**Evaluating a Conversational Agent for Open Government Data Quality Assessment**

Chokki, Abiola Paterne; Clarinval, Antoine; Simonofski, Anthony; Vanderose, Benoit

*Publication date:*
2023

*Document Version*
Early version, also known as pre-print

# Evaluating a Conversational Agent for Open Government Data Quality Assessment

*Completed Research*

**Abiola Paterne Chokki**
Namur Digital Institute, University of Namur
abiola-paterne.chokki@unamur.be

**Antoine Clarinval**
Namur Digital Institute, University of Namur
antoine.clarinval@unamur.be

**Anthony Simonofski**
Namur Digital Institute, University of Namur
anthony.simonofski@unamur.be

**Benoît Vanderose**
Namur Digital Institute, University of Namur
benoit.vanderose@unamur.be

## Abstract

Governments have been publishing Open Government Data (OGD) on online portals to encourage the development of value-added services. The success of these services depends heavily on the quality of OGD and its metadata. Several methods have been proposed to evaluate this quality, but some rely on manual assessments, which can be time-consuming and expensive to perform. Furthermore, these methods focus on the portal rather than the data, ignore user preferences, or do not distinguish between metadata and data quality. This makes it difficult for users to identify data quality issues. This paper proposes a list of OGD quality dimensions for assessing data and metadata quality. The dimensions were identified through a literature review and integrated into a novel conversational agent that incorporates user preferences into the quality assessment. A usability evaluation with 14 users reveals its ease of use and usefulness for obtaining overall and detailed (meta)data quality.

### Keywords

Open Government Data, OGD, metadata, quality, user preference, conversational agent.

## Introduction

Governments around the world have been publishing a vast amount of data on online portals under the Open Government Data (OGD) movement to increase transparency and encourage reuse for the creation of value-added services (Attard et al. 2015) by infomediaries (referred to as users in this article). While the social and economic benefits of reuse are well-known (Berends et al. 2020), using data without considering its quality can diminish the value of the services that rely on it (Crusoe et al. 2019; Kubler et al. 2018). When assessing OGD quality, it is important to consider two elements: the data and its associated metadata (Neumaier et al. 2016). Metadata refers to information about the data, such as authorship, source, or licensing details, etc. Evaluating the quality of the metadata is just as important as evaluating the data itself, as the metadata is crucial for ensuring the data usability.

Many previous works suggest methods to evaluate the quality of metadata, data, or both. This study focuses specifically on those related to OGD to address several of their shortcomings. Firstly, most of the methods (Bhandari et al. 2021; Chokki et al. 2022; Neumaier et al. 2016) focus on evaluating the overall quality of portals instead of data, making it challenging to apply the methods proposed in these studies to evaluate a single dataset. Secondly, some methods (Li et al. 2018; Nogueras-Iso et al. 2021; Vetrò et al. 2016; Wenige et al. 2021) use manual assessment to evaluate quality dimensions, which is time-consuming and expensive to perform. Thirdly, some methods (Raca et al. 2022; Reiche et al. 2014; Wenige et al. 2021) can only evaluate either metadata or data, consider a limited number of dimensions, or combine the dimensions of metadata and data, making it difficult for users to identify data quality issues. Finally, most of the methods (Bhandari et al. 2021; Chokki et al. 2022; Chu and Tseng 2016; Raca et al. 2022) do not consider user preferences in the assessment, even though the level of acceptance of data can vary among users based on the context of data reuse (Even and Shankaranarayanan 2007). Furthermore, to our knowledge no study has explicitly presented a complete list of dimensions applicable separately to OGD and its metadata.

