

Overview of Recent Methodologies for Open Data Quality Assessment

Klara ŽNIDERŠIČ*, Matija MAROLT, Matevž PESEK

University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia

e-mail: klara.znidersic@fri.uni-lj.si, matija.marolt@fri.uni-lj.si, matevz.pesek@fri.uni-lj.si

Received: November 2024; accepted: November 2025

Abstract. The open data movement has led to the widespread sharing of data across all sectors, offering great potential for innovation and informed decision-making. Nevertheless, open data quality remains a key challenge. This study provides a systematic overview of 16 recent methodologies for data quality assessment, emphasizing their alignment with ISO/IEC 25012 and ISO 8000 standards, FAIR principles, 5-Star Linked Open Data System, and DCAT vocabulary. We also highlight foundational work and identify adaptable methods suitable for the Slovenian open data portal. By recommending practical approaches, this work provides a strategic basis for improving data quality in regional and national platforms, supporting improved data utilization and transparency for end users.

Key words: open data, quality assessment, methodologies.

1. Introduction

As technology advances, large amounts of data are produced in formats that can be easily adapted to different standards. With an increasing number of aspects of life being stored in digital formats—from personal data held by government agencies to information collected by businesses—legislation must keep pace by establishing appropriate regulations to ensure a secure and well-organized data marketplace. A crucial step in encouraging the reuse of collected data is the definition and promotion of open data practices (Commission, 2020).

According to The Open Definition (Open Knowledge, 2015), to be truly open, data must be published under an open license that allows anyone to freely access, use, modify and share it for any purpose. It should be structured in a machine-readable open format that imposes no restrictions and allows processing with a free software tool. Open data includes not only government data (OGD), but also data from the private sector, which has even greater potential when combined with shared or personal data. The true value lies not in the data itself, but in the products, services and content that are enhanced or enabled by open data. Therefore, the reuse of data is not only encouraged in public services, but also in private companies.

*Corresponding author.

The open data market is experiencing steady growth, with various studies indicating a steep rise in its value worldwide (Huyer, 2020; Manyika *et al.*, 2013). In the European context, the availability of open data is strongly encouraged and there are several regulations and directives in place to ensure the highest possible quality of data. Most European countries, as well as some major European cities, have their own national portals for the publication of public service data—for example, France (Interministerial Digital Directorate, 2011), Paris (Ville de Paris, 2023), Italy (Agenzia per L'Italia Digitale, 2022), Piemonte, (Consorzio per il Sistema Informativo Piemonte, 2010) and Slovenia (Ministrstvo za javno upravo, 2016). The official portal for European data (European Union, 2021) aggregates information from these national portals and conducts research and comparative analyses of the datasets. In 2020, the value of the open data market in Europe was estimated at 184 billion euros and is expected to reach between 199 and 334 billion euros by 2025 (Huyer, 2020). In this study, the authors suggest that the value of open data can be understood by examining the economic impact of its reuse. They propose a methodological approach for macro- and microeconomic estimates, with macroeconomic estimates involving quantitative analysis based on gross domestic product (GDP) estimates and microeconomic estimates involving quantitative analysis based on company- or individual-level data. The authors have adapted the methodology and applied it to different areas—market size, job market, sectors, efficiency, saved costs and organizations. They show that analysing available data and incorporating the results into services and products based on or enriched by open data can lead to indirect economic benefits, i.e. efficiency gains. The latter also often reflect societal benefits, for example, when it comes to protecting the environment, saving lives or saving time in various tasks such as commuting by public transport or road traffic.

The estimated total value of the open data market in the European Union is increasing significantly every year, which can be understood as a consequence of the accumulation of data on European data portals. At the same time, these developments increase the expectations and requirements of potential data users. Governments often assume that commercial companies will drive innovation and economic growth through the use of open data. However, many companies opt not to engage with open data due to economic considerations and a lack of knowledge (Zuiderwijk *et al.*, 2015). At the same time, despite the apparent ease of use of freely available open data, users face significant obstacles in identifying suitable datasets and preparing them for use, and the obstacles are often related to the quality of the data (Krasikov and Legner, 2023).

The Slovenian open data landscape performs very well in the European Open Data Maturity assessments (Hesteren *et al.*, 2022), which evaluate four key dimensions: policy, portal, impact and quality. However, this assessment does not provide individual assessments for specific datasets. The Slovenian open data portal currently only offers its users the 5-star ratings on the data openness scale (Berners-Lee, 2012), which focuses primarily on the format of datasets and their machine readability. The lack of a more detailed, dataset-specific rating limits potential users' insight into the quality and usability of individual datasets, which hinders effective reuse.

In this paper, we conduct a comprehensive literature review focusing on different approaches to quality analysis of open data. Our review covers a number of methodologies

that have been presented in existing research that could serve as a basis for developing new approaches for open data portals, including those applicable to the Slovenian context. In addition, we review and present relevant standards that have been previously established, as these standards form the basis for many methods and are often the benchmarks against which the quality requirements of the different approaches are measured. With this review, we aim to identify and summarize the best practices and innovative strategies that can be applied to improve the quality and usability of open data. In addition, our review highlights the common challenges and issues associated with working with open data. By addressing these issues in our future work, we aim to provide a solid foundation for improving the overall quality and effectiveness of open data initiatives in Slovenia.

The key research questions (RQs) we aim to address in our work are as follows:

- RQ1: Which of the established standards already serve as benchmarks for assessing the quality requirements of the methods in our overview?
- RQ2: What are the common challenges and problems encountered when working with open data?
- RQ3: Which of the reviewed methodologies can form the basis for the development of a new methodology specifically tailored to the needs of the Slovenian open data portal?

The main contribution of the article is a systematic overview of open-data quality assessment methods—showing the strengths and drawbacks through recent practice—and synthesis of recurrent challenges observed across methods and use-cases. Finally, the article identifies a shortlist of adaptable methodologies and articulates a strategy tailored to the Slovenian open-data portal (OPSI)—providing a practical basis for dataset-level quality assessment beyond current 5-star openness ratings.

The article is structured as follows: In Section 2, we present the methodology of our literature review. We then discuss in Section 3 the most commonly used data standards, which are essential for understanding the scope and relevance of the studies examined. Section 4 provides an analysis of the methods studied and addresses the main challenges identified in the use cases presented by the authors. In Section 5, we summarize the most cited articles that can serve as a basis for researchers who wish to develop their own methods or tools to assess data quality. Finally, Section 6 offers a discussion addressing the initial research questions, and Section 7 concludes the article with a summary of the key findings.

2. Methodology of Our Research

Our literature review methodology also incorporates elements consistent with established frameworks such as the PRISMA 2020 guidelines (Page *et al.*, 2021) and Kitchenham's Procedures for Performing Systematic Reviews (Kitchenham, 2004). Similar to PRISMA, we employed a transparent, replicable process that documents each stage of identification, screening, eligibility, and inclusion of studies, supported by a visual representation of the review pipeline. In line with Kitchenham's framework, our approach follows a structured sequence of planning, conducting, and reporting the review, ensuring methodological rigour, traceability, and reproducibility.

2.1. Search Strategy

The literature review was conducted using a systematic search strategy aimed at identifying scientific articles focused on the assessment and evaluation of the quality of open data. Our search began in March 2024. The primary search terms included “open data”, “quality assessment”, and “quality evaluation”, supplemented by keywords indicating systematic approaches, such as “metrics”, “framework”, “guidelines”, “methodology”, and “measurement”. The search query was therefore structured as follows: (“open data” AND (“quality assessment” OR “quality evaluation”)) AND (“metrics” OR “framework” OR “guidelines” OR “methodology” OR “measurement”).

To ensure a comprehensive and rigorous review, the search was conducted in several leading academic databases with search engines Web of Science, ScienceDirect, Scopus, IEEE Xplore and the ACM Digital Library. These databases were selected for their comprehensive coverage of literature in data science, information technology, public policy and related disciplines.

