What we talk about when we talk about Wikidata quality: a literature survey

Alessandro Piscopo*
University of Southampton
Southampton, United Kingdom
alessandro.piscopo@bbc.co.uk

Elena Simperl
University of Southampton
Southampton, United Kingdom
E.Simperl@soton.ac.uk

ABSTRACT

Launched in 2012, Wikidata has already become a success story. It is a collaborative knowledge graph, whose large community has produced so far data about more than 55 million entities. Understanding the quality of the data in Wikidata is key to its widespread adoption and future development. No study has investigated so far to what extent and which aspects of this topic have been addressed. To fill this gap, we surveyed prior literature about data quality in Wikidata. Our analysis includes 28 papers and categorise by quality dimensions addressed. We showed that a number of quality dimensions has not been yet adequately covered, e.g. accuracy and trustworthiness. Future work should focus on these.

CCS CONCEPTS

• **General and reference → Surveys and overviews**; • **Information systems → Collaborative and social computing systems and tools**; **Wikis**; *Graph-based database models.*

KEYWORDS

Wikidata, data quality, literature survey

ACM Reference Format:
Alessandro Piscopo and Elena Simperl. 2019. What we talk about when we talk about Wikidata quality: a literature survey. In *The 15th International Symposium on Open Collaboration (OpenSym '19), August 20–22, 2019, Skövde, Sweden.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3306446.3340822

*Also with British Broadcasting Corporation.

OpenSym '19, August 20–22, 2019, Skövde, Sweden

ACM ISBN 978-1-4503-6319-8/19/08…$15.00
https://doi.org/10.1145/3306446.3340822

1 INTRODUCTION

Wikidata is a relatively young project—it was launched in 2012—but it is already considered by many a success story. It is a collaborative knowledge graph which has already grown up to include more than 55 million data items[1] and has recently overtaken the English Wikipedia as the most edited Wikimedia website.[2]

Knowledge graphs are graph-based knowledge representations which describe real world entities and the relations between them [39]. Numerous knowledge graphs have been developed prior to Wikidata, with notable examples being DBpedia [10] and YAGO [55]. Whereas Wikidata shares a number of features with these, e.g. releasing all data under an open licence, which allows anyone to share and reuse it, it differs with respect to others. Possibly the most significant is its completely collaborative, bottom-up approach to knowledge engineering—a task typically carried out by trained experts [46]. Anyone can edit Wikidata, either registered or anonymously. These features, combined with a large existing community around the Wikimedia ecosystem and the lessons learned from previous knowledge engineering projects, are likely to be among the determinants of Wikidata's success [46].

The growth of Wikidata in terms of size and visibility has already led to its adoption as a knowledge resource for a variety of purposes. For example, already in 2016 the Finnish Broadcasting Company (Yle) started using Wikidata identifiers to annotate content. [3] It is thus not surprising that substantial efforts around Wikidata have been dedicated to its quality and the approaches to evaluate it. Several community initiatives have attempted to gauge quality of the data in the graph, e.g. the item grading scale used in [44]. Data quality was one of the most debated topics at the first WikidataCon, a conference celebrating the 5th year of Wikidata organised by Wikidata Germany in collaboration with the Wikidata community.[4] More recently, a workshop has been dedicated specifically to Wikidata quality, bringing together

[1]https://www.wikidata.org/wiki/Special:Statistics, accessed 30 March 2019.
[2]https://www.wikidata.org/wiki/Wikidata:News, accessed 30 March 2019.
[3]http://wikimedia.fi/2016/04/15/yle-3-wikidata/, accessed 30 March 2019.
[4]https://www.wikidata.org/wiki/Wikidata:WikidataCon_2017, accessed 30 March 2019.

researchers and practitioners.[5] A large body of scientific literature has addressed the topic under different perspectives, either comparing Wikidata to other projects [17], or assessing a determined aspect of quality on the platform [11]. To the best of our knowledge, a comprehensive review of existing literature about Wikidata quality is still missing. This is needed in order to understand what has been done in terms of evaluating different quality dimensions of Wikidata, which approaches have been used, and which dimensions still need further investigation. This paper seeks to address this gap, by providing the following contributions:

- It carries out the first systematic review of literature about data quality of Wikidata;
- It systematises previous studies according to a common data quality framework;
- It identifies gaps in the existing data quality literature about Wikidata and suggests future studies accordingly.

The next sections provide an overview of the features of Wikidata and of the initiatives regarding data quality that have been carried out by the community. Subsequently, Section 3 provides a definition of data quality and of its dimensions. These are used to classify the papers found in our survey.

2 WIKIDATA

The Wikidata knowledge graph is composed of two main building blocks: *items* and *properties*. Items represent concrete or abstract entities, e.g. *William Shakespeare*, *art*, or *Stratford-upon-Avon*. On the other hand, properties state items' relations, such as *place of birth* or *has part*. These relations are used to create *claims*, item-property-value triples where the value can be either an item or a literal. Items and properties are identified by alphanumeric Unique Resource Identifiers (URIs). The form of these URIs is `Qx` for items (e.g. `Q692` for *William Shakespeare*) and `Px` for properties (e.g. `P19` for *place of birth*). Items and properties can have a human-readable label in any of the languages used in Wikidata. For example, the following claim would be used to state that Ophelia's father is Polonius (in Shakespeare's Hamlet):

Ophelia (`Q1800888`) – *father* (`P22`) – *Polonius* (`Q780191`)

Claims can be enriched through *qualifiers* and *references*. In Wikidata terms, the conjunct of a claim and its related qualifiers and/or references—if any—is called *statement* and it is the minimal unit to state facts about a resource. Qualifiers add contextual information (e.g. specifying a limitation in the validity of a statement), whereas references link to a source. For example, the statement reported above is enhanced by a link to the source that supports it, allowing anyone to verify its truth (Figure 1).

Conversely to other knowledge graphs (e.g. DBpedia [31]) which rely on a formally defined ontology, Wikidata follows another approach. Its conceptual structure is determined by loosely-defined relations between items [16]. Furthermore, classes—sets of individuals sharing some properties, e.g. *humans* or *cities* [46]—are not distinct from other items and as such can be added and edited by any user. However, prior work (e.g. [11], [33], and [46]) relied on the properties `P31` (*instance of*) and `P279` (*subclass of*) to define classes and study the consistency of Wikidata, considering as such all items that are object of `P31` or subject/object of `P279`.

The features outlined above enable to manage and aggregate the data in Wikidata to provide tailored information to users with little or no effort. A query to Wikidata can return a list with all Shakespeare's plays written before *Hamlet*, whereas the same list must have already been manually compiled by someone to achieve the same result in Wikipedia. This data is accessible in various ways, e.g. through a query interface[6] or as Linked Data.[7] As a part of the Linked Open Data (LOD) cloud, i.e. the set of intelinked datasets published on the web following Linked Data practices [9], Wikidata is connected to—and its data can be integrated and expanded with content from—numerous other resources.

