A Framework to Assess Knowledge Graphs Accountability

Jennie Andersen, Sylvie Cazalens, Philippe Lamarre, Pierre Maillot

A Framework to Assess Knowledge Graphs Accountability

Jennie Andersen
Univ. Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205
Villeurbanne, France
jennie.andersen@insa-lyon.fr

Sylvie Cazalens
Univ. Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205
Villeurbanne, France
sylvie.cazalens@insa-lyon.fr

Philippe Lamarre
Univ. Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205
Villeurbanne, France
philippe.lamarre@insa-lyon.fr

Pierre Maillot
Univ. Cote d'Azur, Inria, CNRS, I3S
Sophia Antpolis, France
pierre.maillot@inria.fr

***Abstract*—Knowledge Graphs (KGs), and Linked Open Data in particular, enable the generation and exchange of more and more information on the Web. In order to use and reuse these data properly, the presence of accountability information is essential. Accountability requires specific and accurate information about people's responsibilities and actions. In this article, we define KGAcc, a framework dedicated to the assessment of RDF graphs accountability. It consists of accountability requirements and a measure of accountability for KGs. Then, we evaluate KGs from the LOD cloud and describe the results obtained. Finally, we compare our approach with data quality and FAIR assessment frameworks to highlight the differences.**

***Index Terms*—Dataset accountability, RDF graphs, Evaluation Framework, Data Quality**

I. Introduction

Knowledge Graphs (KGs), and Linked Open Data in particular, enable the generation and exchange of more and more information on the web. This abundance of easily accessible data on the web offers many opportunities for researchers, companies or ordinary citizens. However, in order to share and use these data properly and legally, it is important to know some information about a knowledge graph, such as for what purpose it was created, by whom, etc.

Among the meta-information that contributes to the correct use of KGs, we focus on dataset accountability. It requires to provide information about actions on the dataset, "descriptive information and information on the people responsible for it" [1]. As a concept very close to transparency [2], it requires information to be easily accessible [3]. For the semantic web in particular, where software agents are particularly present, meta-information about KGs "needs to be available in a machine-readable format" [4]. These two things combined, we consider that this information should be present and searchable within the data of the KG itself.

Consider the GDPR (General Data Protection Regulation) for instance, that aims to protect personal data. To do so, Article 17 about the right to be forgotten states that the "subject shall have the right to obtain from the controller the erasure of personal data concerning him or her"[1] as soon as it does not fall under the right of freedom of expression and information. Therefore, every KG holding personal information, such as Wikidata, should provide contact information of this controller, i.e. a person responsible for the data, and ideally allow users to access it directly via its SPARQL endpoint. As another example, to avoid misinterpretation and to improve the (re)use of the data, it is often necessary to know for what purpose they were created, and for whom the data are intended. For instance, in some database mainly dedicated to teaching purposes, such as the MONDIAL Database[2], some inaccuracies can be tolerated (or even desired). However, it cannot be reused without precaution for other purposes. Therefore, it should indicate its intended audience or its expected usage. However, when querying its SPARQL endpoint, this information is not available. Accountability ensures that this kind of information of major interest is effectively available. Several studies are already looking for meta-information, either as some particular aspects of data quality [4], [5], [6], or as some requirements of the FAIR principles (Findability, Accessibility, Interoperability, Reproducibility) [7], [8], [9], [10]. Yet, they neither take into account the information used in the two previous examples nor several other accountability information. This highlights the importance of accountability as a specific and distinct characteristic of RDF datasets and the importance of their evaluation w.r.t. this aspect.

Hence, in this paper, we propose a new framework, KGAcc, dedicated to the assessment of RDF graphs accountability. It consists of organized accountability requirements and a measure of accountability. We experiment it on many KGs offering a publicly available SPARQL endpoint. Our accountability measure gives an indication of the accountability of KGs to dataset users and providers. It aims to guide users in their choice of one KG rather than another and to help providers to identify ways to improve their datasets.

To define such a measure, several questions arise, such as what meta-information is required? How to evaluate heterogeneous KGs? First, to define requirements, we rely on the LiQuID metadata model which focuses on dataset accountability [1] in general. It provides an explicit list of accountability requirements expressed in natural language. The problem, then, is to adapt this model to the specificities of knowledge

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

[1]https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04

[2]https://www.semwebtech.org/mondial/10/

graphs and to define the requirements in terms of SPARQL queries. To evaluate the KGs, we use the SPARQL-based test suite of the IndeGx framework [11]. We observe that most of them do not provide any easily accessible accountability information. However, as some KGs do answer some questions, it shows that our demand is reasonable and that KGs have a lot of room for improvement. In addition, to illustrate the specificities of this measure, we compare it with several assessment frameworks for data quality and FAIRness.

The rest of the article is organized as follows. Section II describes the state of the art. We define the KGAcc framework in Section III. Then, Section IV is devoted to the description of the methodology for evaluating RDF graphs and the results obtained. We then compare our accountability measure with the existing assessments of knowledge graphs in Section V. Finally, we conclude in the last section.