2.2. Sources Evaluation

Our initial survey in this area revealed some extensive literature reviews of earlier studies (Zaveri *et al.*, 2013; Ehrlinger and Wöß, 2022). To ensure that the research is aligned with the latest developments, our review was limited to articles published between 2019 and 2024. The starting point of 2019 was deliberately chosen because it coincides with the implementation of the Open Data Directive in the European Union. This directive represents a policy shift aimed at promoting the availability and reuse of public sector information, and its implementation marks the beginning of a new era in open data research and practice. By focusing on the period following the implementation of the Directive, the report captures the most recent research reflecting the possible impact of this regulatory change on the open data landscape. The articles were retrieved in May 2024. To be eligible for selection, articles had to be written in English, peer-reviewed and accessible in the aforementioned libraries. In addition, the full text of each article had to be either freely available or accessible through library access. This criterion was important in order to provide a detailed insight into the methods, results and discussions presented in the articles.

2.3. Selection Criteria

The initial search and title screening (depicted in Fig. 1) resulted in 30 relevant records for further analysis. These articles were then subjected to a more in-depth review, where each article was read in full to assess its suitability for inclusion in the final analysis. At this stage, we assessed each article’s focus on assessing or evaluating the quality of open data, taking into account the methods, frameworks and tools proposed by the authors. By carefully selecting the articles included in the analysis, we have ensured that the review provides valuable insights into existing methods and highlights areas where further research is needed, particularly in relation to the application of these methods to specific cases.

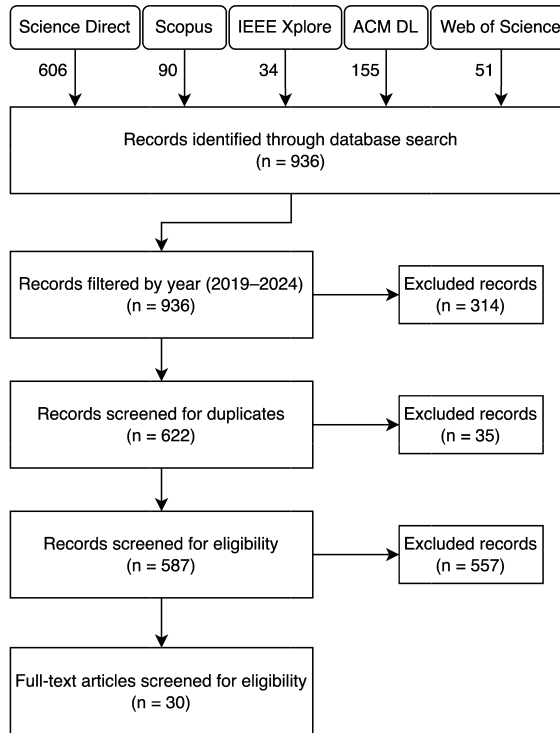


Fig. 1. Literature review pipeline. Source: authors' own elaboration.

3. Data Quality Standards

Addressing the first research question, we review the commonly used standards for data quality.

3.1. *Standard ISO/IEC 25012*

The series of standards designed by the International Organization for Standardization (ISO) to guide the development of software products by defining quality requirements and evaluation criteria is known as SQuaRE (Systems and Software Quality Requirements and Evaluation) (ISO/IEC, 2014). This ISO/IEC 25000 series is composed of:

- ISO/IEC 2500n: Quality Management Division,
- ISO/IEC 2501n: Quality Model Division,
- ISO/IEC 2502n: Quality Measurement Division,
- ISO/IEC 2503n: Quality Requirements Division,
- ISO/IEC 2504n: Quality Evaluation Division,
- ISO/IEC 25050-25099: Extension Division.

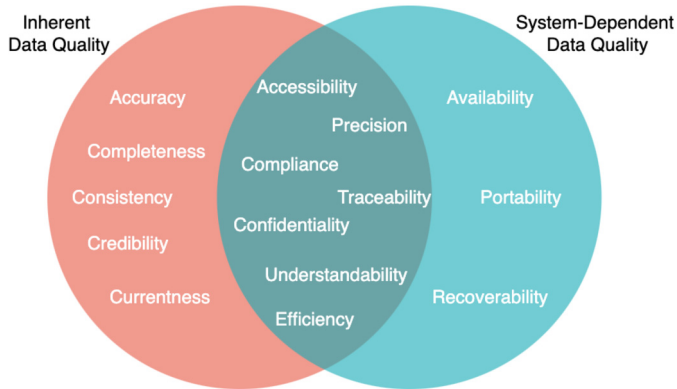


Fig. 2. Fifteen ISO/IEC 25012 characteristics.

For data retained in a structured format within a computer system, the ISO/IEC 25012 : 2008 standard defines a general data quality model that can be used to establish data quality requirements, define data quality measures and plan or perform data quality evaluations (ISO/IEC, 2008). According to the standard, the quality of a data product can be interpreted as the extent to which the data fulfills the requirements of the fifteen characteristics presented in the model (Fig. 2). These characteristics are divided into two main categories: inherent and system-dependent data quality. The first concerns the data values, the relationships between these values and the metadata, while the second is influenced by the capabilities of the computer system components (hardware and software). The importance and priority of these data quality characteristics can vary depending on the use case.

3.2. Standard ISO 8000

While ISO/IEC 25012 defines the characteristics of data quality, the ISO 8000 series focuses on exchange of quality data and information. A key principle of ISO 8000 is that data should be portable, i.e. it must not be tied to the software application from which it is exported. This is achieved through the use of standardized formats such as XML, where the meaning of the data is preserved through structured tags and code values. Decoupling the data from the software applications is not only important for the long-term preservation of the data, but also for minimizing problems such as duplication, inconsistency and inaccuracy. The ISO 8000 series is designed to help organizations define what constitutes quality data, request such data using standard conventions and verify the quality of received data against those same standards. It provides guidelines for improving data quality and includes frameworks for data governance, data quality management (processes, roles, responsibilities, maturity assessment), data quality assessment (profiling, data rules), quality of master data (exchange of characteristic data and identifiers) and quality of industrial data (ISO, 2022).

3.3. 5-Star Linked Open Data System

In 2010, Tim Berners-Lee introduced a 5-star rating system for Linked Data to encourage government data owners in particular to enhance the quality and connectivity of their data (Berners-Lee, 2006). The system relies on standard web technologies such as HTTP, RDF and URIs, which facilitate the sharing of information in a machine-readable format. This allows data from different sources to be linked together and expands the possibilities for semantic queries.

For the implementation of open data, Berners-Lee proposed a five-star scheme where each level must be met to fulfill the requirements of the next level:

1. The data is available on the web with an open license.
2. The data is provided in a machine-readable format (e.g. an Excel file instead of a scanned table).
3. The data is accessible in a non-proprietary format (e.g. CSV instead of Excel).
4. The data follows open standards of the W3C organization (RDF and SPARQL) and uses Uniform Resource Identifiers (URIs), which enable identification.
5. The data is linked to other available data to provide context.

3.4. FAIR Guiding Principles

The guiding principles were designed primarily as guidelines, rather than a formal standard, to support best practice in data management. These principles ensure that digital resources are findable (F), accessible (A), interoperable (I) and reusable (R). They outline the features, attributes and behaviours that enhance data management making it more effective for both machines and humans (Wilkinson *et al.*, 2016). The principles are listed in Table 1, where the term “(meta)data” applies to both data and metadata levels.

3.5. DCAT and DCAT-AP

The Data Catalog Vocabulary (DCAT) for the RDF data model, developed by the W3C (World Wide Web Consortium), provides guidelines and specifications for describing datasets in data catalogs to improve the visibility and interoperability of datasets on the web. It is a globally applicable standard, independent of any political or regulatory framework. DCAT is supplemented with application profiles tailored to the specific needs of various data portals. The DCAT-AP Application Profile for European Data Portals is used as a specification for describing linked public data in Europe. This profile, specially designed for European public administrations, is based on the DCAT standard and offers additional guidelines and constraints to ensure compliance with European requirements. It is also aligned with the INSPIRE directive, which defines the legal framework for establishing and operating the European Spatial Information Infrastructure, ensuring consistency in the description of geospatial datasets. DCAT-AP includes specific guidelines for capturing legal and licensing information related to datasets, aligning with European legal requirements for open data and enhancing interoperability between EU member states.