Wikidata quality from the eyes of Wikidatians

The Wikidata community has developed policies and put in place a number of strategies to uphold quality, adopting consensus-based strategies from its elder sister Wikipedia and inheriting some of its policies. In the following, we describe some Wikidata community-based initiatives to uphold and assess quality—the list is not exhaustive though and it only aims to provide relevant examples of what has been done so far.

Item quality. Items represent entities in the real world and are seen by editors as clearly-defined concepts [45]. The community has undertaken several initiatives to measure quality of items. Showcase items [61] are a set of items selected by the community as outstanding examples of the capabilities of the system. The number of showcase items varies, but has been so far in the order of the few dozens. Showcase items must meet a number of criteria covering the different elements composing items, i.e. statements, human-readable labels, and links to other Wikimedia projects. Yapinus et al. [63] relied upon the showcase item's criteria to devise, in close-collaboration with the community of Wikidata, a single-grading scale which assigns labels to Items from A

[5]https://fardamariam.wixsite.com/wikidatadqworkshop, accessed 30 March 2019.

[6]https://query.wikidata.org/

[7]Linked Data and Linked Open Data refer to a set of best practices to publish structured data on the web [6].

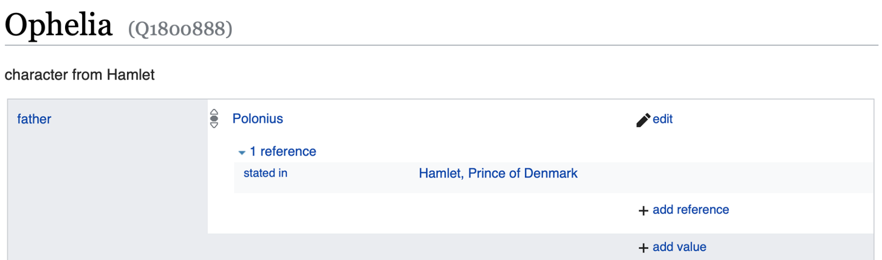

Figure 1: Item Q1800888 (*Ophelia*) with one of its statements

(the highest) to E. The grading scale covers the completeness of an item, described as the number of relevant statements; the number of the sources used to support the statements; the labels and descriptions in an appropriate number of languages; links to other wiki projects; and possibly whether media files are attached.[8]

Reference quality. References are among the features that set Wikidata apart from similar projects. Provenance, i.e. the specification of where a piece of information is derived from, facilitates the reuse of data by improving error-detection and the selection of pieces of information based on their source [30]. The lack of provenance or the use of poor sources may affect trustworthiness of the data and hinder its reuse for business and other purposes [20]. Additionally, the availability of provenance can increase trust in the project, as noted in Wikipedia [32]. The verifiability policy [60] specifies which statements need to be supported by a reference and sets the quality requirements for that. Statements must be verifiable by consulting a referenceable primary source. This must be accessible 'by at least some' Wikidata contributors to confirm the source firsthand [60]. A good reference must also be relevant—it must provide evidence for the claim it is linked to. Additionally, good references must be authoritative or 'deemed trustworthy, up-to-date, and free of bias for supporting a particular statement' [60].

Constraint violations. Properties in Wikidata may include constraints, i.e. restrictions that define how properties should be used and the relations that should exist—or not exist—for the classes they apply to. For instance, property P26 (*spouse*) has the *symmetric* constraints, meaning that if Item A is linked to Item B through P26, B must link back to A using the same property. Property constraints are not enforced in Wikidata, meaning that editors are not prevented from adding content that may violate constraints. However, constraint violations are used to spot potential errors of different types, e.g. involving inconsistencies or inaccuracies in the

[8]https://www.wikidata.org/wiki/Wikidata:Item_quality, accessed 30 March 2019.

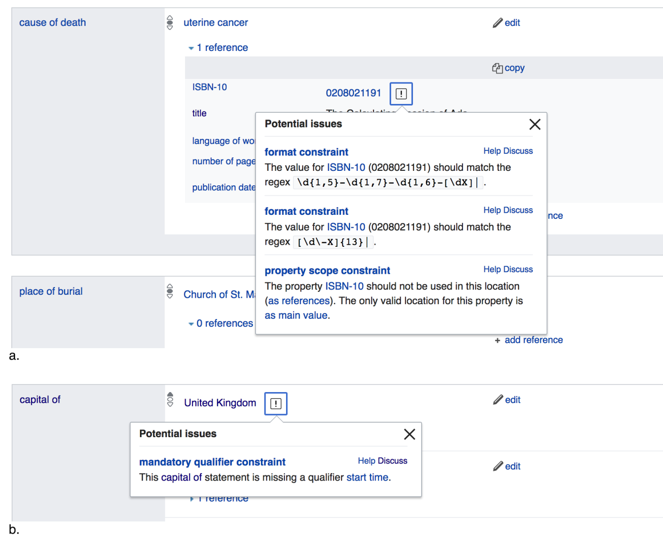

Figure 2: Examples of Property constraints violations. Figure a. is taken from Item Q7259 (*Ada Lovelace*). The *format* constraint checks whether the value used as an object matches a regular expression, whereas the *property scope* constraint refers to a specificity of Wikidata's knowledge representation model, i.e. the type of statement where a Property can be used. Figure b. shows a violation for Q84 (*London*), suggesting that the information in a statement may be incomplete.

data. Property constraint violations are continuously monitored on Wikidata and users may be notified if they are violating one (see Figure 2).

3 DATA QUALITY

The existing literature on data quality is extensive and commonly follows Juran's[25] definition of quality as "fitness for use" for its object of study (e.g. [5, 18, 42, 59, 64]). Different perspectives hinge upon this definition, which stresses either data consumers' needs [42, 59] or the suitability of data for the task to be performed [50]. Both points of view are underpinned by an empirical approach, in which data attributes are rooted in user/task requirements [59]. As Bizer et al. [8] note, this approach implies two aspects: (*i.*) data quality is **task-dependent**, i.e. the same piece of data may be considered of sufficient quality for one task, but insufficient for another; (*ii.*) it is **subjective**, whereas a user may find a piece of data appropriate for a task, another may not deem it suitable for the same task. Regardless of the perspective adopted, data quality is typically a multi-dimensional construct, each dimension being a set of attributes that measure a single aspect of quality [59]. Different sets of dimensions may be relevant for a user, depending on the task at hand [8]. The literature diverges in the dimensions included in each quality framework. The seminal work of Wang and Strong [59] classifies eight dimensions into four categories: *intrinsic*, *contextual*, *representational*, and *accessibility*. Intrinsic dimensions refer to those that are "independent of the user's context" [64] and implies that data has quality in its own right [59]. Contextual dimensions are dependent on the task at hand and on the context of the data consumer [59]. Representational and accessibility dimensions refer to the form in which the data is available and to how it can be accessed [17]. Färber et al. [17] comparatively evaluated several knowledge graphs using a selection of dimentions derived from the framework in [64], which in turn is based on [59], among others. Because [17]'s framework has been used as an evaluation framework in practice and is solidly grounded in prior literature, we rely on it in the present work.

