KGHeartBeat: An Open Source Tool for Periodically Evaluating the Quality of Knowledge Graphs

Maria Angela Pellegrino[1(✉)], Anisa Rula[2], and Gabriele Tuozzo[1]

[1] Dipartimento di Informatica, Unviersità degli Studi di Salerno, Fisciano, SA, Italy
{mapellegrino,gtuozzo}@unisa.it

[2] Department of Information Engineering, University of Brescia, Brescia, Italy
anisa.rula@unibs.it

Abstract. Knowledge Graphs are an extraordinary source of data due to their vastness, the topics heterogeneity and the presence of sources curated by companies, research groups, volunteers, and dedicated communities. Identifying high-quality Knowledge Graphs requires supporting developers and end-users in comparing and assessing data quality of publicly available Knowledge Graphs. However, no fully working and maintained Knowledge Graph quality assessment tool was found during the review of related research.

This article fully describes KGHeartBeat, a community shared open-source knowledge graph quality assessment tool designed to periodically perform quality analysis on a wide range of freely available knowledge graphs registered on the LOD Cloud and DataHub. Users can either visually explore the quality assessment report and compare knowledge graphs via a freely available web-based interface or download data analysis results for further analysis. Moreover, KGHeartBeat is also released as APIs so developers can easily integrate them into any quality management tool. As a proof of concept, we discuss different use cases to show KGHeartBeat in practice, demonstrating how it can be used to compare multiple Knowledge Graphs, assess quality dimensions over time, and report performance analysis in terms of execution time.

Resource type	Community Shared Software Framework
License	MIT
Web-app	http://www.isislab.it:12280/kgheartbeat
Permanent URL	https://zenodo.org/records/10990547
Pypi package	https://pypi.org/project/kgheartbeat

Keywords: Quality assessment · Comparison · Software · Knowledge Graph · Web-interface · Automatic · Open source · Evaluation framework

1 Introduction

The last decades witnessed an unprecedented volume of datasets structured according to the Semantic Web technologies [20,21], published as Knowledge Graphs (KGs). Today, more than 10,000 datasets are available online following

G. Demartini et al. (Eds.): ISWC 2024, LNCS 15233, pp. 40–58, 2025.
https://doi.org/10.1007/978-3-031-77847-6_3

Linked Data (LD) standards [32]. Publicly available KGs are highly heterogeneous in terms of covered topics [31], the maintenance mechanisms ranging from the ones internally curated by companies, such as Google, Microsoft, Apple, and Amazon, to openly available and maintained by research groups or dedicated communities, like DBpedia [22] and Wikidata [36]. They are also heterogeneous in terms of the generation mechanism, ranging from those that are manually created to the ones that are (semi-)automatically authored. This heterogeneity is also reflected in data quality variance, ranging from extensively curated to relatively low-quality datasets [17,20].

Assessing and monitoring KG quality challenge users to choose the best KG according to use case requirements, as data quality concerns the assessment of fitness for use [16]. Data quality assessment is a multidimensional problem encompassing multiple quality dimensions including but not limited to accessibility, interlinking, performance, syntactic validity, and completeness [40,43], and the relevance of each dimension is up to the domain or specific use cases.

Moreover, heterogeneity is also evident in the design choice at the tool level. The process of measuring data quality can be supported by quality-related metadata as well as data itself [43]. Concerning the automation level, tools may require manual configuration of the quality assessment process, such as Sieve [24] and RDFUnit [20] (formerly DataBugger) that require users to manually set up test cases by giving them the possibility to define domain-specific data quality test cases going beyond conventional quality heuristics. In most cases, tools support semi-automated assessment, such as LiQuate [30], DaCura [10]. Fully automated tools, such as LinkQA [13], are pretty rare. Besides efforts over the years, there is no working and maintained KG quality assessment tool as a reference in the Semantic Web community, as made evident in Sect. 2.

This resource paper extends [28] and fully describes `KGHeartBeat`, an open-source community-shared tool that supports developers and lay users to assess and compare several KG quality dimensions. While lay users are provided with a web-based visualization interface to explore results, developers are provided with APIs to integrate the quality dimensions metric computation in any data management workflow. `KGHeartBeat` periodically and fully automatically performs the assessment of all the KGs publicly available on LOD cloud[1] and DataHub[2]. Results are automatically accessible via a freely available web application where they can be either downloaded as a CSV file or visually explored via interactive data visualizations. `KGHeartBeat` lets users explore updated metrics computation outcomes, explore its evolution over time, and enable the comparison among KGs. Hence, it supports the data quality assessment of publicly available KGs and selecting the ones that best fit the use case of interest.

The rest of the article is structured as follows. Section 2 overviews related work. Section 3 presents `KGHeartBeat` in terms of implemented metrics, offered interfaces, and implementation details. Section 4 shows `KGHeartBeat` in practice exploring different use cases and reporting a performance evaluation regarding execution time. The article concludes with final remarks concerning limitations, the maintenance plan, the tool applicability, and future direction.

[1] LOD cloud: https://lod-cloud.net.
[2] DataHub: https://datahub.io.

2 Background and Related Work

Data quality has long been studied, and many definitions and assessment measures are proposed [2,3,34,38,40]. One generally accepted definition for data quality is "fit for use" [3,37], which means that data quality assessment is highly subjective and context-dependent. The dimensions to choose depend on the data consumer and the downstream task.

Quality dimensions and their definitions overlap in well-known evaluation frameworks used for assessing data quality, such as the one defined by Stvila et al. [34] or the one proposed by Wang et al. [38]. For instance, FAIR-Checker [11] is based on the FAIR principles[3] that focus on the data reuse. Other assessments are based on compliance with a set of conditions defined via SHACL[4], as the framework proposed by Spahiu et al. [33]. Still, the assessment can be performed via SPARQL queries, as in IndeGX [23], requiring working SPARQL endpoints.

