Wikidata and DBpedia: A Comparative Study

D. Abián[1,2], F. Guerra[3], J. Martínez-Romanos[1(✉)], and Raquel Trillo-Lado[1]

[1] University of Zaragoza, María de Luna, 1,
Ada Byron building, 50018 Zaragoza, Spain
{702089,raqueltl}@unizar.es
[2] Wikimedia España, Vega Sicilia, 2, 47008 Valladolid, Spain
davidabian@wikimedia.es
[3] University of Modena and Reggio Emilia, Modena, Italy
francesco.guerra@unimore.it

Abstract. DBpedia and Wikidata are two online projects focused on offering structured data from Wikipedia in order to ease its exploitation on the Linked Data Web. In this paper, a comparison of these two widely-used structured data sources is presented. This comparison considers the most relevant data quality dimensions in the state of the art of the scientific research. As fundamental differences between both projects, we can highlight that Wikidata has an open centralised nature, whereas DBpedia is more popular in the Semantic Web and the Linked Open Data communities and depends on the different linguistic editions of Wikipedia.

1 Introduction

Since its creation in 2001, Wikipedia has become one of the most important sources of reliable information on the Web. The English version of Wikipedia contains more than five million English articles and there are more than 280 active versions of Wikipedia in different languages that evolve independently. Wikipedia articles are typically split into two parts: (1) a body of unstructured text with details on the article subject and (2) an optional semistructured box (i.e., the *infobox*), that summarizes the most important facts about the article.

A number of big and important projects, such as Google's Knowledge Graph, Microsoft's Satori, and DBpedia [1], have exploited infoboxes for their purposes. In particular, DBpedia, created in 2007 by Free University of Berlin and Leipzig University in collaboration with the company OpenLink Software, extracts data from Wikipedia and builds an RDF graph. This project aims to extract structured information from Wikipedia and to make it available on the Linked Data Web. More recently (in 2012), Wikimedia Deutschland proposed the Wikidata project [2]. Its main goal is providing high-quality structured data acquired and maintained collaboratively to be directly used by Wikipedia to enrich its contents. Both projects, DBpedia and Wikidata, have become important in the current Linked Data Web. Thus, according to [3], DBpedia is the second node

J. Szymański and Y. Velegrakis (Eds.): IKC 2017, LNCS 10546, pp. 142–154, 2018.
https://doi.org/10.1007/978-3-319-74497-1_14

with more incoming links on the Linked Data Web, whereas Wikidata has been continuously increasing its popularity since its creation [4].

In an ideal situation, DBpedia and Wikidata should contain equivalent content, since Wikidata aims to improve the maintenance of structured data in Wikipedia and avoid data inconsistencies among the editions of Wikipedia in different languages, while DBpedia was created upon Wikipedia to make data from Wikipedia available on the Linked Data Web. However, currently, the content of these data sources is not equivalent, and through this paper we aim to provide a comparison of the quality of the data collected by these sources.

The structure of this paper is as follows. In the next section, a brief overview of the main features of Wikidata and DBpedia is introduced. After that, the criteria or dimensions considered to compare these data sources and the comparative analysis are presented in Sect. 3. Finally, some related work is analyzed in Sect. 4, and conclusions and future work are depicted in Sect. 5.

2 Overview and Evolution

On one hand, DBpedia defines itself as "a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web" to allow users "to ask sophisticated queries against Wikipedia and to link the different data sets on the Web to Wikipedia data"[1]. Thus, DBpedia uses different software components to extract different parts of Wikipedia articles (mainly, infoboxes) and to translate them into RDF statements. For every Wikipedia article representing a resource, DBpedia defines a URI with the following pattern: `http://dbpedia.org/resource/<Title_of_Wikipedia_Article>`. For example, the English Wikipedia article about the city of Zaragoza titled *Zaragoza*[2] corresponds to the resource identified by the URI http://dbpedia.org/resource/Zaragoza in DBpedia. After that, different statements in RDF describing the resource identified by that URI are extracted from the article. For example, the statement (http://dbpedia.org/resource/Zaragoza, http://dbpedia.org/property/website, http://www.zaragoza.es/) which indicates the address of the Zaragoza city website.