Data quality dimensions

Intrinsic Dimensions

Accuracy. Several definitions have been given of accuracy. For [59] it is the extent to which data is accepted as true and free of error, whilst [41] defines it as the extent to which the data value v reflects the correct value v'. Others, such as [4] and [18], assert that data values must correspond to a state of things in the real-world, i.e. a reality existing objectively and independently from the observer, in order to be accurate. Some researchers [5] distinguish between *syntactic* and *semantic* accuracy. Whereas semantic accuracy corresponds to the definitions of accuracy mentioned above, syntactic accuracy refers to the closeness of a value v to any of the possible values in a definition domain D. Unless specified otherwise, this work uses the term accuracy in the sense of semantic accuracy.

Trustworthiness. It indicates the extent to which the user deems data as 'true' [42] and depends on both the trustworthiness of the data producers and the judgement of the data consumer [15]. [17] notes that this dimension subsumes other four, i.e. believability, reputation, objectivity, and verifiability. Hence, in order to be trustworthy data must be accepted as real and credible (believable) [59], its source and content must be highly regarded (reputable) [59], it must be impartial and free of bias (objective) [59], and its correctness must be easy to check (verifiable) [36].

Consistency. The definitions of consistency take into account various characteristics. According to [5], a consistent dataset is free from "violations of semantic rules defined over a set of data items." [64] focuses instead on aspects related to the Semantic Web and sees consistency as the conformity with a particular knowledge representation and inference model. Finally, [34] argues that "a dataset is consistent if it is free of conflicting information." We adopt this definition in the current work.

Contextual Dimensions

Relevancy. This dimension concerns how useful and important data is for the task at hand [59, 64]. It is a highly context-dependent dimension, especially on the web, where a user may be faced with a large amount of potentially relevant information from various sources [17].

Completeness. [5] includes completeness among the set of basic data quality dimensions, defining it as the extent to which a dataset represent a corresponding collection of real-world objects. Other points of view take into account the context in which data is used. For Wang and Strong [59] completeness is "the extent to which data are of sufficient breadth, depth, and scope for the task at hand." These three features are articulated by some into sub-dimensions, i.e. schema, column, and population completeness [17].

Timeliness. According to Zaveri et al. [64], "timeliness measures how up-to-date data is relative to a specific task." Whereas data sources may vary and be updated at different times, these changes may not always reflect those occurring to the objects they represent. As a result, data may lose currency and become outdated for the task at hand of data consumers.

Representation Dimensions

Ease of understanding. In order to facilitate use, data must be unambiguous and understandable by its consumers [59]. As regards Linked Data, whereas software agents rely on URIs to unambiguously communicate between them, humans require labels and descriptions to visualise and browse RDF data [24].

Interoperability. The previous dimension focuses on the representational characteristics of data from the point of view of human users. Interoperability concerns instead representation under a technical perspective, referring to the extent to which machines can obtain a consistent and clear interpretation of data which allows them to exchange and process information without ambiguities [17]. The definition we follow here has been formulated in [64]: interoperability is the extent to which the data conforms with previous sources in terms of format and structure.

Accessibility Dimensions

Accessibility. Data sources on the web need to be timely available in order to be integrated with other sources to produce tailored information for users [36]. Accessibility concerns this aspect and is defined in [59] as "the extent to which data is available or easily and quickly retrievable."

Interlinking. On the Linked Data web, datasets need to be interconnected to enable data integration. The interlinking dimension refers to that. It is the "degree to which entities that represent the same concept are linked to each other, be it within or between two or more data sources" [64].

Licence. The links between datasets on the Linked Data web may be useful to discover new information. However, some data sources may not be suitable for reuse for determined tasks. Therefore, it is important for consumers to provide datasets with a licence clearly expressing the terms for reuse and sharing [24].

4 METHOD

We performed our survey by querying Google Scholar, DBLP, and Semantic Scholar for the terms "Wikidata quality" and for the word Wikidata combined with each of the dimensions listed in Table 1, e.g. "Wikidata accuracy", "Wikidata completeness", etc. These search engines were chosen to include widely used scholarly search engines known to provide very broad (especially Google Scholar and Semantic Scholar) and/or accurate results (DBLP [29]). After a manual check, we limited the results of Google Scholar and Semantic Scholar for each query to the first 100 in order to avoid having an extremely large number of results which would contain numerous non-relevant papers. We included only English-language articles from peer-reviewed conferences and journals. After collecting the results, we removed all duplicates. We aimed to keep in our survey papers which in their study either:

(1) evaluate one or multiple data quality dimensions of Wikidata;
(2) develop an approach to identify quality issues tested on Wikidata.

Therefore, we left out all papers which develop tools not directly related to quality evaluation, although they may be used to uphold quality, those which use Wikidata as a knowledge resource to support their experiment (e.g. [52]), and those which investigate other aspects, such as collaborative dynamics (e.g. [45]). However, papers falling the second category may propose an evaluation approach, without actually gauging Wikidata quality within their study. We selected the paper to be included in our survey by reading their abstract and assessing their suitability to the requirements set above.

5 RESULTS

Our queries yielded a total of 2272 articles. The results from Google Scholar and Semantic Scholar gave several hundreds results, whereas DBLP only a few. Table 2 shows the number of results by search engine. Our final selection included 28 papers, the majority of them published in 2017 (Figure 3). Most papers were aimed at Wikidata alone, more commonly covering only one quality dimension. Other works took a comparative stance, contrasting Wikidata's quality to that of other knowledge graphs or other types of datasets on the web. We present the selected papers by the quality dimension(s) they evaluate. Because of this choice, a paper may be covered in more than one dimension.

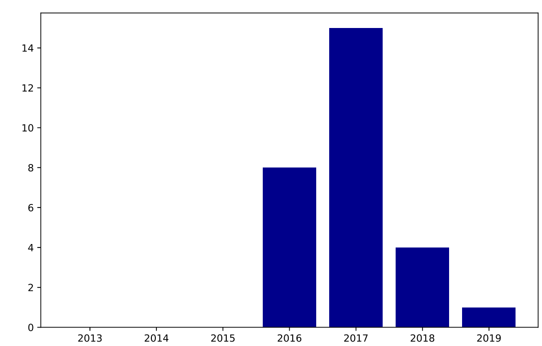

Figure 3: Number of selected papers per year.

