



## RESEARCH ARTICLE

# **REVISED** An empirical study on Resource Description Framework reification for trustworthiness in knowledge graphs [version 2; peer review: 2 approved]

Sini Govindapillai <sup>1</sup>, Lay-Ki Soon <sup>2</sup>, Su-Cheng Haw <sup>1</sup>

<sup>1</sup>Faculty of Computing Informatics, Multimedia University, Cyberjaya, Selangor, 63100, Malaysia

<sup>2</sup>School of Information Technology, Monash University Malaysia, Bandar Sunway, Selangor, 47500, Malaysia

**V2** First published: 02 Sep 2021, 10:881  
<https://doi.org/10.12688/f1000research.72843.1>

Latest published: 29 Nov 2021, 10:881  
<https://doi.org/10.12688/f1000research.72843.2>

## Abstract

Knowledge graph (KG) publishes machine-readable representation of knowledge on the Web. Structured data in the knowledge graph is published using Resource Description Framework (RDF) where knowledge is represented as a triple (subject, predicate, object). Due to the presence of erroneous, outdated or conflicting data in the knowledge graph, the quality of facts cannot be guaranteed. Trustworthiness of facts in knowledge graph can be enhanced by the addition of metadata like the source of information, location and time of the fact occurrence. Since RDF does not support metadata for providing provenance and contextualization, an alternate method, RDF reification is employed by most of the knowledge graphs. RDF reification increases the magnitude of data as several statements are required to represent a single fact. Another limitation for applications that uses provenance data like in the medical domain and in cyber security is that not all facts in these knowledge graphs are annotated with provenance data. In this paper, we have provided an overview of prominent reification approaches together with the analysis of popular, general knowledge graphs Wikidata and YAGO4 with regard to the representation of provenance and context data. Wikidata employs qualifiers to include metadata to facts, while YAGO4 collects metadata from Wikidata qualifiers. However, facts in Wikidata and YAGO4 can be fetched without using reification to cater for applications that do not require metadata. To the best of our knowledge, this is the first paper that investigates the method and the extent of metadata covered by two prominent KGs, Wikidata and YAGO4.

## Keywords

Wikidata, YAGO, RDF reification, Knowledge Graph, provenance data

## Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
<b>version 2</b>		
(revision)		
29 Nov 2021	report	report
<b>version 1</b>		
02 Sep 2021		
	report	report

1. **Kok-Chin Khor**, Tunku Abdul Rahman University, Kajang, Malaysia

2. **Weiguo Zheng**, Fudan University, Shanghai, China

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Research Synergy**  
**Foundation gateway.**

**Corresponding author:** Su-Cheng Haw ([sucheng@mmu.edu.my](mailto:sucheng@mmu.edu.my))

**Author roles:** **Govindapillai S:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation; **Soon LK:** Project Administration, Resources, Supervision, Writing – Review & Editing; **Haw SC:** Project Administration, Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work is partially funded by Fundamental Research Grant Scheme (FRGS) by Malaysia Ministry of Higher Education (Ref: FRGS/2/2013/ICT07/MMU/02/2) and Multimedia University Internal Fund (MMUI/180006).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2021 Govindapillai S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Govindapillai S, Soon LK and Haw SC. **An empirical study on Resource Description Framework reification for trustworthiness in knowledge graphs [version 2; peer review: 2 approved]** F1000Research 2021, 10:881 <https://doi.org/10.12688/f1000research.72843.2>

**First published:** 02 Sep 2021, 10:881 <https://doi.org/10.12688/f1000research.72843.1>

**REVISED Amendments from Version 1**

In version 2, we have revised the title to ensure a clearer reflection of the paper contents. In addition, for improving lucidity we have revised the structure of the abstract so that the analyses of Wikidata and YAGO4 are grouped together and moved towards the latter half of the abstract. For clarity, we have added the definition of trustworthiness in the introduction section. Besides, some of the applications where trustworthiness is crucial are also included as part of the abstract and introduction. In the methods section, the analysis of Wikidata is re-organized into the description of the knowledge graph, querying method, the data model, RDF reification, how Wikidata caters for trustworthiness using qualifiers for contextualization and provenance data and by distinguishing between unknown value and no value for a relation, inconsistency in the data model when metadata is included and finally discusses the need for a balance in contextual information for easy inference of trustworthiness. Future work is specified at the end of the discussion section. Lastly, we have elaborated the conclusion section to include additional details from the discussion section.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

