Quality Assessment for Linked Data: A Survey

A Systematic Literature Review and Conceptual Framework

Amrapali Zaveri** [a,*], Anisa Rula** [b], Andrea Maurino [b], Ricardo Pietrobon [c], Jens Lehmann [a] and Sören Auer [d]

[a] *Universität Leipzig, Institut für Informatik, D-04103 Leipzig, Germany,*
E-mail: (zaveri, lehmann)@informatik.uni-leipzig.de
[b] *University of Milano-Bicocca, Department of Computer Science, Systems and Communication (DISCo),*
Innovative Techonologies for Interaction and Services (Lab), Viale Sarca 336, Milan, Italy
E-mail: (anisa.rula, maurino)@disco.unimib.it
[c] *Associate Professor and Vice Chair of Surgery, Duke University, Durham, NC, USA.,*
E-mail: rpietro@duke.edu
[d] *University of Bonn, Computer Science Department, Enterprise Information Systems and Fraunhofer IAIS,*
E-mail: auer@cs.uni-bonn.de

Abstract. The development and standardization of semantic web technologies has resulted in an unprecedented volume of data being published on the Web as Linked Data (LD). However, we observe widely varying data quality ranging from extensively curated datasets to crowdsourced and extracted data of relatively low quality. In this article, we present the results of a systematic review of approaches for assessing the quality of LD. We gather existing approaches and analyze them qualitatively. In particular, we unify and formalize commonly used terminologies across papers related to data quality and provide a comprehensive list of 18 quality dimensions and 69 metrics. Additionally, we qualitatively analyze the 30 core approaches and 12 tools using a set of attributes. The aim of this article is to provide researchers and data curators a comprehensive understanding of existing work, thereby encouraging further experimentation and development of new approaches focused towards data quality, specifically for LD.

Keywords: data quality, Linked Data, assessment, survey

1. Introduction

The development and standardization of Semantic Web technologies has resulted in an unprecedented volume of data being published on the Web as *Linked Data* (LD). This emerging Web of Data comprises of close to 188 million facts represented as Resource Description Framework (RDF) triples (as of 2014[1]). Although gathering and publishing such massive amounts of data is certainly a step in the right direction, data is only as useful as its quality. Datasets published on the Data Web already cover a *diverse set of domains* such as media, geography, life sciences, government etc[2]. However, data on the Web reveals a large *variation in data quality*. For example, data extracted from semi-structured sources, such as DBpedia [40,49], often contains inconsistencies as well as misrepresented and incomplete information.

Data quality is commonly conceived as *fitness for use* [3,35,62] for a certain application or use case. Even datasets with quality problems might be useful for certain applications, as long as the quality is in the required range. For example, in the case of DBpedia the data quality is perfectly sufficient for enriching Web search with facts or suggestions about general information, such as entertainment topics. In such a scenario, DBpedia can be used to show related movies and personal information, when a user searches for an actor. In this case, it is rather neglectable, when in relatively few cases, a related movie or some personal facts are missing.

The quality of data is crucial when it comes to taking far-reaching decisions based on the results of querying multiple datasets. For developing a medical application, for instance, the quality of DBpedia is probably insufficient, as shown in [64], since data is extracted via crowdsourcing of a semi-structured source. It should be noted that even the traditional, document-oriented Web has content of varying quality but is still perceived to be extremely useful by most people. Consequently, a key challenge is to determine the quality of datasets published on the Web and make this quality information explicit. Assuring data quality is particularly a challenge in LD as the underlying data stems from a set of multiple autonomous and evolving data sources. Other than on the document Web, where information quality can be only indirectly (e.g. via page rank) or vaguely defined, there are more concrete and measurable data quality metrics available for structured information. Such data quality metrics include correctness of facts, adequacy of semantic representation and/or degree of coverage.

There are already many methodologies and frameworks available for assessing data quality, all addressing different aspects of this task by proposing appropriate methodologies, measures and tools. Data quality is an area researched long before the emergence of Linked Data, some quality issues are unique to Linked Data others were looked at before. In particular, the database community has developed a number of approaches [7,39,53,61]. In this survey, we focus on Linked Data Quality and hence, we primarily cited works dealing with Linked Data quality. Where the authors adopted definitions from general data quality literature, we indicated this in the survey. The novel data quality aspects original to Linked Data include, for example, coherence via links to external datasets, data representation quality or consistency with regard to implicit information. Furthermore, inference mech-

anisms for knowledge representation formalisms on the Web, such as OWL, usually follow an open world assumption, whereas databases usually adopt closed world semantics. Additionally, there are efforts focused on evaluating the quality of an ontology either in the form of user reviews of an ontology, which are ranked based on inter-user trust [44] or (semi-) automatic frameworks [59]. However, in this article we focus mainly on the quality assessment of instance data.

Despite LD quality being an essential concept, few efforts are currently in place to standardize how quality tracking and assurance should be implemented. Moreover, there is no consensus on how the data quality dimensions and metrics should be defined. Furthermore, LD presents new challenges that were not handled before in other research areas. Thus, adopting existing approaches for assessing LD quality is not straightforward. These challenges are related to the openness of the linked data, the diversity of the information and the unbounded, dynamic set of autonomous data sources and publishers.

Therefore, in this paper, we present the findings of a systematic review of existing approaches that focus on assessing the quality of LD. We would like to point out to the readers that a comprehensive survey done by Batini et. al. [6] already exists which focuses on data quality measures for other structured data types. Since there is no similar survey specifically for LD, we undertook this study. After performing an exhaustive survey and filtering articles based on their titles, we retrieved a corpus of 118 relevant articles published between 2002 and 2014. Further analyzing these 118 retrieved articles, a total of 30 papers were found to be relevant for our survey and form the core of this paper. These 30 approaches are compared in detail and unified with respect to:

– commonly used terminologies related to data quality,
– 18 different dimensions and their formalized definitions,
– 69 total metrics for the dimensions and an indication of whether they are measured quantitatively or qualitatively and
– comparison of the 12 proposed tools used to assess data quality.

Our goal is to provide researchers, data consumers and those implementing data quality protocols specifically for LD with a comprehensive understanding of the existing work, thereby encouraging further experimentation and new approaches.

This paper is structured as follows: In Section 2, we describe the survey methodology used to conduct this systematic review. In Section 3, we unify and formalize the terminologies related to data quality and in Section 4 we provide (i) definitions for each of the 18 data quality dimensions along with examples and (ii) metrics for each of the dimensions. In Section 5, we compare the selected approaches based on different perspectives such as, (i) dimensions, (ii) metrics, (iii) type of data and also distinguish the proposed tools based on a set of eight different attributes. In Section 6, we conclude with ideas for future work.

2. Survey Methodology

Two reviewers, from different institutions (the first two authors of this article), conducted this systematic review by following the systematic review procedures described in [36,48]. A systematic review can be conducted for several reasons [36] such as: (i) the summarization and comparison, in terms of advantages and disadvantages, of various approaches in a field; (ii) the identification of open problems; (iii) the contribution of a joint conceptualization comprising the various approaches developed in a field; or (iv) the synthesis of a new idea to cover the emphasized problems. This systematic review tackles, in particular, problems (i)−(iii), in that it summarizes and compares various data quality assessment methodologies as well as identifying open problems related to LD. Moreover, a conceptualization of the data quality assessment field is proposed. An overview of our search methodology including the number of retrieved articles at each step is shown in Figure 1 and described in detail below.

Related surveys. In order to justify the need of conducting a systematic review, we first conducted a search for related surveys and literature reviews. We came across a study [3] conducted in 2005, which summarizes 12 widely accepted information quality frameworks applied on the World Wide Web. The study compares the frameworks and identifies 20 dimensions common between them. Additionally, there is a comprehensive review [7], which surveys 13 methodologies for assessing the data quality of datasets available on the Web, in structured or semi-structured formats. Our survey is different since it focuses only on structured (linked) data and on approaches that aim at assessing the quality of LD. Additionally, the prior review (i.e. [3]) only focused on the data quality dimensions identified in the constituent approaches. In our

survey, we not only identify existing dimensions but also introduce new dimensions relevant for assessing the quality of LD. Furthermore, we describe quality assessment metrics corresponding to each of the dimensions and also identify whether they are quantitatively or qualitatively measured.

Research question. The goal of this review is to analyze existing methodologies for assessing the quality of structured data, with particular interest in LD. To achieve this goal, we aim to answer the following general research question:

How can one assess the quality of Linked Data employing a conceptual framework integrating prior approaches?

We can divide this general research question into further sub-questions such as:

– *What are the data quality problems that each approach assesses?*
– *Which are the data quality dimensions and metrics supported by the proposed approaches?*
– *What kinds of tools are available for data quality assessment?*

Eligibility criteria. As a result of a discussion between the two reviewers, a list of eligibility criteria was obtained as listed below. The articles *had to* satisfy the first criterion and *one* of the other four criteria to be included in our study.

– Inclusion criteria:
  Must satisfy:

  ∗ Studies published in English between 2002 and 2014.

  and should satisfy any one of the four criteria:

  ∗ Studies focused on data quality assessment for LD
  ∗ Studies focused on trust assessment of LD
  ∗ Studies that proposed and/or implemented an approach for data quality assessment in LD
  ∗ Studies that assessed the quality of LD or information systems based on LD principles and reported issues

– Exclusion criteria:

  ∗ Studies that were not peer-reviewed or published
  ∗ Assessment methodologies that were published as a poster abstract

* Studies that focused on data quality management

* Studies that neither focused on LD nor on other forms of structured data

* Studies that did not propose any methodology or framework for the assessment of quality in LD

Search strategy. Search strategies in a systematic review are usually iterative and are run separately by both members to avoid bias and ensure complete coverage of all related articles. Based on the research question and the eligibility criteria, each reviewer identified several terms that were most appropriate for this systematic review, such as: *data*, *quality*, *data quality*, *assessment*, *evaluation*, *methodology*, *improvement*, or *linked data*. These terms which were used as follows:

– *linked data* and (*quality* OR *assessment* OR *evaluation* OR *methodology* OR *improvement*)

– (*data* OR *quality* OR *data quality*) AND (*assessment* OR *evaluation* OR *methodology* OR *improvement*)

As suggested in [36,48], searching in the *title* alone does not always provide us with all relevant publications. Thus, the *abstract* or *full-text* of publications should also potentially be included. On the other hand, since the search on the full-text of studies results in many irrelevant publications, we chose to apply the search query first on the *title* and *abstract* of the studies. This means a study is selected as a candidate study if its title or abstract contains the keywords defined in the search string.

After we defined the search strategy, we applied the keyword search in the following list of search engines, digital libraries, journals, conferences and their respective workshops:
Search Engines and digital libraries:

– Google Scholar
– ISI Web of Science
– ACM Digital Library
– IEEE Xplore Digital Library
– Springer Link
– Science Direct

Journals:

– Semantic Web Journal (SWJ)
– Journal of Web Semantics (JWS)
– Journal of Data and Information Quality (JDIQ)
– Journal of Data and Knowledge Engineering (DKE)

– Theoretical Computer Science (TCS)
– International Journal on Semantic Web and Information Systems' (IJSWIS) Special Issue on Web Data Quality (WDQ)

Conferences and Workshops:

– International World Wide Web Conference (WWW)
– International Semantic Web Conference (ISWC)
– European Semantic Web Conference (ESWC)
– Asian Semantic Web Conference (ASWC)
– International Conference on Data Engineering (ICDE)
– Semantic Web in Provenance Management (SWPM)
– Consuming Linked Data (COLD)
– Linked Data on the Web (LDOW)
– Web Quality (WQ)
– I-Semantics (I-Sem)
– Linked Science (LISC)
– On the Move to Meaningful Internet Systems (OTM)
– Linked Web Data Management (LWDM)

Thereafter, the bibliographic metadata about the 118 potentially relevant primary studies were recorded using the bibliography management platform Mendeley[3] and the duplicates were removed.

Titles and abstract reviewing. Both reviewers independently screened the titles and abstracts of the retrieved 118 articles to identify the potentially eligible articles. In case of disagreement while merging the lists, the problem was resolved either by mutual consensus or by creating a list of articles to go under a more detailed review. Then, both the reviewers compared the articles and based on mutual agreement obtained a final list of 73 articles to be included.

Retrieving further potential articles. In order to ensure that all relevant articles were included, the following additional search strategies were applied:

– Looking up the references in the selected articles
– Looking up the article title in Google Scholar and retrieving the "Cited By" articles to check against the eligibility criteria
– Taking each data quality dimension individually and performing a related article search
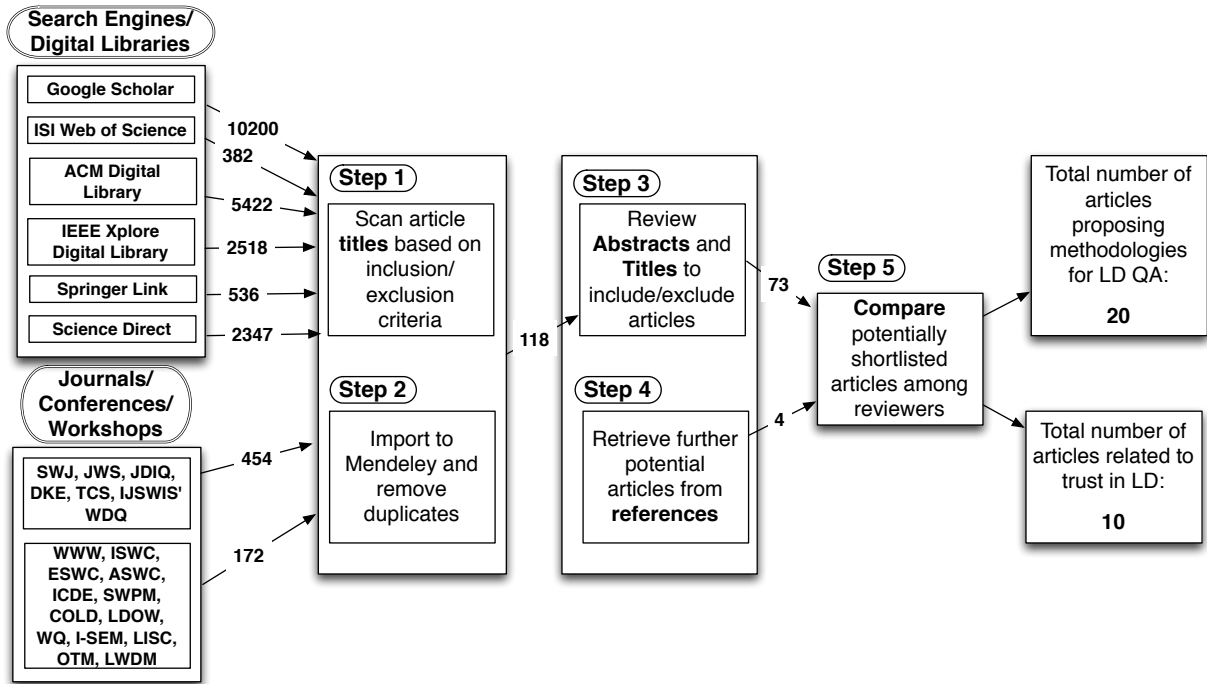
---

[3] http://mnd.ly/1k2Dixy

Fig. 1. Number of articles retrieved during the systematic literature search.