II. Related Work

Generally speaking, accountability requires that there is sufficient information to describe the data [1], the actions on the data, from its creation [12] to its use [2], and the people responsible for these data as well as these actions [1], [2]. It may concern different levels of the information system, such as information accountability [2], [12], systems [13], and dataset accountability [1]. To evaluate the accountability of knowledge graphs, we focus on this latter point. The accountability of knowledge graphs may be considered as part of their data quality, in the broad sense. These last years, several studies have highlighted the many facets of the notion [4], [5], [6]. In addition, general monitoring tools such as SPARQLES [14] and YummyData [15] have been proposed, enabling to assess and draw profiles of SPARQL endpoints.

As a matter of fact, measuring the accountability of KGs is a special case of assessing metadata completeness, which is defined as "the degree to which metadata properties and values are not missing in a dataset for a given task" [16]. Many works consider the presence of meta-information to evaluate knowledge graphs. Studies about the data quality of KGs [4], [5], [6] include many different metrics, among which a few focus on meta-information. For instance, provenance information is required by a metric on trustworthiness. The FAIR principles [10] are also interested in meta-information. One of the principles of findability states that "data [must be] described with rich metadata", and reusability requires that "meta(data) are richly described with a plurality of accurate and relevant attributes", including a license, and provenance information. Therefore, the required meta-information may overlap between accountability, data quality and FAIRness while having their own specificities. Because of the high variability of the actual implementations of these metrics and principles, we confront them with our own requirements at the scale of the RDF properties in section V.

In order to define the KGAcc framework to measure the accountability of KG, we base our work on the LiQuID metadata model [1] which considers datasets in general. It offers a way for datasets to represent accountability meta-information throughout their life cycle. The model has been validated based on a real-world workload that relies on existing regulations (such as the GDPR) and an expert survey. To our knowledge, it is the only one to provide such a precise and explicit list of accountability requirements, presented in the form of questions that describe the model. Our framework adapts the hierarchy and the associated questions of LiQuID, taking into account the expressiveness of the most common vocabularies. It provides the requirements as a set of SPARQL queries corresponding to the questions. Our very first experiments are shortly reported in [17]. The work described in this paper relies on (*i*) new experiments that enable to distinguish between each dataset of a SPARQL endpoint, and *ii*) improved queries, taking into account more vocabularies. In addition, we provide a thorough comparison with other evaluation frameworks [4], [5], [6], [7], [9].

Finally, in order to conduct our experiments and to query KGs with our own set of queries, several frameworks can be used. Luzzu [5] and Sieve [18] enable users to choose metrics among those defined and to declare new ones. Monitoring tools, such as SPARQLES [14], also enable assessing some quality aspects. Instead of these, we choose the IndeGx framework [11] because it relies entirely on SPARQL queries, unlike Sieve and Luzzu, and is easily extendable. The IndeGx framework enables querying many KGs, with multiple queries, and storing the results in RDF. Its primary use case is to build an index of KGs and thus to extract and compute various information about them using a SPARQL-based test suite. To evaluate KG accountability, we use it as an engine to submit our own queries to KGs and to store the evaluation results in RDF.

III. Accountability Requirements and Metric

In this section, we define the KGAcc framework. We define the requirements of knowledge graphs accountability, i.e. the precise information that KGs must contain. To be as unambiguous as possible, we go one step further in expressing these requirements using SPARQL queries. Finally, we formally define the metric of accountability. Our proposal is based on the LiQuID metadata model [1] that enables the representation of information related to the accountability of datasets. To illustrate the use of their model, they provide precise questions that a dataset must answer to be considered accountable. LiQuID is not specific to any type of dataset, so it is necessary to adapt it.

A. The LiQuID Metadata Model of Accountability

The LiQuID metadata model relies on a hierarchical structure. First, it covers all steps of a dataset's life cycle: data collection, processing, maintenance and usage. Then, each life cycle step is structured according to different question types: why, who, when, where, how and what. Finally, each question type is divided into different fields of information level: description, explanation, legal and ethical considerations and limitations. The authors provide an exhaustive list of questions to describe each aspect of this hierarchy. For instance, for "data

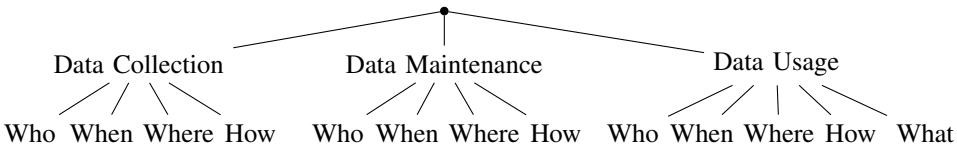


Fig. 1. The KGAcc Hierarchy of Requirements of Accountability, adapted from LiQuID [1] to Fit the Context of Knowledge Graphs