The RDF model uses subject-predicate-object triples to represent information and DCAT defines a set of classes and properties to describe data catalogs. The seven main classes of the vocabulary are:

Table 1
The FAIR Guiding principles.

To be Findable:	
F1	(meta)data are assigned a globally unique and persistent identifier.
F2	data are described with rich metadata (defined by R1 below).
F3	metadata clearly and explicitly include the identifier of the data it describes.
F4	(meta)data are registered or indexed in a searchable resource.
To be Accessible:	
A1	(meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1	the protocol is open, free and universally implementable.
A1.2	the protocol allows for an authentication and authorization procedure, where necessary.
A2	metadata are accessible, even when the data are no longer available.
To be Interoperable:	
I1	(meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
I2	(meta)data use vocabularies that follow FAIR principles.
I3	(meta)data include qualified references to other (meta)data.
To be Reusable:	
R1	(meta)data are richly described with a plurality of accurate and relevant attributes.
R1.1	(meta)data are released with a clear and accessible data usage license.
R1.2	(meta)data are associated with detailed provenance.
R1.3	(meta)data meet domain-relevant community standards.

- `dcat:Catalog` contains metadata about databases and data services;
- `dcat:Dataset` represents a collection of data that can be available in different formats (numbers, text, pixels, images, sound and other multimedia content);
- `dcat:Distribution` represents the accessible format of the database (download file);
- `dcat:DataService` represents a collection of operations accessible through the API interface;
- `dcat:Resource` is the parent class of the `dcat:Dataset`, `dcat:DataService` and `dcat:Catalog` classes and represents an extension point for the definition of any type of catalog;
- `dcat:DatasetSeries` is a dataset for representing separate collections with common characteristics;
- `dcat:CatalogRecord` is a catalog record that refers to data registration information.

Compliance with the DCAT-AP application profile requires adherence to the guidelines set out in the Fragkou (2023) documentation, but despite the limitations, the profile leaves users with a lot of freedom to adapt it to their needs (Fig. 3).

4. Overview of the Existing Open Data Assessment Approaches

4.1. Methodologies

The 16 selected papers, listed in Table 2, propose different methodologies, frameworks and/or metrics. While most of these methods focus exclusively on data assessment, some go beyond this by introducing complementary phases that contribute to data preparation

```
@prefix dcat: <http://www.w3.org/ns/dcat#>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix example-ds: <https://podatki.gov.si/dataset/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

example-ds:IntangibleCulturalHeritageRegister a dcat:Dataset;
dct:title "Register of Intangible Cultural Heritage"@en;
dct:description "A central database on
                intangible heritage in Slovenia."@en;
dct:publisher <https://podatki.gov.si/publisher
              /ministrstvo_za_kulturo>.
```

Fig. 3. An example of a dataset description in Turtle format.

Table 2
 Overview of methodologies for assessing data, metadata and schema quality, as identified in the reviewed literature.

Reference	Method	Scope	Data quality dimensions	Use case examples
Wentzel <i>et al.</i> (2023)	assess	metadata	5	EU portal data.europa.eu
Lämmel <i>et al.</i> (2020)	harvest, assess	metadata	3	OGD (Open government data) in Germany
Šlibar and Mu (2022)	assess	metadata	2	OGD in Canada, USA, New Zealand
Nogueras-Iso <i>et al.</i> (2021)	assess	metadata	6	OGD in Spain
Hafidz <i>et al.</i> (2023)	assess	metadata	3	OGD of Indonesian Local Governments
Krasikov and Legner (2023)	screen, assess, prepare	metadata, schema, data	7	/
Fadlallah <i>et al.</i> (2023)	prepare, assess	data	8	Radiation dataset from Lebanese Atomic Energy Commission
Yan <i>et al.</i> (2023)	assess	data, monitoring, processed data	15	/
Wang <i>et al.</i> (2020)	assess	data	6	OGD in China and USA
Álvarez Sánchez <i>et al.</i> (2019)	assess	data	8	health data from Northern Ireland
Kusnirakova <i>et al.</i> (2022)	assess	data, schema	5	Czech Open Data
Alogaiel and Alrwais (2023)	assess	data	9	OGD in Saudi Arabia
Bouchelouche <i>et al.</i> (2022)	assess	data	1	OGD in USA
Raca <i>et al.</i> (2021)	prepare, assess	data	5	6 OGD portals in Western Balkans
Ferradji and Benchikha (2022)	assess	data	2	Wikidata
Molodtsov and Nikiforova (2024)	assess	portal	9	33 national portals in EU and GCC

and monitoring. For example, Krasikov and Legner (2023) not only presented a methodology for data assessment, but also introduced a method for screening and preparing the assessed data for later use, which adds significant value to the process. In addition, Fadlallah *et al.* (2023) and Raca *et al.* (2021) included a preparation phase to ensure that the data are adequately prepared for the assessment.

The scope of these methods primarily revolves around the assessment of metadata and data content. This focus ensures that both the structure and content of the datasets are considered in the assessment. However, one paper, Molodtsov and Nikiforova (2024), stands out for directing attention to the design of portals. The design of data portals can have a significant impact on accessibility, usability and the overall experience of data users. Ensuring that the portal itself meets high standards of design and accessibility can have a far-reaching impact on how users interact with the data. In addition, Krasikov and Legner (2023) and Kusnirakova *et al.* (2022) have proposed specific metrics for evaluating the schema of datasets, which adds an important layer to the structural integrity of the data.

A recurring theme in many methodologies is the use of three levels of abstraction: categories, dimensions and metrics. According to Debattista *et al.* (2016), a category represents a group of qualitative dimensions where a common type of information serves as an indicator. This categorization enables organizing and simplifying the review of all aspects of data quality, especially when dealing with a large number of dimensions. Grouping dimensions into categories makes the review process more manageable and ensures that no aspect of data quality is overlooked. Metrics, on the other hand, serve as concrete measures of quality. Each metric is usually linked to a specific measurement method that provides a value—either numeric or Boolean—that can be used to assess the quality of an individual indicator. It is important to note that a single dimension can encompass multiple metrics, allowing for a more nuanced and detailed assessment of data quality.

Despite the common goal of providing tangible information about the state of data portals, usually in the form of a numerical or descriptive score, methodologies differ significantly in the way they define and name the dimensions they use for assessment. This lack of standardization poses a challenge for comparing the results of different methods and makes it difficult to integrate the findings from multiple studies. Nonetheless, some works attempt to overcome these challenges by refining existing approaches. For example, Ferradji and Benchikha (2022) proposed an improved version of two formulas introduced in earlier methods.