The aim of this paper is to address these gaps by (1) identifying a list of quality dimensions for evaluating metadata and data separately, (2) integrating these dimensions into a tool that incorporates user preferences, and (3) evaluating the tool's effectiveness. Our research question is "How to support automated OGD quality assessment in a way that distinguishes data and metadata and incorporates user preferences?" To answer this question, a literature review was conducted to identify an exhaustive set of quality dimensions for evaluating both metadata and data, which were then incorporated into a novel goal-oriented conversational agent (referred to as "prototype" in this article), meaning that the agent has a programmed structure for the conversation. The implementation as a conversational agent was favored for three reasons. Firstly, it offers an intuitive and user-friendly way to input user preferences and requires minimal to no training due to its natural language communication and integration into popular instant messaging apps (Cantador et al. 2021). Secondly, it has been demonstrated to be effective in related studies focused on open data search and exploration (Cantador et al. 2021; Keyner et al. 2019; Porreca et al. 2018; Wang et al. 2023). Lastly, to our knowledge, the use of conversational agents in assessing data quality has not been previously explored, which makes its potential relevancy unexplored. The prototype works by first asking users to specify a dataset to work on and their preferences in terms of data attributes and quality dimensions they deem relevant for their reuse. Then, the conversational agent returns a fully automated assessment of the data and metadata associated to the chosen dataset. To evaluate the ease of use and the usefulness of the prototype, we conducted user testing with 14 participants. The evaluation reveals that the prototype is easy to use and useful to get an overview of (meta)data quality but also to easily identify where data quality issues lie.

The remainder of the paper is organized as follows. We discuss the existing methods for OGD quality assessment and their limitations. Then, we explain our research methodology, and present in the results section the list of quality dimensions, our conventional agent and its evaluation. Contributions, limitations, and future avenues are discussed in the conclusion section.

## Background

### *Relevance of Metadata and Data Quality for Open Government Data*

Open Government Data refers to data made available online by governments for free reuse and redistribution by anyone (Attard et al. 2015). These data are published on online portals and consist of two parts: the data in itself and its metadata. Data refers to the resource or distribution available for access or download in various formats (e.g., CSV, PDF, Microsoft Excel spreadsheet, etc.) (Neumaier et al. 2016). On the other hand, metadata refers to data that describes other data (Duval 2001), making it easier to retrieve, use, or manage the data (Milic et al. 2021). In the OGD context, it typically includes information such as the title, description, related topic, keywords, source, license, publisher, contact details, creation, and modification date, among others (Milic et al. 2021; Neumaier et al. 2016). One goal of OGD is to drive the creation of social and economic value through the development of OGD-based innovative products and services. This reuse of OGD is a process that can be structured as four steps (Crusoe et al. 2019). One of these steps is the search and evaluation of data. In this step, good (meta)data quality is necessary since it can facilitate easier discovery of the desired information by users (Attard et al. 2015). Later in the reuse process, data is acquired, processed, and transformed into a product or service. In this step, poor data quality can obstruct the efficient reuse of the data, escalate the costs associated with accessing and interpreting the data, and potentially result in incorrect conclusions being drawn (Zhang and Xiao 2020).

As such, data of good quality can be defined as data that is "fit for use by data consumers [i.e., users]" (Wang and Strong 1996). In other words, it is a measurement of the ability of data to meet the specific needs of users. As for the metadata quality, it can be defined as the fitness to describe the data, supporting the tasks of finding, identifying, selecting, and eventually obtaining the data (Reiche et al. 2014). Previous studies (Kubler et al. 2018; Vetrò et al. 2016) have evaluated (meta)data quality by combining the values of various (meta)data quality dimensions, which are sets of attributes that represent specific aspects of data quality (Batini et al. 2009; Wang and Strong 1996). Examples of data quality dimensions include completeness, accuracy, consistency, and timeliness (Bhandari et al. 2021). Each dimension is described by one or more metrics (also called sub-dimensions) such as the percentage of complete cells (i.e., cells that are not null nor empty) for the completeness dimension (Vetrò et al. 2016).

### *Existing Quality Assessment Studies*

Many studies have investigated the assessment of the quality of (meta)data, but this study focuses specifically on those related to OGD, excluding linked open data (Zaveri et al. 2012). To make a fair comparison between these studies, the criteria established by Zhang and Xiao (2020) have been utilized and supplemented with additional criteria (see those underlined). The considered key criteria are: *indicator* (which quality dimensions were taken into consideration), *data type* (whether the focus was on metadata, data, or both), <u>*clarity of distinction*</u> (whether there was a clear separation between dimensions for data and metadata), <u>*application*</u> (which portals or datasets the study has been applied to), *operation* (whether the assessment was automated or manual), and <u>*user preferences*</u> (whether user can setup the dimensions, metrics, or data attributes to be taken into account in the quality assessment). Table 1 presents the previous studies described by the mentioned criteria.