TABLE I
ACCOUNTABILITY REQUIREMENTS CONCERNING DATA USAGE:
ORIGINAL QUESTIONS FROM LIQUID AND THE ADAPTED ONES IN THE KGACC FRAMEWORK

	Questions from LiQuID	**KGAcc Questions**	**Weight**
Usage. Who	Who publishes this data set?	Who publishes this KG?	1
	Who has used/ can use the published data set?	Who has the right to use the published KG?	1/2
		Who is intended to use the published KG?	1/2
Usage. When	When can/ was the published data set be used?	Since when was the KG available?	1
	When is it available?	Until when is the KG available?	1
	Until what point in time is it valid?	Until when is the KG valid?	1
Usage. Where	Where is the data set published/ available?	What is the webpage presenting the KG and/or allowing to gain access to it?	1/2
		Where to access the KG (either through a dump or a SPARQL endpoint)?	1/2
	Where (place, geographically) can the published data set be used?	In what physical location can the KG be used?	1

processing", question type "when", the question associated with the field "description" is "On what date(s) or time frame(s) has the data been processed?". The LiQuID approach proceeds in a very systematic way and requires a large amount of very detailed information, representing what data sources should expose to be as accountable as possible.

B. Adaptation of LiQuID for Knowledge Graphs

Ideally, to assess the accountability of a KG, we should consider all LiQuID questions. However, it is not possible for all questions to be adapted for KGs and translated into SPARQL queries.

Indeed, as shown by Oppold and Herschel [1], the two general metadata models Dublin Core[3] and PROV [19], cannot cover all the fields proposed by LiQuID. According to them, both models "contain few fields, some of them too general to be mapped to specific LiQuID fields". We make the same observation with other general metadata models used in KGs, especially if the task is not to provide the information required by the model, but to query it. As a consequence of this lack of expressiveness, some questions cannot be translated into queries. As an example, some fields of the information level require too specific information, such as "Why is it lawful to collect this kind of data?", which, to our knowledge, cannot be expressed in a KG using existing vocabularies. As another example, two questions result in the same query for different steps of the life cycle, this is in particular the case for the collection and processing steps.

Faced with these difficulties, we opt for a soft strategy in which the maximum score of accountability seems attainable to us. It consists in keeping only questions compatible with the most common vocabularies of the semantic web. Therefore, we make the following adaptations: *(i)* only the field "description" of the information level is considered, *(ii)* the data processing step of the life cycle level is merged into the data collection step, *(iii)* the question types "why", "what" in "data collection" and "what" in "data maintenance" are not considered, and *(iv)* two questions concerning the exact methods and tools used for creation and maintenance are not considered in favor of more flexible questions concerning the methodology or procedure only. The resulting hierarchy is shown in Figure 1. As for the rest of the paper, we omit the last level, as it only contains the "description" element.

This definition of the accountability requirements, guided by the desire to ask reasonable questions, leads to a core set of 23 LiQuID questions (out of 207). We then define the KGAcc requirements by adapting these questions to the context of KGs. We make them more precise, and divide them into smaller parts, so they focus on only one element each. This precision is made as faithfully as possible, with the aforementioned limitations. Table I illustrates this adaptation. Therefore, the KGAcc framework results in 30 questions: 5 for Data Collection, 5 for Data Maintenance, and 20 for Data Usage. The totality of the questions is available on GitHub[4].

C. SPARQL Implementation of the Questions

Once the questions have been defined, each of them is translated into a SPARQL query or a succession of SPARQL queries. We use more than ten vocabularies of reference, chosen regarding their relevance to describe datasets and concepts around: VoID [20] is used to express metadata about

[3]https://dublincore.org/specifications/dublin-core/dcmi-terms/

[4]https://github.com/Jendersen/KG_accountability/tree/v2.0/docs

RDF datasets. DCAT[5] and DataID[6] allow the description of datasets and catalogs of datasets. SPARQL-SD [21] enables to describe SPARQL endpoints. These vocabularies rely on other general vocabularies, the Dublin Core, FOAF[7] and SKOS[8]. We also use PROV-O and PAV [22] for provenance issues. DQV[9] is used to describe the quality of datasets. Finally, we use schema.org a very general and widely used vocabulary, and some specific vocabularies for licenses, such as Creative Commons[10]. Each query uses all coherent properties and classes of these vocabularies to be as complete as possible. Listing 1 shows an example of a question translated into a query, where `?kg` must be replaced by the IRI of the knowledge graph at hand.

Listing 1. Extended 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 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 .} }
```

As queries are associated with questions requiring answers, they are "ASK" queries. The answer TRUE is considered a success, as it means that the KG contains the desired information. On the opposite, the answer FALSE or an error (e.g., timeout exception) is a failure, as it means the KG is unable to provide the wanted information.

Finally, notice that to express our precise requirements, we can either use the queries in their extended version, including all possible ways of expressing the required information, as in Listing 1. Alternatively, we can express the requirements in the form of a compact query, as in Listing 2, completed with a set of equivalences between properties (and between more complex graph patterns if necessary).

Listing 2. Compact query associated with "Who publishes this dataset?"