This article grounds on the framework proposed by Zaveri et al. [43]. As a result, we consider the following quality dimension clusters:

- *Accessibility dimensions* involves aspects related to the access, authenticity, and retrieval of data to obtain either the entire or some portion of the data (or from another linked dataset) for a particular use case. It includes availability, licensing, security, and performance.
- *Intrinsic dimensions* are those that are independent of the user's context. It includes semantic accuracy, consistency, and conciseness. These dimensions focus on whether the information (syntactically and semantically) represents the real world correctly and compactly, and whether the information is logically consistent.
- *Contextual dimensions* highly depend on the context of the task at hand, assessing the amount of published data, their relevancy, trustworthiness, understandability, and timeliness. This dimension can be further refined by considering the following aspects as separate dimension clusters:
  - *Trust dimensions* focusing on trustworthiness in terms of verifiability, reputation, and believability;
  - *Dataset dynamicity* focusing on the currency and timeliness.
- *Representational dimensions* capture aspects related to the design of the data, such as representational conciseness, interoperability, interpretability, and versatility.

While quality dimensions are rather abstract, they can be measured via quality assessment metrics, which rely on quality indicators. An assessment score is computed from these indicators using a scoring function.

In the following, we provide an overview of (open source) tools supported by peer-reviewed contributions focusing on the KG quality assessment. In particular, the considered tools have been identified by examining peer-reviewed

[3] FAIR principles: https://www.go-fair.org/fair-principles.
[4] SHACL: https://www.w3.org/TR/shacl.

Table 1. Comparison of Open Source KG quality assessment tools, ranked by last update. White rows correspond to working tools, dark gray rows highlight not working tools, while light gray rows correspond to tools declared not maintained.

Tool	Last update	Ref	Input	Quality dim.						GUI	Over time	KGs
				A	C	D	I	R	T			
tRDF/tSPARQL	2014	[14]	RDF data						✓	✓		
Sieve	2014	[24]	metadata	*user defined*						✓		✓
SPARQLES	2016	[35]	SE	✓						✓	✓	✓
Loupe API	2017	[26]	SE		✓			✓				✓
LD Sniffer	2017	[25]	KG resources	✓						✓		
SemQuire	2018	[21]	SE, input file	✓	✓		✓	✓		✓		
DYLDO	2019	[18]	a fixed set of LD docs	✓	✓						✓	✓
DistQualityAssessment	2020	[32]	RDF data	✓	✓		✓			✓		
LODLaundromat	2021	[4]	dumps		✓	✓		✓		✓		
Luzzu	2021	[8]	SE, input file	✓	✓		✓	✓		✓	✓	
ABECTO	2023	[19]	SE				✓					
Roomba	2023	[1]	(meta)data	✓	✓		✓					✓
RDFUnit	2023	[20]	KG URL, SE				✓			✓		
YummyData	2024	[41]	SE, VoID file	✓	✓			✓		✓	✓	✓
KGHeartBeat	2024	*this*	SE, metadata, VoID file	✓	✓	✓	✓	✓	✓	✓	✓	✓

articles published in semantic web conferences and journals and further exploring their related works. Analyzed tools have been summarized in Table 1, which compares them based on several dimensions, reporting in dark gray rows not working tools, while in light gray not maintained tools. The `input` column of Table 1 compares tools in terms of used input, underlining if single or multiple input sources are used and clarifying if the assessment relies on actual data or metadata published by data curators or made available by dataset aggregators. For each tool, Table 1 schematically reports the covered quality dimensions, i.e., accessibility (shortened as **A**), contextual (shortened as **C**), data dynamicity (shortened as **D**), intrinsic (shortened as **I**), representational (shortened as **R**) and trust (shortened as **T**). It is worth mentioning that we consider relevant for the comparison and explicitly report it in the Table as recent as possible working frameworks. Moving back to up to 10 years ago, we only reported working tools, removing those referred to in the literature but not in a fully working status. In fact, many of the tools referenced in the LD quality surveys [6,43], as well as those mentioned in the context of KG quality assessment tools, such as Luzzu [8], are associated with not working links. This includes tools such as LiQuate [30], triple check mate [42], Trellis [12], WIQA [7], LinkQA [13].

The analysis reveals several trends regarding the LD quality tools. These tools typically utilize no more than two data sources and prefer analysis over the data rather than metadata. More recent tools support multiple dimensions. However, there is no single dimension assessed by all the tools or a tool supporting all the dimensions, as already observed in the survey performed by Zaveri et al. [43]. Accessibility emerges as the most covered dimension, followed by contextual (primarily focusing on completeness [15]), intrinsic (primarily focusing on semantic accuracy [5]), and representational dimensions. Conversely, dimensions such as dataset dynamicity and trust received less attention. It is a common practice to provide quality assessment tools with GUIs, but it should not be taken for granted because it is not always guaranteed by surveyed tools. All the tools return a textual report, most of the time made accessible and graphically represented via the GUI. As reported in column `over time`, most of the tools focus on one-shot analysis, necessitating periodic manual interventions for ongoing analysis over time. Moreover, most of them perform the analysis on a single KG at a time, with limited support for automatic KG comparison (see column `KGs` in Table 1). As already noticed in Debattista et al. [8], tools mainly focus on quality assessment, with Sieve being a notable exception that handles data quality improvement. However, the gray rows in Table 1 show that half of the tools are not accessible anymore, and the `last update` column shows that most of them are discontinued. It clearly demonstrates the need for a reference tool for the assessment of the LD quality, maintained and widely used by the Semantic Web Community.

3 KGHeartBeat

This section extends [28] and deeply presents `KGHeartBeat`, a fully automatic KG quality assessment framework detailing the implemented metrics, the proposed workflow, and implementation details.

3.1 Metrics and Implementation Details

`KGHeartBeat` implements a large set of well-known quality metrics proposed in the literature belonging to different quality dimensions. Starting from the metrics defined by the survey authored by Zaveri et al. [43], `KGHeartBeat` implements most of the metrics that can be automatically and objectively computed without requiring a gold standard. Figure 1 provides an overview of the covered dimensions, grouped according to the classification introduced in Sect. 2, along with the corresponding percentage coverage. If all the dimensions in a cluster are covered, as in the case of the representation dimension, Fig. 1 indicates a 100% coverage rate. Furthermore, Fig. 1 also reports the metric coverage for each dimension. We ensure coverage of at least two-thirds of the dimensions in each dimension cluster, and we cover at least 40% of the dimension metrics, except for two dimensions. It is worth stating that the tool is designed for easy extension, allowing for the implementation of other desired metrics.