On the other hand, Wikidata, as Wikipedia, is an open, collaborative project hosted and supported by the Wikimedia Foundation[3], whose main goals are: (1) to act as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikisource, etc., and (2) to provide data to other third-party projects and initiatives. Every resource or entity in Wikidata is represented by a URI that follows the pattern `http://www.wikidata.org/entity/<QX>` where X is an integer. For example, the city of Zaragoza is represented by the URI http://www.wikidata.org/entity/Q10305. Analogously to DBpedia, Wikidata allows the definition of property-value pairs to provide descriptions of

[1] http://wiki.dbpedia.org/about.
[2] http://en.wikipedia.org/wiki/Zaragoza.
[3] https://wikimediafoundation.org.

entities/resources. Moreover, Wikidata also allows to specify *qualifiers*. Qualifiers refer to an assertion of a property-value pair for an entity; i.e. qualifiers allow people to define additional subordinate property-value pairs referred to assertions of a property-value pair for an entity. For example, the property-value pair (http://www.wikidata.org/wiki/Property:P1082, 732.765) of the entity with URI http://www.wikidata.org/entity/Q10305 is qualified with the pair (http://www.wikidata.org/wiki/Property:P585, 2013), indicating that the population of Zaragoza city was 732,765 habitants in the year 2013. Nevertheless, the value of the same property in 2012 was 679,624 habitants.

In contrast to URIs created in DBpedia, URIs in Wikidata do not depend upon a specific language. Besides, all data in Wikidata is international, i.e., although its display may be language-dependent, data are processed and stored in a format independent from the language adopted. For example, the number 1,003.5 is written "1.003,5" in Spanish but it is written "1 003.5" or "1,003.5" in English. However, different versions or chapters of DBpedia are maintained for each language (English DBpedia, Spanish DBpedia, etc.), so inconsistencies among the different chapters can appear. Thus, for example, in the English DBpedia, the values of the property http://dbpedia.org/ontology/leader for the resource about Spain are Felipe VI of Spain and Mariano Rajoy; whereas the values of an equivalent property in the Spanish version (http://es.dbpedia.org/property/dirigentesNombres) are Juan Carlos I of Spain (the previous king) and Mariano Rajoy[4]. Finally, DBpedia uses the standard RDF to model and store its content, whereas Wikidata uses a non standard custom model. Nevertheless, Wikidata offers web services to export data in several standard and commonly adopted formats and models such as JSON and RDF.

Summarizing, both projects, DBpedia and Wikidata, are strictly related to Wikipedia. As a consequence of this, Wikipedia, DBpedia and Wikidata form a bulky network of data flows, some of them directly assumed by the vast community of Wikipedia and Wikidata editors, open to the general public; and some of them algorithmically planned by the more-restricted community of DBpedia contributors. In more detail, DBpedia periodically retrieves information from the different chapters of Wikipedia by using statistic and data mining techniques, whereas Wikidata provides structured data to Wikipedia in real time (see Fig. 1). Moreover, DBpedia and Wikidata are different concerning their nature, structure and functioning. The maintenance of the different chapters of DBpedia is spread along a number of organizations; for example, Spanish DBpedia is maintained mainly by the Ontology Engineering Group of the Technical University of Madrid (Spain). Nevertheless, the different DBpedia chapters are coordinated by a common committee called DBpedia Internationalisation Committee. In contrast, Wikidata is maintained by an open community, hosted and supported by the Wikimedia Foundation, and actively developed by Wikimedia Deutschland.

[4] This data was obtained the May 15th, 2017.

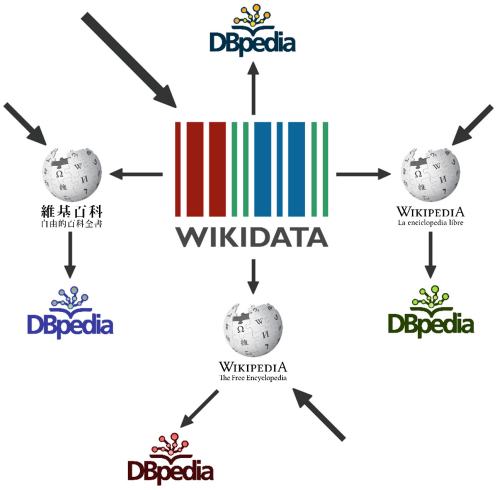

Fig. 1. Wikipedia, DBpedia and Wikidata create a bulky network of data flows, where incoming arrows represent users editing the resources.