```
PREFIX dct: <http://purl.org/dc/terms/>
ASK { ?kg dct:publisher ?publisher . }
```

D. Definition of the Metric

First, we define the score obtained for each question. Then it is possible to determine the score of each node of the KGAcc hierarchy defined in Figure 1, from the bottom to the top. The score at the top of the hierarchy is the overall accountability score.

[5]https://www.w3.org/ns/dcat
[6]http://dataid.dbpedia.org/ns/core
[7]http://xmlns.com/foaf/spec/
[8]https://www.w3.org/TR/skos-reference/
[9]https://www.w3.org/TR/vocab-dqv/
[10]http://creativecommons.org/ns

A successful query gives a score of 1 to its associated question, while a failure gives 0. The score of a question associated with a succession of queries is the average of the score given by each query. The accountability score of a leaf of the KGAcc hierarchy (e.g. "data usage - who") is the weighted average of the scores obtained for its associated questions. Notice that LiQuID does not weight its questions, which suggests they are of equal importance. To stay close to this, we use to the following rule. When m ($m \geq 1$) KGAcc questions come from a same LiQuID question, a weight of $1/m$ is associated to each of them. Table I illustrates these weights. For instance, "data usage - who" has three questions, coming from two LiQuID questions. The first leads to one question, so its weight is 1, and the second leads to two questions, therefore their weight is $1/2$ each. The accountability score of "data usage - who" is the weighted average of these three questions, with the weights of 1, $1/2$, and $1/2$. For the other elements of the hierarchy, we determine their score by computing the (non-weighted) average of the scores of the elements underneath.

Formally, let g be a knowledge graph, ℓ a leaf node of the KGAcc hierarchy (e.g. "data usage - who"), and let $Q(\ell)$ denote all questions associated with ℓ. With $score$ a function giving the score of g for a given question q, w_q the weight of question q, the accountability score of g w.r.t. ℓ is:

$$accountability(g, \ell) = \frac{\sum_{q \in Q(\ell)} w_q \cdot score(g, q)}{\sum_{q \in Q(\ell)} w_q} \quad (1)$$

and the score of a given node n of the KGAcc hierarchy which is not a leaf is:

$$accountability(g, n) = \frac{\sum_{n' \; child \; of \; n} accountability(g, n')}{number \;\; of \;\; children \;\; of \;\; n} \quad (2)$$

In particular, the global accountability score is the score for the upper node in the hierarchy.

IV. Experimentation and Results

In this section, we describe our experiments. First, we detail the method employed to conduct an evaluation campaign of several KGs. Then, we discuss two aspects of the results. First, we examine the capabilities of knowledge graphs with regard to accountability. Second, we discuss the measure itself and the relevance of the KGAcc questions. All our queries and results are publicly available on our GitHub repository[11].

A. Processing and Tool

To evaluate a knowledge graph, we first need to identify its IRI within its own data. Then, we proceed in several stages to evaluate how the KG answers the queries defined in the previous section.

An important prerequisite of all our queries is to identify the IRI that the studied KG uses to refer to itself, or more precisely, the IRIs of the datasets it contains. Indeed, this IRI is the subject of at least one triple in all our queries, as illustrated

[11]https://github.com/Jendersen/KG_accountability/tree/v2.0

in Listing 1. Therefore, a query looking for the IRI is defined and presented in Listing 3, where `$rawEndpointUrl` is replaced by the URL of the endpoint during evaluation. If the KG does not provide an answer to this query, it will not answer any of our queries.

Listing 3. Query to identify the IRI of the studied KG