Due to space constraints, we cannot detail all the implemented metrics. However, metrics can be naturally clustered according to the used input sources. In fact, `KGHeartBeat` combines metadata returned by the data aggregators where KGs have been published, the VoID file (if specified), and a SPARQL endpoint (if correctly retrieved and working). These inputs can be used in isolation or combined according to the metrics to be computed. Table 2 summarizes a metric

Table 2. Exemplary metrics details where metrics are clustered according to the used input sources. For each metric computation, we report the definition and implementation details in terms of input, algorithm, output range, output interpretation, output best value (if any), and the implementation source as `Ref`.

Metric	Ref				
Interlinking					
Centrality	[13]	*Detection of centrality through sameAs chains by using network measures* [43]			
		Input	Metadata		
		Algorithm	– Weighted graph via `networkx` with KGs as nodes and external links as edges, where weights correspond to the number of triples connected to the other KG. – Node centrality of the evaluated KG in this weighted graph		
		Output	[0,1]		The evaluated KG is poorly (value close to 0) or highly (value close to 1) linked to other KGs
Licensing					
Human-readable	[9]	*Detection of a license in the documentation of the dataset* [43]			
		Input	(working) SPARQL endpoint		
		Algorithm	Labels, descriptions, or comments matching the regex .∗(`licensed`?|`copyrighte`?`d`?).∗(`under`|`grante`?`d`?|`rights`?)		
		Output	{0,1}	1	
Amount of data					
Number of triples	[43]	**Input**	metadata, (working) SPARQL endpoint		
		Algorithm	– Number of triples declared in the metadata – Count of all the triples returned via the SPARQL endpoint – Number of triples are returned to the end-user		
		Output	1	✓	Number of triples has been correctly retrieved
			0		Unable to retrieve the number of triples
Verifiability					
Dataset authenticity	[9]	*Whether the dataset uses a provenance vocabulary* [43]			
		Input	VoID file, (working) SPARQL endpoint		
		Algorithm	The list of declared vocabularies is retrieved as follow: – `void:vocabulary` in the VoID file – `void:vocabulary` as a triple predicate in the SPARQL endpoint – all the used namespaces are retrieved and compared with the list of declared vocabularies		
		Output	[0,1]	1	Highest is the number of matches, highest the score.
Availability					
SPARQL endpoint	[9]	*Checking whether the server responds to a SPARQL query* [43]			
		Input	metadata, VoID file		
		Algorithm	– `api/sparql` in metadata – `void:sparqlEndpoint` as a predicate in the VoID file – SELECT ?s WHERE {?s ?p ?o} LIMIT 1		
		Output	1	✓	available & working
			0		available, but not working
			–1		not available
RDF dump	[9]	*Checking whether an RDF dump is provided and can be downloaded* [43]			
		Input	Metadata, VoID file, SPARQL endpoint		
		Algorithm	– tag in metadata, e.g, `text/n3`, `rdf` – `void:dataDump` as a predicate in the VoID file – SELECT ?s WHERE {?s void:dataDump ?o} – check its status via a HEAD request.		
		Output	1	✓	available & working
			0		available, but not working
			–1		not available

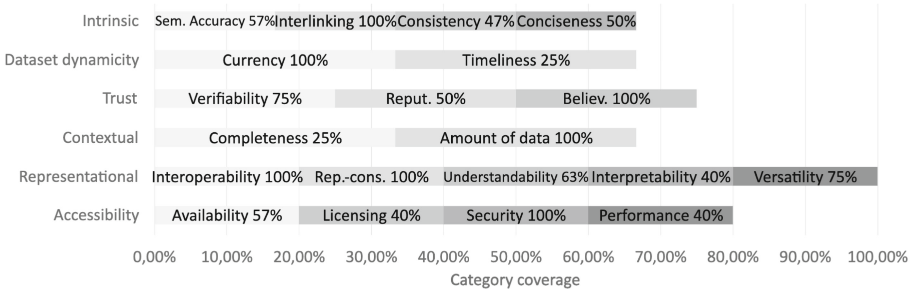


Fig. 1. `KGHeartBeat` dimensions coverage by reporting i) the dimension coverage for each dimension cluster, i.e., those reported on the y-axis, and ii) the metric coverage for each dimension.

for each input scenario, including the adopted metric definition, implementation reference, input source(s), performed input, and output details in terms of range, the best value, and result interpretation. It is worth noting that `KGHeartBeat` always combines the VoID file with metadata or actual data retrieved via a working SPARQL endpoint for all the computed metrics, belonging to one of these input scenarios. Hence, Table 2 can be used as a valuable reference for understanding `KGHeartBeat`'s implementation logic, while the complete list of metrics is freely accessible online[5]. Furthermore, the online documentation[6] shows that we cover the VoID[7], DCAT[8] and DoublinCore[9] standards.

3.2 `KGHeartBeat` Workflow

Following the steps recommended in the literature to perform KG quality assessment, `KGHeartBeat` workflow (Fig. 2) is structured around the following phases:

– *Selection of KGs of interest* (optional step). Users can select KG(s) of interest to perform the evaluation.
– *KGs retrieval phase*: `KGHeartBeat` retrieves metadata of KGs of interest by exploiting a publicly available API[10] that currently relies on LODCloud and DataHub as aggregators. For each KG, `KGHeartBeat` computes quality metrics by retrieving data from metadata provided by aggregators, the VoID file, or by querying the SPARQL endpoint (if available and functional).

[5] Metric details: https://isislab-unisa.github.io/KGHeartbeat.
[6] Vocabularies used by `KGHeartBeat`: https://isislab-unisa.github.io/KGHeartbeat/vocabularies.
[7] VoID vocabulary: https://www.w3.org/TR/void.
[8] Data Catalog Vocabulary (DCAT): https://www.w3.org/TR/vocab-dcat-2.
[9] DublinCore: https://www.w3.org/wiki/DublinCore.
[10] KG search engine: http://www.isislab.it:12280/kgsearchengine.

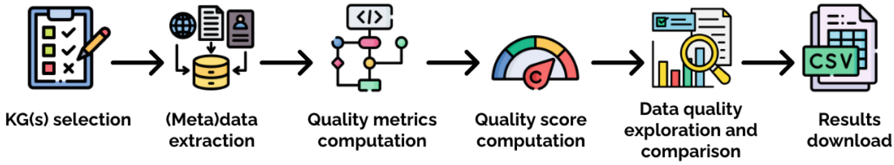


Fig. 2. `KGHeartBeat` quality assessment workflow.