3 Data Quality Analysis

The dimensions adopted in the literature [5–7] (see Sect. 4 for more details) to evaluate the quality of the data available in Wikipedia, DBpedia and Wikimedia are typically grouped in categories as it follows:

- *Intrinsic.* This category includes dimensions which do not depend on either the context or the task in which the data source is used. Generally, the dimensions in this category focus on whether the data properly represents the real world and on whether this representation is consistent.
- *Contextual.* This category refers to dimensions which depend on the task that users have at hand in a specific environment and time.
- *Representational.* This category refers to the design features of the data sources. It mainly focuses on how the data are stored, the ease of interpreting the data, and the interfaces and roles of the systems to interact with the data sources.
- *Accessibility.* This category refers to features related to how to obtain and access the data sources (typically via the Web) and to which extent data are sufficiently interlinked to other resources.

In this paper, we compare the data sources by using the mentioned categories and, for each of them, the dimensions proposed by at least two papers (among the ones in the literature considered as a reference) in the same category.

3.1 Intrinsic Category

Intrinsic category includes the accuracy, objectivity, consistence and reputation dimensions.

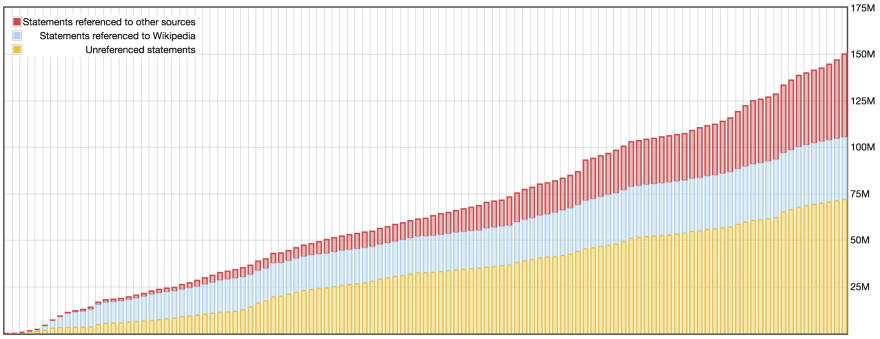

Fig. 2. Evolution of the number of references in Wikidata from 4th February 2013 to 15th May 2017. Extracted from: https://tools.wmflabs.org/wikidata/-todo/stats.php.

Accuracy. The generally adopted definition for this dimension is "the degree of correctness and precision with which information represents states of the real world" [5], i.e., to which extent data is error-free and represents the real world.

The statements in Wikidata should be verifiable and should include qualifiers that refer to trusted sources which they are based on, such as books, scientific papers, articles, etc. In particular, there exist two types of references: (1) *stated in*, to indicate the text in a media or publication where the statement is presented, and (2) *reference URL*, to refer to web sites and databases. However, if a statement represents general knowledge (for example, "the Earth is a planet") or if there exists an external identifier that immediately corroborates the statement (for example, ""Don Quixote", with ISBN-10: 0060934344, is a book") or if the statement is the information source (for example, "the name of "Miguel de Cervantes Saavedra" is "Miguel""), then references are not required. Therefore, we consider that a statement in Wikidata is precise and correct when it is conformed to the references. When references are not included, a statement is considered precise and correct when it represents the real world. According to the Fig. 2, nowadays (May 2017), almost 50% of data in Wikidata do not include references. Hence, some of these data could be considered imprecise or incorrect.

On the other hand, DBpedia incorporated N-Quads [9] from its version 3.5 (released on April 2010) onwards[5]. N-Quads allows to extend RDF triplets with an optional value in the 4th position of the statement. This value is used in DBpedia to include a provenance URI to indicate the exact origin in Wikipedia from which the associated triplet was extracted. In particular, the provenance URI is composed of the URI of the article from Wikipedia and several parameters indicating the source line from where the statement was extracted [10].

Therefore, Wikidata allows to refer to primary sources of information and its editors foster the inclusion of reliable sources and criticize the absence of these. In contrast, DBpedia can not indicate primary sources, but intermediate pages of Wikipedia which Wikidata does not consider as valid sources.

[5] http://wiki.dbpedia.org/changelog.

Objectivity Objectivity is typically defined as "the degree to which data is unbiased, unprejudiced and impartial" [6].