There were five papers that dealt exclusively with metadata and emphasized its crucial role in ensuring the usability of data. Wentzel *et al.* (2023) assessed the quality of metadata within the DCAT-AP standard. They defined their assessment dimensions based on the four FAIR principles and added a fifth dimension, contextuality, to better capture the impact of metadata. These dimensions were used together with the adapted FAIR and 5-star principles to define metrics. They also developed a scalable metrics pipeline and implemented their methodology in the form of the Piveau Metrics service. In their comparison with other tools (Sem-Quire, Open Data Portal Watch, FAIR Evaluator, FAIR Checker and F-UJI), Piveau Metrics was identified as the only tool that combines data validation with support for FAIR principles, the 5-star model and DCAT-AP, while also offering a user interface, API access, export functionality, notifications and score comparison (Wentzel *et al.*, 2023). Lämmel *et al.* (2020) described their process for collecting metadata and evaluating the collected metadata. Their assessment focused on the completeness of the required information, the availability of the URL and the overall conformance to the schema. They emphasized that findability, an important FAIR principle, can

be significantly improved by the completeness and quality of metadata. While their quality assurance approach is presented in technical detail, it has not been released as a finalized or publicly available tool. Šlibar and Mu (2022) examined the compliance of metadata with the publication guidelines for OGD portals and calculated scores for the completeness and consistency of metadata to assess conformity. Nogueras-Iso *et al.* (2021) evaluated the quality of geographic metadata using six dimensions from the ISO 19157 standard: completeness, logical consistency, temporal accuracy, thematic accuracy, positional accuracy and the quality of free text. They used the Data Quality RDF Vocabulary for the representation of evaluation results. Hafidz *et al.* (2023) assessed the Open Data portal quality by relying on the Open Data Portal Quality (ODPQ) framework (Kubler *et al.*, 2018), focusing on two main quality categories: data openness and transparency. These categories were broken down into three key measurement dimensions: existence, conformance and open data, each divided into relevant sub-dimensions based on the DCAT standards. The existence dimension covers aspects of access, discovery and preservation; conformance includes accessURL, license and file format elements; and open data focuses on open formats and machine readability.

A particularly valuable contribution in the larger scope came from Krasikov and Legner (2023), who proposed a set of metrics based on a comprehensive research review. Their methodology included steps such as use case ideation, identification of relevant open data, high-level metadata assessment, schema-level assessment, content analysis of datasets, semantic documentation and integration of open datasets with internal data. Their three-tiered approach included both traditional data quality dimensions and context-aware assessments to ensure that open data can be used effectively for predefined purposes. Their proposed dimensions also overlap with the ISO characteristics in the completeness and compliance dimensions.

Fadlallah *et al.* (2023) presented the BIGQA model, which was developed for quality assessment of large datasets based on 8 ISO characteristics (accuracy, completeness, consistency, credibility, currentness, compliance, precision and understandability). BIGQA uses parallel processing to efficiently process large data files and the model was demonstrated with custom data quality reports in an application. In the area of multi-source open data, Yan *et al.* (2023) went beyond data quality and proposed evaluation indicators for data monitoring and processing methods to ensure that the value of the data is preserved during processing. They assessed data quality based on five key dimensions: integrity, relevance, availability, timeliness and legitimacy. Wang *et al.* (2020) examined methods for assessing the security, openness, comprehensiveness, sustainability and availability of data and emphasized that high-quality metadata contributes to the discoverability and usability of open government data. Álvarez Sánchez *et al.* (2019) developed the TAQIH tool for assessing the quality of tabular data, which focuses on the dimensions of accuracy, completeness, accessibility, consistency, redundancy, readability, usefulness and trust, the first four of which are also in line with ISO standards. Kusnirakova *et al.* (2022) adapted an existing method and applied it to a specific use case, using five quality categories, accuracy, completeness (of data and schema) and consistency. These categories overlapped with ISO standards, as well as the formal requirements of the Czech government's

open formal standards. Alogaiel and Alrwais (2023) introduced nine dimensions for assessing OGD in Saudi Arabia: completeness, granularity, timeliness, machine-readability, reusability, consistency, accuracy, understandability, usage and redundancy. Each dimension is thoroughly explained, including its sub-dimensions, and accompanied by equations for score calculation. The final dataset score is presented on a 0–100 scale, with weighted values assigned to each dimension. RapidMiner software was used to process algorithms and compute scores for certain dimensions. Although Raca *et al.* (2021) indicate a focus on two dimensions, the term “dimensions” is used to denote specific components of the dataset—namely, the format and the data content itself. In fact, the study assesses five distinct quality dimensions related to the openness of OGD. They evaluate availability, accessibility, discoverability and timeliness, as well as the openness of formats according to the Berners-Lee scale.

On the other hand, Bouchelouche *et al.* (2022) focused on the evaluation of only one characteristic. They proposed a percentage grading scale for metrics to assess the accessibility of OGD portals, examining access options, available formats, licensing and timeliness. Similarly, Ferradji and Benchikha (2022) focused only on an upgrade for time-related metrics, namely currency and volatility, to make them more efficient and suitable for evaluating linked data.

Compared to others, focused on data or metadata, Molodtsov and Nikiforova (2024) proposed a unique approach by emphasizing portal evaluation. This is important, as the design and functionality of the portal can be equally crucial for enabling effective data reuse. The proposed framework comprises 72 sub-dimensions organized into nine key dimensions: multilingualism, navigation, general performance, data understandability, data quality, data findability, public engagement, feedback mechanisms and service quality, and portal sustainability and collaboration. Most sub-dimensions are scored using a binary method, while accessibility is assessed using the Accessibility Checker web tool. 16 sub-dimensions are evaluated on a sample basis with a 70% threshold required to score 1 point. If sorting by both relevance and modification date is supported, the sample includes the first 4 and last 3 records; otherwise, it includes the first 8 and last 6 records. For update accuracy, at least 70% of the records must match the declared update frequency (e.g. monthly). Because of overlaps between dimensions, a priority-based weighting system is applied, assigning importance levels of low, medium and high weights of 1, 2 and 3, respectively, according to its importance in relation to the central concepts of the framework.

Across these methodologies, completeness proves to be the most frequently used dimension, followed by consistency, accuracy, timeliness and findability. Accessibility, compliance and understandability are also frequently included. The five most frequently used dimensions shared across the analysed studies are summarized and defined in Table 3.

4.2. Use Cases

In the selected studies, numerous methods were applied to specific datasets, which are listed in Table 2. In the process, the authors identified several challenges and proposed practical solutions.

Table 3
 Definitions of most frequently used dimensions based on ISO/IEC (2008), Fadlallah *et al.* (2023).

Dimension	Definition
Completeness	Data completeness refers to the degree to which an entity has values for all expected attributes and related entity instances in a specific context of use.
Consistency	Refers to the degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use. It can be either or both among data regarding one entity and across similar data for comparable entities.
Accuracy	Refers to the degree to which data correctly reflects the true value of an intended attribute in a specific context. It has two main aspects: syntactic and semantic accuracy. Syntactic accuracy refers to the syntactical correctness of the values themselves. Semantic accuracy refers to the closeness of the data values to a set of values defined in a domain considered semantically correct.
Timeliness	(also Currentness) Refers to the degree to which data has attributes that are of the right age in a specific context of use.
Accessibility	Refers to the degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.

A common challenge in many studies is insufficient data quality, which can lead to increased costs and loss of time in scientific projects. Wentzel *et al.* (2023) found that despite efforts to improve the FAIRness of the portals on data.europa.eu by integrating their Piveau Metrics tool, progress over the course of a year was minimal. One of the biggest challenges is the need for better engagement from data providers. Problems with the completeness of metadata were also identified as a major challenge. Šlibar and Mu (2022) also pointed out the problems with the American OGD portal, where almost 20% of the required fields were missing from the published datasets, and the New Zealand portal performed poorly in terms of consistency of required and optional fields. A major obstacle to improving metadata quality is the time-consuming assessment of compliance with publication guidelines.