- *Quality metrics computation.* All the metrics are automatically computed for each KG without requiring any manual interventions. Each metric is computed independently, and the output is then mapped to a quality dimension score representing the fulfillment of the related quality aspect.
- *Quality score computation.* All quality dimension scores are then linearly combined into an overall quality assessment score. It is a numeric value between 0.0 and 100.0. A higher score indicates better quality.
- *Results exploration.* Results are stored in CSV format, while log activities are stored in a TXT file to take note of any raised exceptions. At the current stage, results are returned as CSV as it is compatible with commonly used software to perform data analysis, such as Excel, and it minimizes technical skills required to make results interpretable by a wide range of end-users. However, we explored the possibility of automatically transforming CSV results in RDF triples relying on the Data Quality Vocabulary[11] via ChatGPT. Results are available in the GitHub repository. Users can download and analyze those files manually or visually explore computed metrics through tables and charts within a freely available web application. `KGHeartBeat` web application enables the quality assessment and the comparison of KGs of interest. It lets users perform KG assessment by visualizing metric details, inspecting their changes over time, and facilitating easy comparisons. Users have access both to individual quality dimension scores and the overall quality score.

3.3 `KGHeartBeat` Implementation Details

`KGHeartBeat`'s back-end is implemented in Python, while the web application is implemented in Javascript. The source code is freely available with the MIT license, developed following the best practice of modularity and loosely coupled components. The resource has been designed by using the abstract factory pattern to generalize the input module to deal with heterogeneous KG sources and the output module to return results in different formats. The logging activity implements the decorator design pattern.

Figure 3 shows the architecture of `KGHeartBeat` reflecting the workflow described above. Users can optionally perform the configuration to ask for KGs

[11] Data Quality Vocabulary: https://www.w3.org/TR/vocab-dqv.

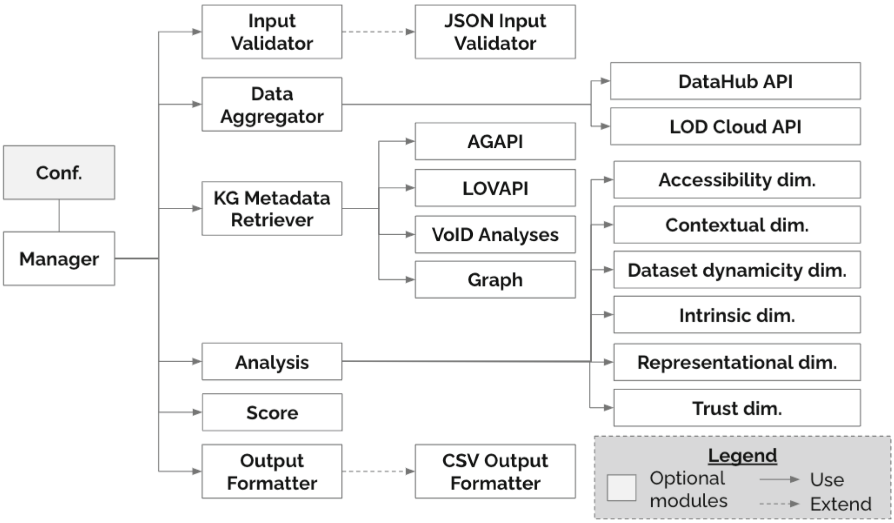


Fig. 3. `KGHeartBeat` architecture

of interest by using keywords or KG IDs; otherwise, all available KGs are considered. The tool then performs the analysis and stores metrics results in CSV format which are used in the visualization application. A dedicated chart graphically represents each metric for users interested in assessing the KG quality, while others can inspect the KGs comparison and identify the best KGs according to quality dimensions.

`KGHeartBeat` is released in different ways: i) It is released on a GitHub page, and it periodically (weekly since November 2023) performs the quality assessment on all the available KGs. Results are then redirected to the web application. It is worth noting that it is an ongoing project which involves continuous refinement of metrics and visualizations. This evolution has affected the regularity of data points, making some of the quality assessments computed in the past incompatible with the current interface. To address this, we uploaded all the performed assessments to the GitHub repository, ensuring transparency and consistency. All assessments compatible with the current interface are directly accessible via the web application. ii) A web application for enabling users to explore KG quality and perform comparisons visually. For each KG, users can look at the graphical representation of the metric score as a chart. As an alternative, users can select multiple KGs and look at their comparison. iii) A set of implemented metrics released as APIs, developed in Python and freely available as a pip package. It enables the integration of `KGHeartBeat` in any workflow to assess KGs under production or perform a comparison. Each metric is accessible via an independent function.

As documented in the GitHub repository, `KGHeartBeat` can be easily extended or customized as follows:

– Developers can easily customize the input and output by replacing the concrete classes `JSON Input Validator` and `CSV Output Formatter` by proposing alternative implementations coherent with the `Input Validator` and `Output Formatter` abstract classes.
– Developers can easily introduce further KG aggregators by customizing or replacing the `KG search engine` module, represented by the `Data Aggregator` class. It provides developers with the possibility of adding custom KGs. However, as a design principle, we opt for querying well-known and commonly used aggregators to rely on already validated sources.
– Developers can easily introduce either novel dimensions or metrics by extending and customizing the `Analysis` class.

3.4 Web Application and Visualization Options

Accessing the web application (visible in Fig. 4), users have access to all the performed analyses for all the available KGs. Users can either explore results aggregated by dimensions or select a desired number of KGs to perform the comparison. When metric assessments might change over time, such as the SPARQL endpoint availability, users can inspect the metric assessment outcome on an interactive line chart, as the one reported in Fig. 4, customizing the time interval of interest. The reported chart demonstrates the discontinuity of the WASABI RDF SPARQL endpoint[12] over time.