After performing these search strategies, we retrieved four additional articles that matched the eligibility criteria.

Compare potentially shortlisted articles among reviewers. At this stage, a total of 73 articles relevant for our survey were retrieved. Both reviewers then compared notes and went through each of these 73 articles to shortlist potentially relevant articles. As a result of this step, 30 articles were selected.

Extracting data for quantitative and qualitative analysis. As a result of this systematic search, we retrieved 30 articles from 2002 to 2014 as listed in Table 1, which are the core of our survey. Of these 30, 20 propose generalized quality assessment methodologies and 10 articles focus on trust related quality assessment.

Comparison perspective of selected approaches. There exist several perspectives that can be used to analyze and compare the selected approaches, such as:

– the definitions of the core concepts
– the dimensions and metrics proposed by each approach
– the type of data that is considered for the assessment

– the comparison of the tools based on several attributes

This analysis is described in Section 3 and Section 4.

Quantitative overview. Out of the 30 selected approaches, eight (26%) were published in a journal, particularly in the Journal of Web Semantics, International Journal on Semantic Web and Information Systems and Theoretical Computer Science. On the other hand, 21 (70%) approaches were published in international conferences such as WWW, ISWC, ESWC, I-Semantics, ICDE and their associated workshops. Only one (4%) of the approaches was a master thesis. The majority of the papers were published in an even distribution between the years 2010 and 2014 (four papers on average each year $-$ 73%), two papers were published in 2008 and 2009 (6%) and the remaining six between 2002 and 2008 (21%).

3. Conceptualization

There exist a number of discrepancies in the definition of many concepts in data quality due to the contextual nature of quality [7]. Therefore, we first describe and formally define the research context terminology

Table 1
List of the selected papers.

Citation	Title
Gil et al., 2002 [23]	Trusting Information Sources One Citizen at a Time
Golbeck et al., 2003 [25]	Trust Networks on the Semantic Web
Mostafavi et al., 2004 [50]	An ontology-based method for quality assessment of spatial data bases
Golbeck, 2006 [24]	Using Trust and Provenance for Content Filtering on the Semantic Web
Gil et al., 2007 [22]	Towards content trust of Web resources
Lei et al., 2007 [42]	A framework for evaluating semantic metadata
Hartig, 2008 [28]	Trustworthiness of Data on the Web
Bizer et al., 2009 [9]	Quality-driven information filtering using the WIQA policy framework
Böhm et al., 2010 [11]	Profiling linked open data with ProLOD
Chen et al., 2010 [15]	Hypothesis generation and data quality assessment through association mining
Flemming, 2010 [19]	Assessing the quality of a Linked Data source
Hogan et al.,2010 [31]	Weaving the Pedantic Web
Shekarpour et al., 2010 [58]	Modeling and evaluation of trust with an extension in semantic web
Fürber et al.,2011 [20]	SWIQA — a semantic web information quality assessment framework
Gamble et al., 2011 [21]	Quality, Trust and Utility of Scientific Data on the Web: Towards a Joint Model
Jacobi et al., 2011 [33]	Rule-Based Trust Assessment on the Semantic Web
Bonatti et al., 2011 [12]	Robust and scalable linked data reasoning incorporating provenance and trust annotations
Ciancaglini et al., 2012 [16]	Tracing where and who provenance in Linked Data: a calculus
Guéret et al., 2012 [26]	Assessing Linked Data Mappings Using Network Measures
Hogan et al., 2012 [32]	An empirical survey of Linked Data conformance
Mendes et al., 2012 [47]	Sieve: Linked Data Quality Assessment and Fusion
Rula et al., 2012 [56]	Capturing the Age of Linked Open Data: Towards a Dataset-independent Framework
Acosta et al., 2013 [1]	Crowdsourcing Linked Data Quality Assessment
Zaveri et al., 2013 [64]	User-driven Quality evaluation of DBpedia
Albertoni et al., 2013 [2]	Assessing Linkset Quality for Complementing Third-Party Datasets
Feeney et al., 2014 [18]	Improving curated web-data quality with structured harvesting and assessment
Kontokostas et al., 2014 [37]	Test-driven Evaluation of Linked Data Quality
Paulheim et al., 2014 [52]	Improving the Quality of Linked Data Using Statistical Distributions
Ruckhaus et al., 2014 [55]	Analyzing Linked Data Quality with LiQuate
Wienand et al., 2014 [63]	Detecting Incorrect Numerical Data in DBpedia

(in this section) as well as the LD quality dimensions along with their respective metrics in detail (in Section 4).

Data Quality. Data quality is commonly conceived as a multi-dimensional construct with a popular definition "'fitness for use' [35]". Data quality may depend on various factors (dimensions or characteristics) such as accuracy, timeliness, completeness, relevancy, objectivity, believability, understandability, consistency, conciseness, availability and verifiability [62].

In terms of the Semantic Web, there exist different means of assessing data quality. The process of measuring data quality is supported by quality related metadata as well as data itself. On the one hand, provenance (as a particular case of metadata) information, for example, is an important concept to be considered when assessing the trustworthiness of datasets [41]. On the other hand, the notion of link quality is another important aspect that is introduced in LD, where it is automatically detected whether a link is useful or not [26]. It is to be noted that *data* and *information* are interchangeably used in the literature.

Data Quality Problems. Bizer et al. [9] relates data quality problems to those arising in web-based information systems, which integrate information from different providers. For Mendes et al. [47], the problem of data quality is related to values being in conflict between different data sources as a consequence of the diversity of the data. Flemming [19], on the other hand, implicitly explains the data quality problems in terms of *data diversity*. Hogan et al. [31,32] discuss about *errors*, *noise*, *difficulties* or *modelling issues*, which are

prone to the non-exploitations of the data from the applications.

Thus, the term *data quality problem* refers to a set of issues that can affect the potentiality of the applications that use the data.

Data Quality Dimensions and Metrics. Data quality assessment involves the measurement of quality *dimensions* or *criteria* that are relevant to the consumer. The dimensions can be considered as the characteristics of a dataset. A data quality assessment *metric*, *measure* or *indicator* is a procedure for measuring a data quality dimension [9]. These metrics are heuristics that are designed to fit a specific assessment situation [43]. Since dimensions are rather abstract concepts, the assessment metrics rely on quality *indicators* that allow the assessment of the quality of a data source w.r.t the criteria [19]. An assessment score is computed from these indicators using a scoring function.

There are a number of studies which have identified, defined and grouped data quality dimensions into different classifications [7,9,34,51,54,60,62] For example, Bizer et al. [9], classified the data quality dimensions into three categories according to the type of information that is used as a quality dimension: (i) Content Based − information content itself; (ii) Context Based − information about the context in which information was claimed; (iii) Rating Based − based on the ratings about the data itself or the information provider. However, we identify further dimensions (defined in Section 4) and classify the dimensions into the (i) Accessibility (ii) Intrinsic (iii) Contextual and (iv) Representational groups.

Data Quality Assessment Methodology. A data quality assessment methodology is defined as the process of evaluating if a piece of data meets the information consumers need in a specific use case [9]. The process involves measuring the quality dimensions that are relevant to the user and comparing the assessment results with the user's quality requirements.

4. Linked Data quality dimensions

After analyzing the 30 selected approaches in detail, we identified a core set of 18 different data quality dimensions that can be applied to assess the quality of LD. We grouped the identified dimensions according to the classification introduced in [62]:

– Accessibility dimensions
– Intrinsic dimensions

– Contextual dimensions
– Representational dimensions

We further re-examine the dimensions belonging to each group and change their membership according to the LD context. In this section, we unify, formalize and adapt the definition for each dimension according to LD. For each dimension, we identify metrics and report them too. In total, 69 metrics are provided for all the 18 dimensions. Furthermore, we classify each metric as being quantitatively or qualitatively assessed. Quantitatively (QN) measured metrics are those that are quantified or for which a concrete value (score) can be calculated. Qualitatively (QL) measured metrics are those which cannot be quantified and depend on the users perception of the respective metric.

In general, a group captures the same essence for the underlying dimensions that belong to that group. However, these groups are not strictly disjoint but can partially overlap since there exist trade-offs between the dimensions of each group as described in Section 4.5. Additionally, we provide a general use case scenario and specific examples for each of the dimensions. In certain cases, the examples point towards the quality of the information systems such as search engines (e.g. performance) and in other cases, about the data itself.

Use case scenario. Since data quality is conceived as "fitness for use", we introduce a specific use case that will allow us to illustrate the importance of each dimension with the help of an example. The use case is about an intelligent flight search engine, which relies on aggregating data from several datasets. The search engine obtains information about airports and airlines from an airline dataset (e.g. *OurAirports*[4], *OpenFlights*[5]). Information about the location of countries, cities and particular addresses is obtained from a spatial dataset (e.g. *LinkedGeoData*[6]). Additionally, aggregators pull all the information related to flights from different booking services (e.g., *Expedia*[7]) and represent this information as RDF. This allows a user to query the integrated dataset for a flight between any start and end destination for any time period. We will use this scenario throughout as an example to explain each quality dimension through a quality issue.

---

[4] http://thedatahub.org/dataset/ourairports
[5] http://thedatahub.org/dataset/open-flights
[6] http://linkedgeodata.org
[7] http://www.expedia.com/

4.1. Accessibility dimensions

The dimensions belonging to this category involve 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. There are five dimensions that are part of this group, which are *availability, licensing, interlinking, security* and *performance*. Table 2 displays metrics for these dimensions and provides references to the original literature.

4.1.1. Availability.

Flemming [19] referred to availability as the proper functioning of all access methods. The other articles [31,32] provide metrics for this dimension.

Definition 1 (Availability). *Availability of a dataset is the extent to which data (or some portion of it) is present, obtainable and ready for use.*

Metrics. The metrics identified for availability are:

– A1: checking whether the server responds to a SPARQL query [19]
– A2: checking whether an RDF dump is provided and can be downloaded [19]
– A3: detection of dereferenceability of URIs by checking:

  * for dead or broken links [31], i.e. that when an HTTP-GET request is sent, the status code `404 Not Found` is not returned [19]
  * that useful data (particularly RDF) is returned upon lookup of a URI [31]
  * for changes in the URI, i.e. compliance with the recommended way of implementing redirections using the status code `303 See Other` [19]

– A4: detect whether the HTTP response contains the header field stating the appropriate content type of the returned file, e.g. `application/rdf+xml` [31]
– A5: dereferenceability of all forward links: all available triples where the local URI is mentioned in the subject (i.e. the description of the resource) [32]

Example. Let us consider the case in which a user looks up a flight in our flight search engine. She requires additional information such as car rental and hotel booking at the destination, which is present in another dataset and interlinked with the flight dataset.

However, instead of retrieving the results, she receives an error response code `404 Not Found`. This is an indication that the requested resource cannot be dereferenced and is therefore unavailable. Thus, with this error code, she may assume that either there is no information present at that specified URI or the information is unavailable.

4.1.2. Licensing.

Licensing is a new quality dimensions not considered for relational databases but mandatory in the LD world. Flemming [19] and Hogan et al. [32] both stated that each RDF document should contain a license under which the content can be (re)used, in order to enable information consumers to use the data under clear legal terms. Additionally, the existence of a machine-readable indication (by including the specifications in a VoID[8] description) as well as a human-readable indication of a license are important not only for the permissions a license grants but as an indication of which requirements the consumer has to meet [19]. Although both these studies do not provide a formal definition, they agree on the use and importance of licensing in terms of data quality.

Definition 2 (Licensing). *Licensing is defined as the granting of permission for a consumer to re-use a dataset under defined conditions.*

Metrics. The metrics identified for licensing are:

– L1: machine-readable indication of a license in the VoID description or in the dataset itself [19, 32]
– L2: human-readable indication of a license in the documentation of the dataset [19,32]
– L3: detection of whether the dataset is attributed under the same license as the original [19]

Example. Since our flight search engine aggregates data from several existing data sources, a clear indication of the license allows the search engine to re-use the data from the airlines websites. For example, the LinkedGeoData dataset is licensed under the Open Database License[9], which allows others to copy, distribute and use the data and produce work from the data allowing modifications and transformations. Due to the presence of this specific license, the flight search engine is able to re-use this dataset to pull geo-spatial information and feed it to the search engine.