Addressing challenges of working with different metadata models, Hafidz *et al.* (2023) developed a mapping between CKAN and DCAT so that their methodology works across formats in the context of Indonesian local government open data. They used the Open Data Portal Quality (ODPQ) framework (Kubler *et al.*, 2018), which is based on the Analytic Hierarchy Process and integrates multiple quality dimensions and user preferences. This approach combines metadata collection (via Open Data Portal Watch (Neumaier *et al.*, 2016)) with a web-based dashboard and RESTful APIs to create quality rankings. Extending practical quality assessment techniques, Nogueras-Iso *et al.* (2021) has implemented a two-part assessment approach—automated and manual—of a national open data portal. The manual assessment by two metadata experts revealed problems with thematic classification and low-quality dataset titles, while the access URLs and descriptions showed better results. Automated checks, on the other hand, highlighted structural problems, including missing mandatory properties (e.g. dcat:mediaType), incorrect or misused URIs and poor readability of free text fields. In addition to these approaches, Álvarez Sánchez *et al.* (2019) applied their TAQIH tool to two use cases—Northern Ireland General Practitioners Prescription Dataset and a glucose monitoring system dataset. TAQIH detected

issues such as missing values, unnamed columns, redundant variables and outliers. While automated profiling and visualization improved the completeness of the dataset and reduced noise, some steps (e.g. editing combined date and time fields) required manual intervention. Bouchelouche *et al.* (2022), who also focused on evaluation, validated their Assessment Scale of Marks on a selection of datasets from the American OGD portal. Their scoring system, based on the presence of specific accessibility criteria, showed the strengths of the portal in terms of structural openness, but also pointed to the need for improved updating practices to ensure that the data remains current and usable. Finally, Ferradji and Benchikha (2022) used the DBpedia Live Extraction Framework, specifically the Infobox Extractor, to retrieve semi-structured data from Wikipedia using SPARQL queries and APIs, focusing on frequently updated facts. Temporal metadata such as the start time and last modified time were used to analyse the update patterns.

In the field of big data, Fadlallah *et al.* (2023) faced challenges related to the volume, processing speed and variety of data. The architecture of their solution is divided into two main modules: a data preparation module, which is responsible for profiling, storing metadata and creating a quality assessment plan, and a data quality assessment module, which executes these plans in a configurable way. The core of the system includes a design layer (logical quality assessment plan), a mapping layer (validation) and an execution layer. Such a modular structure enables a scalable and automated quality assessment. The authors tested their methodology with both Stack Overflow data on a single machine and a large-scale radiation dataset from the Lebanese Atomic Energy Commission, running the methodology on multiple distributed machines. One of the problems they identified is the generalization of contexts and workflows by leading technology providers in quality assessment, which is often insufficient to meet the unique requirements of big data.

The need for improved metadata standards and policy frameworks was highlighted in several studies. For example, Wang *et al.* (2020) identified numerous problems with US forest data, including unclear data authorization agreements and inadequate data security measures. The authors recommended the development of open source platforms, improved machine-readable data and compliance with international metadata standards. In this way, data quality could be significantly improved, according to the authors.

Furthermore, Kusnirakova *et al.* (2022) applied an evaluation framework that assigns points from 0 to 100 for five quality dimensions: file format, schema accuracy, schema completeness, data type consistency and data completeness. When applied to datasets from six Czech municipalities, the assessment revealed large deficiencies in schema accuracy. For example, while data completeness was generally high, schema accuracy scored poorly due to inconsistent feature naming that often deviated from national standards.

Raca *et al.* (2021) describe their implementation process as a web service that continuously collected and stored metadata from six national OGD portals, which formed the basis for the quality assessment. Due to the structural differences between the portals, a separate data preparation phase was carried out using custom SQL queries to remove duplicates, incomplete records and inconsistent values. This was followed by a validation step to standardize elements such as date formats, license descriptions and character encoding. Their study showed that differences in metadata practices and file formats between portals significantly influenced the comparative quality assessment.

Finally, Alogaiel and Alrwais (2023) examined the effectiveness of existing methods for Saudi Arabia's OGD portal. They found that the current framework is ineffective because there are no clear indicators of what constitutes high-quality data. The authors suggested that continuous monitoring and evaluation, improved search capabilities of the portal, the inclusion of visual representations (such as maps and charts) and better timeliness and comprehensiveness of data could improve the portal's performance.

5. Analysis

In our review, we found many researchers relying on previous works. We therefore include the five most frequently cited papers related to the ones under our review, since they can serve as a base for many more specifically oriented methodologies.

In their 2016 article, Vetrò *et al.* (2016) presented an indicator framework designed to assess the quality of open government data by examining various dimensions of data quality at the most granular level. They argued that a comprehensive theoretical framework was lacking as most evaluations focused on open data platforms rather than individual datasets. The authors emphasized that poor data quality increases the cost and complexity of accessing and interpreting information, often leading to decentralized and uncoordinated efforts by data re-users to independently verify and improve the quality of datasets. Their aim was to improve the meaningfulness and reusability of public sector information by ensuring that datasets are easily accessible, queryable, processable and linkable to other data without restrictions. The assessment framework was based on a thorough analysis of existing methodologies. They identified the most appropriate data quality model as a theoretical basis, defined the methodology for the selection of quality characteristics and metrics, and determined the characteristics and metrics to be used. The authors selected a subset of quality characteristics from the Square-Aligned Portal Data Quality Model (Moraga *et al.*, 2009) and defined 14 metrics for seven key characteristics: traceability, timeliness, expiration, completeness, compliance, understandability and accuracy—each focused on the dataset and cell levels. The framework adhered to several key principles. First, the metrics had to be normalized and at least interval scaled. It was also important that the metrics were interpretable, i.e. they had to be clearly defined so that users could easily understand them. The framework also needed to support aggregation, i.e. the quantification of data quality at multiple levels (individual attributes, tuples, datasets and databases) while maintaining consistency across all levels. Finally, the metrics had to be based on determinable input parameters and be automatable for practical implementation. The strength of this methodology lies in its clarity and concreteness, as it defines variables and provides formulas for measurement. Its universality is also a major advantage, as it is not tailored to a specific use case and thus enables broad applicability across different sectors and datasets.

Neumaier *et al.* (2016) are known for its comprehensive evaluation of over 260 open data portals, which together contain more than 110,000 datasets and over 2 million resource URLs. This extensive analysis required a methodology adaptable to a variety of

formats to address the common challenge of unification among these portals. However, the study focused primarily on the metadata of these portals, emphasizing that inadequate descriptions or classifications of datasets directly impact the usability and searchability of resources. To tackle this large-scale assessment, the authors proposed a comprehensive set of objective quality metrics, based on the W3C metadata schema DCAT, designed to monitor the quality of open data portals in a systematic and automated way. To achieve this, they have introduced a versatile abstraction of web-based data portals that enables the integration of a significant number of existing portals in a scalable and extensible manner. Building on previous approaches to metadata homogenization, they have mapped the metadata of major open data publishing systems—such as CKAN, Socrata and OpenDataSoft—to the DCAT schema. In addition to proposing this method, they also described their Open Data Portal Watch Framework (Neumaier *et al.*, 2016), which, at the time of writing, is not available.

Debattista *et al.* (2016) presented Luzzu, a comprehensive framework for assessing the quality of linked data that addresses the growing challenge of evaluating the suitability of datasets for different applications. With Luzzu, they have introduced a scalable, extensible and interoperable methodology that covers the entire data quality lifecycle, from metrics identification to dataset ranking. The key features of Luzzu are extensibility through the Luzzu Quality Metric Language (LQML) or Java, which enables the inclusion of domain-specific quality criteria, an ontology-driven backend that leverages the Dataset Quality Ontology (daQ) and the Quality Problem Report Ontology (QPRO) to capture and share assessment results, and robust scalability demonstrated by the efficient processing of large datasets with 22 metrics across nine dimensions. Despite its strengths, Luzzu also has its limitations, such as the lack of automatic data repair and the reliance on external resources for some metrics, which could lead to performance issues. In addition, not all metrics could be distributed across processing clusters, which affected scalability. Nevertheless, Luzzu has made an important contribution to Linked Data quality assessment by filling the existing gaps in the literature. Due to its extensibility, ontology-based design and scalable processing, Luzzu is particularly appropriate for domains with rigorous data quality requirements, such as healthcare, where the reliability of datasets directly impacts the integrity of downstream analytical and operational processes.

Zaveri *et al.* (2013) conducted a comprehensive review of data quality dimensions, which resulted in the paper being widely cited, not because it proposed a new methodology, but because of its comprehensive analysis of the state of data quality assessment in the literature at the time. This systematic review was conducted against the backdrop of the rapid growth of Linked Open Data (LOD) as a result of advances in semantic web technologies. The paper organized the different approaches into a unified classification scheme, addressing terminological inconsistencies and providing a comprehensive list of 26 definitions of data quality dimensions, with each dimension classified as either subjective or objective based on the nature of its measurement. In addition, 21 methodologies were qualitatively analysed to improve the understanding of existing work. The authors compared these approaches to identify the most commonly used dimensions, the types of data they support and whether the authors developed tools for implementation or designed their methods for general or specific use cases.