Knowledge regarding the real-world entities in machine-readable format is furnished by Knowledge Graphs (KGs). These large-scale KGs provide both domain-dependent and domain-independent knowledge for many applications like entity linking, information retrieval and several other data mining tasks. KGs can be created by the extraction of structured knowledge from data sources like Wikipedia, collected by Artificial Intelligent (AI) projects, imported from other data sets, or by crowd-sourcing. Regardless of the methods followed, the presence of noise diminishes the quality of the KGs.<sup>1</sup>

Trustworthiness can be defined as “the degree to which the information is accepted to be correct, true, real, and credible”.<sup>2</sup> Trust in the data can be increased by providing additional information like the source of information or contextual information like the time or the location in which this fact was true or any other relevant additional information pertaining to a fact.<sup>3</sup> This extra information would support the authenticity of the data, and in return, can help the machines to extract correct facts for critical applications such as in medical domain and in cyber security.

However, data in KGs are published using Resource Description Framework (RDF)<sup>4</sup> where RDF encodes facts in the form of triples. RDF statement is an ordered set of the form (s, p, o) where s and o describe the subject and object entities and p describes the relationship between these two entities.

RDF triple can be represented as:

$$(s, p, o) \in (I \cup B) \times I \times (I \cup B \cup L)$$

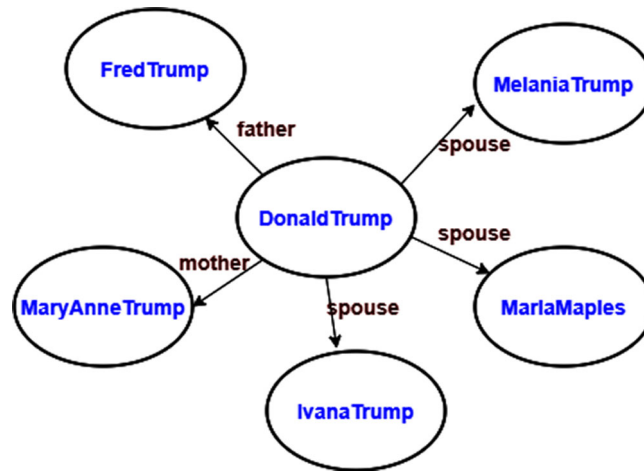
- I is the set of Internationalized Resource Identifiers (IRIs), which identifies a unique resource on the Web
- L is the set of RDF Literals,
- B is the set of blank nodes that are anonymous and cannot be dereferenced globally.

Figure 1 depicts examples of personal facts of the entity Donald Trump expressed as RDF triples. Some facts are eternal truth such as an individual’s parents, while some other facts can be true at a certain point in time. This figure shows that Donald Trump has three wives, which is partially correct. Donald Trump was married to three entities at different points in time. This example shows the importance of contextual data and the inadequacy of RDF in its expression.

However, RDF data model does not support adding metadata to triples and therefore provenance data cannot be attached. Sikos and Philip<sup>5</sup> describes different scales of provenance for datasets. In this paper, we focus on the provenance and contextualization of RDF data at the triple level. We first look at the existing approaches for integrating metadata into RDF. Next, we explore two popular KGs, namely Wikidata<sup>6,7</sup> and YAGO<sup>8</sup> to see how and to what extent they support meta information.

**Reification approaches**

Trustworthiness on the KG level is dependent on the trustworthiness of the facts at the basic level, which is the triple level. Contextual data restricts or gives meaning to a specific situation, location or time where this fact is true. Provenance data add source from which this triple is curated or imported. Therefore, we hypothesize that both contextual and provenance data adds trust to the fact that is described.



**Figure 1. Sample RDF triples of the entity Donald Trump.**

In order to encapsulate this additional information to RDF triple, RDF reification is used. In this section, we will explore some of the available methods for reifying RDF triples. These approaches include metadata of statements as triples in KGs. Additionally, some of the popular reification approaches standard reification, singleton property, n-ary relations and RDF\* will be discussed.