---

[8]`http://vocab.deri.ie/void`
[9]`http://opendatacommons.org/licenses/odbl/`

Table 2

Data quality metrics related to accessibility dimensions (type QN refers to a quantitative metric, QL to a qualitative one).

Dimension	Abr	Metric	Description	Type
Availability	A1	accessibility of the SPARQL end-point and the server	checking whether the server responds to a SPARQL query [19]	QN
	A2	accessibility of the RDF dumps	checking whether an RDF dump is provided and can be down-loaded [19]	QN
	A3	dereferenceability of the URI	checking (i) for dead or broken links i.e. when an HTTP-GET request is sent, the status code `404 Not Found` is not be returned (ii) that useful data (particularly RDF) is returned upon lookup of a URI, (iii) for changes in the URI i.e the compliance with the recommended way of implementing redirections using the status code `303 See Other` [19,31]	QN
	A4	no misreported content types	detect whether the HTTP response contains the header field stating the appropriate content type of the returned file e.g. `application/rdf+xml` [31]	QN
	A5	dereferenced forward-links	dereferenceability of all forward links: all available triples where the local URI is mentioned in the subject (i.e. the description of the resource) [32]	QN
Licensing	L1	machine-readable indication of a license	detection of the indication of a license in the VoID description or in the dataset itself [19,32]	QN
	L2	human-readable indication of a license	detection of a license in the documentation of the dataset [19, 32]	QN
	L3	specifying the correct license	detection of whether the dataset is attributed under the same license as the original [19]	QN
Interlinking	I1	detection of good quality inter-links	(i) detection of (a) interlinking degree, (b) clustering coefficient, (c) centrality, (d) open sameAs chains and (e) description richness through sameAs by using network measures [26], (ii) via crowdsourcing [1,64]	QN
	I2	existence of links to external data providers	detection of the existence and usage of external URIs (e.g. using `owl:sameAs` links) [32]	QN
	I3	dereferenced back-links	detection of all local in-links or back-links: all triples from a dataset that have the resource's URI as the object [32]	QN
Security	S1	usage of digital signatures	by signing a document containing an RDF serialization, a SPARQL result set or signing an RDF graph [14,19]	QN
	S2	authenticity of the dataset	verifying authenticity of the dataset based on a provenance vocabulary such as the author and his contributors, the publisher of the data and its sources, if present in the dataset [19]	QL
Performance	P1	usage of slash-URIs	checking for usage of slash-URIs where large amounts of data is provided [19]	QN
	P2	low latency	(minimum) delay between submission of a request by the user and reception of the response from the system [19]	QN
	P3	high throughput	(maximum) no. of answered HTTP-requests per second [19]	QN
	P4	scalability of a data source	detection of whether the time to answer an amount of ten requests divided by ten is not longer than the time it takes to answer one request [19]	QN

4.1.3. Interlinking.

Interlinking is a relevant dimension in LD since it supports data integration. Interlinking is provided by RDF triples that establish a link between the entity identified by the subject with the entity identified by the object. Through the typed RDF links, data items are effectively interlinked. Even though the core articles in this survey do not contain a formal definition for interlinking, they provide metrics for this dimension [26,31,32].

Definition 3 (Interlinking). *Interlinking refers to 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.*

Metrics. The metrics identified for interlinking are:

– I1: detection of:

* interlinking degree: how many hubs there are in a network[10] [26]
* clustering coefficient: how dense is the network [26]
* centrality: indicates the likelihood of a node being on the shortest path between two other nodes [26]
* whether there are open sameAs chains in the network [26]
* how much value is added to the description of a resource through the use of sameAs edges [26]

  – I2: detection of the existence and usage of external URIs (e.g. using `owl:sameAs` links) [31,32]
  – I3: detection of all local in-links or back-links: all triples from a dataset that have the resource's URI as the object [32]

Example. In our flight search engine, the instance of the country "`United States`" in the airline dataset should be interlinked with the instance "`America`" in the spatial dataset. This interlinking can help when a user queries for a flight, as the search engine can display the correct route from the start destination to the end destination by correctly combining information for the same country from both datasets. Since names of various entities can have different URIs in different datasets, their interlinking can help in disambiguation.

4.1.4. Security.

Flemming [19] referred to security as "the possibility to restrict access to the data and to guarantee the confidentiality of the communication between a source and its consumers". Additionally, Flemming referred to the verifiability dimension as the means a consumer is provided with to examine the data for correctness. Thus, security and verifiability point towards the same quality dimension i.e. to avoid alterations of the dataset and verify its correctness.

Definition 4 (Security). *Security is the extent to which data is protected against alteration and misuse.*

Metrics. The metrics identified for security are:

  – S1: using digital signatures to sign documents containing an RDF serialization, a SPARQL result set or signing an RDF graph [19]

  – S2: verifying authenticity of the dataset based on provenance information such as the author and his contributors, the publisher of the data and its sources (if present in the dataset) [19]

Example: In our use case, if we assume that the flight search engine obtains flight information from arbitrary airline websites, there is a risk for receiving incorrect information from malicious websites. For instance, an airline or sales agency website can pose as its competitor and display incorrect flight fares. Thus, by this spoofing attack, this airline can prevent users to book with the competitor airline. In this case, the use of standard security techniques such as digital signatures allows verifying the identity of the publisher.

4.1.5. Performance.

Performance is a dimension that has an influence on the quality of the information system or search engine, not on the dataset itself. Flemming [19] states "the performance criterion comprises aspects of enhancing the performance of a source as well as measuring of the actual values". Also, response-time and performance point towards the same quality dimension.

Definition 5 (Performance). *Performance refers to the efficiency of a system that binds to a large dataset, that is, the more performant a data source is the more efficiently a system can process data.*

Metrics. The metrics identified for performance are:

  – P1: checking for usage of slash-URIs where large amounts of data is provided[11] [19]
  – P2: low latency[12]: (minimum) delay between submission of a request by the user and reception of the response from the system [19]
  – P3: high throughput: (maximum) number of answered HTTP-requests per second [19]
  – P4: scalability: detection of whether the time to answer an amount of ten requests divided by ten is not longer than the time it takes to answer one request [19]

Example. In our use case, the performance may depend on the type and complexity of the query by a large number of users. Our flight search engine can perform well by considering response-time when deciding which sources to use to answer a query.

---

[10]In [26], a network is described as a set of facts provided by the graph of the Web of Data, excluding blank nodes.

[11]`http://www.w3.org/wiki/HashVsSlash`

[12]Latency is the amount of time from issuing the query until the first information reaches the user [51].

4.1.6. Intra-relations

The dimensions in this group are related with each other as follows: performance (response-time) of a system is related to the availability dimension. A dataset can perform well only if it is available and has low response time. Also, interlinking is related to availability because only if a dataset is available, it can be interlinked and these interlinks can be traversed. Additionally, the dimensions security and licensing are related since providing a license and specifying conditions for re-use helps secure the dataset against alterations and misuse.

4.2. Intrinsic dimensions

Intrinsic dimensions are those that are independent of the user's context. There are five dimensions that are part of this group, which are *syntactic validity, semantic accuracy, consistency, conciseness* and *completeness*. These dimensions focus on whether information correctly (syntactically and semantically), compactly and completely represents the real world and whether information is logically consistent in itself. Table 3 provides metrics for these dimensions along with references to the original literature.

4.2.1. Syntactic validity.

Fürber et al. [20] classified accuracy into syntactic and semantic accuracy. He explained that a "value is syntactically accurate, when it is part of a legal value set for the represented domain or it does not violate syntactical rules defined for the domain". Flemming [19] defined the term validity of documents as "the valid usage of the underlying vocabularies and the valid syntax of the documents". We thus associate the validity of documents defined by Flemming to syntactic validity. We similarly distinguish between the two types of accuracy defined by Fürber et al. and form two dimensions: *Syntactic validity* (syntactic accuracy) and *Semantic accuracy*. Additionally, Hogan et al. [31] identify syntax errors such as RDF/XML syntax errors, malformed datatype literals and literals incompatible with datatype range, which we associate with syntactic validity. The other articles [1,18,37,63,64] provide metrics for this dimension.

Definition 6 (Syntactic validity). *Syntactic validity is defined as the degree to which an RDF document conforms to the specification of the serialization format.*

Metrics. The metrics identified for syntactic validity are:

– SV1: detecting syntax errors using (i) validators [19,31], (ii) via crowdsourcing [1,64]
– SV2: detecting use of:

  * explicit definition of the allowed values for a certain datatype, (ii) syntactic rules [20], (iii) detecting whether the data conforms to the specific RDF pattern and that the "types" are defined for specific resources [37], (iv) use of different outlier techniques and clustering for detecting wrong values [63]
  * syntactic rules (type of characters allowed and/or the pattern of literal values) [20]

– SV3: detection of ill-typed literals, which do not abide by the lexical syntax for their respective datatype that can occur if a value is (i) malformed, (ii) is a member of an incompatible datatype [18, 31]

Example. In our use case, let us assume that the ID of the flight between Paris and New York is A123 while in our search engine the same flight instance is represented as A231. Since this ID is included in one of the datasets, it is considered to be syntactically accurate since it is a valid ID (even though it is incorrect).

4.2.2. Semantic accuracy.

Fürber et al. [20] classified accuracy into syntactic and semantic accuracy. He explained that values are semantically accurate when they represent the correct state of an object. Based on this definition, we also considered the problems of *spurious annotation* and *inaccurate annotation* (inaccurate labeling and inaccurate classification) identified in Lei et al. [42] related to the semantic accuracy dimension. The other articles [1,9,11,15,18,37,52,64] provide metrics for this dimension.

Definition 7 (Semantic accuracy). *Semantic accuracy is defined as the degree to which data values correctly represent the real world facts.*

Metrics. The metrics identified for semantic accuracy are:

– SA1: detection of outliers by (i) using distance-based, deviation-based and distribution-based methods [9,18], (ii) using the statistical distributions of a certain type to assess the statement's correctness [52]

Linked Data Quality

Table 3

Data quality metrics related to intrinsic dimensions (type QN refers to a quantitative metric, QL to a qualitative one).

Dimension	Abr	Metric	Description	Type
Syntactic validity	SV1	no syntax errors of the documents	detecting syntax errors using (i) validators [19,31], (ii) via crowdsourcing [1,64]	QN
	SV2	syntactically accurate values	by (i) use of explicit definition of the allowed values for a datatype, (ii) syntactic rules [20], (iii) detecting whether the data conforms to the specific RDF pattern and that the "types" are defined for specific resources [37], (iv) use of different outlier techniques and clustering for detecting wrong values [63]	QN
	SV3	no malformed datatype literals	detection of ill-typed literals, which do not abide by the lexical syntax for their respective datatype that can occur if a value is (i) malformed, (ii) is a member of an incompatible datatype [18,31]	QN
Semantic accuracy	SA1	no outliers	by (i) using distance-based, deviation-based and distribution-based methods [9,18], (ii) using the statistical distributions of a certain type to assess the statement's correctness [52]	QN
	SA2	no inaccurate values	by (i) using functional dependencies between the values of two or more different properties [20], (ii) comparison between two literal values of a resource [37], (iii) via crowdsourcing [1,64]	QN
	SA3	no inaccurate annotations, labellings or classifications	$1 - \frac{\text{inaccurate instances}}{\text{total no. of instances}} * \frac{\text{balanced distance metric}}{\text{total no. of instances}}$ [42]	QN
	SA4	no misuse of properties	by using profiling statistics, which support the detection of discordant values or misused properties and facilitate to find valid formats for specific properties [11]	QN
	SA5	detection of valid rules	ratio of the number of semantically valid rules to the number of nontrivial rules [15]	QN
Consistency	CS1	no use of entities as members of disjoint classes	$\frac{\text{no. of entities described as members of disjoint classes}}{\text{total no. of entities described in the dataset}}$ [19,31,37]	QN
	CS2	no misplaced classes or properties	using entailment rules that indicate the position of a term in a triple [18,31]	QN
	CS3	no misuse of `owl:DatatypeProperty` or `owl:ObjectProperty`	detection of misuse of `owl:DatatypeProperty` or `owl:ObjectProperty` through the ontology maintainer [31]	QN
	CS4	members of `owl:DeprecatedClass` or `owl:DeprecatedProperty` not used	detection of use of members of `owl:DeprecatedClass` or `owl:DeprecatedProperty` through the ontology maintainer or by specifying manual mappings from deprecated terms to compatible terms [18,31]	QN
	CS5	valid usage of inverse-functional properties	(i) by checking the uniqueness and validity of the inverse-functional values [31], (ii) by defining a SPARQL query as a constraint [37]	QN
	CS6	absence of ontology hijacking	detection of the re-definition by third parties of external classes/properties such that reasoning over data using those external terms is affected [31]	QN
	CS7	no negative dependencies/correlation among properties	using association rules [11]	QN
	CS8	no inconsistencies in spatial data	through semantic and geometric constraints [50]	QN
	CS9	correct domain and range definition	the attribution of a resource's property (with a certain value) is only valid if the resource (domain), value (range) or literal value (rdfs ranged) is of a certain type - detected by use of SPARQL queries as a constraint [37]	QN
	CS10	no inconsistent values	detection by the generation of a particular set of schema axioms for all properties in a dataset and the manual verification of these axioms [64]	QN
Conciseness	CN1	high intensional conciseness	$\frac{\text{no. of unique properties/classes of a dataset}}{\text{total no. of properties/classes in a target schema}}$ [47]	QN
	CN2	high extensional conciseness	(i) $\frac{\text{no. of unique objects of a dataset}}{\text{total number of objects representations in the dataset}}$ [47], (ii) $1 - \frac{\text{total no. of instances that violate the uniqueness rule}}{\text{total no. of relevant instances}}$ [20,37,42]	QN
	CN3	usage of unambiguous annotations/labels	$1 - \frac{\text{no. of ambiguous instances}}{\text{no. of instances contained in the semantic metadata set}}$ [42,55]	QN
Completeness	CM1	schema completeness	$\frac{\text{no. of classes and properties represented}}{\text{total no. of classes and properties}}$ [20,47]	QN
	CM2	property completeness	(i) $\frac{\text{no. of values represented for a specific property}}{\text{total no. of values for a specific property}}$ [18,20], (ii) exploiting statistical distributions of properties and types to characterize the property and then detect completeness [52]	QN
	CM3	population completeness	$\frac{\text{no. of real-world objects are represented}}{\text{total no. of real-world objects}}$ [18,20,47]	QN
	CM4	interlinking completeness	(i) $\frac{\text{no. of instances in the dataset that are interlinked}}{\text{total no. of instances in a dataset}}$ [26,55], (ii) calculating percentage of mappable types in a datasets that have not yet been considered in the linksets when assuming an alignment among types [2]	QN

- SA2: detection of inaccurate values by (i) using functional dependencies[13] [20] between the values of two or more different properties [20], (ii) comparison between two literal values of a resource [37], (iii) via crowdsourcing [1,64]
- SA3: detection of inaccurate annotations[14], labellings[15] or classifications[16] using the formula:
$1 - \frac{\text{inaccurate instances}}{\text{total no. of instances}} * \frac{\text{balanced distance metric}}{\text{total no. of instances}}$ 17 [42]
- SA4: detection of misuse of properties[18] by using profiling statistics, which support the detection of discordant values or misused properties and facilitate to find valid values for specific properties [11]
- SA5: ratio of the number of semantically valid rules [19] to the number of nontrivial rules[20] [15]

Example. Let us assume that the ID of the flight between Paris and New York is A123, while in our search engine the same flight instance is represented as A231 (possibly manually introduced by a data acquisition error). In this case, the instance is semantically inaccurate since the flight ID does not represent its real-world state i.e. A123.

