytics start with accountable data: the liquid metadata model. In: ER Forum/Posters/Demos, pp. 59–72 (2020)
5. Rosnet, T., de Lamotte, F., Devignes, M.D., Lefort, V., Gaignard, A.: FAIR-checker - supporting the findability and reusability of digital life science resources (2021)
6. Weitzner, D.J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., Sussman, G.J.: Information accountability. Commun. ACM **51**(6), 82–87 (2008)
7. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: a survey. Semant. Web **7**(1), 63–93 (2016)