Both Wikipedia and Wikidata are edited by open communities, while DBpedia only admits editions of invited members and extracts most of its data from Wikipedia. Despite the fact that users can erroneously or intentionally include biased or partial data in Wikipedia and Wikidata, incorrect data are usually detected and fixed by other users. When no agreement among users editing an article is achieved, privileged users block the edition of conflicting data and keep only fact-based or well-referenced information. Other mechanisms to avoid vandalism are based on: (1) blocking ranges of IP addresses that have recently included false data and (2) developing bots to automatically detect and revert vandalism act[6].

Due to the fact that DBpedia automatically extracts most of its content from Wikipedia infoboxes, some noise could be introduced during this process. On the other hand, data in Wikidata is directly edited by users or by adhoc processes specifically designed to incorporate data from primary information sources (e.g., different national institutes of statistics). For this reason, we can consider the objectivity of Wikidata slightly higher than the one of DBpedia.

Reputation There is not a widely adopted definition of reputation of a data source. In this paper, we consider it as the degree of trustworthiness/reliability of the data and the extent to which third parties use or recommend such source.

Both Wikidata and DBpedia allow to include references to the sources where data is extracted. Besides, Wikidata and Wikipedia log the editions performed by the different users. Hence, analyzing the activities of the users along time and consulting the state of the data source in a certain time is possible. Regarding the extent to which third parties use both projects, we can consider DBpedia as the most important/referred node of the Linked Open Data Web according to the LOD cloud diagram [11] (a diagram that shows the relationships among datasets that have been published in Linked Data format based on metadata provided by Data Hub https://datahub.io). On the other hand, despite the fact that six entries for the query "Wikidata" are retrieved by the Data Hub's data source search engine, Wikidata is still not represented in the LOD cloud diagram.

Therefore, it can be considered that DBpedia is definitely more popular than Wikidata in the Linked Open Data community. Nevertheless, it must be emphasized that the number of users of, and requests to, Wikidata is increasing since its creation in 2012 [4], and that, recently, it has been announced that Wikidata and DBpedia will be federated [8].

Consistency The definition generally adopted for this dimension is the degree to which data in a data source "is free of (logical/formal) contradictions with respect to a particular knowledge representation and inference mechanisms" [6]. So, the conceptual and formal model of the data source must consistently represent the domain knowledge and no contradictions should be found in the model and data of the data source.

[6] https://wiki.data.org/wiki/Wikidata:ORES.

Both DBpedia and Wikidata use a model based on RDF which allows hierarchies of classes and properties, and instances of classes. However, the DBpedia model is semantically richer than the Wikidata model as it has been defined by means of OWL, while the core of the semantic constraints of Wikidata is defined in RDFS [12]. DBpedia does not apply semantic reasoners to effectively detect and reject data not satisfying constraints in ontologies. This is motivated by the fact that DBpedia reflects the Wikipedia data, where inconsistencies are introduced by users (e.g., in the infobox templates) [14]. On the other hand, Wikidata allows users to define ad hoc constraints for properties concerning a wide range of aspects (the most used constraints are listed in https://www.wikidata.org/wiki/Template:Constraint). Moreover, Wikidata also provides mechanisms to detect and remove data that does not satisfy those constraints. Finally, we would like to remark that, currently, DBpedia is working on the development of mechanisms to improve the data consistency [13].

3.2 Contextual Category

Contextual category includes the following dimensions: timelessness, completeness, relevancy and appropriate amount of data.

Timelessness Timelessness refers to the degree to which data are actual at the current moment. So, the update rate is one of the best metrics for this dimension.

The update rate of Wikidata depends on the activity of its users. Users can edit the content of Wikidata collaboratively by means of different interfaces, or/and they can program and execute bots to edit data. According to [4], in 2014, the edition frequency in Wikidata was up to 500 editions/minute. In contrast, updates in DBpedia are less frequent. DBpedia is updated approximately once per month[7]. Finally, it must be noticed that DBpedia also offers a live synchronization module (DBpedia Live System) [3] that allows the changes in Wikipedia to be propagated to DBpedia when an article is updated. However, this module is not running in all DBpedia chapters (e.g., it is not deployed in the Spanish chapter).