Janssen *et al.* (2012) explored the benefits and barriers associated with open data and improve understanding of the promises and challenges associated with open government initiatives. Through expert interviews, the authors collected and categorized a comprehensive list of benefits in terms of political, social, economic, operational and technical aspects, highlighting both the potential and the sometimes unrealistic expectations of the impact of open data. Their research found that there is a significant gap between the expected benefits of open data and the barriers to its adoption. The barriers identified were also grouped into the following categories: institutional barriers, task complexity, use and participation, legislation, information quality and technical issues. These barriers are often interrelated, adding to the overall complexity. Based on their research into the advantages and disadvantages of open data, the authors have formulated five myths and examined them in more detail. One common misconception is that publishing data will automatically yield benefits. Another myth suggests that all information should be published without restriction, overlooking the importance of privacy. There is also the belief that it is merely a matter of publishing public data without any processing. The assumption that every constituent can make use of open data fails to recognize the varying levels of data literacy and data access among different groups. Finally, the idea that open data will result in open government suggests that the provision of data per se leads to greater government transparency and accountability. The article serves as an important resource for developing innovative strategies for dealing with open data by clearly stating the challenges that need to be addressed. It underscores the essential truth that open data has no intrinsic value; its true potential is only realized when it is actively used.

5.1. Additional Relevant Literature

In addition to the most frequently cited articles mentioned above, there is a wider range of literature that provides valuable contributions to the understanding of quality, maturity and governance of open data. Although they did not fulfill the inclusion criteria for the main literature review, they are worth highlighting for their practical and conceptual value.

At the European level, the Open Data Maturity (ODM) assessment provides a structured approach for evaluating the progress of open data initiatives in the EU member states (Page *et al.*, 2024). The assessment covers four core dimensions—policy, portal, quality and impact—and is carried out annually using a detailed questionnaire that is reviewed by experts. Since 2018, the methodology has been consistent in its structure but is regularly updated to reflect changes in policy, practice and data governance and to reinforce its status as an important benchmarking tool in the EU open data landscape.

At a global level, the Global Open Data Index (GODI), launched by the Open Knowledge Foundation in 2013, provided one of the first comparative indicators of the openness of government data in different countries (Open Knowledge, 2012). Although it was discontinued after the 2016/17 cycle, GODI provided a transparent and standardized assessment methodology. It assessed certain categories of data—such as land ownership, government budgets and environmental quality—each based on certain characteristic criteria. Datasets that fully met all criteria were scored at 100% and underwent an additional

review step before being officially classified as open data, while others were classified as public or access-controlled.

In addition to these large-scale assessments, two comprehensive books provide broadly applicable suggestions for managing data quality:

- *The DAMA-DMBOK2 (Data Management Body of Knowledge)* presents a detailed reference for data governance and quality, including principles, roles, dimensions and best practices (DAMA International, 2017). Chapter 13 in particular focuses on data quality, outlining key concepts and providing practical guidance on implementation and improvement.
- *Executing Data Quality Projects* by McGilvray (2021) describes a structured ten-step method for improving and maintaining data quality in a wide range of organizational settings. The book provides templates, examples and best practices for each step and guides users in selecting and customizing the steps to their specific needs.

6. Discussion

We reflect on the key findings and insights related to standards, challenges and methodological frameworks in assessing the quality of open data, structured according to the research objectives of this study.

Addressing RQ1: In assessing existing standards, we note that several prominent benchmarks serve as established foundations for assessing open data quality. The ISO/IEC 25012 standard is notable for its robust framework of 15 data quality characteristics and has been used in studies by Fadlallah *et al.* (2023), Álvarez Sánchez *et al.* (2019) and Kusnirakova *et al.* (2022). Although Krasikov and Legner (2023) does not explicitly refer to it, two of the quality dimensions he proposes coincide with the ISO characteristics, which underline the widespread use of this ISO standard. In addition, the 5-star model for linked open data is widely used to assess data format and accessibility. Researchers such as Raca *et al.* (2021), Wang *et al.* (2020) and Alogaiel and Alrwais (2023) emphasize its adaptability for open data applications, and Wentzel *et al.* (2023) extend this approach by incorporating FAIR principles for data findability, accessibility, interoperability and reusability. DCAT, a structured RDF vocabulary for metadata, also appears frequently, specifically in the studies by Wentzel *et al.* (2023) and Hafidz *et al.* (2023), where it facilitates the organization and retrieval of metadata.

Addressing RQ2: Common challenges in the quality of open data arise primarily from a lack of standardization. Although standards such as ISO/IEC 25012 provide structured guidelines, many data publishers are unwilling or unable to strictly adhere to them due to resource constraints, low enforcement and the time-consuming process of data adaptation. This inconsistency is a significant barrier to creating a consistent framework for quality assessment and often results in contextual adjustments rather than a universally applicable methodology. The frequent lack of clear data definitions, inconsistent data formats and missing fields further complicate the reusability of the data and limit the possibilities for comprehensive analysis and informed decision-making. Technical and logistical

issues such as incomplete data access, registration requirements for downloading data and download size restrictions in datasets further compound the barriers. This limited accessibility can limit the potential for interoperability, a core goal of open data, especially when technical factors such as legacy systems, data fragmentation and lack of supporting infrastructure come into play. Although the basic idea of open data promotes reusability and integration, it therefore remains a challenge to achieve these goals.

Data quality gaps have recently been addressed also in a paper by Kiran *et al.* (2024). The authors investigate the relationship between open data quality and smart mobility systems, emphasizing that while open data is vital for transparency and innovation in urban transport, its quality issues (accuracy, timeliness, interoperability, etc.) pose major challenges to effectiveness. They scrutinized six fundamental open data quality parameters for smart mobility and performed a comparative analysis of eight open data maturity models, revealing gaps and assessing the coverage of identified parameters. They empirically examined these identified gaps through an analysis of 54 real-world datasets. From the comparative analysis, the authors find that existing open data maturity frameworks—such as those from the European Data Portal, OECD, and other governmental or institutional sources—only partially address the data quality parameters essential for smart mobility. While most frameworks include general indicators like completeness, accuracy, and timeliness, they often neglect parameters specific to mobility contexts, such as interoperability, usability, and governance structures ensuring continuous data updates and cross-agency coordination. The analysis thus exposes a misalignment between theoretical models and the practical demands of smart mobility ecosystems, where data is dynamic, sensor-driven, and highly interdependent.

Although the papers we examined were recently published, they include hardly any direct references to the newest European Open Data Directive (EU, 2019). This is due to the fact that several methods were developed outside the European Union and originate from regions such as the United States, Lebanon, the Middle East, Indonesia and China. Nevertheless, methodologies that include FAIRness principles are indirectly linked to the Directive, as these four principles are part of the requirements of the said EU Directive. Although some studies do not explicitly refer to the Directive, they are nevertheless well aligned with the EU legal framework. The work of Wentzel *et al.* (2023), for example, comprehensively assesses the European Data Portal and draws on the EU Data Catalog Vocabulary (DCAT) application profile, which is also an important focus in the methods of Lämmel *et al.* (2020) and Nogueras-Iso *et al.* (2021). When assessing the maturity of open data, the leading EU portals perform consistently well, as highlighted in the EU Open Data Maturity Reports (Page *et al.*, 2024). Molodtsov and Nikiforova (2024) argue that this high performance indicates some alignment between EU standards and established open data benchmarks. However, their study introduces unique assessment dimensions that provide a new perspective by comparing the EU portals with those of the Gulf Cooperation Council (GCC) countries. In particular, the portals in Saudi Arabia, Qatar and Bahrain excel and sometimes even lead in certain sub-dimensions compared to their EU counterparts.