Based on Table 1, we can note that no study was able to fully cover all the criteria mentioned. For example, Raca et al. (2022) considered the coverage of metadata and data metrics as well as automation, but their work is applicable to the portal rather than the dataset, there is no clear distinction between the metrics to be used on metadata or data, and there is no option for users to define their preferences. This article aims to provide a data quality assessment that can cover all the listed criteria. It therefore differs from previous studies in several aspects. It defines a list of dimensions that distinguishes between data and metadata and focuses on the dataset level rather than the portal level. It provides a tool that supports fully automated data and metadata quality assessment, and it allows users to set their preferences at the dimension, metric, and attribute levels. Finally, it is the first work that documents the necessary features to be integrated into a tool for assessing OGD quality.

### *Conversational Agents for Open Government Data*

According to Nuseibeh (2018), conversational agents are software programs that interpret and respond to natural language statements made by users. Nuseibeh also categorizes conversational agents into two types: chatbots and goal-oriented conversational agents. Chatbots are designed to simulate conversations with human users, with ChatGPT being a well-known example. In the context of OGD, several studies (Cantador et al. 2021; Keyner et al. 2019; Porreca et al. 2018; Wang et al. 2023) have utilized chatbots for OGD search and exploration.

Goal-oriented conversational agents, on the other hand, have a pre-programmed structure for the conversation, controlling the conversation flow by asking questions and ignoring user inputs that do not answer the question. Moreau et al. (2019) propose a tool named SemanticBot, a semi-interactive ontology mapping tool that utilizes goal-oriented conversational agents to provide an easy-to-use interface for mapping ontologies on OpenDataSoft datasets.

Despite the popularity of conversational agents in OGD research, there is a lack of studies that utilize them as interfaces for data quality assessment. This paper aims to address this gap by proposing a goal-oriented conversational agent that infomediaries can use for assessing (meta)data quality. In addition to filling this gap, the implementation of a conversational agent was preferred for two reasons. Firstly, it offers an intuitive and user-friendly way for inputting user preferences, requiring minimal to no training due to its natural language communication and integration into popular instant messaging apps (Cantador et al. 2021). Secondly, it has been shown to be effective in related OGD studies (Cantador et al. 2021; Keyner et al. 2019; Moreau et al. 2019; Porreca et al. 2018; Wang et al. 2023).

## Research Methodology

To address the research question "How to support automated OGD quality assessment in a way that distinguishes data and metadata and incorporates user preferences?", this study follows a research design comprising a literature review, prototyping, and an evaluation through user testing with follow-up interviews.

| **Studies** | **Indicator** | **Data type** | **Clarity of distinction** | **Application** | **Operation** | **User preferences** |
|---|---|---|---|---|---|---|
| (Chu and Tseng 2016) | Accessibility, Primary, Timely, Accuracy, Integrity, and Abundance | Metadata & Data | No | OGD platform of the central government of Taiwan (including open data of nine agencies) | Not Mentioned | No |
| (Reiche and Hofig 2013) | Completeness, Weighted Completeness, Accuracy, Richness of Information and Accessibility | Metadata | N/A | 3 public government data repositories, namely GovData.de, data.gov.uk and publicdata.eu | Automated | No |
| (Chokki et al. 2022) | Existence of following attributes: title, description, language, theme, keywords, license, publisher, references, and release date | Metadata | N/A | 6 Gulf Cooperation Council national portals: Bahrain, Kuwait, Oman, Qatar, KSA, and UAE | Automated | No |
| (Li et al. 2018) | Completeness, Accuracy, Consistency, Timeliness, Uniqueness, Understandability, Openness | Metadata & Data | No | Beijing, Guangzhou, and Harbin data platforms | Semi-automated | No |
| (Raca et al. 2022) | Openness, Availability, Accessibility, Discoverability, Timeless, Completeness, Uniqueness, Consistency, Validity | Metadata & Data | No | 6 Western Balkan National Open Data Portals: Albania, Bosnia and Herzegovina, Kosovo, North Macedonia, Montenegro, and Serbia | Automated | No |
| (Bhandari et al. 2021) | Completeness | Data | N/A | OGD of South Korea: National and 3 municipalities (Incheon, Seoul, and Gyeonggi) | Automated | No |
| (Kubler et al. 2018) | Existence, Conformance, Retrievability, Accuracy, Open data | Metadata | N/A | 250 open data portals, powered by organizations across 43 different countries | Automated | Yes |
| (Umbrich et al. 2015) | Retrievability, Usage, Completeness, Accuracy, Openness, Contactability | Metadata | N/A | 82 CKAN portals | Automated | No |
| (Neumaier et al. 2016) | Existence, Conformance, Retrievability, Accuracy, Open Data | Metadata | N/A | 260 Open Data portals | Automated | No |
| (Kubler et al. 2016) | Usage, Completeness, Openness, Addressability, Retrievability | Metadata | N/A | 146 CKAN portals across 44 countries | Automated | Yes |
| (Reiche et al. 2014) | Completeness, Weighted Completeness, Accuracy, Richness of Information, Readability, Availability, Misspelling | Metadata | N/A | 10 CKAN OGD portals | Automated | No |
| (Vetrò et al. 2016) | Traceability, Currentness, Expiration, Completeness, Compliance, Understandability, Accuracy | Data | N/A | Italian portals: National level and municipality level (Torino, Roma, Milano, Firenze, Bologna) | Semi-automated | No |