Intrinsic dimensions

Accuracy. The number of studies addressing explicitly accuracy on Wikidata is low, compared to other dimensions (e.g. completeness). One of these is the comparative study of Färber et al. [17]. They propose a framework to select the most suitable knowledge graph for a given task. Their work compares five projects—DBpedia, Wikidata, YAGO, OpenCyc, and Freebase—along a set of dimensions derived from [59], [7], and [64]. They perform a high-level evaluation, i.e. providing for each dimension an overall score for each knowledge graph. Some of their metrics are expressed as a ratio between the number of correct instances to the total of instances. With respect to accuracy evaluation, they compare values to a gold standard (semantic accuracy) or verify whether values match an expected patterns (syntactic accuracy). Wikidata is the second best performer within the set of knowledge graphs studied [17]. 100% of Wikidata statements and literals evaluated in [17] are syntactically correct,

CATEGORY	**Intrinsic**	**Contextual**	**Representational**	**Accessibility**
DIMENSIONS	Accuracy Trustworthiness *Consistency*	Relevancy Completeness Timeliness	Ease of understanding Interoperability	Accessibility *Interlinking* *Licence*

Table 1: Data quality dimensions used in [17]. In italics the dimensions not originally in [59].

	Google Scholar	**Semantic Scholar**	**DBLP**
Total papers	1090	1177	5
Unique papers	790	557	5
Selected papers	24	25	5

Table 2: Results by search engine. Searches were performed on 1 April 2019.

whilst > 90% match those in the gold standard. Abián et al. performed another comparative multi-dimensional evaluation of Wikidata, against DBpedia [1]. Their approach is rather speculative though and consists of a series of assumptions based on the characteristics of the knowledge graphs taken into consideration. For example, with regard to accuracy they consider as imprecise or inaccurate all statements not including a reference.

A number of studies look at vandalism in Wikidata [12, 22, 23, 48, 53, 62, 65]. These were submitted to a challenge at the 2017 WSDM conference [21]. Although these aim at developing automated approaches to detect edits carried out by vandals on the platform, they may somewhat be assimilated to the accuracy dimension. The approaches used were various, with different degrees of involvement from human users, and using various algorithms and features.

Trustworthiness. Trustworthiness is among the dimensions evaluated in the comparative study in [17]. Three different metrics are extracted (*trustworthiness on knowledge graph level*, *t. on statement level*, and *indicating unknown and empty values*). Wikidata achieves the highest scores in the trustworthiness dimension, due to being manually curated, to the possibility to add references for each statement, and to the support of empty and unknown values [17]. However, these metrics only measure the extent to which a determined feature is present or supported. Hence, whilst they provide an indication of the capabilities of Wikidata compared to other knowledge graphs, they give little insights about its data quality. Thakkar et al. [56] adopt a data consumer viewpoint, choosing a task—i.e. question answering around a number of domains—and comparing the quality of Wikidata and DBpedia over several dimensions. Although provenance richness varies in Wikidata, i.e. the extent to which sources are specified for the piece of data in the graph, this feature—i.e. references—is simply missing in DBpedia [56].

Other work focuses on Wikidata sources. In [43] we have devised an approach to evaluate Wikidata external sources, i.e. those linking to resources without the knowledge graph. The approach consists of two steps: in the first, a sample of references is evaluated by means of microtask crowsourcing; these references are fed to a machine learning algorithm in the second step, which allows its application on a large scale over the whole project. The results, which include only English language sources, show that Wikidata references are of good quality overall. Using the quality criteria set by the Wikidata community (see Section 2), almost 70% of Wikidata external references are relevant and around 80% are authoritative. Another study targets specifically Wikidata external references, comparing them to those in Wikipedia [47]. According to that, Wikidata has a more diverse pool of sources, in terms of country of provenance, and employs a larger percentage of external databases and reference sources, such as library catalogues, compared to the online encyclopaedia.

Consistency. This dimension is included in the comparative study in [56], with Wikidata and DBpedia obtaining equal scores for the observed slices of the datasets. The metrics used include whether properties from external datasets are reused (*ontology hijacking*) and the proportion of misused OWL[9] properties (*Misused OWL Datatype/Object Property*). [17] looks at the existence of schema restrictions checks at the time of statement creation, a feature implemented in Wikidata editing interface. Other aspects of consistency considered in [17] are hard to measure in Wikidata, due to its approach to expressing ontological knowledge—e.g. compared to the other KGs, Wikidata does not use OWL. Furthermore, the evaluation estimates the number of inconsistent axioms by checking disjoint statements via `owl:disjointWith`. This property is not used in Wikidata, therefore no inconsistent axioms are found. Brasileiro et al. [11] explore common issues in the Wikidata taxonomy by applying multi-level modelling theory. They highlight three anti-patterns, attributable to the misuse of `P31` and `P279`. This generally consists of using a type or a subclass relation in a statement instead of the correct one. Other quality issues involve an incorrect object item or cause redundancies (rather than inconsistencies).

[9]The Web Ontology Language (OWL) is a formal language commonly used to describe logical relations in ontologies.

For example, these occur when an item is a sub-class of two items, one of which is an instance of the other [11].

We include under this dimension also some studies which assess characteristics related to the Wikidata data model, although not strictly its consistency. [46] addresses Wikidata's conceptual structure. This work devises a framework assessing structural features of the Wikidata ontology, such the number of classes, its average depth, and the average number of subclasses and instances per class. This framework is then applied to evaluate the Wikidata ontology over time, which appears to be of uneven quality. Next to a curated core with deep taxonomies, large swathes of the ontology are flat, with numerous classes likely to be the product of the misuse of `P31` and `P279`, confirming what noted in [11].

Whereas [54] is not a systematic evaluation of any quality dimension of Wikidata, it reports about the issues encountered by its authors in using Wikidata. Among these issues, a number of inconsistencies are found in Wikidata taxonomic structure and in the definitions of some of its properties. Wikidata is also included in a study that compares how eight ontologies define units of measurement [28]. The findings show that in this area Wikidata has broader coverage and a larger number of labels and links to other ontologies than other projects within the study. The work in [57] discusses the suitability of the Wikidata data model, i.e. its properties and classes, to build a registry in the domain of digital preservation. The authors highlight a number of potential issues, e.g. concerning the ambiguity of some classes and the misuses of some properties. Finally, Voss [58] evaluates the fitness of Wikidata to create taxonomies of knowledge organisation systems and ways to improve its consistency.