4.2.3. Consistency.

Hogan et al. [31] defined consistency as "no contradictions in the data". Another definition was given by Mendes et al. [47] that "a dataset is consistent if it is free of conflicting information". The other articles [11,18,19,31,37,50,64] provide metrics for this dimension. However, it should be noted that for some languages such as OWL DL, there are clearly defined semantics, including clear definitions of what inconsistency means. In description logics, model based semantics are used: A knowledge base is a set of axioms.

A model is an interpretation, which satisfies all axioms in the knowledge base. A knowledge base is consistent if and only if it has a model [5].

Definition 8 (Consistency). *Consistency means that a knowledge base is free of (logical/formal) contradictions with respect to particular knowledge representation and inference mechanisms.*

Metrics. A straightforward way to check for consistency is to load the knowledge base into a reasoner and check whether it is consistent. However, for certain knowledge bases (e.g. very large or inherently inconsistent ones) this approach is not feasible. Moreover, most OWL reasoners specialize in the OWL (2) DL sublanguage as they are internally based on description logics. However, it should be noted that Linked Data does not necessarily conform to OWL DL and, therefore, those reasoners cannot directly be applied. Some of the important metrics identified in the literature are:

- CS1: detection of use of entities as members of disjoint classes using the formula:
$\frac{\text{no. of entities described as members of disjoint classes}}{\text{total no. of entities described in the dataset}}$ [19,31, 37]
- CS2: detection of misplaced classes or properties[21] using entailment rules that indicate the position of a term in a triple [18,31]
- CS3: detection of misuse of `owl:DatatypeProperty` or `owl:ObjectProperty` through the ontology maintainer[22] [31]
- CS4: detection of use of members of `owl:DeprecatedClass` or `owl:DeprecatedProperty` through the ontology maintainer or by specifying manual mappings from deprecated terms to compatible terms [18,31]
- CS5: detection of bogus `owl:InverseFunctionalProperty` values (i) by checking the uniqueness and validity of the inverse-functional values [31], (ii) by defining a SPARQL query as a constraint [37]
- CS6: detection of the re-definition by third parties of external classes/properties (ontology hijacking) such that reasoning over data using those external terms is not affected [31]

---

[13]Functional dependencies are dependencies between the values of two or more different properties.

[14]Where an instance of the semantic metadata set can be mapped back to more than one real world object or in other cases, where there is no object to be mapped back to an instance.

[15]Where mapping from the instance to the object is correct but not properly labeled.

[16]In which the knowledge of the source object has been correctly identified by not accurately classified.

[17]Balanced distance metric is an algorithm that calculates the distance between the extracted (or learned) concept and the target concept [45].

[18]Properties are often misused when no applicable property exists.

[19]Valid rules are generated from the real data and validated against a set of principles specified in the semantic network.

[20]The intuition is that the larger a dataset is, the more closely it should reflect the basic domain principles and the semantically incorrect rules will be generated.

---

[21]For example, a URI defined as a class is used as a property or vice-a-versa.

[22]For example, attribute properties used between two resources and relation properties used with literal values.

- CS7: detection of negative dependencies/correlation among properties using association rules [11]
- CS8: detection of inconsistencies in spatial data through semantic and geometric constraints [50]
- CS9: the attribution of a resource's property (with a certain value) is only valid if the resource (domain), value (range) or literal value (rdfs ranged) is of a certain type - detected by use of SPARQL queries as a constraint [37]
- CS10: detection of inconsistent values by the generation of a particular set of schema axioms for all properties in a dataset and the manual verification of these axioms [64]

Example. Let us assume a user looking for flights between Paris and New York on the 21st of December, 2013. Her query returns the following results:

```
Flight From   To        Arrival Departure
A123   Paris  NewYork   14:50   22:35
A123   Paris  London    14:50   22:35
```

The results show that the flight number A123 has two different destinations[23] at the same date and same time of arrival and departure, which is inconsistent with the ontology definition that one flight can only have one destination at a specific time and date. This contradiction arises due to inconsistency in data representation, which is detected by using inference and reasoning.

4.2.4. Conciseness.

Mendes et al. [47] classified conciseness into schema and instance level conciseness. On the schema level (intensional), "a dataset is concise if it does not contain redundant attributes (two equivalent attributes with different names)". Thus, intensional conciseness measures the number of unique schema elements (i.e. properties and classes) of a dataset in relation to the overall number of schema elements in a schema. On the data (instance) level (extensional), "a dataset is concise if it does not contain redundant objects (two equivalent objects with different identifiers)". Thus, extensional conciseness measures the number of unique objects in relation to the overall number of objects in the dataset. This definition of conciseness is very similar to the definition of 'uniqueness' defined by Fürber et al. [20] as the "degree to which data is free of redundancies, in breadth, depth and scope". This comparison shows that uniqueness

and conciseness point to the same dimension. Redundancy occurs when there are *equivalent* schema elements with different names/identifiers (in case of intensional conciseness) and when there are *equivalent* objects (instances) with different identifiers (in case of extensional conciseness) present in a dataset [42]. Kontokostas et al.[37] provide metrics for this dimension.

Definition 9 (Conciseness). *Conciseness refers to the minimization of redundancy of entities at the schema and the data level. Conciseness is classified into (i) intensional conciseness (schema level) which refers to the case when the data does not contain redundant schema elements (properties and classes) and (ii) extensional conciseness (data level) which refers to the case when the data does not contain redundant objects (instances).*

Metrics. The metrics identified for conciseness are:

- CN1: intensional conciseness measured by $\frac{\text{no. of unique properties/classes of a dataset}}{\text{total no. of properties/classes in a target schema}}$ [47]
- CN2: extensional conciseness measured by:
  * $\frac{\text{no. of unique instances of a dataset}}{\text{total number of instances representations in the dataset}}$ [47],
  * $1 - \frac{\text{total no. of instances that violate the uniqueness rule}}{\text{total no. of relevant instances}}$ [20, 37,42]

- CN3: detection of unambiguous annotations using the formula:
  $1 - \frac{\text{no. of ambiguous instances}}{\text{no. of instances contained in the semantic metadata set}}$[24] [42, 55]

Example. In our flight search engine, an example of intensional conciseness would be a particular flight, say A123, being represented by two different properties in the same dataset, such as http://flights.org/airlineID and http://flights.org/name. This redundancy ('airlineID' and 'name' in this case) can ideally be solved by fusing the two properties and keeping only one unique identifier. On the other hand, an example of extensional conciseness is when both these identifiers of the same flight have the same information associated with them in both the datasets, thus duplicating the information.

---

[23]Under the assumption that we can infer that NewYork and London are different entities or, alternatively, make the unique name assumption.

[24]Detection of an instance mapped back to more than one real world object leading to more than one interpretation.

4.2.5. Completeness.

Fürber et al. [20] classified completeness into: (i) Schema completeness, which is the degree to which classes and properties are not missing in a schema; (ii) Column completeness, which is a function of the missing property values for a specific property/column; and (iii) Population completeness, which refers to the ratio between classes represented in an information system and the complete population. Mendes et al. [47] distinguished completeness on the schema and the data level. On the schema level, a dataset is complete if it contains all of the attributes needed for a given task. On the data (i.e. instance) level, a dataset is complete if it contains all of the necessary objects for a given task. The two types of completeness defined in Mendes et al. can be mapped to the two categories (i) Schema completeness and (iii) Population completeness provided by Fürber et al. Additionally, we introduce the category interlinking completeness, which refers to the degree to which instances in the dataset are interlinked [26]. Albertoni et al. [2] define interlinking completeness as "linkset completeness as the degree to which links in the linksets are not missing." The other articles [18,52,55] provide metrics for this dimension.

Definition 10 (Completeness). *Completeness refers to the degree to which all required information is present in a particular dataset. In terms of LD, completeness comprises of the following aspects: (i) Schema completeness, the degree to which the classes and properties of an ontology are represented, thus can be called "ontology completeness", (ii) Property completeness, measure of the missing values for a specific property, (iii) Population completeness is the percentage of all real-world objects of a particular type that are represented in the datasets and (iv) Interlinking completeness, which has to be considered especially in LD, refers to the degree to which instances in the dataset are interlinked.*

Metrics. The metrics identified for completeness are:

– CM1: schema completeness $\frac{\text{no. of classes and properties represented}}{\text{total no. of classes and properties}}$ [20,47]
– CM2: property completeness (i) $\frac{\text{no. of values represented for a specific property}}{\text{total no. of values for a specific property}}$ [18,20], (ii) exploiting statistical distributions of properties and types to characterize the property and then detect completeness [52]
– CM3: population completeness $\frac{\text{no. of real-world objects are represented}}{\text{total no. of real-world objects}}$ [18,20,47]

– CM4: interlinking completeness
  (i) $\frac{\text{no. of instances in the dataset that are interlinked}}{\text{total no. of instances in a dataset}}$ [26,55], (ii) calculating percentage of mappable types in a datasets that have not yet been considered in the linksets when assuming an alignment among types [2]

It should be noted that in this case, users should assume a closed-world-assumption where a gold standard dataset is available and can be used to compare against the converted dataset.

Example. In our use case, the flight search engine contains complete information to include all the airports and airport codes such that it allows a user to find an optimal route from the start to the end destination (even in cases when there is no direct flight). For example, the user wants to travel from Santa Barbara to San Francisco. Since our flight search engine contains interlinks between these close airports, the user is able to locate a direct flight easily.

4.2.6. Intra-relations

The dimensions in this group are related to each other as follows: Data can be semantically accurate by representing the real world state but still can be inconsistent. However, if we merge accurate datasets, we will most likely get fewer inconsistencies than merging inaccurate datasets. On the other hand, being syntactically valid does not necessarily mean that the value is semantically accurate. Moreover, if a dataset is complete, tests for syntactic validity, semantic accuracy and consistency checks need to be performed to determine if the values have been completed correctly. Additionally, the conciseness dimension is related to the completeness dimension since both point towards the dataset having all, however unique (non-redundant) information. However, if data integration leads to duplication of instances, it may lead to contradictory values thus leading to inconsistency [10].

4.3. Contextual dimensions

Contextual dimensions are those that highly depend on the context of the task at hand. There are four dimensions that are part of this group, namely *relevancy, trustworthiness, understandability* and *timeliness*. These dimensions along with their corresponding metrics and references to the original literature are presented in Table 4 and Table 5.

Table 4

Data quality metrics related to contextual dimensions (type QN refers to a quantitative metric, QL to a qualitative one).