6.1. *Threats to Validity*

Certain limitations should be acknowledged. First, the search was limited to English-language publications and to the period 2019–2024, which may have excluded relevant studies published in other languages or outside this timeframe. Second, although the inclusion and exclusion criteria were systematically applied, some degree of subjectivity in judgment may persist when interpreting the focus of borderline cases. Third, the search relied on a fixed set of keywords without iterative refinement, which could have constrained the breadth of retrieved studies. Finally, as our synthesis draws on published literature, potential publication bias toward positive findings cannot be fully excluded. These factors were mitigated through processes inspired by PRISMA and Kitchenham’s guidelines.

6.2. *Lessons Learned*

When evaluating suitable methods for adapting open data quality standards to the specific requirements in Slovenia, the methods of Kusnirakova *et al.* (2022), Alogaiel and Alrwais (2023) and Krasikov and Legner (2023) prove to be particularly promising. They each offer a targeted approach that could be efficiently adapted to the needs of the Slovenian Open Data Portal. The first is relevant as it was developed for the data of an EU country that is in a similar context to Slovenia, so it lends itself well to local adaptation with minimal changes. The second also provides an appropriately structured approach to assessing data quality. However, certain score weights would need to be adjusted to exclude dimensions that are not directly applicable to the Slovenian data environment.

Addressing RQ3: One of the most comprehensive is the proposal by Krasikov and Legner (2023). It integrates a large collection of findings from previous studies and proposes detailed measurement procedures that could significantly improve the assessment framework for Slovenia. Currently, the Slovenian Open Data Portal uses a 5-star scale that focuses exclusively on the openness of datasets, which results in most datasets receiving a uniform three-star rating, which limits the value of such a rating for re-users. For a more robust assessment of metadata quality, the inclusion of methods such as those proposed by Lämmel *et al.* (2020) and Nogueras-Iso *et al.* (2021) could provide a more refined rating system. These methods go beyond basic openness and provide a more nuanced analysis of metadata that could better meet the information needs of data users. Given the below-average DCAT-AP compliance score assessed in the recent European Open Data Maturity Report (Page *et al.*, 2023), it would also be beneficial for publishers to include a DCAT-AP compliance assessment section in the development of a comprehensive methodology. Building on this review and in response to the growing volume of open data, we developed a qualitative assessment methodology specifically designed to evaluate datasets published on the Slovenian Open Data Portal (OPSI). The initial version of the methodology is published in Pesek *et al.* (2025). In addition, we are continuously reiterating and refining the accompanying assessment tool to support the partial automation of the evaluation process. While the tool is not yet publicly available, it is currently being tested and improved based on internal use and feedback.

7. Conclusion

This study provides an overview of the newest methods for assessing the quality of open data and also focuses on the applicability of these methods to the Slovenian Open Data Portal. Our research highlights the importance of established standards that have proven their usefulness in developing effective quality assessment frameworks. The key findings show that standards such as ISO/IEC 25012, the 5-star model for linked open data and DCAT serve as essential reference points for assessing the quality of open data.

In addition, we highlighted several common challenges in the open data landscape, including inconsistent application of established standards by data providers, inconsistent data formats and various barriers to data accessibility. These hinder the potential for reusability and interoperability, which are fundamental to the field of open data.

In identifying suitable methods for adapting quality standards for open data to the specific needs of Slovenia, we found that several frameworks offer promising approaches. Each of these methods can be tailored with minimal customization, allowing for a more effective quality assessment tailored to the Slovenian environment. Furthermore, the inclusion of additional methods could significantly improve the assessment of metadata quality and go beyond the current use of a simple 5-star rating.

In summary, our findings emphasize the need to use verified standards and ensure compliance with European directives to improve the assessment of government data capabilities in the EU context. When developing a new methodology or tool to assess data quality, these results should be taken into account to promote better data practices and ultimately increase the value of open data. Our future work will therefore continue with the development of a new methodology that takes into account all the findings to date and seeks to further improve the quality and usability of open data initiatives.

Acknowledgements

The authors acknowledge the financial support from the state budget by the Slovenian Research and Innovation Agency (project No. V2-2388).

References

- Agenzia per L'Italia Digitale (2022). *I dati aperti della pubblica amministrazione*. <https://www.dati.gov.it>.
- Alogaiel, N.F., Alrwais, O.A. (2023). An assessment of the quality of Open Government Data in Saudi Arabia. *IEEE Access*, 11, 61560–61599. <https://doi.org/10.1109/access.2023.3285611>.
- Álvarez Sánchez, R., Beristain Iraola, A., Epelde Unanue, G., Carlin, P. (2019). TAQIH, a tool for tabular data quality assessment and improvement in the context of health data. *Computer Methods and Programs in Biomedicine*, 181, 104824. <https://doi.org/10.1016/j.cmpb.2018.12.029>.
- Berners-Lee, T. (2006). *Linked Data – Design Issues*. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, T. (2012). *5-star Open Data*. <https://5stardata.info/en/>.
- Bouchelouche, K., Ghomari, A.R., Zemmouchi-Ghomari, L. (2022). Enhanced analysis of Open Government Data: proposed metrics for improving data quality assessment. In: *2022 5th International Symposium on Informatics and Its Applications (ISIA)*. IEEE, pp. 1–6. <https://doi.org/10.1109/isia55826.2022.9993482>.

- Commission, E. (2020). *A European strategy for data*. Technical report.
- Conorzio per il Sistema Informativo Piemonte (2010). *Il portale degli Open Data della Regione Piemonte*. <https://www.dati.piemonte.it/#/home>.
- DAMA International (2017). *DAMA-DMBOK: Data Management Body of Knowledge*, 2nd edition. Technics Publications, Basking Ridge, NJ, USA.
- Debattista, J., Auer, S., Lange, C. (2016). Luzzu—a methodology and framework for linked data quality assessment. *Journal of Data and Information Quality*, 8(1), 1–32. <https://doi.org/10.1145/2992786>.
- Ehrlinger, L., Wöß, W. (2022). A survey of data quality measurement and monitoring tools. *Frontiers in Big Data*, 5. <https://doi.org/10.3389/fdata.2022.850611>.
- EU (2019). Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). OJ L 172, 26.6.2019, pp. 56–83. <http://data.europa.eu/eli/dir/2019/1024/oj>.
- European Union (2021). *The official portal for European data*. <https://data.europa.eu>.
- Fadlallah, H., Kilany, R., Dhayne, H., El Haddad, R., Haque, R., Taher, Y., Jaber, A. (2023). BIGQA: declarative Big Data Quality Assessment. *Journal of Data and Information Quality*, 15(3), 1–30. <https://doi.org/10.1145/3603706>.
- Ferradji, M.A., Benchikha, F. (2022). Enhanced metrics for temporal dimensions toward assessing Linked Data: a case study of Wikidata. *Journal of King Saud University – Computer and Information Sciences*, 34(8), 4983–4992. <https://doi.org/10.1016/j.jksuci.2021.05.010>.
- Fragkou, P. (2023). DCAT-AP 3.0. <https://semiceu.github.io/DCAT-AP/releases/3.0.0/>.
- Hafidz, I., Adzanni, G.A., Aini Rakhmawati, N. (2023). Open Data Portal Quality (ODPQ) framework based metric for assessing the quality of open data portals in Indonesian Local Governments. In: *2023 International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*. IEEE, pp. 127–132. <https://doi.org/10.1109/icsgteis60500.2023.10424389>.
- Hesteren, D., Weyzen, R., Knippenberg, L. (2022). *Open data best practices in Europe – Estonia, Slovenia and Ukraine*. Publications Office of the European Union. <https://doi.org/10.2830/277405>.
- Huyer, E. (2020). *The Economic Impact of Open Data – Opportunities for Value Creation in Europe*. Publications Office of the European Union. <https://doi.org/doi/10.2830/63132>.
- Interministerial Digital Directorate (2011). *Plateforme ouverte des données publiques françaises*. <https://www.data.gouv.fr>.
- ISO (2022). Data quality. ISO 8000, International Organization for Standardization, Geneva, Switzerland.
- ISO/IEC (2008). Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Data quality model. ISO/IEC 25012, International Organization for Standardization, Geneva, Switzerland.
- ISO/IEC (2014). Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE). ISO/IEC 25000, International Organization for Standardization, Geneva, Switzerland.
- Janssen, M., Charalabidis, Y., Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <https://doi.org/10.1080/10580530.2012.716740>.
- Kiran, S., Donnellan, B., Helfert, M. (2024). Addressing data quality gaps in open data maturity models: a comparative study and real-world dataset analysis. In: *ECIS 2024 Proceedings*. https://aisel.aisnet.org/ecis2024/track10_dmds_ecosystems/track10_dmds_ecosystems/11.
- Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*, 33. Keele, UK, Keele University.
- Krasikov, P., Legner, C. (2023). A method to screen, assess, and prepare open data for use. *Journal of Data and Information Quality*, 15(4), 1–25. <https://doi.org/10.1145/3603708>.
- Kubler, S., Robert, J., Neumaier, S., Umbrich, J., Le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1), 13–29. <https://doi.org/10.1016/j.giq.2017.11.003>.
- Kusnirakova, D., Ge, M., Wallezky, L., Buhnova, B. (2022). Interoperability-oriented quality assessment for Czech Open Data. In: *Proceedings of the 11th International Conference on Data Science, Technology and Applications*. SCITEPRESS – Science and Technology Publications, pp. 446–453. <https://doi.org/10.5220/0011291900003269>.
- Lämmel, P., Dittwald, B., Bruns, L., Tcholtchev, N., Glikman, Y., Cuno, S., Flügge, M., Schieferdecker, I. (2020). Metadata harvesting and quality assurance within open urban platforms. *Journal of Data and Information Quality*, 12(4), 1–20. <https://doi.org/10.1145/3409795>.