We evaluated the chart for each metric to facilitate the quality assessment or comparison. Figure 5 reports an overview of some of the selected charts. More in detail, Fig. 5(a) shows URIs length by box plots to highlight extreme situations and common behavior. By Fig. 5(a) is evident that the majority of the considered KGs satisfies the W3C recommendation of keeping URIs shorter than 80 characters [9]. Figure 5(c), coverage, and level of detail are depicted using (stacked)

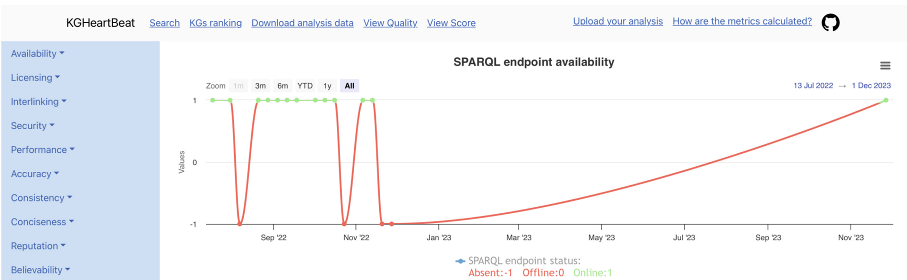
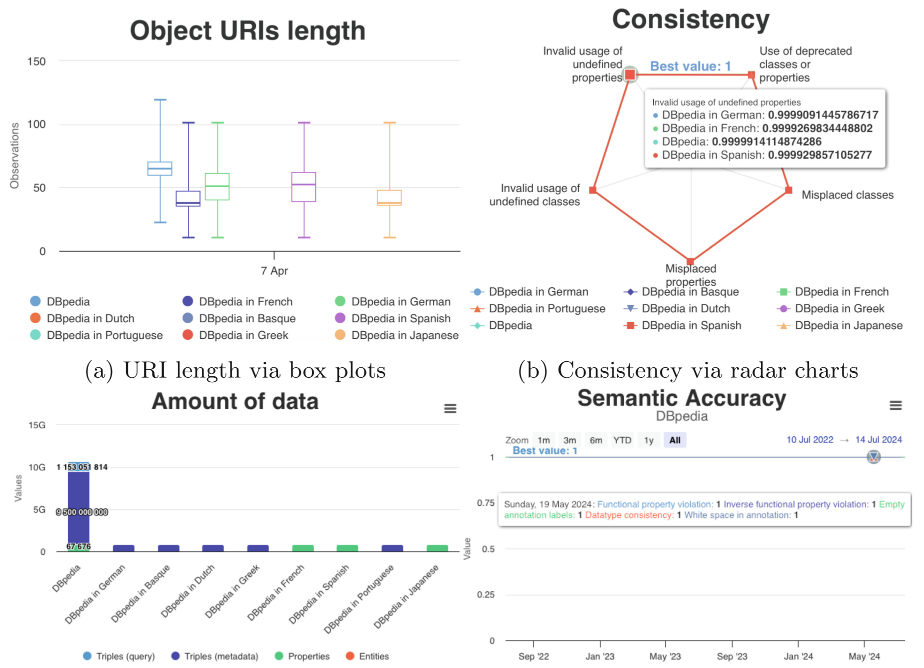


Fig. 4. `KGHeartBeat` web application while focusing on the assessment of the availability dimension in terms of the SPARQL endpoint.

[12] WASABI RDF SPARQL endpoint: https://wasabi.inria.fr.



(a) URI length via box plots

(b) Consistency via radar charts

(c) Amount of data via stacked bar charts

(d) Semantic accuracy via line charts

Fig. 5. Charts used in `KGHeartBeat` to support KGs assessment and comparison.

column charts. Fig. 5(b) reports the consistency values via a radar chart, demonstrating that it is close to one in all the cases and it corresponds to the best value for this dimension. Fig. 5(c) reports only the number of triples from metadata or the number of properties by querying a working SPARQL endpoint in most of the cases. DBpedia is the only case among the DBpedia versions where all the values are correctly computed. Triples are represented with human-readable labels via pie charts to highlight the percentage of triples with labels relative to the overall number of triples, while Timelines are suitable for tracking the behavior of the metric over time, such as metrics related to semantic accuracy (see Fig. 5(d)). If any visualization fits, values are listed in data tables by default. When metrics can be computed using metadata or actual data, the interface reports both, as in the licensing dimension.

From the interface, users can perform the following customization:

- customize the KG set to be assessed and compared;
- customize the metrics to be rendered and used in the score computation;
- upload the analysis computed on the desired KG(s) according to the format returned by the `manager` module. Exemplary CSV are available in the GitHub repository.

4 Use Case and Evaluation

This section first overviews some use cases that show `KGHeartBeat` and the practical application of quality metrics. These scenarios, which pertain to projects and topics led by the authors, illustrate our commitment to maintaining and concretely adopting the proposed toolkit in internal projects. Additionally, we encourage project exploitation in external projects by curating the GitHub repository and the published APIs detailing tutorials, maintaining and updating the code, and documenting use cases. Following the use cases, we present a performance evaluation test to document the time required to compute the overall analysis on single KGs and each dimension sequentially.

Linguistic Linked Open Data in the context of Thematic KGs. The ISISLab[13], the research group of some of the authors of this contribution, collaborates with several departments of the University of Salerno, joining projects concerning learning, cultural heritage, and linguistics, among others. One of the current collaborations focuses on Linguistic Linked Open Data (LLOD). According to the LLOD Cloud[14], remarkable care should be dedicated to the `licensing`, `availability`, and `interlinking` quality dimensions. We can query the LLOD sub-cloud by typing relevant keywords in the `KGHeartBeat` web application and configure only the dimensions of interest to compute the final score. While the complete analysis is available online on GitHub, Table 3 reports the top 7 KGs. The cut to seven is justified by the aspect that `Apertium RDF CA-IT` occupies the 7th position, followed by several KGs focusing on different language pairs attached to exactly the same scores as the 7th one. For this reason, this set of KGs is rephrased as `Apertium RDF x-y` in Table 3. Quality dimension scores are relatively low, negatively affecting the overall scores. Linguistic KGs span from controlled vocabularies, such as the `language name authority list` maintained by the publications office of the European Union, to lexical and morphological data, such as `Open Bantu isiXhosa Lexicon` that link English terms

Table 3. Top-7 KGs concerning the LLOD according to availability, licensing, and interlinking. Each dimension score ranges from 0 to 1, while the score ranges from 0 to 100. The higher, the better. KGs are sorted via the final score.

KG	Availability	Licensing	Interlinking	Score
Language Name Authority List	0.18	0.18	0.012	54
Dutch DBpedia	0.15	0.15	4.6e–6	46
Lexicon of Syntactic and Semantic Framework	0.15	0.15	0.15	38
PreMOn	0.15	0.15	24.8e–6	37
Open Bantu isiXhosa Lexicon	0.12	0.12	0.00054	31
Private correspondences	0.10	0.10	0	28
Apertium RDF x-y	0.10	0.10	1.03	28

[13] ISISLab https://www.isislab.it.
[14] LLOD Cloud: https://linguistic-lod.org.

to WordNet, and ontologies, such as the Predicate Model for Ontologies (PreMOn). `Apertium RDF` is the case of KGs automatically generated by an open-source framework named Apertium, which is based on a rule-based machine translation platform of pairs of languages.