Dimension	Abr	Metric	Description	Type
Relevancy	R1	relevant terms within meta-information attributes	obtaining relevant data by (i) ranking (a numerical value similar to PageRank), which determines the centrality of RDF documents and statements [12], (ii) via crowd-sourcing [1,64]	QN
	R2	coverage	measuring the coverage (i.e. number of entities described in a dataset) and level of detail (i.e. number of properties) in a dataset to ensure that the data retrieved is appropriate for the task at hand [19]	QN
Trustworthiness	T1	trustworthiness of statements	computing statement trust values based on: (i) provenance information which can be either unknown or a value in the interval $[-1,1]$ where 1: absolute belief, -1: absolute disbelief and 0: lack of belief/disbelief [18, 28] (ii) opinion-based method, which use trust annotations made by several individuals [23,28] (iii) provenance information and trust annotations in Semantic Web-based social-networks [24] (iv) annotating triples with provenance data and usage of provenance history to evaluate the trustworthiness of facts [16]	QN
	T2	trustworthiness through reasoning	using annotations for data to encode two facets of information [12]: (i) blacklists (indicates that the referent data is known to be harmful) (ii) authority (a boolean value which uses the Linked Data principles to conservatively determine whether or not information can be trusted)	QN
	T3	trustworthiness of statements, datasets and rules	using trust ontologies that assigns trust values that can be transferred from known to unknown data using: (i) content-based methods (from content or rules) and (ii) metadata-based methods (based on reputation assignments, user ratings, and provenance, rather than the content itself) [33]	QN
	T4	trustworthiness of a resource	computing trust values between two entities through a path by using: (i) a propagation algorithm based on statistical techniques (ii) in case there are several paths, trust values from all paths are aggregated based on a weighting mechanism [58]	QN
	T5	trustworthiness of the information provider	computing trustworthiness of the information provider by: (i) construction of decision networks informed by provenance graphs [21] (ii) checking whether the provider/contributor is contained in a list of trusted providers [9]	QN
			(iii) indicating the level of trust for the publisher on a scale of $1-9$ [22,25]	QL
	T6	trustworthiness of information provided (content trust)	checking content trust based on associations (e.g. anything having a relationship to a resource such as author of the dataset) that transfers trust from content to resources [22]	QL
	T7	reputation of the dataset	assignment of explicit trust ratings to the dataset by humans or analyzing external links or page ranks [47]	QL
Understandability	U1	human-readable labelling of classes, properties and entities as well as presence of metadata	detection of human-readable labelling of classes, properties and entities as well as indication of metadata (e.g. name, description, website) of a dataset [18,19,32]	QN
	U2	indication of one or more exemplary URIs	detect whether the pattern of the URIs is provided [19]	QN
	U3	indication of a regular expression that matches the URIs of a dataset	detect whether a regular expression that matches the URIs is present [19]	QN
	U4	indication of an exemplary SPARQL query	detect whether examples of SPARQL queries are provided [19]	QN
	U5	indication of the vocabularies used in the dataset	checking whether a list of vocabularies used in the dataset is provided [19]	QN
	U6	provision of message boards and mailing lists	checking the effectiveness and the efficiency of the usage of the mailing list and/or the message boards [19]	QL

Table 5

Data quality metrics related to contextual dimensions (continued) (type QN refers to a quantitative metric, QL to a qualitative one).

Dimension	Abr	Metric	Description	Type
Timeliness	TI1	freshness of datasets based on currency and volatility	$$\max\{0, 1 - \frac{currency}{volatility}\}$$ [29], which gives a value in a continuous scale from 0 to 1, where score of 1 implies that the data is timely and 0 means it is completely outdated thus unacceptable. In the formula, volatility is the length of time the data remains valid [20] and currency is the age of the data when delivered to the user [18,47,56]	QN
	TI2	freshness of datasets based on their data source	detecting freshness of datasets based on their data source by measuring the distance between last modified time of the data source and last modified time of the dataset [20, 46]	QN

4.3.1. Relevancy.

Flemming [19] defined amount-of-data as the "criterion influencing the usability of a data source". Thus, since the amount-of-data dimension is similar to the relevancy dimension, we merge both dimensions. Bonatti et al. [12] provides a metric for this dimension. The other articles [1,64] provide metrics for this dimension.

Definition 11 (Relevancy). *Relevancy refers to the provision of information which is in accordance with the task at hand and important to the users' query.*

Metrics. The metrics identified for relevancy are:

– R1: obtaining relevant data by: (i) ranking (a numerical value similar to PageRank), which determines the centrality of RDF documents and statements [12]), (ii) via crowdsourcing [1,64]
– R2: measuring the coverage (i.e. number of entities described in a dataset) and level of detail (i.e. number of properties) in a dataset to ensure that the data retrieved is appropriate for the task at hand [19]

Example. When a user is looking for flights between any two cities, only relevant information i.e. departure and arrival airports, starting and ending time, duration and cost per person should be provided. Some datasets, in addition to relevant information, also contain much irrelevant data such as car rental, hotel booking, travel insurance etc. and as a consequence a lot of irrelevant extra information is provided. Providing irrelevant data distracts service developers and potentially users and also wastes network resources. Instead, restricting the dataset to only flight related information simplifies application development and increases the likelihood to return only relevant results to users.

4.3.2. Trustworthiness.

Trustworthiness is a crucial topic due to the availability and the high volume of data from varying sources on the Web of Data. Jacobi et al. [33], similar to Pipino et al., referred to trustworthiness as a subjective measure of a user's belief that the data is "true". Gil et al. [22] used reputation of an entity or a dataset either as a result from direct experience or recommendations from others to establish trust. Ciancaglini et al. [16] state "the degree of trustworthiness of the triple will depend on the trustworthiness of the individuals involved in producing the triple and the judgement of the consumer of the triple." We consider reputation as well as objectivity are part of the trustworthiness dimension. Other articles [12,16,18,21,23,24,25,28,47,58] provide metrics for assessing trustworthiness.

Definition 12 (Trustworthiness). *Trustworthiness is defined as the degree to which the information is accepted to be correct, true, real and credible.*

Metrics. The metrics identified for trustworthiness are:

– T1: computing statement trust values based on:

  * provenance information which can be either unknown or a value in the interval $[-1,1]$ where 1: absolute belief, -1: absolute disbelief and 0: lack of belief/disbelief [18,28]
  * opinion-based method, which use trust annotations made by several individuals [23,28]
  * provenance information and trust annotations in Semantic Web-based social-networks [24]
  * annotating triples with provenance data and usage of provenance history to evaluate the trustworthiness of facts [16]

– T2: using annotations for data to encode two facets of information:

  * blacklists (indicates that the referent data is known to be harmful) [12] and
  * authority (a boolean value which uses the Linked Data principles to conservatively determine whether or not information can be trusted) [12]

– T3: using trust ontologies that assigns trust values that can be transferred from known to unknown data [33] using:

  * content-based methods (from content or rules) and
  * metadata-based methods (based on reputation assignments, user ratings, and provenance, rather than the content itself)

– T4: computing trust values between two entities through a path by using:

  * a propagation algorithm based on statistical techniques [58]
  * in case there are several paths, trust values from all paths are aggregated based on a weighting mechanism [58]

– T5: computing trustworthiness of the information provider by:

  * construction of decision networks informed by provenance graphs [21]
  * checking whether the provider/contributor is contained in a list of trusted providers [9]
  * indicating the level of trust for the publisher on a scale of $1 - 9$ [22,25]

– T6: checking content trust[25] based on associations (e.g. anything having a relationship to a resource such as author of the dataset) that transfer trust from content to resources [22]

– T7: assignment of explicit trust ratings to the dataset by humans or analyzing external links or page ranks [47]

Example. In our flight search engine use case, if the flight information is provided by trusted and well-known airlines then a user is more likely to trust the information when it is provided by an unknown travel agency. Generally information about a product or ser-

---

[25]Content trust is a trust judgement on a particular piece of information in a given context [22].

vice (e.g. a flight) can be trusted when it is directly published by the producer or service provider (e.g. the airline). On the other hand, if a user retrieves information from a previously unknown source, she can decide whether to believe this information by checking whether the source is well-known or if it is contained in a list of trusted providers.

4.3.3. Understandability.

Flemming [19] related understandability to the comprehensibility of data i.e. the ease with which human consumers can understand and utilize the data. Thus, comprehensibility can be interchangeably used with understandability. Hogan et al. [32] specified the importance of providing human-readable metadata "for allowing users to visualize, browse and understand RDF data, where providing labels and descriptions establishes a baseline". Feeney et al. [18] provide a metric for this dimension.

Definition 13 (Understandability). *Understandability refers to the ease with which data can be comprehended without ambiguity and be used by a human information consumer.*

Metrics. The metrics identified for understandability are:

– U1: detection of human-readable labelling of classes, properties and entities as well as indication of metadata (e.g. name, description, website) of a dataset [18,19,32]
– U2: detect whether the pattern of the URIs is provided [19]
– U3: detect whether a regular expression that matches the URIs is present [19]
– U4: detect whether examples of SPARQL queries are provided [19]
– U5: checking whether a list of vocabularies used in the dataset is provided [19]
– U6: checking the effectiveness and the efficiency of the usage of the mailing list and/or the message boards [19]

Example. Let us assume that a user wants to search for flights between Boston and San Francisco using our flight search engine. From the data related to Boston in the integrated dataset for the required flight, the following URIs and a label is retrieved:

– `http://rdf.freebase.com/ns/m.049jnng`
– `http://rdf.freebase.com/ns/m.043j22x`

– "Boston Logan Airport"@en

For the first two items no human-readable label is available, therefore the machine is only able to display the URI as a result of the users query. This does not represent anything meaningful to the user besides perhaps that the information is from Freebase. The third entity, however, contains a human-readable label, which the user can easily understand.

4.3.4. Timeliness.

Gamble et al. [21] defined timeliness as "a comparison of the date the annotation was updated with the consumer's requirement". The timeliness dimension is motivated by the fact that it is possible to have current data that is actually incompetent because it reflects a state of the real world that is too old for a specific usage. According to the timeliness dimension, data should ideally be recorded and reported as frequently as the source values change and thus never become outdated. Other articles [18,20,29,47,56] provide metrics for assessing timeliness.

Definition 14. *Timeliness measures how up-to-date data is relative to a specific task.*

Metrics. The metrics identified for timeliness are:

– TI1: detecting freshness of datasets based on currency and volatility using the formula:

$$\max\{0, 1 - \frac{currency}{volatility}\}$$

[29], which gives a value in a continuous scale from 0 to 1, where a score of 1 implies that the data is timely and 0 means it is completely outdated and thus unacceptable. In the formula, currency is the age of the data when delivered to the user [18,47,56] and volatility is the length of time the data remains valid [20]
– TI2: detecting freshness of datasets based on their data source by measuring the distance between the last modified time of the data source and last modified time of the dataset [20]

Example. Consider a user checking the flight timetable for her flight from city A to city B. Suppose that the result is a list of triples comprising of the description of the resource A such as the connecting airports, the time of departure and arrival, the terminal, the gate, etc. This flight timetable is updated every 10 minutes (volatility). Assume there is a change of the flight departure time, specifically a delay of one hour. However, this information is communicated to the con-

trol room with a slight delay. They update this information in the system after 30 minutes. Thus, the timeliness constraint of updating the timetable within 10 minutes is not satisfied which renders the information out-of-date.

4.3.5. Intra-relations

The dimensions in this group are related to each other as follows: Data is of high relevance if data is current for the user needs. The timeliness of information thus influences its relevancy. On the other hand, if a dataset has current information, it is considered to be trustworthy. Moreover, to allow users to properly understand information in a dataset, a system should be able to provide sufficient relevant information.

4.4. Representational dimensions

Representational dimensions capture aspects related to the design of the data such as the *representational-conciseness, interoperability, interpretability* as well as *versatility*. Table 6 displays metrics for these four dimensions along with references to the original literature.

4.4.1. Representational-conciseness.

Hogan et al. [31,32] provide benefits of using shorter URI strings for large-scale and/or frequent processing of RDF data thus encouraging the use of concise representation of the data. Moreover, they emphasized that the use of RDF reification should be avoided "as the semantics of reification are unclear and as reified statements are rather cumbersome to query with the SPARQL query language".

Definition 15 (Representational-conciseness). *Representational-conciseness refers to the representation of the data, which is compact and well formatted on the one hand and clear and complete on the other hand.*

Metrics. The metrics identified for representational-conciseness are:

– RC1: detection of long URIs or those that contain query parameters [18,32]
– RC2: detection of RDF primitives i.e. RDF reification, RDF containers and RDF collections [18, 32]

Example. Our flight search engine represents the URIs for the destination compactly with the use of the airport codes. For example, LEJ is the airport code for Leipzig, therefore the URI is `http://airlines.org/LEJ`. Such short representation of the URIs helps users share and memorize them easily.

Table 6

Data quality metrics related to representational dimensions (type QN refers to a quantitative metric, QL to a qualitative one).

Dimension	Abr	Metric	Description	Type
Representational-conciseness	RC1	keeping URIs short	detection of long URIs or those that contain query parameters [18,32]	QN
	RC2	no use of prolix RDF features	detection of RDF primitives i.e. RDF reification, RDF containers and RDF collections [18,32]	QN
Interoperability	IO1	re-use of existing terms	detection of whether existing terms from all relevant vocabularies for that particular domain have been reused [32]	QL
	IO2	re-use of existing vocabularies	usage of relevant vocabularies for that particular domain [19]	QL
Interpretability	IN1	use of self-descriptive formats	identifying objects and terms used to define these objects with globally unique identifiers [18]	QN
	IN2	detecting the interpretability of data	detecting the use of appropriate language, symbols, units, datatypes and clear definitions [19,53]	QL
	IN3	invalid usage of undefined classes and properties	detection of invalid usage of undefined classes and properties (i.e. those without any formal definition) [31]	QN
	IN4	no misinterpretation of missing values	detecting the use of blank nodes [32]	QN
Versatility	V1	provision of the data in different serialization formats	checking whether data is available in different serialization formats [19]	QN
	V2	provision of the data in various languages	checking whether data is available in different languages [4,19,38]	QN

4.4.2. Interoperability.

Hogan et al. [32] state that the re-use of well-known terms to describe resources in a uniform manner increases the interoperability of data published in this manner and contributes towards the interoperability of the entire dataset. The definition of "uniformity", which refers to the re-use of established formats to represent data as described by Flemming [19], is also associated to the interoperability of the dataset.

Definition 16 (Interoperability). *Interoperability is the degree to which the format and structure of the information conforms to previously returned information as well as data from other sources.*

Metrics. The metrics identified for interoperability are:

– IO1: detection of whether existing terms from all relevant vocabularies for that particular domain have been reused [32]
– IO2: usage of relevant vocabularies for that particular domain [19]

Example. Let us consider different airline datasets using different notations for representing the geo-cordinates of a particular flight location. While one dataset uses the WGS 84 geodetic system, another one uses the GeoRSS points system to specify the location. This makes querying the integrated dataset difficult, as it requires users and the machines to understand the heterogeneous schema. Additionally, with the differ-

ence in the vocabularies used to represent the same concept (in this case the co-ordinates), consumers are faced with the problem of how the data can be interpreted and displayed.