- Manyika, J., Chui, M., Farrell, D., Van Kuiken, S., Groves, P., Doshi, E.A. (2013). Open Data: Unlocking Innovation and Performance with Liquid Information. https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf.
- McGilvray, D. (2021). *Executing Data Quality Projects*, 2nd ed. Academic Press, San Diego, CA.
- Ministrstvo za javno upravo (2016). *Odpri podatki Slovenije*. <https://podatki.gov.si>.
- Molodtsov, F., Nikiforova, A. (2024). An integrated usability framework for evaluating Open Government Data Portals: comparative analysis of EU and GCC countries. In: *Proceedings of the 25th Annual International Conference on Digital Government Research*. ACM, pp. 899–908. <https://doi.org/10.1145/3657054.3657159>.
- Moraga, C., Moraga, M., Calero, C., Caro, A. (2009). SQuaRE-aligned Data Quality Model for Web Portals. In: *2009 Ninth International Conference on Quality Software*, Vol. 31. IEEE, pp. 117–122. <https://doi.org/10.1109/qsic.2009.23>.
- Neumaier, S., Umbrich, J., Polleres, A. (2016). Automated quality assessment of metadata across Open Data Portals. *Journal of Data and Information Quality*, 8(1), 1–29. <https://doi.org/10.1145/2964909>.
- Nogueras-Iso, J., Lacasta, J., Urena-Camara, M.A., Ariza-Lopez, F.J. (2021). Quality of metadata in Open Data Portals. *IEEE Access*, 9, 60364–60382. <https://doi.org/10.1109/access.2021.3073455>.
- Open Knowledge (2012). *Global Open Data Index*. <http://index.okfn.org/>.
- Open Knowledge (2015). *The Open Definition – Version 2.1*. <https://opendefinition.org/od/2.1/en/>.
- Page, M., Behrooz, A., Moro, M. (2024). *Open Data Maturity Report 2024*. Publications Office of the European Union. <https://doi.org/doi/10.2830/8656811>.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>.
- Page, M., Hajduk, E., Lincklaen Arriëns, E.N., Cecconi, G., Brinkhuis, S. (2023). *Open Data Maturity Report 2023*. Publications Office of the European Union. <https://doi.org/doi/10.2830/384422>.
- Pesek, M., Juvan, J., Žnideršič, K., Marolt, M. (2025). Metodologija za kvalitativno vrednotenje odprtih podatkov. In: *International Conference on Organizational Science Development: Human Being, Artificial Intelligence and Organization, Conference Proceedings*, Vol. 44. Univerzitetna založba Univerze v Mariboru, 729–742. <https://press.um.si/index.php/ump/catalog/book/962/chapter/320>.
- Raca, V., Velinov, G., Cico, B., Kon-Popovska, M. (2021). Measuring the government openness using an assessment tool: case study of six western Balkan countries. In: *2021 10th Mediterranean Conference on Embedded Computing (MECO)*, Vol. 27. IEEE, pp. 1–5. <https://doi.org/10.1109/meco52532.2021.9460163>.
- Šlibar, B., Mu, E. (2022). OGD metadata country portal publishing guidelines compliance: a multi-case study search for completeness and consistency. *Government Information Quarterly*, 39(4), 101756. <https://doi.org/10.1016/j.giq.2022.101756>.
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C.O., Iemma, R., Morando, F. (2016). Open data quality measurement framework: definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325–337. <https://doi.org/10.1016/j.giq.2016.02.001>.
- Ville de Paris (2023). *Paris Data*. <https://opendata.paris.fr>.
- Wang, B., Wen, J., Zheng, J. (2020). Research on assessment and comparison of the forestry Open Government Data Quality between China and the United States. In: *Data Science, ICDS 2019, Communications in Computer and Information Science*, Vol. 1179. Springer Singapore, pp. 370–385. 9789811528101. https://doi.org/10.1007/978-981-15-2810-1_36.
- Wentzel, B., Kirstein, F., Jastrow, T., Sturm, R., Peters, M., Schimmler, S. (2023). *An Extensive Methodology and Framework for Quality Assessment of DCAT-AP Datasets*. Springer Nature Switzerland, pp. 262–278. 9783031411380. https://doi.org/10.1007/978-3-031-41138-0_17.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>.

- Yan, T., You, Z., Zhang, Y., Hua, R. (2023). Multi-source Open Data Quality Evaluation Model in the Web 3.0 Era. In: *2023 6th International Conference on Data Science and Information Technology (DSIT)*, Vol. 264. IEEE, pp. 203–207. <https://doi.org/10.1109/dsit60026.2023.00038>.
- Zaveri, A., Rula, A., Maurino, A., c, R., Lehmann, J., Auer, S. (2013). Quality assessment methodologies for Linked Open Data. *Semantic Web Journal*.
- Zuiderwijk, A., Janssen, M., Poulis, K., van de Kaa, G. (2015). Open data for competitive advantage: insights from open data use by companies. In: *Proceedings of the 16th Annual International Conference on Digital Government Research*. Association for Computing Machinery. <https://doi.org/10.1145/2757401.2757411>.

K. Žnideršič has been a member of the Laboratory for Computer Graphics and Multimedia at Faculty of Computer and Information Science at the University of Ljubljana since 2023. Her research interests include modern approaches in music pedagogy and qualitative evaluation of open data collections.

M. Marolt is a full professor at the Faculty of Computer and Information Science, where he is head of Laboratory for Computer Graphics and Multimedia. His research interests are in the areas of music/audio information retrieval, computer graphics and visualization. He focuses on problems such as music transcription, audio segmentation and classification, and organization, search and visualization of music collections.

M. Pesek is an associate professor and researcher at the Faculty of Computer Science and Informatics at the University of Ljubljana, where he graduated in 2012 and received his PhD in 2018. He has been a member of the Laboratory for Computer Graphics and Multimedia since 2009. His research interests are information retrieval, e-learning, and the design of open data applications aimed at generating meaningful social impact.