Information Disorder and Automatic Fact-Checking. The SERICS project[15] focuses on security and rights in cyberspace, with one of its key objectives being the detection and mitigation of information disorder. Information disorders encompasses a wide range of misinformation [39] and the Semantic Web play a crucial role when the information content must be debunk by comparing it with external sources [44]. As reported in [28], `KGHeartBeat` can be configured to prioritize the dimensions of trust and dataset dynamicity in computing the final score and identify most promising KGs for information debunking.

Automatic Consumption of Knowledge Graphs. When KGs must be automatically queried by any consumption tool, `accessibility` enables the identification of sources that can be freely and easily accessible. Some previous carried out by ISISLab focused on the quality assessment of Cultural Heritage KGs to be automatically queried by Virtual Assistants and information retrieval tools [27,29]. In that context, ISISLab (manually) performed a quality assessment in terms of `accessibility`, `amount of data`, and `understandability`. The performed analysis demonstrated that a plethora of cultural heritage data had been published according to the Semantic Web technologies [27]. However, there is a discontinuous effort in maintaining live SPARQL endpoints [27,29], and human-readable labels are mainly attached to classes rather than to predicates and

KG name	2024-02-04	2024-01-28	2024-01-21	2024-01-14	2024-01-07	2023-12-31	2023-12-24	2023-12-17	2023-12-10	2023-11-
Art & Architecture Thesaurus	-	-	-	-	-	-	-	-	-	-
Thesaurus W for Local Archives	-	-	-	-	-	-	-	-	-	-
English Heritage Places	-	-	-	-	-	-	-	-	-	-
World Loanword Database	-	-	-	-	-	-	-	-	-	-
Linked Logainm	Online	Online	Online	Online	Online	Online	Online	Online	Online	Online
Thesaurus BNCF	Online	Online	Online	Online	Online	-	-	Online	Online	Online
UNESCO Thesaurus	Online	Online	Online	Online	Online	-	-	Online	Online	Online
Lista de Encabezamientos de Materia as Linked Open Data	-	-	-	-	-	-	-	-	-	-

Fig. 6. SPARQL endpoints status of some Cultural Heritage KGs as rendered in the `KGHeartBeat` web application.

[15] The SERICS project: https://serics.eu.

other resources [29]. The same results are still confirmed by KGHeartBeat. By randomly checking the SPARQL endpoint status of some Cultural Heritage KGs represented in Fig. 6, we observed that most of them seem to be offline for long periods and occasionally move to online status, not guaranteeing continuity.

Knowledge Graphs in Agri-Food Industry. NEXTCART is a National Italian project in the Agri-Food Industry that addresses challenges related to consumer food education and sustainability within Food Supply Chains (FSCs). The tool for measuring data quality, such as KGHeartBeat, can simplify the comparison and integration of diverse data sources. In particular, by utilizing KGHeartBeat, decision-makers gain confidence in the reliability and accuracy of the integrated data within the KG. Quality assessment ensures that data from heterogeneous sources meets predefined standards. This, in turn, enhances visibility for FSC

Table 4. Execution time per KG and dimension, in seconds if not differently indicated. *Legend:* Q stands for quartiles, SD stands for Standard Deviation.

Dimension	Min	Q1	Median	Q3	Max	Mean	SD
Accessibility							
Availability	4.84e–05	0.07	0.9	2.07	25h	61.96	496.09
Licensing	0.13	1.28	30.94	60.76	8’	43.53	67.10
Security	0.06	0.15	0.27	0.48	37	0.68	1.18
Performance	1.13	11.29	12.14	13.05	29.59	12.82	3.80
Contextual							
Amount of data	0.17	0.85	3.15	48.02	5’	19.98	38.95
Completeness	0.02	0.03	0.25	0.87	4.67’	1.61	11.18
Dataset dynamicity							
Currency	0.17	0.46	0.70	1.47	5’	2.54	18.38
Timeliness	0.07	0.17	0.26	0.41	2.06	0.32	0.34
Intrinsic							
Conciseness	0.00	0.38	1.16	3.12	106.36	2.41	8.16
Consistency	2.43	8.72	62.03	137.04	6.28h	3.70’	28.27’
Interlinking	0.02	0.02	0.02	0.02	4’	0.29	7.03
Semantic Accuracy	0.31	1.50	3.78	9.27	11’	15.02	55.58
Representational							
Interoperability	0.00	1.63	3.28	30.08	3’	18.64	31.96
Interpretability	0.16	0.57	0.89	2.92	4.93’	10.97	39.84
Representational Conciseness	0.33	3.07	44.98	109.39	10’	87.21	109.29
Understandability	0.29	1.43	2.62	10.27	5’	9.93	33.69
Versatility	0.17	1.42	60.53	120.61	10’	88.33	109.49
Trust							
Believability	0.02	0.53	0.84	2.78	2’	0.42	0.15
Verifiability	0.22	0.58	1.22	1.98	26’	9.42	78.04
Reputation	0.02	0.02	0.02	0.02	2’	0.02	0.06
KG	1.71	3.68	8.89	23.42	7h	3.86’	22.11’

actors, such as producers, distributors, and regulatory bodies, fostering transparency and accountability. Further, the efficient and automated assessment of KGHeartBeat streamlines the integration process.

Evaluation. The tool undergoes comprehensive testing to ensure reliability and effectiveness. Testing encompasses both unit testing, where individual components are examined for correctness, and integration testing, ensuring seamless collaboration among different modules. This section reports the execution time required to perform the quality assessment via KGHeartBeat, where all the KGs, dimensions, and metrics are computed sequentially, according to evaluations performed in April 2024. The overall analysis takes 89.40h ($\sim$ approx. four days) to assess 1,882 KGs. Of these, 8% (156) are provided only with a SPARQL endpoint, 3% (55) are only provided with a VoID file, while 18% are provided both with a SPARQL endpoint and a VoID file. Table 4 reports quartiles (shortened as Q. in the table) to perform the assessment per KG and each covered dimension per KG. It can be observed a high variability in execution times, spanning from dimensions that can be computed in a few minutes in all cases, such as Believability, Reputation, and Timeliness, to those that require a few (milli)seconds in some cases to hours in others, such as Availability and Consistency. Further analysis is required to verify if execution time correlates with queried sources, amount of data, SPARQL endpoint performance, or any other influencing factor.