**Standard reification**

Standard reification<sup>4</sup> annotates facts using RDF built-in vocabulary and the concept of blank nodes as described in RDF Primer. In standard reification, four triples are used for representing a fact. Blank nodes of type *Statement* together with properties subject, predicate and object are used for representing the triple. Additional provenance data can then be added to these triples using the same blank node.

For example, the fact *Serena Williams is born in Saginaw* can be represented as RDF statement (SerenaWilliams, birthPlace, Saginaw). The source of the triple is given as the book *History of Tennis by Bud Collins*. Standard reification employing blank node labelled *\_:x* is used for reification in Table 1. Provenance information is added as the last row in the table using the same blank node *\_:x*. Any number of metadata can be added with this blank node as the subject.

Here for representing a single fact, four triples are needed. This will magnify the size of the KG by four times. At the same time, queries for retrieving these metadata facts will become complicated. This simple method of reification does not have formal semantics to link reified and original triples, and it is not commonly used.

**Singleton property**

In order to overcome issues with standard reification, *singleton property*<sup>9</sup> was proposed. Singleton property provided formal semantics to RDF reification and the number of triples describing fact is reduced. This reification is based on the idea that the relation between two specific entities is unique. This unique statement-level property can be an instance of the general predicate using *singletonPropertyOf* property, which is a sub-property of *rdf:type*. This unique predicate is used to attach additional information to the fact like contextualization or provenance information.

**Table 1. Standard reification.**

Subject	Predicate	Object
_:x	rdf:type	rdf:Statement
_:x	rdf:subject	SerenaWilliams
_:x	rdf:predicate	birthPlace
_:x	rdf:object	Saginaw
_:x	dc:source	HistoryOfTennis

**Table 2. Singleton property.**

Subject	Predicate	Object
Serena Williams	birthplace_1	Saginaw
birthplace_1	rdf:singleton-PropertyOf	birthPlace
birthplace_1	dc:source	HistoryOfTennis

In the example given in [Table 2](#), birthplace\_1 is an instance of *singleton property* describing the relation birthPlace between entities Serena Williams and Saginaw. Each of these relations has a unique IRI. Any number of metadata can be added to this fact with singleton property birthplace\_1 as the subject. This will introduce a large number of unique predicates in the KG. This can be problematic for indexing techniques adopted by current triplestores.

**n-ary relations**

Another alternative to achieve reification is the use of the *n-ary* relations.<sup>6,10</sup> In this model, the subject has a relationship with some qualifiers and values. Resource node is used for this, but unlike standard reification, it is attached to the original triple.

Example in [Table 3](#) depicts *n-ary* relation of the triple (SerenaWilliams, birthPlace, Saginaw) where subject Serena Williams is related to a resource node, which cannot be dereferenced on the Web. This resource birthPlace\_1 is related to the actual object Saginaw of the original triple. This is done using properties which are of *statementProperty* and *valueProperty* and are constructed from the property birthProperty of original triple. Here five triples are required to depict a fact and any number of complex meta-information can be attached to the resource node.

**RDF\***

Another data model RDF\*<sup>11</sup> was introduced by Hartig in 2017. Unlike earlier approaches, which introduced metadata as triples, this data model follows embedded triple. The fact triple is enclosed within double brackets and is assigned as the subject or object of the triple as shown in [Table 4](#). In the table, the embedded triple, <<s,p,o>> refers to <<Serena Williams, birthplace, Saginaw>>. Any number of metadata can be added as the predicate and object of the triple. RDF\* is extended to have nested triples.