4.4.3. Interpretability.

Hogan et al. [31,32] specify that the ad-hoc definition of classes and properties as well use of blank nodes makes the automatic integration of data less effective and forgoes the possibility of making inferences through reasoning. Thus, these features should be avoided in order to make the data much more interpretable. The other articles [18,19] provide metrics for this dimension.

Definition 17 (Interpretability). *Interpretability refers to technical aspects of the data, that is, whether information is represented using an appropriate notation and whether the machine is able to process the data.*

Metrics. The metrics identified for interpretability are:

– IN1: identifying objects and terms used to define these objects with globally unique identifiers [18]
– IN2: detecting the use of appropriate language, symbols, units, datatypes and clear definitions [19]
– IN3: detection of invalid usage of undefined classes and properties (i.e. those without any formal definition) [31]

– IN4: detecting the use of blank nodes[26] [32]

Example. Consider our flight search engine and a user that is looking for a flight from `Mumbai` to `Boston` with a two day stop-over in `Berlin`. The user specifies the dates correctly. However, since the flights are operated by different airlines, thus different datasets, they have a different way of representing the date. In the first leg of the trip, the date is represented in the format `dd/mm/yyyy` whereas in the other case, the date is represented as `mm/dd/yy`. Thus, the machine is unable to correctly interpret the data and cannot provide an optimal result for this query. This lack of consensus in the format of the date hinders the ability of the machine to interpret the data and thus provide the appropriate flights.

4.4.4. Versatility.

Flemming [19] defined versatility as the "alternative representations of the data and its handling."

Definition 18 (Versatility). *Versatility refers to the availability of the data in different representations and in an internationalized way.*

Metrics. The metrics identified for versatility are:

– V1: checking whether data is available in different serialization formats [19]
– V2: checking whether data is available in different languages [19]

Example. Consider a user who does not understand English but only Chinese and wants to use our flight search engine. In order to cater to the needs of such a user, the dataset should provide labels and other language-dependent information in Chinese so that any user has the capability to understand it.

4.4.5. Intra-relations

The dimensions in this group are related as follows: Interpretability is related to the interoperability of data since the consistent representation (e.g. re-use of established vocabularies) ensures that a system will be able to interpret the data correctly [17]. Versatility is also related to the interpretability of a dataset as the more different forms a dataset is represented in (e.g. in different languages), the more interpretable a dataset is. Additionally, concise representation of the data allows the data to be interpreted correctly.

4.5. Inter-relationships between dimensions

The 18 data quality dimensions explained in the previous sections are not independent from each other but correlations exist among them. In this section, we describe the inter-relations between the 18 dimensions, as shown in Figure 2. If some dimensions are considered more important than others for a specific application (or use case), then favouring the more important ones will result in downplaying the influence of others. The inter-relationships help to identify which dimensions should possibly be considered together in a certain quality assessment application. Hence, investigating the relationships among dimensions is an interesting problem, as shown by the following examples.

First, relationships exist between the dimensions trustworthiness, semantic accuracy and timeliness. When assessing the trustworthiness of a LD dataset, the semantic accuracy and the timeliness of the dataset should be assessed. Frequently the assumption is made that a publisher with a high reputation will produce data that is also semantically accurate and current, when in reality this may not be so.

Second, relationships occur between timeliness and the semantic accuracy, completeness and consistency dimensions. On the one hand, having semantically accurate, complete or consistent data may require time and thus timeliness can be negatively affected. Conversely, having timely data may cause low accuracy, incompleteness and/or inconsistency. Based on quality preferences given by an application, a possible order of quality can be as follows: timely, consistent, accurate and then complete data. For instance, a list of courses published on a university website might be first of all timely, secondly consistent and accurate, and finally complete. Conversely, when considering an e-banking application, first of all it is preferred that data is accurate, consistent and complete as stringent requirements and only afterwards timely since delays are allowed in favour of correctness of data provided.

The representational-conciseness dimension (belonging to the representational group) and the conciseness dimension (belonging to the intrinsic group) are also closely related with each other. On the one hand, representational-conciseness refers to the conciseness of *representing* the data (e.g. short URIs) while conciseness refers to the compactness of the *data itself* (no redundant attributes and objects). Both dimensions thus point towards the compactness of the data. Moreover, representational-conciseness not only allows users to understand the data better but also
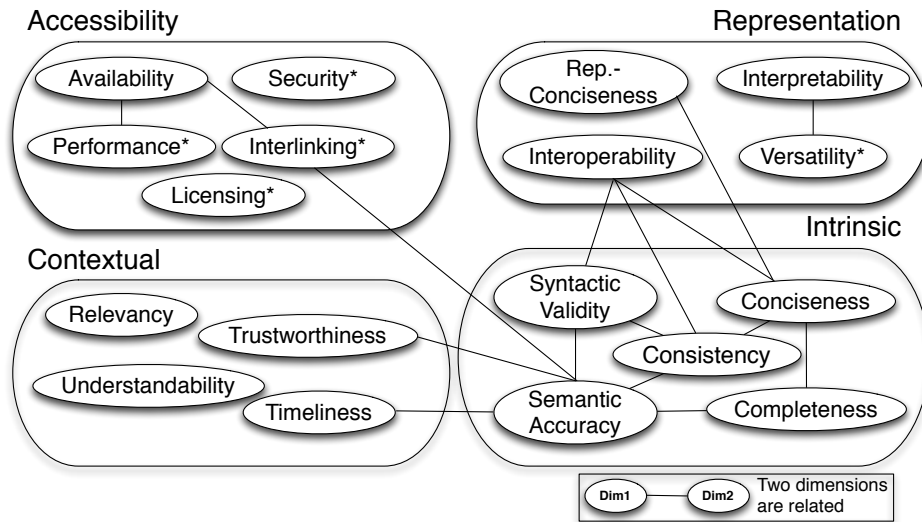
Fig. 2. Linked Data quality dimensions and the relations between them. The dimensions marked with '*' are specific for Linked Data.

provides efficient processing of frequently used RDF data (thus affecting performance). On the other hand, Hogan et al. [32] associated performance to the issue of "using prolix RDF features" such as (i) reification, (ii) containers and (iii) collections. These features should be avoided as they are cumbersome to represent in triples and can prove to be expensive to support in data intensive environments.

Additionally, the interoperability dimension (belonging to the representational group) is inter-related with the consistency dimension (belonging to the intrinsic group), because the invalid re-usage of vocabularies (mandated by the interoperability dimension) may lead to inconsistency in the data. The versatility dimension, also part of the representational group, is related to the accessibility dimension since provision of data via different means (e.g. SPARQL endpoint, RDF dump) inadvertently points towards the different ways in which data can be accessed. Additionally, versatility (e.g. providing data in different languages) allows a user to understand the information better, thus also relates to the understandability dimension. Furthermore, there exists an inter-relation between the conciseness and the relevancy dimensions. Conciseness frequently positively affects relevancy since removing redundancies increases the proportion of relevant data that can be retrieved.

The interlinking dimension is associated with the semantic accuracy dimension. It is important to choose the correct similarity relationship such as *same, matches, similar* or *related* between two entities to capture the most appropriate relationship [27] thus contributing towards the semantic accuracy of the data. Additionally, interlinking is directly related to the interlinking completeness dimension. However, the interlinking dimension focuses on the quality of the interlinks whereas the interlinking completeness focus on the presence of *all* relevant interlinks in a dataset.

These sets of non-exhaustive examples of inter-relations between the dimensions belonging to different groups indicates the interplay between them and show that these dimensions are to be considered differently in different data quality assessment scenarios.

5. Comparison of selected approaches

In this section, we compare the 30 selected approaches based on the different perspectives discussed in Section 2. In particular, we analyze each approach based on (i) the dimensions (Section 5.1), (ii) their respective metrics (Section 5.2), (iii) types of data (Section 5.3), and (iv) compare the proposed tools based on several attributes (Section 5.4).

5.1. Dimensions

Linked Data touches upon three different research and technology areas, namely the *Semantic Web* to generate semantic connections among datasets, the *World Wide Web* to make the data available, preferably under an open access license, and *Data Management*

for handling large quantities of heterogeneous and distributed data. Previously published literature provides a thorough classification of the data quality dimensions [13,34,51,54,60,62]. By analyzing these classifications, it is possible to distill a core set of dimensions, namely accuracy, completeness, consistency and timeliness. These four dimensions constitute the focus of most authors [57]. However, no consensus exists which set of dimensions defines data quality as a whole or the exact meaning of each dimension, which is also a problem occurring in LD.

As mentioned in Section 3, data quality assessment involves the measurement of data quality dimensions that are relevant for the user. We therefore gathered all data quality dimensions that have been reported as being relevant for LD by analyzing the 30 selected approaches. An initial list of data quality dimensions was obtained from [8]. Thereafter, the problem addressed by each approach was extracted and mapped to one or more of the quality dimensions. For example, the problems of dereferenceability, the non-availability of structured data and misreporting content types as mentioned in [32] were mapped to the availability dimension.

However, not all problems related to LD could be mapped to the initial set of dimensions, such as the problem of the alternative data representation and its handling, i.e. the dataset versatility. Therefore, we obtained a further set of quality dimensions from [19], which was one of the first few studies focusing specifically on data quality dimensions and metrics applicable to LD. Still, there were some problems that did not fit in this extended list of dimensions such as incoherency of interlinking between datasets, security against alterations etc. Thus, we introduced new dimensions such as *licensing, interlinking, security, performance and versatility* in order to cover all the identified problems in all of the included approaches, while also mapping them to at least one of these five dimensions.

Table 7 shows the complete list of 18 LD quality dimensions along with their respective occurrence in the included approaches. This table can be intuitively divided into the following three groups: (i) a set of approaches focusing only on trust [12,16,21,22,23,24,25, 28,33,58]; (ii) a set of approaches covering more than four dimensions [9,18,19,20,31,32,37,47,64] and (iii) a set of approaches focusing on very few and specific dimensions [1,11,15,26,31,42,50,52,55,56,63].

Overall, it is observed that the dimensions trustworthiness, consistency, completeness, syntactic valid-

ity, semantic accuracy and availability are the most frequently used. Additionally, the categories intrinsic, contextual, accessibility and representational groups rank in descending order of importance based on the frequency of occurrence of dimensions. Finally, we can conclude that none of the approaches cover all the data quality dimensions.

5.2. Metrics

As defined in Section 3, a data quality metric is *a procedure for measuring an information quality dimension*. We notice that most of the metrics take the form of a ratio, which measures the occurrence of observed entities out of the occurrence of the desired entities, where by entities we mean properties or classes [42]. For example, for the interoperability dimension, the metric for determining the re-use of existing vocabularies takes the form of the ratio:

$$\frac{\text{no. of reused resources}}{\text{total no. of resources}}$$

Other metrics, which cannot be measured as a ratio, can be assessed using algorithms. Table 2, Table 3, Table 4, Table 5 and Table 6 provide the metrics for each of the dimensions.

For some of the surveyed articles, the problem, its corresponding metric and a dimension were clearly mentioned [9,19]. However, for other articles, we first extracted the problem addressed along with the way in which it was assessed (i.e. the metric). Thereafter, we mapped each problem and the corresponding metric to a relevant data quality dimension. For example, the problem related to keeping URIs short (identified in [32]) measured by the presence of long URIs or those containing query parameters, was mapped to the representational-conciseness dimension. On the other hand, the problem related to the re-use of existing terms (also identified in [32]) was mapped to the interoperability dimension.

Additionally, we classified the metrics as being either quantitatively (QN) or qualitatively (QL) assessable. Quantitative metrics are those for which a concrete value (score) can be calculated. For example, for the completeness dimension, the metrics such as schema completeness or property completeness are quantitatively measured. The ratio form of the metrics is generally applied to those metrics, which can be measured objectively (quantitatively). Qualitative metrics are those which cannot be quantified but depend on

the users perception of the respective metric (e.g. via surveys). For example, metrics belonging to the trustworthiness dimension, detection of the trustworthiness of a publisher or a data source can be measured subjectively.

It is worth noting that for a particular dimension there are several metrics associated with it but each metric is only associated with one dimension. Additionally, there are several ways of measuring one dimension either individually or by combining different metrics.

5.3. Type of data

The goal of a data quality assessment activity is the analysis of data in order to measure the quality of datasets along relevant quality dimensions. Therefore, the assessment involves the comparison between the obtained measurements and the references values, in order to enable a diagnosis of quality. The assessment considers different types of data that describe real-world objects in a format that can be stored, retrieved, and processed by a software procedure and communicated through a network.

Thus, in this section, we distinguish between the types of data considered in the various approaches in order to obtain an overview of how the assessment of LD operates on such different levels. The assessment is associated with small-scale units of data such as assessment of RDF triples to the assessment of entire datasets, which potentially affect the whole assessment process.

In LD, we distinguish the assessment process operating on three types of data:

- RDF triples, which focus on individual triple assessment.
- RDF graphs, which focus on entities assessment (where entities are described by a collection of RDF triples [30]).
- Datasets, which focus on dataset assessment where a dataset is considered as a set of default and named graphs.

In Table 8, we can observe that most of the methods are applicable at the triple or graph level and to a lesser extent on the dataset level. Additionally, it can be seen that 12 approaches assess data on both a triple and graph level [11,12,15,19,20,23,37,42,50, 52,55,56], two approaches assess data both at graph and dataset level [21,26] and five approaches assess data at triple, graph and dataset levels [9,24,31,32,64].

There are seven approaches that apply the assessment only at triple level [1,2,16,18,28,47,63] and four approaches that only apply the assessment at the graph level [22,25,33,58].

In most cases, if the assessment is provided at the triple level, this assessment can usually be propagated at a higher level such as the graph or dataset level. For example, in order to assess the rating of a single source, the overall rating of the statements associated to the source can be used [23].