5 Conclusions and Future Directions

Data published with the Semantic Web technologies are a valuable resource, but quality issues remain a well-known problem that limits their exploitation. Besides ongoing efforts over the years in designing and developing quality assessment tools, there is currently no working and maintained KG quality assessment tool as a reference in the Semantic Web community. To address this gap, we propose KGHeartBeat, a community-shared software framework designed to support developers and lay users in performing periodic KG assessment and comparison over a wide range of quality dimensions. Lay users can access a freely available web application to visually inspect the quality assessment outcomes, while developers are provided with API to integrate the proposed workflow into any quality management tool.

The intention is not to present a one-size-fits-all definitive quality assessment tool. KGHeartBeat follows all the best practices in software *design* (e.g., abstraction and modularity) to guarantee *technical quality* and make the framework fully extensible in terms of implemented metrics, input formats, and returned output files. The GitHub repository reports instructions to extend the framework, catering to the evolving and distinct needs of different consumers. It ensures that the framework is not static but can dynamically adapt to accommodate a wide range of quality dimensions, thereby fostering a more inclusive and versatile approach to quality assessment. Furthermore, the design is guided by an extensive analysis of tools and approaches proposed in the literature. KGHeartBeat is *available* on both GitHub and Zenodo with an open-source license. The ISISLab research lab maintains the code and drives the evolution.

Limitation and Future Directions. The performance evaluation indicates a high variability in execution times for different metrics. It might be a limitation for users needing quick assessments, especially for large datasets. As a future direction, we are considering improving the scalability of KGHeartBeat, e.g., by introducing mechanisms similar to the ones tested and introduced in the scalable framework developed by Sejdiu et al. [32].

In terms of findability of resources, KGHeartBeat relies on external aggregators like LOD Cloud and DataHub. Any changes or issues with these sources could impact the tool's functionality and accuracy. To partially mitigate this dependence, we stored a backup version of the queried KGs for temporary issues. However, we have to explicitly report it as a maintenance issue.

Besides the reported evaluation, we plan to perform an in-depth comparative analysis with related work. At the moment, on GitHub there is a comparison with SPAQLES [35] and Luzzu [8].

KGHeartBeat is intended to be a quality assessment tool for KG users. However, we are interested in exploring quality improvement in the future mainly investigating (LLM-based) suggested corrective actions.

Acknowledgments.
This work was partially supported by the SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU. This communication reflects only the author's view, and the Commissions are not responsible for any use that may be made of the information it contains.

References

1. Assaf, A., Troncy, R., Senart, A.: Roomba: an extensible framework to validate and build dataset profiles. In: Gandon, F., Guéret, C., Villata, S., Breslin, J., Faron-Zucker, C., Zimmermann, A. (eds.) ESWC 2015. LNCS, vol. 9341, pp. 325–339. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25639-9_46
2. Avogadro, R., Cremaschi, M., Jiménez-Ruiz, E., Rula, A.: A framework for quality assessment of semantic annotations of tabular data. In: Hotho, A., et al. (eds.) ISWC 2021. LNCS, vol. 12922, pp. 528–545. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88361-4_31
3. Batini, C., Scannapieco, M.: Data and Information Quality. DSA, Springer, Cham (2016). https://doi.org/10.1007/978-3-319-24106-7
4. Beek, W., Rietveld, L., Bazoobandi, H.R., Wielemaker, J., Schlobach, S.: LOD laundromat: a uniform way of publishing other people's dirty data. In: Mika, P., et al. (eds.) ISWC 2014. LNCS, vol. 8796, pp. 213–228. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11964-9_14

5. Behkamal, B., Bagheri, E., Kahani, M., Sazvar, M.: Data accuracy: what does it mean to LOD? In: 4th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 80–85. IEEE (2014). https://doi.org/10.1109/ICCKE.2014.6993457
6. Ben Ellefi, M.: RDF dataset profiling-a survey of features, methods, vocabularies and applications. Semantic Web **9**(5), 677–705 (2018). https://doi.org/10.3233/SW-180294
7. Bizer, C., Cyganiak, R.: Quality-driven information filtering using the WIQA policy framework. J. Web Semant. **7**(1), 1–10 (2009). https://doi.org/10.1016/j.websem.2008.02.005
8. Debattista, J., Auer, S., Lange, C.: Luzzu-a methodology and framework for linked data quality assessment. J. Data Inf. Qual. **8**(1) (2016). https://doi.org/10.1145/2992786
9. Debattista, J., Lange, C., Auer, S., Cortis, D.: Evaluating the quality of the LOD cloud: an empirical investigation. Semant. Web **9**(6), 859–901 (2018). https://doi.org/10.3233/SW-180306
10. Feeney, K.C., O'Sullivan, D., Tai, W., Brennan, R.: Improving curated web-data quality with structured harvesting and assessment. Int. J. Semant. Web Inf. Syst. (IJSWIS) **10**(2), 35–62 (2014). https://doi.org/10.4018/ijswis.2014040103
11. Gaignard, A., Rosnet, T., De Lamotte, F., Lefort, V., Devignes, M.D.: Fair-checker: supporting digital resource findability and reuse with knowledge graphs and semantic web standards. J. Biomed. Semant. **14**(1), 7 (2023)
12. Gil, Y., Ratnakar, V.: TRELLIS: an interactive tool for capturing information analysis and decision making. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) EKAW 2002. LNCS (LNAI), vol. 2473, pp. 37–42. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45810-7_6
13. Guéret, C., Groth, P., Stadler, C., Lehmann, J.: Assessing linked data mappings using network measures. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V. (eds.) ESWC 2012. LNCS, vol. 7295, pp. 87–102. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30284-8_13
14. Hartig, O.: Trustworthiness of data on the web. In: Proceedings of the STI Berlin & CSW PhD Workshop (2008)
15. Issa, S., Adekunle, O., Hamdi, F., Cherfi, S.S.S., Dumontier, M., Zaveri, A.: Knowledge graph completeness: a systematic literature review. IEEE Access **9**, 31322–31339 (2021). https://doi.org/10.1109/ACCESS.2021.3056622
16. Juran, J.: Juran's Quality Control Handbook. Mcgraw-Hill, New York (1988)
17. Käfer, T., Abdelrahman, A., Umbrich, J., O'Byrne, P., Hogan, A.: Observing linked data dynamics. In: Cimiano, P., Corcho, O., Presutti, V., Hollink, L., Rudolph, S. (eds.) ESWC 2013. LNCS, vol. 7882, pp. 213–227. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38288-8_15
18. Käfer, T., Umbrich, J., Hogan, A., Polleres, A.: Towards a dynamic linked data observatory. In: Proceedings of the Linked Data on the Web (LDOW) Workshop at The Web Conference (WWW) (2012). http://events.linkeddata.org/ldow2012/papers/ldow2012-paper-14.pdf
19. Keil, J.M.: ABECTO: an ABox evaluation and comparison tool for ontologies. In: Harth, A., et al. (eds.) ESWC 2020. LNCS, vol. 12124, pp. 140–145. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62327-2_24
20. Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R.: Databugger: a test-driven framework for debugging the web of data. In: 23rd International World Wide Web Conference, WWW. pp. 115–118. ACM (2014). https://doi.org/10.1145/2567948.2577017