Accuracy	Abián et al. [1], Crescenzi et al. [12], Färber et al. [17], Heindorf et al. [22], Heindorf et al. [23], Potthast et al. [48], Sarabadani et al.[53], Yamazaki et al. [62], Zhu et al. [65]
Trustworthiness	Färber et al. [17], Piscopo et al. [43, 47], Thakkar et al. [56]
Consistency	Brasileiro et al. [11], Färber et al. [17], Keils and Schneider [28], Piscopo et al. [46], Spitz et al. [54], Thornton et al. [57], Voss [58]

Table 3: Papers covering intrinsic dimensions

Contextual Dimensions

Relevancy. Besides Abián et al. [1], who briefly discuss the characteristics of DBpedia and Wikidata related to this dimension, only Färber et al. [17] include relevancy. The metric adopted checks only if ranking of statements is allowed in a knowledge graph. Wikidata is the only one where this feature is present among the projects evaluated.

Completeness. A large number of papers targets completeness. [17] measures *schema completeness*, *column completeness*, and *population completeness*. Whereas Wikidata outperforms the other knowledge graphs in the evaluation of schema (i.e. completeness with respect to classes and relations) and population completeness (i.e. referring to all entities in the graph), its completeness of relations for each entity in the graph (column completeness) is worse. Ringler and Paulheim [51] look at the coverage of five projects (DBpedia, YAGO, Wikidata, OpenCyc, NELL) across different areas, e.g. people, places, songs, or events. Wikidata is the most complete resource when it comes to people, albums and movies, whilst it trails with respect to organisations, places, and events. [49] builds a tool to measure Wikidata's completeness called COOL-WD[10]. Combining crowdsourced work, information extraction techniques, and entailments from the Wikidata RDF graph [13, 14], COOL-WD creates completeness statements, describing whether an item, a statement, or parts of the graph are complete. Human users can manage and add new completeness statements, therefore adding a further manual check to the system. Ahmeti et al. [2] and Balaraman et al. [3] build another tool to evaluate *relative completeness* of Wikidata items, i.e. the extent to which they contain all relevant information in comparison with similar entities. A similar approach is evaluated on YAGO and Wikidata by Galárraga et al. [19]. Their rule mining approach is able to predict completeness with precision up to 100% for some properties in the graph. The authors of [38] devise an approach to automatically create scholarly profiles based on Wikidata. As a part of their study, they evaluate the completeness of scientometrics information in Wikidata, finding that coverage levels change by type of information, with gaps regarding e.g. individual articles, individual researchers, and citations between scientific articles. [37] reports about the creation of an app to visualise places in literary works, which uses Wikidata as a back-end. Incidentally, it also reports about the issues encountered in working with this knowledge graph, mentioning types of missing items and properties that were unsuitable for the task at hand. Completeness is also discussed in [1], comparing the number of instances and statements in DBpedia and Wikidata, concluding that "both projects have the same order of magnitude of wideness", with Wikidata describing a larger number of concepts.

Timeliness. Three aspects of timeliness are evaluated in [17]: frequency of updates; the capability to specify validity for statements; the capability to specify modification date for statements. Wikidata is maintained by a community of contributors, thus it allows more frequent updates with respect

[10] The tool can be found at https://cool-wd.inf.unibz.it/, accessed 30 March 2019.

to most of knowledge graphs in the study, which are automatically extracted from a source. Concerning specifying validity and modification date for statements, Wikidata uses qualifiers to achieve the first—a feature missing in other projects, such as YAGO or DBpedia—but has no way to add modification dates [17].

Relevancy	Abián et al. [1], Färber et al. [17]
Completeness	Abián et al. [1], Ahmeti et al. [2], Balaraman et al. [3], Darari et al. [13], Darari et al. [14], Färber et al. [17], Galárraga et al. [19], Nielsen [37], Nielsen et al. [38], Prasojo et al. [49], Ringler and Paulheim [51]
Timeliness	Abián et al. [1], Färber et al. [17]

Table 4: Papers covering contextual dimensions

Representation Dimensions

Ease of understanding. [27] examines the availability of labels in different languages for seven datasets, including samples from the LOD cloud, datasets published by museums and governments, and Wikidata. This has the most comprehensive coverage in terms of proportion of entities with human-readable labels. Furthermore, it is the most diverse, supporting the largest variety of languages and having the least unequal distribution of coverage across languages. Abián et al. [1], who use the term 'understandability' to refer to ease of understanding, compare the number of labels and languages available in Wikidata and DBpedia, with the first having more labels in more languages. These knowledge graphs are compared also in [56], where this dimension is called 'data diversity'. The evaluation is carried out by comparing the percentage of entities with a human-readable label available and the number of languages supported. Whereas Wikidata outperforms DBpedia with regard to the latter, it score worse in the second. This is a result of the approach used to represent the Wikidata data model in RDF, which creates an entities representing each statement (i.e. reification [16]). All the entities generated this way have no label, which leads [56]'s score to go down. In [17] ease of understanding encompasses metrics regarding the use of human-readable labels, such as their number and availability in multiple languages, and other Linked Data-specific features, e.g. an RDF serialisation and the presence of URIs. Wikidata achieves high scores in each of the metrics gauged.

Interoperability. This dimension subsumes interpretability, representational consistency, and concise representation. They are evaluated in [1] and [17]. The latter checks whether a determined feature is present and the extent to which external vocabularies, i.e. relations and classes from other knowledge graphs, are used. Compared to other projects in [17], Wikidata is more complex to query, due to its use of reification. Furthermore, it is available in various serialisations (like YAGO and DBpedia) and employs a large extent of external vocabularies. [1] discusses the characteristics of DBpedia and Wikidata, although without employing any metrics.

Ease of understanding	Abián et al. [1], Färber et al. [17], Kaffee et al. [26, 27], Thakkar et al. [56]
Interoperability	Färber et al. [17]

Table 5: Papers covering representation dimensions

Accessibility Dimensions

Accessibility, Interlinking, and Licence. These dimensions are covered by a small number of studies and their names may vary in the literature, hence we cover them together. [17] is the only one to assess all three accessibility dimensions. Its authors gauge various metrics for accessibility, finding out that Wikidata has high percentage of availability times, although its query system has some limitations (the max query execution time is 30 seconds), and many of its resources cannot be dereferenced. Regarding interlinking and licence, Wikidata has a very large number of connection to other resources and provides machine-readable information about its licence [17]. Thakkar et al. [56] use similar metrics to those in [17] to gauge accessibility, which they refer to as "availability", and interlinking. They compare again DBpedia and Wikidata, finding that this has a larger number of links to other resources and performs slightly better (in contrast to [17]) when it comes to dereferencing entities. These two projects are compared also in [1] with respect to interlinking. However, the evaluation does not rely upon any metrics and consists only of a reflection about how their respective features may lead to better or worse interlinking. Finally, Mountantonakis et al. [35] devise an approach to measure the extent to which LOD datasets are connected. They test their approach—which relies on the `owl:sameAs` property—on several datasets, including Wikidata. Their findings highlight the need to increase the number of real world objects that are described across LOD resources.