Completeness Although there is not a widely adopted definition for completeness, this dimension usually refers to the degree of data available in a specific context, i.e., the degree to which information is not missing. Different authors also define different types of completeness: schema vs column vs population completeness, completeness on schema level vs on data level, schema vs population completeness, etc. [6]. In this paper, we consider two perspectives: *completeness in depth* and *completeness in wideness*, being *wideness* the amount of entities represented in the data source and *depth* the average extent of the entities, i.e., the relation between the total of declarations and the total of entities of the data source.

[7] http://wiki.dbpedia.org.

In December, 2016, in Wikidata, there were 125 millions of statements distributed in 24.5 millions of items, which could be translated into one or more RDF triplets. So, Wikidata had an average depth of 5.1 declarations per element. In contrast, in the release published in April 2016 of DBpedia, 6 millions of entities were defined by means of 9500 millions of RDF triplets. So, this version contains 28.66 millions of instances with an average of 331.5 triplets per instance. The completeness of Wikidata and DBpedia cannot be compared in a non-biased way considering the previous data as the measurements for the data sources are not equivalent. Nevertheless, it can be concluded that both projects have the same order of magnitude of wideness, being Wikidata probably wider since it contains entities that does not correspond to a Wikipedia article. As well, it can be concluded that the depth of DBpedia is probably higher than the depth of Wikidata, although DBpedia resources are less structured than Wikidata entities.

Relevancy The most popular definition is "the extent to which information is applicable and helpful for the task at hand" [6]. Therefore, relevancy largely depends on the context, on the user performing the task and on the mechanisms provided by the data source to analyze and retrieve relevant resources.

Both DBpedia and Wikidata are generic data sources that provide tools for: (1) navigating and visualizing data, (2) formally querying data via SPARQL endpoints, (3) performing keyword-based searches, and (4) retrieving information about how to use the data and the different Key Performance Indicators (KPIs) that can help users to do their tasks.

Appropriate Amount of Data This dimension depends on the particular task and on the context where the data source is used. It is usually defined as the amount of data available and appropriate for a particular task.

Three SPARQL queries requesting for old, recent and very recent data, respectively, were used to provide a preliminary view of this dimension in Wikidata and in the English version of DBpedia (the most complete one) as, unfortunately, it is not possible to query all DBpedia chapters together. In particular, the following queries were submitted to both data sources on 4th January 2017:

- *Query 1*: Obtaining Spanish people born in the 20th century. DBpedia retrieved 1086 different resources, while Wikidata provided 39854 different entities.
- *Query 2*: Obtaining people who died in December, 2015. DBpedia retrieved 341 different resources, while Wikidata provided 1184 different entities.
- *Query 3*: Obtaining people who died in January, 2017. DBpedia retrieved 2 different wrong resources, while Wikidata provided 1184 different entities.

In all cases, Wikidata returns significantly more results than the English DBpedia. Other queries provided similar results. It can be concluded that Wikidata offers more appropriate data than the English DBpedia and, consequently, more appropriate data than any isolated version of DBpedia.

3.3 Representation Category

The dimensions considered in this category are: interpretability, ease of understanding, representational consistency and concise representation.

Interpretability There is not a widely adopted definition of interpretability of a data source. Nevertheless, all definitions of this dimension refer to the technical representation of the data in the data source.

Both DBpedia and Wikidata use self-descriptive formats including metadata. Moreover, resources are identified by means of global unique identifiers. However, the URIs stability is not the same in both projects. In DBpedia, the titles of the Wikipedia articles are used as identifiers. So, different versions of DBpedia can have different identifiers for the same resources (e.g., the city of London is identified by http://dbpedia.org/resource/London in the English DBpedia and by http://es.dbpedia.org/resource/Londres in the Spanish DBpedia). Moreover, the same resource can change its identifier in different versions of DBpedia (e.g., the Russian city St. Petersburg was previously named Leningrad). On the other hand, Wikidata uses auto-numeric identifiers, which are semantically independent from the resources that they represent. Thanks to this independence, URIs are stable and are not updated when even all data of the resource that they are identifying is edited. There are some rare exceptions to this rule of randomness in Wikidata IDs: Q1–Q5, Q13 and Q666 which are kept to represent specific resources.

Understandability Understandability is generally defined as the degree to "which data is easily comprehended by the information consumer, without ambiguity" [6]. So, the language of the data source labels and the representation model adopted are key for this dimension.