RDF\* data model is much more compact than other reification approaches and does not introduce any extra predicates like in singleton properties. It is also backward compatible with the RDF data model and other reification approaches. Although retrieving queries are simple compared to other approaches, various groups of annotations of the same fact (for example, the same incident happened at two different points in time) cannot be interpreted correctly. Frey et al. provide an alternative to overcome this by adding resource nodes for each annotation group.<sup>12</sup>

**Table 3. n-ary relations.**

Subject	Predicate	Object
Serena Williams	birthplaceSub	birthplace_1
birthplace_1	birthPlaceVal	Saginaw
birthplaceSub	statementProperty	birthPlace
birthPlaceVal	valueProperty	birthPlace
birthplace_1	rdf:type	rdf: Statement
birthplace_1	dc:source	HistoryOfTennis

**Table 4. RDF\*.**

Subject	Predicate	Object
<< s, p, o>>	dc:source	HistoryOfTennis

**Methods**

The trustworthiness of a KG is categorized into knowledge graph level, statement level and whether the statement indicates it does not have value or the value is not known.<sup>4</sup> In our research, we focus on trustworthiness defined at the statement level. Wikidata and YAGO provide statement-level provenance and contextual data, which is the focus of this paper and hence further explored in this section.

**Wikidata**

Wikidata is one of the largest KG in machine-readable format on the Web. This curated and crowd-sourced dataset has a rich set of entity information. Wikidata uses its own data model and data is inherently stored in JavaScript Object Notation (JSON) format in the BlazeGraph database. But in order to effortlessly query Wikidata like other KGs, Wikidata provides RDF dumps and has introduced Wikidata SPARQL query service for querying Wikidata in RDF form. A large number of facts in Wikidata are annotated with metadata and therefore, widely uses reification to describe the facts.<sup>2</sup>

Wikidata encodes facts in the form of item, label, description, alias and statement. Item describe an entity and different pieces of information about the entities are stored as primary relations (property-value pairs) of statements. This approach is very similar to the concept of representing facts as triple in RDF. Metadata of these facts are added by annotating statements with qualifiers, references and ranks.

Wikidata employs n-ary relations to cater for RDF annotations.<sup>6,7</sup> Each entity is linked to the statement node and the statement node is linked to the value nodes. Any number of references, contextualization data and ranks can be added to statement nodes. For applications that do not require the complexity of accessing these n-ary relations, data in plain RDF without annotations is provided by Wikidata. A simplified relation between subject and object is provided as RDF triples for top rank statements.

The most popular qualifier for describing contextual data is temporal data.


Wikidata resolves the awkward interpretation of many wives in [Figure 1](#) by attaching temporal information to supplement the facts as displayed in [Figure 2](#).

Wikidata allows property-value pairs to represent two special cases, one is if a property has a value, which is unknown and another is if a property does not have a value. *Some* is employed to represent an unknown value and *none* is employed to represent no value. This increases the trustworthiness of Wikidata.





The source of the statements is referred by using predicates reference URL (P854), imported from (P143), stated in (P248), title (P1476) or publisher (P123). Wikidata references are optional and may be Wikidata items, links to websites,

spouse	 <a href="#">Ivana Trump</a>	
	start time	7 April 1977
	end time	22 March 1992
	end cause	divorce
	▶ 1 reference	
	 <a href="#">Maria Maples</a>	
	end time	8 June 1999
	start time	19 December 1993
	end cause	divorce
	▶ 1 reference	
 <a href="#">Melania Trump</a>		
start time	22 January 2005	
place of marriage	Mar-a-Lago	
▶ 1 reference		

**Figure 2.** Qualifiers of statements of Donald Trump in Wikidata.

place of birth	 <b>Saginaw</b>						
	<div style="background-color: #e0e0e0; padding: 2px;">                 ▼ 2 references             </div> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #e0e0e0;">stated in</td> <td>The Bud Collins History of Tennis (2nd ed.)</td> </tr> <tr> <td style="background-color: #e0e0e0;">page(s)</td> <td>681</td> </tr> <tr> <td style="background-color: #e0e0e0;">reference URL</td> <td><a href="http://espn.go.com/tennis/player/results/_/id/394/serena-williams">http://espn.go.com/tennis/player/results/_/id/394/serena-williams</a></td> </tr> </table>	stated in	The Bud Collins History of Tennis (2nd ed.)	page(s)	681	reference URL	<a href="http://espn.go.com/tennis/player/results/_/id/394/serena-williams">http://espn.go.com/tennis/player/results/_/id/394/serena-williams</a>
stated in	The Bud Collins History of Tennis (2nd ed.)						
page(s)	681						
reference URL	<a href="http://espn.go.com/tennis/player/results/_/id/394/serena-williams">http://espn.go.com/tennis/player/results/_/id/394/serena-williams</a>						