| | | | | | | |
|---|---|---|---|---|---|---|
| (Data.europa.eu 2020) | Completeness, Findability, Accessibility, Interoperability (Conformity, Machine Readability, Openness), Reusability (Timeliness, Consistency, Accuracy, Relevance, Understandability, Credibility), Contextuality | Metadata | N/A | Applicable to individual and overall datasets on Data Europa portal | Automated | No |

**Table 1: Existing OGD quality assessment studies.**

The literature review (Webster and Watson 2002) was conducted using the "Scopus," "Association for Computing Machinery," and "Google Scholar" databases. It helped to identify dimensions and associated metrics related to (meta)data quality, using as the following search queries: ("metadata quality" OR "data quality") AND (evaluation OR assessment OR assess OR evaluate) AND ("data" OR "open data" OR "open government data" OR "public data" OR "public government data" OR "government data" OR "public sector information"). An automated search returned 311 articles, out of which 13 relevant ones were selected through a three-stage process that evaluated type, domain, and title, examined the abstract, and scanned the content. As the focus was on automating quality assessment, only objective (i.e., computable/assessable without human intervention) dimensions and metrics were considered, which were grouped and categorized into metadata or data quality dimensions/metrics based on conceptual similarities.

After identifying the dimensions/metrics, we utilized a prototyping approach (Budde et al. 1992) to propose a practical tool for users that can effortlessly integrate them into their quality assessment process. User requirements for the prototype were collected from existing tools from the literature review and integrated along with the dimensions/metrics. The backend of the prototype is developed with Django and PostgreSQL, while the frontend is built using Angular (code available at https://github.com/chokkipaterne/qualityogd).

After implementing the prototype, we recruited 14 users through an open call for participation in a "data analytics" course at the University of the first author and conducted user testing to evaluate the prototype ease of use and usefulness. Additionally, we administered an online survey (available online[1]) to gather further feedback. To minimize errors associated with survey research, the survey was pretested with two users (Grimm 2010). The survey comprised items measured on a 5-point Likert scale (from "strongly disagree" to "strongly agree") based on the System Usability Scale (SUS) questionnaire (Brooke 1986) with 5 additional questions to evaluate ease of use and usefulness. In addition to these questions, there were free-text questions to collect general opinions and suggestions for future versions and 3 additional questions to collect demographic data (data manipulation skills, age, and education). Participants were invited to test the prototype with their preferred datasets from the OpenDataSoft (ODS) data portal[2] to assess their (meta)data quality. During the test, we used an exploratory approach (Rubin and Chisnell 2008) and only provided guidance when participants asked for assistance. After completing the survey, users were asked what features should be *kept, improved, removed, or added* to facilitate (meta)data quality assessment. We calculated the median, mean, and standard deviation for the Likert scale questions, and coded verbal thoughts and responses from the free-text questions using short sentences to retain context and conceptual relations.

# Results

### *A Conversational Agent for (Meta)data Quality Assessment*

Using existing tools features, we gather initial features for the OGD quality assessment tool, which include (F1) an interface that is easy to use and intuitive, (F2) an automated assessment process, (F3) the ability to differentiate between metadata and data quality assessments, (F4) direct dataset selection from portals,

[1] https://forms.gle/Ht4neqYdhUV7YVPG6

[2] https://data.opendatasoft.com/

(F5) user preference selection for dimensions, metrics, and attributes during quality assessment, (F6) data quality visualization, and (F7) an explanation of the reasons behind each score for each dimension.