```
SELECT ?kg WHERE {
  ?kg ?endpointLink $rawEndpointUrl .
  { ?kg a dcat:Dataset }
  UNION { ?kg a void:Dataset }
  UNION { ?kg a dcmitype:Dataset }
  UNION { ?kg a schema:Dataset }
  UNION { ?kg a sd:Dataset }
  UNION { ?kg a dataid:Dataset }
}
```

In order to reduce the complexity of the queries sent to KGs, we focus on the accountability requirements in their compact form (cf. Listing 2). So, we proceed according to the following steps. For each KG, we first extract the triples corresponding to its metadata. As explained before, we use queries that begin with the lines described in Listing 3 to identify them. Then, we saturate this description of the KG by adding equivalent properties and classes, as defined by the requirements. Finally, we evaluate a KG according to this saturated metadata, using compact queries.

To carry out all these steps, we use the framework IndeGx[12]. It relies on a SPARQL-based test suite and can pre-process some steps for a better scalability. To use it, we provide SPARQL queries and configure the actions to be taken based on their results, i.e. which triples to write or update in the resulting RDF graph. So, we embed a set of queries into the framework, following the steps previously detailed, and declare how to store the result (True or False) for each evaluation query and for each KG using the DQV Vocabulary.

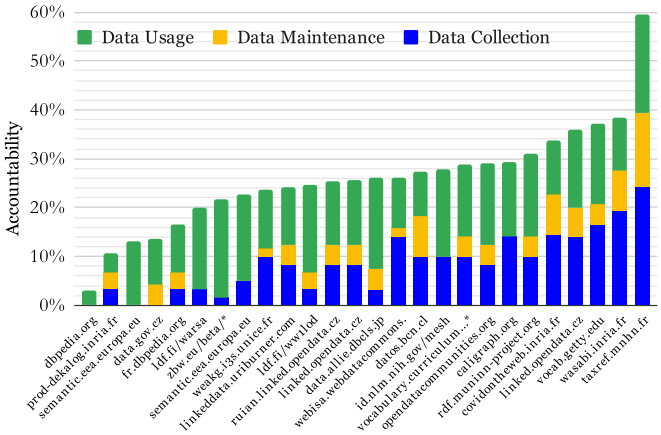
B. Querying of the SPARQL Endpoints

Our experiments query the 336 SPARQL endpoints already identified by IndeGx, extracted from LOD Cloud, Wikidata, SPARQLES, Yummy Data, and Linked Wiki in February 2023. These endpoints were queried at three different time points in June 2023. For each endpoint, only the results of an experiment for which it was available are kept. In this way, KGs are not penalized if they were unavailable at a given time. All endpoints not succeeding the query of Listing 3 are assigned an accountability score of 0, as no triple concerning its KG could be extracted.

Finally, given the results obtained for each query and thus each question, the accountability score can be computed. As defined in subsection III-D, an average is used to calculate the score for each aspect of the KGAcc hierarchy of Figure 1, until the overall accountability score is obtained.

C. Analysis of the Results

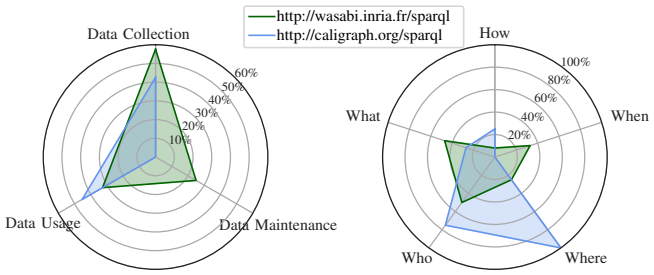
Among the 336 endpoints tested, only 26 successfully provide accountability information (Listing 3). The others were unavailable or did not provide easily accessible meta-information within their data. Among the 26 endpoints, 166 different datasets were identified (in the sense of `dcat:Dataset` or `void:Dataset...`), with accountability scores varying between 3.1% and 59%, with an average score of 26%. Even though most of the KGs do not provide any accountability information, the distribution of the values shows that this measure allows to discriminate between the datasets and the 26 endpoints left. On average, KGs are more accountable concerning "data usage" (41%) than "data collection" (25%), and twice better on "data collection" than on "data maintenance" (12%).



All URLs start with http(s):// and, except for *, end with /sparql.

Fig. 2. Accountability Score Obtained by the Best Dataset of Each Evaluated Endpoint, Detailed according to the Three Life Cycle Steps.

Figure 2 shows the accountability score of the best dataset of each endpoint. This score is divided according to the three life cycle steps "data collection", "data maintenance" and "data usage". It is possible to compare two datasets in more detail. As an example, Figure 3 shows the strengths and weaknesses of the main dataset of http://caligraph.org/sparql and of http://wasabi.inria.fr/sparql according to the life cycle steps (3a) and more precisely on the question types of the "data usage" step (3b).



(a) Accountability of Two KGs w.r.t. the Life Cycle Steps

(b) Accountability of Two KGs w.r.t. the Question Types of "Data Usage"

Fig. 3. Accountability of Two KGs w.r.t. Different Elements of the Hierarchy

The small number of KGs having a non-zero accountability score is not surprising. This observation is in line with other

[12]https://github.com/Wimmics/dekalog

results [23] showing that less than 10% of KGs provide self-descriptions within their data. However, for some KGs, it is possible that some meta-information may be present outside of the KG itself, for instance on their web page, or inside the KG but not findable in the way shown in Listing 3. While not taking them into account may penalize some KGs, it points out the fact that they are less transparent because the information is less accessible.