**Figure 3. References of entity Serena Williams in Wikidata.**

position held	 <b>Lucasian Professor of Mathematics</b>									
	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #e0e0e0;">start time</td> <td>1979</td> </tr> <tr> <td style="background-color: #e0e0e0;">end time</td> <td>2009</td> </tr> <tr> <td style="background-color: #e0e0e0;">replaces</td> <td>James Lighthill</td> </tr> <tr> <td style="background-color: #e0e0e0;">replaced by</td> <td>Michael Green</td> </tr> </table>	start time	1979	end time	2009	replaces	James Lighthill	replaced by	Michael Green	
start time	1979									
end time	2009									
replaces	James Lighthill									
replaced by	Michael Green									
	<div style="background-color: #e0e0e0; padding: 2px;">                 ▶ 2 references             </div>									
	+ add value									
employer	 <b>Gonville and Caius College</b>									
	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #e0e0e0;">start time</td> <td>1965</td> </tr> <tr> <td style="background-color: #e0e0e0;">end time</td> <td>1969</td> </tr> <tr> <td style="background-color: #e0e0e0;">position held</td> <td>research fellow</td> </tr> </table>	start time	1965	end time	1969	position held	research fellow			
start time	1965									
end time	1969									
position held	research fellow									
	<div style="background-color: #e0e0e0; padding: 2px;">                 ▶ 1 reference             </div>									

**Figure 4. Excerpts of the position held by Stephen Hawking in Wikidata.**

classical citations or datasets. More than one references are possible as depicted in Figure 3, where the references of the birthplace of American female tennis player Serena Williams is provided and serves as the provenance information.

If there are many statements involved for the same relations Wikidata provides ranks for filtering the result. *Preferred*, *normal* and *deprecated* are provided as ranks in the descending order of precedence.

Wikidata uses the same predicate/property as part of primary relations and qualifiers. In Figure 4, the same predicate position held (P39) is used in primary relation (Stephen Hawking, position held (P39), Lucasian Professor of Mathematics) and as contextual information of the fact (Stephen Hawking, employer (P108), Gonville and Caius College). As we can see, here there is an inconsistency as to how to include the information of position held in a workplace.

Wikidata has more than a hundred and fifty qualifiers. Being a dynamic KG, qualifiers can keep getting added to the KG. Therefore Patel-Schneider<sup>13</sup> states that if the number of qualifiers attached to a fact is more, it is not trivial work to infer the trustworthiness of a fact. Therefore, limiting the context information to specific needs might be the smarter way for balancing the complexity of inferring trustworthiness.

### YAGO

YAGO4 employs RDF\* data model to annotate facts in RDF format. YAGO4 is a combination of rich instance data from Wikidata and the ontology from [schema.org](http://schema.org). YAGO is constructed with very stringent measures to control noise. Facts which does not abide by these strict constraints are rejected. Semantic constraints like disjointness, domain and range

restrictions, functional constraints on relations, cardinality constraint of objects are enforced on the dataset and has an accuracy of 95%. For the case in [Figure 1](#), YAGO4 keeps only statements with the best rank from Wikidata, thus avoiding stale facts in KG.

YAGO4 provides temporal information to the facts by employing nesting of triples using RDF\* data model. RDF\* provides SPARQL\* for easy accessing data on the Web. YAGO4 provides temporal dimension using two general predicates *startDate* and *endDate*, which can be applied to any class.

The previous version, YAGO2,<sup>14,15</sup> built from Wikipedia, GeoNames and WordNet is furnished with more contextualization and provenance data. YAGO2 employed a tuple type, sextuples, which is called SPOTLX. SPOTLX representation uses six tuples, with time, location and context added to the triple (s, p, o), which is of the form (Subject, Predicate, Object, Time, Location, conteXt).

## Discussion

Based on the investigation, Wikidata has better coverage of contextualization and provenance data of real-world entities. For applications which is less susceptible to noise and more sensitive to specific contexts, the role of this meta information is really tremendous. YAGO4 provides coverage for temporal information. YAGO2 is more enriched with spatial and temporal facts and is provided with many predicates for including meta information.