On the other hand, if the assessment is performed at the graph level, it is further propagated either to a more fine-grained level, that is, the RDF triple level or to a more generic one, that is, the dataset level. For example, the evaluation of trust of a data source (graph level) is propagated to the statements (triple level) that are part of the Web source associated with that trust rating [58]. However, there are no approaches that perform an assessment only at the dataset level (see. Table 8). A reason is that the assessment of a dataset always involves the assessment of a fine-grained level (such as triple or entity level) and this assessment is then propagated to the dataset level.

5.4. Comparison of tools

Out of the 30 core articles, 12 provide tools (see Table 8). Hogan et al. [31] only provide a service for validating RDF/XML documents, thus we do not consider it in this comparison. In this section, we compare these 12 tools based on eight different attributes (see Table 9).

Accessibility/Availability. In Table 9, only the tools marked with a ✔ are available to be used for quality assessment. The URL for accessing each tool is available in Table 8. The other tools are either available only as a demo or screencast (Trellis, ProLOD) or not available at all (TrustBot, WIQA, DaCura).

Licensing. Most of the tools are available using a particular software license, which specifies the restrictions with which they can be redistributed. The Trellis and LinkQA tools are open-source and as such by default they are protected by copyright, which is *All Rights Reserved.* Also, WIQA, Sieve, RDFUnit and TripleCheckMate are all available with open-source license: the Apache Version 2.0[27] and Apache licenses. tSPARQL is distributed under the GPL v3 license[28]. However, no licensing information is available

---

[27] http://www.apache.org/licenses/LICENSE-2.0
[28] http://www.gnu.org/licenses/gpl-3.0.html

Approaches / Dimensions	Wïenand et al.2014 [63]	Ruckhaus et al.2014 [55]	Paulheim et al.2014 [52]	Kontokostas et al.2014 [37]	Feeney et al.2014 [18]	Albertoni et al.2013 [2]	Zaveri et al.2013 [64]	Acosta et al.2013 [1]	Bonatti et al.2011 [12]	Jacobi et al., 2011 [33]	Gil et al., 2007 [22]	Golbeck et al., 2003 [25]	Gil et al., 2002 [23]	Golbeck, 2006 [24]	Shekarpour et al., 2010 [58]	Gamble et al., 2011 [21]	Hartig, 2008 [28]	Rula et al. 2012 [56]	Fürber et al.2011 [20]	Mostafavi et al., 2004 [50]	Mendes et al., 2012 [47]	Lei et al.2007 [42]	Hogan et al.2012 [32]	Hogan et al.2010 [31]	Guéret et al.2012 [26]	Ciancaglini et al.2012 [16]	Chen et al.2010 [15]	Böhm et al.2010 [11]	Flemming.2010 [19]	Bizer et al.2009 [9]
Availability																							✔	✔					✔	
Licensing																							✔						✔	
Interlinking					✔	✔																	✔		✔				✔	
Security																													✔	
Performance																													✔	
Syntactic validity				✔	✔		✔	✔											✔					✔					✔	
Semantic accuracy	✔		✔	✔	✔		✔	✔											✔			✔					✔	✔		✔
Consistency				✔	✔		✔													✔				✔			✔	✔		
Conciseness		✔		✔															✔		✔	✔								
Completeness		✔	✔		✔	✔													✔		✔				✔					
Relevancy							✔		✔																				✔	
Trustworthiness					✔			✔	✔	✔	✔	✔	✔	✔	✔	✔			✔						✔					✔
Understandability					✔																		✔						✔	
Timeliness					✔													✔	✔		✔									
Rep.-conciseness					✔																		✔							
Interoperability																							✔						✔	
Interpretability					✔																		✔	✔						
Versatility																													✔	

Table 7: Occurrences of the 18 data quality dimensions in each of the included approaches.

for TrustBot, ProLOD, Flemming's tool, DaCura and LiQuate.

Automation. The automation of a system is the ability to automatically perform its intended tasks thereby reducing the need for human intervention. In this context, we classify the 12 tools into semi-automated and automated approaches. As seen in Table 9, all the tools are semi-automated except for LinkQA, which is completely automated, as there is no user involvement. LinkQA automatically selects a set of resources, information from the Web of Data (i.e. SPARQL endpoints and/or dereferenceable resources) and a set of triples as input and generates the respective quality assessment reports.

On the other hand, the WIQA, Sieve and RDFUnit require a high degree of user involvement. Specifically in Sieve, the definition of metrics has to be done by creating an XML file, which contains specific configurations for a quality assessment task. In case of RDFUnit, the user has to define SPARQL queries as constraints based on SPARQL query templates, which are instantiated into concrete quality test queries. Although it gives the users the flexibility of tweaking the tool to match their needs, it requires much time for understanding the required XML file structure and specification as well as the SPARQL language.

The other semi-automated tools, Trellis, TrurstBot, tSPARQL, ProLOD, Flemming's tool, DaCura, TripleCheckMate and LiQuate require a minimum amount of user involvement. TripleCheckMate provides evaluators with triples from each resource and they are required to mark the triples, which are incorrect as well as map it to one of the pre-defined quality problem. Even though the user involvement here is higher than the other tools, the user-friendly interface allows a user to evaluate the triples and map them to corresponding problems efficiently.

For example, Flemming's Data Quality Assessment Tool requires the user to answer a few questions regarding the dataset (e.g. existence of a human-readable license) or they have to assign weights to each of the pre-defined data quality metrics via a form-based interface.

Collaboration. Collaboration is the ability of a system to support co-operation between different users of the system. From all the tools, Trellis, DaCura and TripleCheckMate support collaboration between different users of the tool. The Trellis user interface allows several users to express their trust value for a data source. The tool allows the users to add and store their observations and conclusions. Decisions made by

users on a particular source are stored as annotations, which can be used to analyze conflicting information or handle incomplete information.

In case of DaCura, the data-architect, domain expert, data harvester and consumer collaborate together to maintain a high-quality dataset. TripleCheckMate, allows multiple users to assess the same Linked Data resource and therefore allowing to calculate the inter-rater agreement to attain a final quality judgement.

Customizability. Customizability is the ability of a system to be configured according to the users' needs and preferences. In this case, we measure the customizability of a tool based on whether the tool can be used with any dataset that the user is interested in. Only LinkQA and LiQuate cannot be customized since the user cannot add any dataset of her choice. The other ten tools can be customized according to the use case. For example, in TrustBot, which is an IRC bot that makes trust recommendations to users (based on the trust network it builds), the users have the flexibility to submit their own URIs to the bot at any time while incorporating the data into a graph. Similarly, Trellis, tSPARQL, WIQA, ProLOD, Flemming's tool, Sieve, RDFUnit, DaCura and TripleCheckmate can be used with any dataset.

Scalability. Scalability is the ability of a system, network, or process to handle a growing amount of work or its ability to be enlarged to accommodate that growth. Out of the 12 tools only five, the tSPARQL, LinkQA, Sieve, RDFUnit and TripleCheckMate tools are scalable, that is, they can be used with large datasets. Flemming's tool and TrustBot are reportedly not scalable for large datasets [19,25]. Flemming's tool, on the one hand, performs analysis based on a sample of three entities whereas TrustBot takes as input two email addresses to calculate the weighted average trust value. Trellis, WIQA, ProLOD, DaCura and LiQuate do not provide any information on the scalability.

Usability/Documentation. Usability is the ease of use and learnability of a human-made object, in this case the quality assessment tool. We assess the usability of the tools based on the ease of use as well as the complete and precise documentation available for each of them thus enabling users to find help easily. We score them based on a scale from 1 (low usability) to 5 (high usability). TripleCheckMate is the easiest tool with a user-friendly interface and a screencast explaining its usage. Thereafter, TrustBot, tSPARQL and Sieve score high in terms of usability and documentation followed by Flemming's tool and RDFUnit. Trel-

Table 8

Qualitative evaluation of the 30 core frameworks included in this survey.

Paper	Goal	Type of data			Tool support	
		RDF Triple	RDF Graph	Dataset	Tool implemented	URL
Gil et al., 2002 [23]	Approach to derive an assessment of a data source based on the annotations of many individuals	✔	✔	—	✔	`http://www.isi.edu/ikcap/ trellis/demo.html`
Golbeck et al., 2003 [25]	Trust networks on the semantic web	—	✔	—	✔	`http://trust.mindswap.org/ trustMail.shtml`
Mostafavi et al.,2004 [50]	Spatial data integration	✔	✔	—	—	—
Golbeck, 2006 [24]	Algorithm for computing personalized trust recommendations using the provenance of existing trust annotations in social networks	✔	✔	✔	—	
Gil et al., 2007 [22]	Trust assessment of web resources	—	✔	—	—	—
Lei et al., 2007 [42]	Assessment of semantic metadata	✔	✔	—	—	—
Hartig, 2008 [28]	Trustworthiness of Data on the Web	✔	—	—	✔	`http://trdf.sourceforge.net/ tsparql.shtml`
Bizer et al., 2009 [9]	Information filtering	✔	✔	✔	✔	`http://wifo5-03.informatik. uni-mannheim.de/bizer/wiqa/`
Böhm et al., 2010 [11]	Data integration	✔	✔	—	✔	`http://tinyurl.com/prolod-01`
Chen et al., 2010 [15]	Generating semantically valid hypothesis	✔	✔	—	—	—
Flemming, 2010 [19]	Assessment of published data	✔	✔	—	✔	`http://linkeddata.informatik. hu-berlin.de/LDSrcAss/`
Hogan et al., 2010 [31]	Assessment of published data by identifying RDF publishing errors and providing approaches for improvement	✔	✔	✔	✔	`http://swse.deri.org/RDFAlerts/`
Shekarpour et al., 2010 [58]	Method for evaluating trust	—	✔	—	—	—
Fürber et al., 2011 [20]	Assessment of published data	✔	✔	—	—	—
Gamble et al., 2011 [21]	Application of decision networks to quality, trust and utility assessment	—	✔	✔	—	—
Jacobi et al., 2011 [33]	Trust assessment of web resources	—	✔	—	—	—
Bonatti et.al., 2011 [12]	Provenance assessment for reasoning	✔	✔	—	—	—
Ciancaglini et al., 2012 [16]	A calculus for tracing where and who provenance in Linked Data	✔	—	—	—	—
Guéret et al., 2012 [26]	Assessment of quality of links	—	✔	✔	✔	`https://github.com/LATC/ 24-7-platform/tree/master/ latc-platform/linkqa`
Hogan et al., 2012 [32]	Assessment of published data	✔	✔	✔	—	—
Mendes et al., 2012 [47]	Data integration	✔	—	—	✔	`http://sieve.wbsg.de/`
Rula et al., 2012 [56]	Assessment of time related quality dimensions	✔	✔	—	—	—
Acosta et al., 2013 [1]	Crowdsourcing Linked Data Quality Assessment	✔	—	—	—	—
Zaveri et al., 2013 [64]	User-driven Quality evaluation of DBpedia	✔	✔	✔	✔	`nl.dbpedia.org: 8080/TripleCheckMate/`
Albertoni et al., 2013 [2]	Assessing Linkset Quality for Complementing Third-Party Datasets	✔	—	—	—	—
Feeney et al., 2014 [18]	Improving curated web-data quality with structured harvesting and assessment	✔	—	—	✔	`http: //dacura.cs.tcd.ie/pv/info.html`
Kontokostas et al., 2014 [37]	Test-driven Evaluation of Linked Data Quality	✔	—	✔	✔	`http://databugger.aksw.org: 8080/rdfunit/`
Paulheim et al., 2014 [52]	Improving the Quality of Linked Data Using Statistical Distributions	✔	—	✔	—	—
Ruckhaus et al., 2014 [55]	Analyze the quality of data and links in the LOD cloud using Bayesian Networks	✔	—	✔	✔	`http://liquate.ldc.usb.ve`
Wienand et al., 2014 [63]	Detecting Incorrect Numerical Data in DBpedia using outlier detection and clustering	✔	—	—	—	—

Table 9

Comparison of quality assessment tools according to several attributes.

	Trellis, Gil et al., 2002 [23]	TrustBot, Golbeck et al., 2003 [25]	tSPARQL, Hartig, 2008 [28]	WIQA, Bizer et al., 2009 [9]	ProLOD, Böhm et al., 2010 [11]	Flemming, 2010 [19]	LinkQA, Gueret et al., 2012 [26]	Sieve, Mendes et al., 2012 [47]	RDFUnit, Kontokostas et al.,2014 [37]	DaCura, Feeney et al., 2014 [18]	Triple Check-Mate, Zaveri et al., 2013 [64]	LiQuate, Ruckhaus et al., 2014 [55]
Accessibility/ Availability	–	–	✔	–	–	✔	✔	✔	✔	–	✔	✔
Licensing	Open-source	–	GPL v3	Apache v2	–	–	Open-source	Apache	Apache	–	Apache	–
Automation	Semi-automated	Semi-automated	Semi-automated	Semi-automated	Semi-automated	Semi-automated	Automated	Semi-automated	Semi-automated	Semi-automated	Semi-automated	Semi-automated
Collaboration	Yes	No	No	No	No	No	No	No	No	Yes	Yes	No
Customizability	✔	✔	✔	✔	✔	✔	No	✔	✔	✔	✔	No
Scalability	–	No	Yes	–	–	No	Yes	Yes	Yes	No	Yes	No
Usability	2	4	4	2	2	3	2	4	3	1	5	1
Maintenance (Last updated)	2005	2003	2012	2006	2010	2010	2011	2012	2014	2013	2013	2013

lis, WIQA, ProLOD and LinkQA rank lower in terms of ease of use since they do not contain useful documentation of how to use the tool. DaCura and LiQuate do not provide any documentation except for a description in the paper.