Several reasons may explain the scores obtained on the different life cycle steps. The "data usage" step covers general description elements that are widely used such as a description, a publisher, or a license, that more than half of the 26 endpoints provide. Furthermore, it encompasses all questions involving VoID vocabulary, which are each answered at least once by more than 50% of the endpoints on average. "Data usage" also requires a link to the endpoint or a dump, so having an answer to this question is expected considering how we identify the IRI of the KG. "Data maintenance" usually has bad scores. This may be due to the fact that half of the questions have only one possible property, with no alternative. For instance, the modification frequency can only be obtained with the property `accrualPeriodicity` from the Dublin Core vocabulary. This lack of alternative solutions to express this concept makes it more difficult to answer the query and highlights the fact that the question is more unusual. "Data creation" has various results with very common requirements, such as the creator and the creation date, and more difficult questions to answer such as the creation method.

D. Discussion about the KGAcc Framework

As far as the framework is concerned, at least two questions can be asked: are the requirements relevant? Are they too demanding? Figure 4 provides some answers. It represents the distribution of the values of accountability of the best dataset of each endpoint w.r.t. each aspect of the KGAcc hierarchy. Each box represents the first quartile (Q1), the median, and the third quartile (Q3) of the values obtained on the different aspects and the whiskers indicate the minimum and the maximum values obtained. It shows that the question types of "data usage" usually have good values in the different KGs. It also shows that half of the aspects can be fully covered, including "data collection - who", "data collection - when" and "data collection - how" for instance.

On the one hand, as many of the aspects sometimes get the maximum score, it shows that these queries are relevant and that KGs have a good margin to improve themselves. On the other hand, some of the low scores observed on that figure may be explained by too demanding queries. Indeed, it is important to notice that 6 out of 30 queries never succeed. For instance, in "data usage - when" the end date of availability of the dataset may be difficult for providers to specify, as they may consider that their KGs will be available indefinitely. Other questions concerning locations may not be in line with the current practices. Indeed, they are especially important for KGs that hold private information, which is generally not the case for public SPARQL endpoints. As other frameworks, depending on the context, KGAcc may be discussed and improved with experts and KG providers.

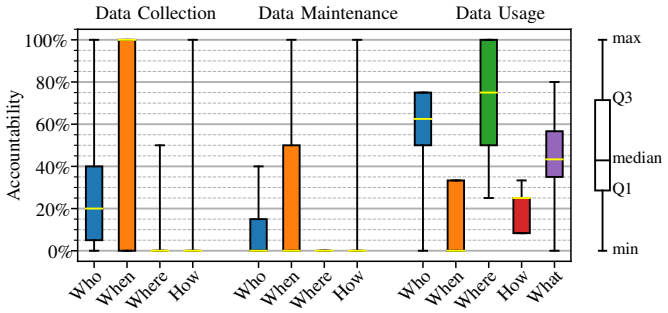


Fig. 4. Box Plot Showing the Distribution of the Values of Accountability w.r.t. Each Aspect of the Hierarchy.

V. Comparison with some Evaluation Frameworks

To take the analysis of our framework a step further, we compare it in detail with several data quality and FAIRness assessment frameworks. The comparison is made at the level of the required properties: **we aim to verify to what extent other studies require the properties demanded by the KGAcc framework**. To do so, we focus on studies that consider RDF properties and RDF datasets, and that either provide an open access implementation or that describe the metrics in sufficient detail to allow comparison. This is why works such as F-UJI [8] or Sieve [18] are not considered.

Concerning data quality, Zaveri et al. [6] provide an organized list of metrics obtained by a systematic literature review. This work is theoretical and does not implement these metrics, therefore, we do not take it into account. However, we focus on two major studies of data quality inspired by Zaveri et al. First, Färber et al. [4] evaluate the data quality of five cross-domain KGs, namely DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. While the implementation of their metrics is not available, each metric is richly described. Secondly, Debattista et al. [5] provide a more generic set of data quality metrics enabling the evaluation of any KG. Their implementation is available online but the article [5] describing them is more detailed and understandable. Therefore, for all these studies, we base our comparison solely on the referenced article.

For FAIRness, FAIR-checker [9] is interested in RDF triples as embedded metadata in web pages. For comparison, we rely on the specifications provided by the online tool[13] when evaluating a resource. We also consider O'FAIRe [7] which focuses on RDF ontologies. It provides an online tool[14] to see the results obtained by a list of ontologies. To compare with them, we consider the complete list of questions and their required properties[15]. In total, this leads us to consider two data quality studies and two FAIRness ones.

[13]fair-checker.france-bioinformatique.fr/check Accessed: 10 April 2023
[14]agroportal.lirmm.fr/landscape#fairness_assessment
[15]https://github.com/agroportal/fairness/blob/master/doc/results/FAIR-questions.md Accessed: 10 April 2023

TABLE II
COMPARISON BETWEEN THE KGACC QUERIES AND THE METRICS PROPOSED BY THE WORKS ON DATA QUALITY AND FAIR

Lifecycle step	Question type	Accountability query	Data Quality		FAIR	
			Debattista [5]	**Färber** [4]	**FAIR-Checker** [9]	**O'FAIRe** [7]
Data Collection	Who	Creator	✓		⊂	✓
		- Creator's info.				
	When	Creation date			⊂	⊂
	Where	Source		≈	⊂	✓
		Creation location			⊂	
	How	Creation method			⊂	✓
Data Maintenance	Who	Contributor			⊂	✓