The discussion of metadata in KGs is not complete without mentioning the downside. These provenance and contextual data come with a cost. The size of the KGs, which are already in humongous size, will further increase with the addition of metadata. Besides, the size depends on the approach used by each KG and the type and the number of provenance and contextualization data employed. The complexity of data retrieval is also another issue to be considered.

In order to reduce the effect of these downsides, the data in KGs are provided without annotations too, that is without employing reification. Both Wikidata and YAGO4 are available in RDF format, without metadata information.

In another angle, for applications where provenance data is crucial, all facts in KGs are not furnished with annotations. Nevertheless, this issue is under constant improvement.

The investigation outcome paves the way for further research to apply provenance and contextualization information for identifying fully reliable and truthful facts from knowledge graphs.

## Conclusions

In this paper, we have discussed RDF reification methods for annotating knowledge graphs with metadata. Two prominent knowledge graphs, namely Wikidata and YAGO4 were investigated. The method of incorporating metadata by these KGs and the extent of contextualization and provenance data supported by them are analyzed. Wikidata is equipped with higher level of contextualization and provenance data whereas YAGO4 is equipped with temporal information. For applications which do not require metadata, both Wikidata and YAGO are available in RDF format too, and thus reducing the complexity of the addition of meta-information. The findings obtained will be further incorporated for the detection of errors in KG, in order to increase their trustworthiness.

## Data availability

The data are taken from Wikidata and Yago.

## Author contributions

Sini Govndapillai did the conception of the work, data collection, data analysis and interpretation, drafting the article, and revision to the final version under the guidance of her supervisors, Lay-Ki Soon and Su-Cheng Haw. Su-Cheng Haw is the corresponding author for this paper.

## References

1. Paulheim H: **Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods.** *Semant. Web.* 2017; 489–508. [Publisher Full Text](#)
2. Färber M, Ell B, Menne C, *et al.*: **Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO.** *Semant. Web.* 2017; 9: 77–129. [Publisher Full Text](#)



3. Bienvenu M: **Provenance for Web 2.0 Data.**  
[Publisher Full Text](#)
4. Manola F, Miller E, McBride B: **RDF Primer.** *W3C Recommendation 10 February 2004.* [Online].  
[Reference Source](#)
5. Sikos LF, Philp D: **Provenance-Aware Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs.** *Data Sci. Eng.* 2020; **5**(3): 293–316.  
[Publisher Full Text](#)
6. Erxleben F, Günther M, Krötzsch M, et al.: **Introducing wikidata to the linked data web.** *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).* 2014; vol. **8796**, pp. 50–65.  
[Publisher Full Text](#)
7. Malyshev S, Krötzsch M, González L, et al.: **Getting the Most Out of Wikidata: Semantic Technology Usage in Wikipedia's Knowledge Graph.** *Proc. 17th Int. Semant. Web Conf. (ISWC 2018).* 2018; vol. **11137**, pp. 8–12.
8. Hartig O: **Foundations of RDF\* and SPARQL\* (An Alternative Approach to Statement-Level Metadata in RDF).** *CEUR Workshop Proc.* June, 2017. vol. **1963**, no.
9. Nguyen V, Bodenreider O, Sheth A: **Don't like RDF reification? Making statements about statements using singleton property.** *WWW 2014 - Proc. 23rd Int. Conf. World Wide Web.* 2014; no. April, pp. 759–769.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Hayes P, Wely C, Noy N, et al.: **Defining N-ary Relations on the Semantic Web.**  
[Reference Source](#)
11. Hartig O: **RDF\* and SPARQL\*: An alternative approach to annotate statements in RDF.** *Int. Semant. Web Conf.* 2017.
12. Frey J, Müller K, Hellmann S, et al.: **Evaluation of metadata representations in RDF stores.** *Semant. Web.* 2019; **10**(2): 205–229.  
[Publisher Full Text](#)
13. Patel-Schneider PF: **Contextualization via qualifiers.** *CEUR Workshop Proc.* 2018; vol. 2317.
14. Hoffart J, Lewis-kelham E, Suchanek FM, et al.: **YAGO2: Exploring and Querying World Knowledge in Time , Space, Context, and Many Languages.** *Time.* 2011; **23**(7): 229–232.  
[Publisher Full Text](#)
15. Hoffart J, Suchanek FM, Berberich K, et al.: **YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia.** *IJCAI Int. Jt. Conf. Artif. Intell.* 2013; 3161–3165.  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 01 December 2021