Although English is the default language of the interface of Wikidata, Wikidata is a multilingual project designed to centralize all data from different chapters in different languages of Wikipedia. In 2013, the Wikimedia Foundation's Linguistic Committee decided that Wikidata would accept all languages of the world with three basic considerations[8]: (1) the language has an ISO-639-3 code (a numeric code used in computer systems); (2) historical languages are permitted; and (3) constructed languages are allowed. Currently, in Wikidata, there are labels, descriptions and aliases for more than 400 languages and dialects. On the other hand, DBpedia has a different version for each language. Nevertheless, DBpedia community would like to integrate and merge multi-language instances by linking different versions of DBpedia [3].

In December 2016, the set of elements of Wikidata had a total of 136.85 millions of labels (14 millions in English) and 222.80 millions of descriptions (13.59 millions in English); while DBpedia was available in 125 languages and offered a total of 38 millions of labels and abstracts (abstracts are referred as descriptions in Wikidata).

[8] https://blog.wikimedia.org/2013/11/06/any-language-allowed-in-wikidata.

Representational Consistency Althought there is no consensus definition for this dimension, different authors agree on that it evaluates the degree to which information is represented in an uniform way with the purposes of increasing the interoperability and reusing intra and inter data sources. So, this dimension is extremely related to the dimension *Interlinking*.

As previously mentioned, the stability of the identifiers (URIs) is bigger in Wikidata than in DBpedia. So, we could consider that the representational consistency is bigger in Wikidata. However, if we consider inter data sources relationships, the inter-data-sources interoperability and reuse are bigger for DBpedia than for Wikidata according to the information provided by Data Hub.

Representational Conciseness Few frameworks and authors consider this dimension [6], which is usually defined as the degree to which data is compactly represented being complete. So, representational conciseness is measured by considering the length of the identifiers, and the representational language and format used by the data source.

As previously exposed, the identifiers of DBpedia depend on titles of articles in Wikipedia, while Wikidata uses shorter specific numeric identifiers. With respect to the representational language and format, DBpedia uses Virtuoso as platform to manage content based on RDF and OWL whose representational formats are quite verbose due to the fact that they are mainly based on text codification [15]. In contrast, Wikidata uses Wikibase, a specific software to manage large amounts of structured data that basically consists of two MediaWiki extensions: (1) Wikibase Repository and (2) Wikibase Client. As far as we know, Wikibase content is neither stored nor code in a standard way. Finally, we would like to remark that both, DBpedia and Wikidata, allows to export their contents to different common formats such as JSON, RDF, XML and even CSV.

3.4 Accessibility

Only the interlinking dimension is considered in this category.

Interlinking refers to the degree to which the data are linked with other data in both ways (in-going and out-going links) as it is important that each element of a data source is related to other elements (equivalent or not) defined in other data sources. Different types of interlinking are possible, being the *linked data interlinking* one of the most relevant in the Semantic Web and Web of Data.

Both Wikidata and DBpedia allow their elements to be referred from other data sources and to establish links between their elements and other resources. As previously mentioned, the number of inter-data-sources links in DBpedia is quite higher than in Wikidata. On the other hand, Wikidata has a stronger relationship with Wikimedia projects (e.g. Wikipedia or Wikispecies) than DBpedia.

4 Related Work

Knowledge bases and ontologies as DBpedia, Freebase, OpenCyc, Wikidata, and YAGO have been experimented in a large number of projects and are now adopted in commercial applications. Therefore, it is important, not only for the academic but also for the industrial community, to have a description of the main features of these knowledge sources. To the best of our knowledge, even if a complete description of these sources is available in many papers, an approach providing a critical comparison is still missing.

We started this work by focusing on Wikidata and DBpedia. In particular, we extended [8], where it is shown how Wikidata is incorporated into DBpedia and some relevant statistics about both knowledge bases are provided. This paper extends that analysis by providing a more general and complete framework for evaluating Wikidata and DBpedia. The dimensions adopted in our framework for evaluating the data quality have been selected according to the ones available