		- Contributor's info.				
	When	Modification date		≈	⊂	⊂
		Frequency				✓
	Where	Modification location				
	How	Modification method			⊂	
Data Usage	Who	Publishers	✓			⊂
		Usage rights	✓	✓	✓	✓
		Audience				⊂
	When	Start of availability				
		End of availability				
		End of validity				⊂
	Where	Webpage				⊂
		Access URL		✓		⊂
		Usage location	✓	✓	✓	✓
	How	License	✓	✓	✓	✓
		Access URL		✓		⊂
		- Endpoint's info.				
		Usage information				
		Usage requirements				
	What	Examples				
		Concepts				⊂
		Description				⊂
		Triples				
		Entities prop. classes				
		Serialization	✓			
		Quality				⊂

✓ Property required and must concern the dataset (or the ontology).
≈ Property required but not linked to any particular resource.
⊂ One of the required properties is listed in the metric among other semantically different properties.

Table II summarizes our comparative study. For each KGAcc query, if one of its required properties is also required in a data quality or FAIR metric, a mark is indicated in the table. If this property must not necessarily concern the KG (e.g. the creator of a resource instead of the creator of the dataset), then the mark is ≈, showing that the FAIR or data quality metric is not really related to dataset accountability. Otherwise, if this property is mandatory to obtain a maximum score on the data quality or FAIR metric, the mark is ✓. If the property is listed among other properties and only one or two or n of these properties are necessary for success, then the mark is ⊂. For instance, in O'FAIRe the third question for principle F2 states that to obtain the maximum score, six properties should be used from a list of 37 properties, whatever those six properties may be. Therefore, unlike the ✓ mark, a ⊂ mark does not guarantee that passing the FAIR metric ensures passing the accountability query.

This table highlights several elements concerning data quality metrics. First, as part of the measure of accessibility, all these evaluations require a license to be present using properties such as `dcterms:licence` [4], [5]. They also demand some particular provenance information: the creators or the publishers of the KG [5], or other information not specifically related to the KG such as the source of some data to enhance trustworthiness [4], their modification dates [4], or traceability of the data [5]. Some other meta-information is expected to be provided, such as the serialization formats [5]. Finally, Färber et al. [4] request the provision of KG metadata citing as an example the URI of the SPARQL endpoint or the RDF export URL to indicate where to access the data.

Concerning FAIR metrics, only two metrics are related to accountability in FAIR-checker. The first one measures the 'R1.1' principle that requires a license. The second one measures the provisioning of provenance information (R1.2) by checking that at least one of the 23 listed properties is found (such as `prov:wasDerivedFrom`, `pav:createdBy`, etc.). O'FAIRe offers more similarities with our work. There are mainly two different kinds of metrics of interest compared to our queries. First, the metrics concerning the reusability principles focus on one information each and are mostly also required by accountability (creator, contributor, source, method, periodicity, license and

rights). Secondly, some metrics of findability require that the ontology uses some well-known properties. Indeed, two questions of the principle 'F2' cover properties required by at least 11 accountability queries (such as `dct:created`, `void:dataDump`...).

As a result, both data quality and FAIRness share common interests with accountability. Therefore, improving them may have a positive impact on the assessment of accountability and vice versa. However, neither data quality nor FAIRness focuses specifically on accountability as a whole and does not take into account all the elements it requires. The general studies on data quality only slightly overlap with accountability. FAIRness has more similarities, particularly with regard to the steps of data collection and maintenance as it is mainly interested in questions of provenance. In particular, O'FAIRe seems to have many similarities. However, most of them result solely from the two findability metrics, which are not very informative about the type of metadata present, since they cover no more than 11 of our queries. Therefore, the measure of accountability is much more detailed, precise, and focused than O'FAIRe on our point of interest. And as the result of each query is available, the former provides a much more relevant view of accountability.

VI. Conclusion

In order to evaluate the accountability of RDF graphs, we proposed : *(i)* the KGAcc framework defining requirements concerning the metadata that the KGs should expose and an associated metric, *(ii)* the evaluation of a large set of endpoints, *(iii)* a comparison of our approach with other frameworks that assess data quality or compliance to the FAIR principles.