<https://doi.org/10.5256/f1000research.79774.r101427>

© 2021 Khor K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Kok-Chin Khor

Lee Kong Chian Faculty of Engineering Science, Tunku Abdul Rahman University, Kajang, Malaysia

I have no other comments.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Data Mining, Computer Networks.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 30 November 2021

<https://doi.org/10.5256/f1000research.79774.r101428>

© 2021 Zheng W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Weiguo Zheng

Fudan University, Shanghai, China

The paper introduces the RDF reification methods. I have no more comments.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** knowledge graph

**I confirm that I have read this submission and believe that I have an appropriate level of**

expertise to confirm that it is of an acceptable scientific standard.

---

### Version 1

Reviewer Report 25 October 2021

<https://doi.org/10.5256/f1000research.76450.r95305>

© 2021 Zheng W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



#### Weiguo Zheng

Fudan University, Shanghai, China

#### General Comments:

This paper focuses on RDF reification and introduces several approaches with examples. Then the author investigated the reification methods used by Wikidata and YAGO. However, there is no new method proposed in this paper, and the reification model, the query method and the ranking method are not well organized when analysing Wikidata and YAGO.

#### Strong Points:

1. There are enough examples to make relevant approaches easy to understand. The examples of Donald Trump and Serena Williams are throughout the paper to make relevant approaches easy to understand.
2. Several different methods from 2004 to 2017 are covered.

#### Weak Points:

1. The structure of the abstract is not clear. For example, the analysis of YAGO4 and Wikidata is interspersed in the utility of provenance data.
2. A specific definition of KG trustworthiness and RDF reification are needed.
3. Some applications of trustworthiness need to be introduced to make RDF reification more meaningful.
4. No new method has been proposed in this paper. In the Methods section, the authors only analyse the methods described in the Introduction section used by Wikidata and YAGO.
5. In the sections of Wikidata and YAGO, the contents of the reification method, the query method and the ranking method are organized without clear structures.

#### Detailed Comments:

1. In the abstract, "Therefore, the provenance of knowledge can assist in building up the trust

of these knowledge graphs.” and “Therefore, the reliability of data in knowledge graphs can be increased by provenance data.” are repetitive.

2. In the Introduction section, “These facts can be classified as the provenance of knowledge and can contribute to the trust of these KGs.” is confusing due to the phrase of “facts”. In other parts of this paper the word “fact” refers to a triple in the KG.
3. “Reification approaches” needs to be a section with the same level as “Introduction” and “Methods”.
4. In Wikidata section, “Wikidata uses the same predicate as part of primary relations and qualifiers.” needs more explanation to clarify the difference between predicate and primary relation.
5. The results from the Discussion section can be added to the Conclusions section.

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**Is the study design appropriate and is the work technically sound?**

Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

No source data required

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** knowledge graphs

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Reviewer Report 06 September 2021

<https://doi.org/10.5256/f1000research.76450.r93428>

© 2021 Khor K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### **Kok-Chin Khor**

Lee Kong Chian Faculty of Engineering Science, Tunku Abdul Rahman University, Kajang, Malaysia

This paper provides an analysis on Wikidata and Yago4 related to provenance data and contextualisation. My comments are as follows:

- The finding of comparing Wikidata and Yago4 can be stated in the Abstract section.
- A more detailed explanation on RDF is needed to help the readers to understand the semantic web representation. Please refer to [3].
- The content in the Methods section is more towards analysis.
- The Conclusion section, besides restating the research method and summarise your overall arguments, you may want to summarise your finding and reiterates the important parts for supporting your claim.
- An error in Page 4, second paragraph:  
"Additionally, some of the popular reification approaches standard reification, singleton property..."

#### **Is the work clearly and accurately presented and does it cite the current literature?**

Partly

#### **Is the study design appropriate and is the work technically sound?**

Yes

#### **Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

#### **If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

#### **Are all the source data underlying the results available to ensure full reproducibility?**

No source data required

#### **Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Data Mining, Computer Networks.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**