Table 1. Frameworks to evaluate the quality of data sources

Dimension Categories	Wang and Wrong (1996) [5]	Zaveri et al. (2015) [6]		Wikidata (2016) [7]
Intrinsic	Accuracy Objectivity Reputation Believability	Accuracy Consistency Conciseness Timeless		Accuracy Objectivity Reputation Consistency
Contextual	Timeliness Completeness Relevancy Value-added Amount of data	Amount of data Completeness Relevancy Verifiability Reputation Believability		Timeliness Completeness: Schema Item Population
Representational	Interpretability Understandability Consistency Conciseness	Interpretability Understandability Versatility Consistency Conciseness		Interpretability Understandability
Accessibility	Accessibility Access security	Interlinking Availability Security Performance Licensing		Interlinking
Others	Flexibility Traceability Cost-effectiveness Ease of operation Variety	**Trust**	Objectivity Verifiability Reputation Believability	Believability Relevancy Accessibility Access security Value-added Amount of data Consistency Conciseness
		Dataset Dynamicity	Timeless Currency Volatility	

in the state of the art. In particular, [5,6] classify the dimensions in a number of categories as Table 1 shows.

Starting from these categories, Wikidata opened a "Requests for comment" [7], where users are asked to provide their opinion on a data quality framework for Wikidata. Our proposal takes into account the work done and identified the relevant dimensions as shown in Table 1. The table highlights the dimensions which only appear in one of the proposed frameworks (in blue in Table 1), in two frameworks but in different categories (in green in Table 1) and in three frameworks but in three different categories (in red in Table 1). These dimensions were discarded[9] in our analysis since there is no consensus among different authors. After the filtering process, 11, 13 and 12 dimensions from the frameworks [5–7], respectively, have been considered in our analysis (these dimensions are represented in black color in Table 1).

5 Conclusions and Future Work

The number of structured data sources available on the Web has been increasing during the last decade. Two popular data sources commonly used are DBpedia and Wikidata. So, we have compared them by considering the criteria defined in the main frameworks to evaluate the quality of structured data sources.

As future work, we would like to develop a tool to audit the quality of data sources by considering the standard UNE 178301 Smart Cities and Open Data and the methodologies used in this paper to compare DBpedia and Wikidata. Currently, this kind of evaluations (audits) are made by considering the opinion of experts on the topic, who provide a value between zero and four for each dimension.

Acknowledgments. This work has been partially funded by: Action COST Keystone IC-1302, TIN2016-78011-C4-3-R (AEI/FEDER, UE), TIN2013-46238-C4-4-R. We thank Á. Poc, D. Martínez, X. Pan and F. del Molino for their support on this work.

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., et al. (eds.) ASWC/ISWC 2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_52
2. Vrandečić, D.: Wikidata: a new platform for collaborative data collection. In: 21st International Conference on World Wide Web, pp. 1063–1064. ACM, France (2012)

[9] Dimensions "Cost-effectiveness" and "Flexibility" defined in [5] are considered very related to "Performance" and "Versatility" defined in [6], respectively, so they are represented in color green. Notice also that, despite the fact that "Interlinking" is in green, it is not discarded because it appears in the same category in [6,7].

3. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web J. **6**(2), 167–195 (2015)
4. Vrandečić, D., Krőtzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (2014)
5. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (1996)
6. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: a survey. Semantic Web **7**(1), 63–93 (2015)
7. Wikidata discussion page about Data Quality. https://www.wikidata.org/wiki/Wikidata:Requests_for_comment/Data_quality_framework_for_Wikidata. Accessed May 2017
8. Kontokostas, D., Auer, S., Lehmann, J., Hellmann, S.: Wikidata through the Eyes of DBpedia. CoRR, vol. abs/1507.04180 (2015)
9. Carothers, G., Machina, L.: RDF 1.1 N-Quads. A line-based syntax for RDF datasets. W3C Recommendation (2014)
10. Provenance meta-data in DBpedia Datasets. http://wiki.dbpedia.org/services-resources/datasets/dbpedia-datasets#h434-17. Accessed May 2017
11. The Linking Open Data diagram. http://lod-cloud.net. Accessed June 2017
12. Antoniou, G., Harmelen, F.: A Semantic Web Primer. The MIT Press, Cambridge (2008)
13. Rodríguez, I.: DBpedia Mappings Front-End Administration. Google Summer of Code Project (2017)
14. Rodriguez-Hernandez, I., Trillo-Lado, R., Yus, R.: WikInfoboxer: a tool to create wikipedia infoboxes using DBpedia. XXI Jornadas de Ingeniería del Software y Bases de Datos at Congreso Español De Infomática (2016)
15. Fernández, J.D., Martínez-Prieto, M.A., Gutierrez, C., Polleres, A.: Binary RDF Representation for Publication and Exchange (HDT). W3C Member Submission (2011)