21. Langer, A., Siegert, V., Göpfert, C., Gaedke, M.: SemQuire - assessing the data quality of linked open data sources based on DQV. In: Pautasso, C., Sánchez-Figueroa, F., Systä, K., Murillo Rodríguez, J.M. (eds.) ICWE 2018. LNCS, vol. 11153, pp. 163–175. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03056-8_14
22. Lehmann, J., et al.: DBpedia-a large-scale, multilingual knowledge base extracted from Wikipedia. Semant. Web **6**(2), 167–195 (2015). https://doi.org/10.3233/SW-140134
23. Maillot, P., Corby, O., Faron, C., Gandon, F., Michel, F.: Indegx: a model and a framework for indexing rdf knowledge graphs with sparql-based test suits. J. Web Semant. **76**, 100775 (2023)
24. Mendes, P.N., Mühleisen, H., Bizer, C.: Sieve: linked data quality assessment and fusion. In: Proceedings of the Joint EDBT/ICDT Workshops, pp. 116–123 (2012). https://doi.org/10.1145/2320765.2320803
25. Mihindukulasooriya, N., García-Castro, R., Gómez-Pérez, A.: LD sniffer: a quality assessment tool for measuring the accessibility of linked data. In: Ciancarini, P., et al. (eds.) EKAW 2016. LNCS (LNAI), vol. 10180, pp. 149–152. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58694-6_20
26. Mihindukulasooriya, N., García-Castro, R., Priyatna, F., Ruckhaus, E., Saturno, N.: A linked data profiling service for quality assessment. In: Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., Hartig, O. (eds.) ESWC 2017. LNCS, vol. 10577, pp. 335–340. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70407-4_42
27. Monaco, D., Pellegrino, M.A., Scarano, V., Vicidomini, L.: Linked open data in authoring virtual exhibitions. J. Cult. Herit. **53**, 127–142 (2022). https://doi.org/10.1016/j.culher.2021.11.002
28. Pellegrino, M.A., Rula, A., Gabriele, T.: KGHeartBeat: a knowledge graph quality assessment tool. In: The Semantic Web: ESWC Satellite Events. Springer, Heidelberg (2024)
29. Pellegrino, M.A., Scarano, V., Spagnuolo, C.: Move cultural heritage knowledge graphs in everyone's pocket. Semant. Web **14**(2), 323–359 (2023). https://doi.org/10.3233/SW-223117
30. Ruckhaus, E., Baldizán, O., Vidal, M.-E.: Analyzing linked data quality with LiQuate. In: Demey, Y.T., Panetto, H. (eds.) OTM 2013. LNCS, vol. 8186, pp. 629–638. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41033-8_80
31. Schmachtenberg, M., Bizer, C., Paulheim, H.: Adoption of the linked data best practices in different topical domains. In: Mika, P., et al. (eds.) ISWC 2014. LNCS, vol. 8796, pp. 245–260. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11964-9_16
32. Sejdiu, G., Rula, A., Lehmann, J., Jabeen, H.: A scalable framework for quality assessment of RDF datasets. In: Ghidini, C., et al. (eds.) ISWC 2019. LNCS, vol. 11779, pp. 261–276. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30796-7_17
33. Spahiu, B., Maurino, A., Palmonari, M., et al.: Towards improving the quality of knowledge graphs with data-driven ontology patterns and shacl. In: ISWC (Best Workshop Papers), pp. 103–117 (2018)
34. Stvilia, B., Gasser, L., Twidale, M.B., Smith, L.C.: A framework for information quality assessment. J. Am. Soc. Inf. Sci. Technol. **58**(12), 1720–1733 (2007)

35. Vandenbussche, P.Y., Umbrich, J., Matteis, L., Hogan, A., Buil-Aranda, C.: SPARQLES: monitoring public SPARQL endpoints. Semant. Web **8**(6), 1049–1065 (2017)
36. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledge base. Communications **57**(10), 78–85 (2014). https://doi.org/10.1145/2629489
37. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (1996). https://doi.org/10.1080/07421222.1996.11518099
38. Wang, X., et al.: Knowledge graph quality control: a survey. Fund. Res. **1**(5), 607–626 (2021)
39. Wardle, C., Derakhshan, H.: Information disorder: toward an interdisciplinary framework for research and policy making (2017). https://policycommons.net/artifacts/421935/information-disorder/1392979. Accessed August 2023
40. Xue, B., Zou, L.: Knowledge graph quality management: a comprehensive survey. IEEE Trans. Knowl. Data Eng. **35**(5), 4969–4988 (2023). https://doi.org/10.1109/TKDE.2022.3150080
41. Yamamoto, Y., Yamaguchi, A., Splendiani, A.: Yummydata: providing high-quality open life science data. Database **2018**, bay022 (2018)
42. Zaveri, A., et al.: User-driven quality evaluation of DBpedia. In: Proceedings of the 9th International Conference on Semantic Systems, pp. 97–104 (2013). https://doi.org/10.1145/2506182.2506195
43. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: a survey. Semant. Web **7**(1), 63–93 (2016). https://doi.org/10.3233/SW-150175
44. Zhou, X., Zafarani, R.: A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput. Surv. **53**(5) (2020). https://doi.org/10.1145/3395046