Accessibility	Färber et al. [17], Thakkar et al. [56]
Interlinking	Abián et al. [1], Färber et al. [17], Mountantonakis et al. [35], Thakkar et al. [56]
Licence	Färber et al. [17]

Table 6: Papers covering accessibility dimensions

6 DISCUSSION

Around six years since Wikidata's launch, the quality of its data has already been the object of numerous studies. These could be roughly classified into two groups (Table 7): first, those which perform an evaluation of one or multiple dimensions—this is the largest group; second, those which develop an approach which is evaluated on Wikidata or a tool which serves to gauge its quality. The papers in the second group address one dimension at the time, whereas those in the first include comparative, multi-dimensional studies. These contrast Wikidata against well-established resources, such as DBpedia or YAGO. Besides being a sign of the growing popularity of Wikidata, these comparative studies are the only ones to investigate consistently its quality across all, or a large part of, the dimensions in our framework. It is the case of Färber et al. [17], who aim at providing guidance to choose the most suitable knowledge graph for a user's needs. Abián et al. [1] also cover several of our framework's dimensions, but their analysis is superficial and limited to discussing the respective characteristics of DBpedia and Wikidata with respect to each dimension taken into consideration.

Completeness is the dimension covered by the largest number of papers. Considering the massive editing activity of the Wikidata community—through automated editing software (bots) and tools to perform semi-automate revision—research on this aspect might have been motivated by the intention to understand where the efforts of the community were most required and provide guidance. Other dimensions were less studied. The number of papers targeting Wikidata's accuracy was rather low, if we exclude the works developing automated vandalism detection approaches. This may sound surprising, considered that accuracy is generally deemed as a key quality dimension [59]. Future studies should focus on measuring how accurate Wikidata is, possibly employing existing metrics such as those in [39]. Further research should also address trustworthiness. Whereas the evaluation in [43] points out important issues regarding the quality of references in Wikidata, e.g. the difference between human editors and bots in terms of the sources they contribute, it is limited in scope, as it includes only external references, which are the minority of all Wikidata references. Moreover, [43] evaluates sources only in English, leaving out a crucial aspect of Wikidata—its vast multi-language support. Concerning that, we would have expected to find a larger number of studies about ease of understanding. The papers from Kaffee et al. [26, 27] address the extent to which labels and descriptions have been added across the many languages used in Wikidata. However, it would be desirable to compare correctness of human-readable labels between different languages. All evaluations of Wikidata consistency in our survey investigate that under a limited perspective—in the case of [46], assessing structural aspects of Wikidata's conceptual structure, rather than its actual consistency. The large size of the Wikidata ontology might be an obstacle to that [46], hampering the application of automated reasoning approaches (e.g. [40]). A possible solution may be using dataset slices, similarly to [56]. With regard to Wikidata's features to specify temporal limitation for statements (i.e. qualifiers), the lack of research touching specifically upon this topic is somewhat surprising. Future work should investigate the role of qualifiers in determining Wikidata's timeliness, also in comparison with other knowledge graphs.

Finally, some notes about our methodology and its limitations. The number of results from the search engines utilised had a large variance, with DBLP showing the largest difference. This might be imputable to the search technology and database they rely upon and suggests that future work including a larger number of search engines, digital libraries, and conference and journal websites may lead to different, larger set of papers. Moreover, the current study analysed only papers in English. A follow-up should include a greater pool of languages, given also Wikidata's multi-lingual, multicultural nature.

7 CONCLUSIONS

This is the first paper to perform a systematic review of literature about data quality and Wikidata. Our findings show that, since its launch in 2012, 28 papers in total have addressed this topic. Whereas the majority of them have focused only on Wikidata, some have evaluated that as part of a multi-dimensional, comparative study along other knowledge graphs. Overall, the data quality of Wikidata has been a fruitful area of research so far. Some dimensions have been covered in depth, whilst other have been insufficiently investigated. Completeness and consistency are the most studied dimensions, together with accuracy. However, the majority of papers addressing accuracy develops some type of tool to flag up a particular quality issue, rather than directly measuring quality (Table 7). Other dimensions which require further efforts are consistency, trustworthiness, and timeliness. Future studies should focus on these (and accuracy), possibly providing reusable metrics that could be applied to regularly monitor the evolution of the project.

REFERENCES

[1] D. Abián, F. Guerra, J. Martínez-Romanos, and Raquel Trillo Lado. 2017. Wikidata and DBpedia: A Comparative Study. In *International KEYSTONE Conference (Lecture Notes in Computer Science)*, Vol. 10546. Springer, 142–154.
[2] Albin Ahmeti, Simon Razniewski, and Axel Polleres. 2017. Assessing the Completeness of Entities in Knowledge Bases. In *ESWC (Satellite Events) (Lecture Notes in Computer Science)*, Vol. 10577. Springer, 7–11.
[3] Vevake Balaraman, Simon Razniewski, and Werner Nutt. 2018. Recoin: Relative Completeness in Wikidata. In *WWW (Companion Volume)*.

Evaluations	**All dimensions**: Abián et al. [1], Färber et al. [17] **Accuracy, trustworthiness, ease of understanding, accessibility, interlinking**: Thakkar et al. [56] **Trustworthiness**: Piscopo et al. [43] **Consistency**: Brasileiro et al. [11], Keils and Schneider [28], Piscopo et al. [46], Voss [58] **Completeness**: Nielsen et al. [38], Ringler and Paulheim [51] **Ease of understanding**: Kaffee et al. [26, 27] **Interlinking**: Mountantonakis et al. [35]
Tools and approaches	**Accuracy**: Crescenzi et al. [12], Heindorf et al. [22], Heindorf et al. [23], Potthast et al. [48], Sarabadani et al.[53], Yamazaki et al. [62], Zhu et al. [65] **Trustworthiness**: Piscopo et al. [43] **Completeness**: Ahmeti et al. [2], Balaraman et al. [3], Darari et al. [13], Darari et al. [14], Galárraga et al. [19], Prasojo et al. [49]

Table 7: Classification of the papers analysed into Evaluations and Tools and approaches. Please note that papers discussing aspects of Wikidata quality without applying any metric are not included in this table.

ACM, 1787–1792.