The KGAcc requirements are expressed as SPARQL queries. They are obtained through a meticulous adaptation of an existing hierarchy of natural language questions, proposed for datasets in general [1], to the specific context of KGs. Indeed, the dedicated vocabularies make easy stating some requirements. But their lack of expressiveness makes some questions collapse into a same query or prevents from considering demanding and precise questions about ethical and legal questions. This is why we end up with a relatively small set of queries.

The evaluation of many RDF graphs reveals that most of them do not provide any of the required information within their data, even though some of the required information is very commonly requested. However, there are RDF graphs that provide some of the expected information, showing our demands are reasonable. Our comparative study shows in particular that O'FAIRe is the framework considering the most properties common to accountability. However, it is not so demanding in terms of accountability and cannot be considered as a framework dedicated to this aspect.

In future works, to improve our measure, we will introduce some weights to aggregate the results differently, and we will propose an online visualization of the results. Finally, it could be interesting to automate some tasks, such as defining the equivalences.

References

[1] S. Oppold and M. Herschel, "Accountable data analytics start with accountable data: The LiQuID metadata model." in *ER Forum/Posters/Demos*, 2020, pp. 59–72.

[2] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman, "Information accountability," *Communications of the ACM*, vol. 51, no. 6, pp. 82–87, 2008.

[3] D. Wyatt, "The many dimensions of transparency: A literature review," *Helsinki Legal Studies Research Paper*, no. 53, 2018.

[4] M. Färber, F. Bartscherer, C. Menne, and A. Rettinger, "Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago," *Semantic Web*, vol. 9, no. 1, pp. 77–129, 2018.

[5] J. Debattista, C. Lange, S. Auer, and D. Cortis, "Evaluating the quality of the lod cloud: An empirical investigation," *Semantic Web*, vol. 9, no. 6, pp. 859–901, 2018.

[6] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer, "Quality assessment for linked data: A survey," *Semantic Web*, vol. 7, no. 1, pp. 63–93, 2016.

[7] E. Amdouni, S. Bouazzouni, and C. Jonquet, "O'FAIRe: Ontology FAIRness evaluator in the agroportal semantic resource repository," in *ESWC 2022-19th Extended Semantic Web Conference, Poster and demonstration.*, 2022.

[8] A. Devaraju and R. Huber, "An automated solution for measuring the progress toward fair research data," *Patterns*, vol. 2, no. 11, 2021.

[9] T. Rosnet, F. de Lamotte, M.-D. Devignes, V. Lefort, and A. Gaignard, "FAIR-checker - supporting the findability and reusability of digital life science resources," 2021.

[10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[11] P. Maillot, O. Corby, C. Faron, F. Gandon, and F. Michel, "IndeGx: A model and a framework for indexing RDF knowledge graphs with SPARQL-based test suits," *Journal of Web Semantics*, p. 100775, 2023.

[12] M. Rowe and J. Butters, "Assessing trust: Contextual accountability." in *SPOT@ ESWC*, 2009.

[13] I. Naja, M. Markovic, P. Edwards, and C. Cottrill, "A semantic framework to support ai system accountability and audit," in *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18*. Springer, 2021, pp. 160–176.

[14] P.-Y. Vandenbussche, J. Umbrich, L. Matteis, A. Hogan, and C. Buil-Aranda, "Sparqles: Monitoring public sparql endpoints," *Semantic web*, vol. 8, no. 6, pp. 1049–1065, 2017.

[15] Y. Yamamoto, A. Yamaguchi, and A. Splendiani, "Yummydata: providing high-quality open life science data," *Database*, vol. 2018, 2018.

[16] S. Issa, O. Adekunle, F. Hamdi, S. S.-S. Cherfi, M. Dumontier, and A. Zaveri, "Knowledge graph completeness: A systematic literature review," *IEEE Access*, vol. 9, pp. 31 322–31 339, 2021.

[17] J. Andersen, S. Cazalens, and P. Lamarre, "Assessing knowledge graphs accountability," in *The Semantic Web: ESWC 2023 Satellite Events*. Hersonissos, Greece: Springer, 2023.

[18] P. N. Mendes, H. Mühleisen, and C. Bizer, "Sieve: linked data quality assessment and fusion," in *Proceedings of the 2012 joint EDBT/ICDT workshops*, 2012, pp. 116–123.

[19] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, D. Corsar, J. Cheney, D. Garijo, S. Soiland-Reyes, S. Zednik, and J. Zhao, "PROV-O: The PROV Ontology," *W3C Recommendation. W3C*, 2013. [Online]. Available: https://www.w3.org/TR/prov-o/

[20] K. Alexander, R. Cyganiak, M. Hausenblas, and J. Zhao, "Describing linked datasets with the VoID vocabulary," *W3C Note. W3C*, 2011. [Online]. Available: https://www.w3.org/TR/void/

[21] G. T. Williams, "Sparql 1.1 service description," *W3C Recommendation. W3C*, 2013. [Online]. Available: https://www.w3.org/TR/sparql11-service-description/

[22] P. Ciccarese, S. Soiland-Reyes, K. Belhajjame, A. J. Gray, C. Goble, and T. Clark, "Pav ontology: provenance, authoring and versioning," *Journal of biomedical semantics*, vol. 4, no. 1, pp. 1–22, 2013.

[23] P. Maillot, O. Corby, C. Faron, F. Gandon, and F. Michel, "KartoGraphI: Drawing a Map of Linked Data," in *ESWC 2022 - 19th European Semantic Web Conferences*. Hersonissos, Greece: Springer, May 2022. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03652865