We developed a conversational agent to incorporate the identified features. A demonstration video of the first author evaluating the quality of the data "JCDecaux Bike Stations Data"[3] collected from the ODS portal is accessible at https://rb.gy/89yfpr. As explained in the introduction, conversational agents require minimal to no training due to their natural language communication and integration into popular instant messaging apps. Based on this, the implemented prototype meets requirement (F1). Figure 1 portrays the prototype conversation flow on its left side, where user inputs (in rectangle boxes) and prototype outputs (in parallelogram boxes) are represented by intents. On the right side of the figure, there is a screenshot of the prototype interface.

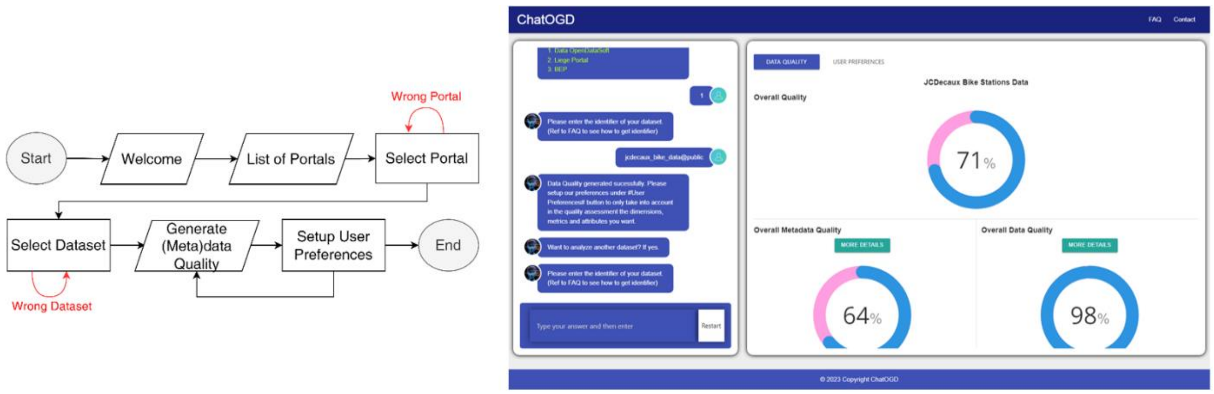


**Figure 1: Conversation flow of the prototype (parallelogram boxes for prototype outputs and rectangle boxes for user inputs) (Left). Screenshot of prototype interface (Right).**

**Welcome**. This intent is automatically triggered at the start of the conversation, during which the prototype introduces itself and greets the user.

**Portals List.** This intent is automatically triggered after the welcome intent or when the user clicks the "Restart" button to initiate the process from the beginning. Its purpose is to provide a list of the portals included in the prototype.

**Portal Selection.** This intent enables users to choose the portal to which the dataset being processed belongs, from the portals list. If the user selects an incorrect portal identifier, the prototype will notify them that the selected identifier does not exist and prompt the user to choose another.

**Dataset Selection.** This intent enables the user to input the identifier of the dataset they wish to assess (*F4*). The prototype then sends this identifier to the backend to verify its existence and the distribution format of the dataset, which must be machine-readable. If the identifier exists and the format is supported, the prototype automatically generates the (meta)data quality of the selected dataset using all dimensions, metrics, and attributes. If the identifier is invalid, the prototype informs the user and prompts them to try again.

**Generate (Meta)data Quality.** This intent enables the user to automatically retrieve the (meta)data quality of the selected dataset (*F2*). The results are presented in three parts with visualizations to facilitate understanding:

– *overall quality*, calculated as the mean of all (meta)data dimensions and displayed using a gauge chart (*F6*).
– *metadata quality (F3),* calculated as the mean of only metadata dimensions and displayed using a gauge chart (*F6*). By clicking on "More details", the gauge chart is replaced with a bar chart (*F6*) where each bar represents one metadata dimension, and its value is computed as the mean of the metrics

[3] https://data.opendatasoft.com/explore/dataset/jcdecaux_bike_data%40public/table/

under that dimension. When the user clicks on a dimension, the prototype displays the bar chart for the metrics under the selected dimension (F7).

– *data quality (F3),* similar to metadata quality but only displays dimensions and metrics related to the data.