[4] Donald P Ballou and Harold L Pazer. 1985. Modeling data and process quality in multi-input, multi-output information systems. *Management science* 31, 2 (1985), 150–162.

[5] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. 2009. Methodologies for data quality assessment and improvement. *ACM Comput. Surv.* 41, 3 (2009), 16:1–16:52.

[6] Tim Berners-Lee. 2006. Linked Data. https://www.w3.org/DesignIssues/LinkedData.html

[7] Christian Bizer. 2007. *Quality-Driven Information Filtering in the Context of Web-Based Information Systems.* Ph.D. Dissertation. Freie Universität Berlin, Germany.

[8] Christian Bizer and Richard Cyganiak. 2009. Quality-driven information filtering using the WIQA policy framework. *J. Web Sem.* 7, 1 (2009), 1–10.

[9] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2009. Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.* 5, 3 (2009), 1–22.

[10] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. DBpedia - A crystallization point for the Web of Data. *J. Web Sem.* 7, 3 (2009), 154–165.

[11] Freddy Brasileiro, João Paulo A. Almeida, Victorio Albani de Carvalho, and Giancarlo Guizzardi. 2016. Applying a Multi-Level Modeling Theory to Assess Taxonomic Hierarchies in Wikidata. In *WWW (Companion Volume).* ACM, 975–980.

[12] Rafael Crescenzi, Marcelo Fernández, Federico A. Garcia Calabria, Pablo Albani, Diego Tauziet, Adriana Baravalle, and Andrés Sebastián D'Ambrosio. 2017. A Production Oriented Approach for Vandalism Detection in Wikidata - The Buffaloberry Vandalism Detector at WSDM Cup 2017. *CoRR* abs/1712.06919 (2017).

[13] Fariz Darari, Radityo Eko Prasojo, Simon Razniewski, and Werner Nutt. 2017. COOL-WD: A Completeness Tool for Wikidata. In *International Semantic Web Conference (Posters, Demos & Industry Tracks) (CEUR Workshop Proceedings)*, Vol. 1963. CEUR-WS.org.

[14] Fariz Darari, Simon Razniewski, Radityo Eko Prasojo, and Werner Nutt. 2016. Enabling Fine-Grained RDF Data Completeness Assessment. In *ICWE (Lecture Notes in Computer Science)*, Vol. 9671. Springer, 170–187.

[15] Mariangiola Dezani-Ciancaglini, Ross Horne, and Vladimiro Sassone. 2012. Tracing where and who provenance in Linked Data: A calculus. *Theor. Comput. Sci.* 464 (2012), 113–129.

[16] Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. 2014. Introducing Wikidata to the Linked Data Web. In *Semantic Web Conference (1) (Lecture Notes in Computer Science)*, Vol. 8796. Springer, 50–65.

[17] Michael Färber, Frederic Bartscherer, Carsten Menne, and Achim Rettinger. 2018. Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web* 9, 1 (2018), 77–129.

[18] Christian Fürber and Martin Hepp. 2011. Swiqa - a semantic web information quality assessment framework. In *ECIS*. 76.

[19] Luis Galárraga, Simon Razniewski, Antoine Amarilli, and Fabian M. Suchanek. 2017. Predicting Completeness in Knowledge Bases. In *WSDM*. ACM, 375–383.

[20] Olaf Hartig and Jun Zhao. 2009. Using Web Data Provenance for Quality Assessment. In *Proceedings of the First International Workshop on the role of Semantic Web in Provenance Management (SWPM 2009), collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington DC, USA, October 25, 2009 (CEUR Workshop Proceedings)*, Vol. 526. CEUR-WS.org.

[21] Stefan Heindorf, Martin Potthast, Gregor Engels, and Benno Stein. 2017. Overview of the Wikidata Vandalism Detection Task at WSDM Cup 2017. *CoRR* abs/1712.05956 (2017).

[22] Stefan Heindorf, Martin Potthast, Benno Stein, and Gregor Engels. 2015. Towards Vandalism Detection in Knowledge Bases: Corpus Construction and Analysis. In *SIGIR*. ACM, 831–834.

[23] Stefan Heindorf, Martin Potthast, Benno Stein, and Gregor Engels. 2016. Vandalism Detection in Wikidata. In *CIKM*. ACM, 327–336.

[24] Aidan Hogan, Andreas Harth, Alexandre Passant, Stefan Decker, and Axel Polleres. 2010. Weaving the Pedantic Web. In *LDOW (CEUR Workshop Proceedings)*, Vol. 628. CEUR-WS.org.

[25] JM Juran. 1962. Quality control handbook. In *Quality control handbook*. McGraw-Hill.

[26] Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. 2017. A Glimpse into Babel: An Analysis of Multilinguality in Wikidata. In *OpenSym*. ACM, 14:1–14:5.

[27] Lucie-Aimée Kaffee and Elena Simperl. 2018. The Human Face of the Web of Data: A Cross-sectional Study of Labels. In *SEMANTICS (Procedia Computer Science)*, Vol. 137. Elsevier, 66–77.

[28] Jan Martin Keil and Sirko Schindler. 2019. Comparison and evaluation of ontologies for units of measurement. *Semantic Web* 10, 1 (2019), 33–51.

[29] Jinseok Kim. 2018. Evaluating author name disambiguation for digital libraries: a case of DBLP. *Scientometrics* 116, 3 (2018), 1867–1886.

[30] Jens Lehmann, Daniel Gerber, Mohamed Morsey, and Axel-Cyrille Ngonga Ngomo. 2012. DeFacto - Deep Fact Validation. In *The Semantic Web - ISWC 2012 - 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I (Lecture Notes in Computer Science)*, Vol. 7649. Springer, 312–327. https://doi.org/10.1007/978-3-642-35176-1_20

[31] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.

[32] Teun Lucassen and Jan Maarten Schraagen. 2010. Trust in Wikipedia: how users trust information from an unknown source. In *Proceedings of the 4th ACM Workshop on Information Credibility on the Web, WICOW 2010, Raleigh, North Carolina, USA, April 27, 2010*. ACM, 19–26.

[33] Maximilian Marx and Markus Krötzsch. 2017. SQID: Towards Ontological Reasoning for Wikidata. In *International Semantic Web Conference (Posters, Demos & Industry Tracks) (CEUR Workshop Proceedings)*, Vol. 1963. CEUR-WS.org.

[34] Pablo N. Mendes, Hannes Mühleisen, and Christian Bizer. 2012. Sieve: linked data quality assessment and fusion. In *EDBT/ICDT Workshops*. ACM, 116–123.

[35] Michalis Mountantonakis and Yannis Tzitzikas. 2016. On Measuring the Lattice of Commonalities Among Several Linked Datasets. *PVLDB* 9, 12 (2016), 1101–1112.