Maintenance/Last updated. With regards to the current status of the tools, while TrustBot, Trellis and WIQA have not been updated since they were first introduced in 2003, 2005 and 2006 respectively, ProLOD and Flemming's tool have been updated in 2010. The recently updated tools are LinkQA (2011), tSRARQL and Sieve (2012), DaCura, TripleCheckMate, LiQuate (2013) and RDFUnit (2014) and are currently being maintained.

6. Conclusions and future work

In this paper, we have presented, to the best of our knowledge, the most comprehensive systematic review of data quality assessment methodologies applied to LD. The goal of this survey is to obtain a clear understanding of the differences between such approaches, in particular in terms of quality dimensions, metrics, type of data and tools available.

We surveyed 30 approaches and extracted 18 data quality dimensions along with their definitions and corresponding 69 metrics. We also classified the metrics into either being quantitatively or qualitatively assessed. We analyzed the approaches in terms of the dimensions, metrics and type of data they focus on. Additionally, we identified tools proposed by 12 (out of the 30) and compared them using eight different attributes.

We observed that most of the publications focusing on data quality assessment in Linked Data are presented at either conferences or workshops. As our literature review reveals, this research area is still in its infancy and can benefit from the possible re-use of research from mature, related domains. Additionally, in most of the surveyed literature, the metrics were often not explicitly defined or did not consist of precise statistical measures. Moreover, only few approaches were actually accompanied by an implemented tool. Also, there was no formal validation of the methodologies that were implemented as tools. We also observed that none of the existing implemented tools covered all the data quality dimensions. In fact, the best coverage in terms of dimensions was achieved by Flemming's data quality assessment tool with 11 covered dimensions.

Our survey shows that the flexibility of the tools, with regard to the level of automation and user involvement, needs to be improved. Some tools required a considerable amount of configuration while some others were easy-to-use but provided results with limited usefulness or required a high-level of interpretation.

Meanwhile, there is much research on data quality being done and guidelines as well as recommendations on how to publish "good" data are currently available. However, there is less focus on how to *use* this "good" data. Moreover, the quality of datasets should be assessed and an effort to increase the quality aspects that are amiss should be performed thereafter. We deem our data quality dimensions to be very useful for data consumers in order to assess the quality of datasets. As a consequence, query answering can be increased in effectiveness and efficiency using data quality criteria [51]. As a next step, we aim to integrate the various

data quality dimensions into a comprehensive methodological framework for data quality assessment comprising the following steps:

1. Requirements analysis,
2. Data quality checklist,
3. Statistics and low-level analysis,
4. Aggregated and higher level metrics,
5. Comparison,
6. Interpretation.

We aim to develop this framework for data quality assessment allowing a data consumer to select and assess the quality of suitable datasets according to this methodology. In the process, we also expect new metrics to be defined and implemented.

References

[1] Maribel Acosta, Amrapali Zaveri, Elena Simperl, Dimitris Kontokostas, Sören Auer, and Jens Lehmann. Crowdsourcing linked data quality assessment. In *12th International Semantic Web Conference, 21-25 October 2013, Sydney, Australia*, pages 260–276, 2013.

[2] Riccardo Albertoni and Asuncion Gomez Perez. Assessing linkset quality for complementing third-party datasets. In *EDBT '13 Proceedings of the Joint EDBT/ICDT 2013 Workshops*, 2013.

[3] Shirlee ann Knight and Janice Burn. Developing a framework for assessing information quality on the World Wide Web. *Information Science*, 8:159 – 172, 2005.

[4] Sören Auer, Matthias Weidl, Jens Lehmann, Amrapali Zaveri, and Key-Sun Choi. I18n of semantic web applications. In *ISWC 2010*, volume 6497 of *LNCS*, pages 1 – 16, 2010.

[5] Franz Baader, Calvanese Diageo, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge, 2003.

[6] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, 41(3), 2009.

[7] Carlo Batini and Monica Scannapieco. *Data Quality: Concepts, Methodologies and Techniques*. Springer-Verlag New York, Inc., 2006.

[8] Christian Bizer. *Quality-Driven Information Filtering in the Context of Web-Based Information Systems*. PhD thesis, Freie Universität Berlin, March 2007.

[9] Christian Bizer and Richard Cyganiak. Quality-driven information filtering using the WIQA policy framework. *Journal of Web Semantics*, 7(1):1–10, 2009.

[10] Jens Bleiholder and Felix Naumann. Data fusion. *ACM Computing Surveys (CSUR)*, 41(1):1, 2008.

[11] Christoph Böhm, Felix Naumann, Ziawasch Abedjan, Dandy Fenz, Toni Grütze, Daniel Hefenbrock, Matthias Pohl, and David Sonnabend. Profiling linked open data with ProLOD. In *ICDE Workshops*, pages 175–178. IEEE, 2010.

[12] Piero A. Bonatti, Aidan Hogan, Axel Polleres, and Luigi Sauro. Robust and scalable linked data reasoning incorporating provenance and trust annotations. *Journal of Web Semantics*, 9(2):165 – 201, 2011.

[13] Matthew Bovee, Rajendra P. Srivastava, and Brenda Mak. A conceptual framework and belief-function approach to assessing overall information quality. *International Journal of Intelligent Systems*, 18(1):51–74, 2003.

[14] Jeremy J Carroll. Signing RDF graphs. In *ISWC*, pages 369–384. Springer, 2003.

[15] Ping Chen and Walter Garcia. Hypothesis generation and data quality assessment through association mining. In *IEEE ICCI*, pages 659–666. IEEE, 2010.

[16] Mariangiola Dezani-Ciancaglini, Ross Horne, and Vladimiro Sassone. Tracing where and who provenance in linked data: A calculus. *Theoretical Computer Science*, 464:113–129, 2012.

[17] Li Ding and Tim Finin. Characterizing the semantic web on the web. In *5th ISWC*, 2006.

[18] Kevin Chekov Feeney, Declan O'Sullivan, Wei Tai, and Rob Brennan. Improving curated web-data quality with structured harvesting and assessment. *International Journal on Semantic Web and Information Systems*, 2014.

[19] Annika Flemming. Qualitätsmerkmale von Linked Data-veröffentlichenden Datenquellen. Diplomarbeit https://cs.uwaterloo.ca/~ohartig/files/ DiplomarbeitAnnikaFlemming.pdf, 2011.

[20] Christian Fürber and Martin Hepp. SWIQA - a semantic web information quality assessment framework. In *ECIS*, 2011.

[21] Matthew Gamble and Carole Goble. Quality, trust, and utility of scientific data on the web: Towards a joint model. In *ACM WebSci*, pages 1–8, June 2011.

[22] Yolanda Gil and Donovan Artz. Towards content trust of web resources. *Web Semantics*, 5(4):227 – 239, December 2007.

[23] Yolanda Gil and Varun Ratnakar. Trusting information sources one citizen at a time. In *ISWC*, pages 162 – 176, 2002.

[24] Jennifer Golbeck. Using trust and provenance for content filtering on the semantic web. In *Workshop on Models of Trust on the Web at the 15th World Wide Web Conference*, 2006.

[25] Jennifer Golbeck, Bijan Parsia, and James Hendler. Trust networks on the semantic web. In *CIA*, 2003.

[26] Christophe Guéret, Paul Groth, Claus Stadler, and Jens Lehmann. Assessing linked data mappings using network measures. In *ESWC*, 2012.

[27] Harry Halpin, Pat Hayes, James P. McCusker, Deborah McGuinness, and Henry S. Thompson. When owl:sameAs isn't the same: An analysis of identity in linked data. In *9th ISWC*, volume 1, pages 53–59, 2010.

[28] Olaf Hartig. Trustworthiness of data on the web. In *STI Berlin and CSW PhD Workshop, Berlin, Germany*, 2008.

[29] Olaf Hartig and Jun Zhao. Using web data provenance for quality assessment. In Juliana Freire, Paolo Missier, and Satya Sanket Sahoo, editors, *SWPM*, volume 526 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.

[30] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*, chapter 2, pages 1 – 136. Number 1:1 in Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan and Claypool, 1st edition, 2011.

[31] Aidan Hogan, Andreas Harth, Alexandre Passant, Stefan Decker, and Axel Polleres. Weaving the pedantic web. In *3rd Linked Data on the Web Workshop at WWW*, 2010.

[32] Aidan Hogan, Jürgen Umbrich, Andreas Harth, Richard Cyganiak, Axel Polleres, and Stefan Decker. An empirical survey of Linked Data conformance. *Journal of Web Semantics*, 2012.

[33] Ian Jacobi, Lalana Kagal, and Ankesh Khandelwal. Rule-based trust assessment on the semantic web. In *RuleML*, pages 227 – 241, 2011.

[34] Matthias Jarke, Maurizio Lenzerini, Yannis Vassiliou, and Panos Vassiliadis. *Fundamentals of Data Warehouses*. Springer Publishing Company, 2nd edition, 2010.

[35] Joseph Juran. *The Quality Control Handbook*. McGraw-Hill, New York, 1974.

[36] Barbara Kitchenham. Procedures for performing systematic reviews. Technical report, Joint Technical Report Keele University Technical Report TR/SE-0401 and NICTA Technical Report 0400011T.1, 2004.

[37] Dimitris Kontokostas, Patrick Westphal, Sören Auer, Sebastian Hellmann, Jens Lehmann, Roland Cornelissen, and Amrapali Zaveri. Test-driven evaluation of linked data quality. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 747–758. International World Wide Web Conferences Steering Committee, 2014.

[38] Jose Emilio Labra Gayo, Dimitris Kontokostas, and Sören Auer. Multilingual linked open data patterns. *Semantic Web Journal*, 2012.

[39] Yang W. Lee, Diane M. Strong, Berverly K. Kahn, and Richard Y. Wang. AIMQ: a methodology for information quality assessment. *Information Management*, 40(2):133 – 146, 2002.

[40] 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. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 2013. Under review.

[41] Yuangui Lei, Andriy Nikolov, Victoria Uren, and Enrico Motta. Detecting quality problems in semantic metadata without the presence of a gold standard. In *Workshop on "Evaluation of Ontologies for the Web" (EON) at the WWW*, pages 51–60, 2007.

[42] Yuangui Lei, Victoria Uren, and Enrico Motta. A framework for evaluating semantic metadata. In *4th International Conference on Knowledge Capture*, number 8 in K-CAP '07, pages 135 – 142. ACM, 2007.

[43] David Kopcso Leo Pipino, Ricahrd Wang and William Rybold. *Developing Measurement Scales for Data-Quality Dimensions*, volume 1. M.E. Sharpe, New York, 2005.

[44] Holger Michael Lewen. *Facilitating Ontology Reuse Using User-Based Ontology Evaluation*. PhD thesis, Karlsruher Instituts für Technologie (KIT), 2010.

[45] Diana Maynard, Wim Peters, and Yaoyong Li. Metrics for evaluation of ontology-based information extraction. In *Workshop on "Evaluation of Ontologies for the Web" (EON) at WWW*, May 2006.

[46] Pablo Mendes, Christian Bizer, Zoltán Miklos, Jean-Paul Calbimonte, Alexandra Moraru, and Giorgos Flouris. D2.1: Conceptual model and best practices for high-quality metadata publishing. Technical report, PlanetData Deliverable, 2012.

[47] Pablo Mendes, Hannes Mühleisen, and Christian Bizer. Sieve: Linked data quality assessment and fusion. In *LWDM*, March 2012.

[48] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, 6(7), 2009.

[49] Mohamed Morsey, Jens Lehmann, Sören Auer, Claus Stadler, and Sebastian Hellmann. DBpedia and the Live Extraction of Structured Data from Wikipedia. *Program: electronic library and information systems*, 46:27, 2012.

[50] MA. Mostafavi, Edwards G., and R. Jeansoulin. Ontology-based method for quality assessment of spatial data bases. In *International Symposium on Spatial Data Quality*, volume 4, pages 49–66, 2004.

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

[52] Heiko Paulheim and Christian Bizer. Improving the quality of linked data using statistical distributions. *International Journal on Semantic Web and Information Systems*, 2014.

[53] Leo L. Pipino, Yang W. Lee, and Richarg Y. Wang. Data quality assessment. *Communications of the ACM*, 45(4), 2002.

[54] Thomas C. Redman. *Data Quality for the Information Age*. Artech House, 1st edition, 1997.

[55] Edna Ruckhaus, Oriana Baldizán, and María-Esther Vidal. Analyzing linked data quality with liquate. In LNCS, editor, *On the Move to Meaningful Internet Systems: OTM 2013 Workshops*, volume 8186, pages 629–638, 2014.

[56] Anisa Rula, Matteo Palmonari, and Andrea Maurino. Capturing the Age of Linked Open Data: Towards a Dataset-independent Framework. In *IEEE International Conference on Semantic Computing*, 2012.

[57] Monica Scannapieco and Tiziana Catarci. Data quality under a computer science perspective. *Archivi & Computer*, 2:1–15, 2002.

[58] Saeedeh Shekarpour and S.D. Katebi. Modeling and evaluation of trust with an extension in semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(1):26 – 36, March 2010.

[59] Denny Vrandecic. *Ontology Evaluation*. PhD thesis, Karlsruher Instituts für Technologie (KIT), 2010.

[60] Yair Wand and Richard Y. Wang. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11):86–95, 1996.

[61] Richard Y. Wang. A product perspective on total data quality management. *Communications of the ACM*, 41(2):58 – 65, Feb 1998.

[62] Richard Y. Wang and Diane M. Strong. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems*, 12(4):5–33, 1996.

[63] Dominik Wienand and Heiko Paulheim. Detecting incorrect numerical data in dbpedia. In *11th ESWC*, 2014.

[64] Amrapali Zaveri, Dimitris Kontokostas, Mohamed A. Sherif, Lorenz Bühmann, Mohamed Morsey, Sören Auer, and Jens Lehmann. User-driven quality evaluation of DBpedia. In *9th International Conference on Semantic Systems, I-SEMANTICS '13*. ACM, 2013.