[36] Felix Naumann. 2002. *Quality-Driven Query Answering for Integrated Information Systems*. Lecture Notes in Computer Science, Vol. 2261. Springer.

[37] Finn Årup Nielsen. 2016. Literature, Geolocation and Wikidata. In *Wiki@ICWSM (AAAI Workshops)*, Vol. WS-16-17. AAAI Press.

[38] Finn Årup Nielsen, Daniel Mietchen, and Egon L. Willighagen. 2017. Scholia, Scientometrics and Wikidata. In *ESWC (Satellite Events) (Lecture Notes in Computer Science)*, Vol. 10577. Springer, 237–259.

[39] Heiko Paulheim. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8, 3 (2017), 489–508.

[40] Heiko Paulheim and Aldo Gangemi. 2015. Serving DBpedia with DOLCE - More than Just Adding a Cherry on Top. In *International Semantic Web Conference (1) (Lecture Notes in Computer Science)*, Vol. 9366. Springer, 180–196.

[41] Barbara Pernici and Monica Scannapieco. 2003. Data quality in web information systems. In *Journal on Data Semantics I*. Springer, 48–68.

[42] Leo Pipino, Yang W. Lee, and Richard Y. Wang. 2002. Data quality assessment. *Commun. ACM* 45, 4 (2002), 211–218.

[43] Alessandro Piscopo, Lucie-Aimée Kaffee, Chris Phethean, and Elena Simperl. 2017. Provenance Information in a Collaborative Knowledge Graph: An Evaluation of Wikidata External References. In *International Semantic Web Conference (1) (Lecture Notes in Computer Science)*, Vol. 10587. Springer, 542–558.

[44] Alessandro Piscopo, Chris Phethean, and Elena Simperl. 2017. What Makes a Good Collaborative Knowledge Graph: Group Composition and Quality in Wikidata. In *SocInfo (1) (Lecture Notes in Computer Science)*, Vol. 10539. Springer, 305–322.

[45] Alessandro Piscopo, Christopher Phethean, and Elena Simperl. 2017. Wikidatians are Born: Paths to Full Participation in a Collaborative Structured Knowledge Base. In *HICSS*. AIS Electronic Library (AISeL).

[46] Alessandro Piscopo and Elena Simperl. 2018. Who Models the World?: Collaborative Ontology Creation and User Roles in Wikidata. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 141:1–141:18.

[47] Alessandro Piscopo, Pavlos Vougiouklis, Lucie-Aimée Kaffee, Christopher Phethean, Jonathon S. Hare, and Elena Simperl. 2017. What do Wikidata and Wikipedia Have in Common?: An Analysis of their Use of External References. In *OpenSym*. ACM, 1:1–1:10.

[48] Martin Potthast, Stefan Heindorf, and Hannah Bast. 2017. Proceedings of the WSDM Cup 2017: Vandalism Detection and Triple Scoring. *CoRR* abs/1712.09528 (2017).

[49] Radityo Eko Prasojo, Fariz Darari, Simon Razniewski, and Werner Nutt. 2016. Managing and Consuming Completeness Information for Wikidata Using COOL-WD. In *COLD@ISWC (CEUR Workshop Proceedings)*, Vol. 1666. CEUR-WS.org.

[50] Thomas C Redman. 2001. *Data quality: the field guide*. Digital press.

[51] Daniel Ringler and Heiko Paulheim. 2017. One Knowledge Graph to Rule Them All? Analyzing the Differences Between DBpedia, YAGO, Wikidata & co. In *KI (Lecture Notes in Computer Science)*, Vol. 10505. Springer, 366–372.

[52] Tomás Sáez and Aidan Hogan. 2018. Automatically Generating Wikipedia Info-boxes from Wikidata. In *WWW (Companion Volume)*. ACM, 1823–1830.

[53] Amir Sarabadani, Aaron Halfaker, and Dario Taraborelli. 2017. Building Automated Vandalism Detection Tools for Wikidata. In *WWW (Companion Volume)*. ACM, 1647–1654.

[54] Andreas Spitz, Vaibhav Dixit, Ludwig Richter, Michael Gertz, and Johanna Geiß. 2016. State of the Union: A Data Consumer's Perspective on Wikidata and Its Properties for the Classification and Resolution of Entities. In *Wiki@ICWSM (AAAI Workshops)*, Vol. WS-16-17. AAAI Press.

[55] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 697–706.

[56] Harsh Thakkar, Kemele M. Endris, José M. Giménez-García, Jeremy Debattista, Christoph Lange, and Sören Auer. 2016. Are Linked Datasets fit for Open-domain Question Answering? A Quality Assessment. In *WIMS*. ACM, 19:1–19:12.

[57] Katherine Thornton, Euan Cochrane, Thomas Ledoux, Bertrand Caron, and Carl Wilson. 2017. Modeling the Domain of Digital Preservation in Wikidata. In *Proceedings of ACM International Conference on Digital Preservation, Kyoto, Japan*.

[58] Jakob Voß. 2016. Classification of Knowledge Organization Systems with Wikidata. In *NKOS@TPDL (CEUR Workshop Proceedings)*, Vol. 1676. CEUR-WS.org, 15–22.

[59] Richard Y. Wang and Diane M. Strong. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. *J. of Management Information Systems* 12, 4 (1996), 5–33.

[60] Wikidata. 2017. Wikidata:Verifiability — Wikidata, the free knowledge base. https://www.wikidata.org/wiki/Wikidata:Verifiability. [Online; accessed 07-April-2017].

[61] Wikidata. 2018. Wikidata:Showcase Items. https://www.wikidata.org/wiki/Wikidata:Showcase_items. [Online; accessed 08-August-2018].

[62] Tomoya Yamazaki, Mei Sasaki, Naoya Murakami, Takuya Makabe, and Hiroki Iwasawa. 2017. Ensemble Models for Detecting Wikidata Vandalism with Stacking - Team Honeyberry Vandalism Detector at WSDM Cup 2017. *CoRR* abs/1712.06921 (2017).

[63] Glorian Yapinus, Amir Sarabadani, and Aaron Halfaker. 2017. Wikidata item quality labels. (5 2017). https://doi.org/10.6084/m9.figshare.5035796.v1

[64] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2016. Quality assessment for Linked Data: A Survey. *Semantic Web* 7, 1 (2016), 63–93.

[65] Qi Zhu, Hongwei Ng, Liyuan Liu, Ziwei Ji, Bingjie Jiang, Jiaming Shen, and Huan Gui. 2017. Wikidata Vandalism Detection - The Loganberry Vandalism Detector at WSDM Cup 2017. *CoRR* abs/1712.06922 (2017).