Table 2 displays the list of dimensions and metrics used in this study. Detailed information on these metrics, including their definitions and associated formulas, can be accessed at https://rb.gy/87udy. To ensure that their assessment remains dynamic, we chose to include only dimensions that can be calculated using formulas, based on previous studies (Data.europa.eu 2020; Reiche et al. 2014; Vetrò et al. 2016).

| **Dimensions** | **Metrics** |
|---|---|
| **Metadata** | |
| Completeness | Percentage of complete fields in metadata |
| Findability | Keywords assigned, Categories assigned, Title given, Description of data given, Temporal information given, Spatial information given |
| Accessibility | Access URL accessible, Download URL given, Download URL accessible without registration |
| Conformity | Conformity of URLs, Conformity of date formats, Conformity of email addresses, DCAT-AP compliance of metadata |
| Machine readability/ processability | Processability of file format and media type |
| Openness | Openness of file format and media type, License information given, Openness of license |
| Timeliness | Update information given, Creation date given, Modification date given |
| Accuracy | File format accuracy, Content size accuracy |
| Understandability | Percentage of columns with metadata |
| Credibility | Contact point given, Dataset publisher given |
| Uniqueness | Title is unique, Description is unique, Identifier is unique |
| **Data** | |
| Completeness | Percentage of complete cells, Percentage of complete rows |
| Accuracy | Percentage of accurate cells |
| Uniqueness | Percentage of unique rows, Percentage of unique columns |

**Table 2: List of dimensions and metrics integrated into the conversational agent.**

**User Preferences.** This intent allows users to configure their preferences (*F5*). In contrast to other studies (Kubler et al. 2018) that limit user choices to dimensions, this study focuses on three aspects of user preferences: dimensions, metrics, and data attributes. For the dimensions and metrics, the user can set a weight for each of them. The default weight is 1, meaning that all dimensions and metrics have the same impact on the quality assessment, but users can assign a greater weight to a specific dimension or metric to indicate that it is more important. Users can also set a weight of 0 to exclude a specific dimension or metric from the quality assessment. In addition, users can choose which data attributes to include in the data quality assessment by unchecking unwanted attributes using checkboxes.

### *Evaluation Results*

We collected opinions from 14 participants (10 undergraduates between the ages of 18 and 29 and 4 doctoral students between the ages of 30 and 49), regarding the ease of use and usefulness of the conversational agent to assess (meta)data quality. This was done through surveys that the participants completed after exploring the prototype.

| | SUS Score (Q1-Q10) | Usefulness for quality assessment | | | | |
|---|---|---|---|---|---|---|
| | | Data + Metadata (Q11) | Metadata (Q12) | Data (Q13) | better highlight errors related to a dataset (Q14) | take into account user preferences (Q15) |
| MD | 85 | 4 | 5 | 5 | 4 | 4 |
| M (SD) | 86.07 (6.10) | 4.42 (0.53) | 4.57 (0.53) | 4.57 (0.53) | 3.57 (0.78) | 3.71 (0.75) |

**Table 3. Median (MD), mean (M) and standard deviation (SD) of survey scores.**

Table 3 presents the median (MD), mean (M), and standard deviation (SD) of the SUS score (which summarizes questions about usability), and five additional questions used to assess the usefulness of the prototype. The results of Table 3 lead to several conclusions. First, based on the interpretability of the SUS score presented in (Lewis and Sauro 2018), the prototype offers excellent usability as both the MD and M scores for the SUS are above 85. Second, participants agree that the prototype is useful for assessing overall quality, metadata quality, and data quality, as the MD and M of these questions are each equal to or greater than 4 with SD < 1. Third, most users agree that the prototype helps them to identify where data quality issues exist (Q14) and set their preferences (Q15), as MD and M are greater than 3.5 with SD≈0.75.

During testing, we found that all participants were able to assess their selected dataset's (meta)data quality with little or no assistance. Participants indicated that the prototype was easy sto use, intuitive, and ergonomic, which explains the excellent SUS score. The user preferences section, particularly at the attribute level, was highly valued by most participants as it allowed them to evaluate only the columns that were relevant to them. All participants appreciated the use of visualizations to express the quality of the (meta)data instead of simple numbers. They also appreciated the hierarchical presentation of quality results (starting with an overview, followed by dimension level, and finally metric level), which allowed them to quickly grasp the overall quality of the (meta)data and then delve deeper to understand the underlying values.

All participants agreed that the current features should be kept, but many of them suggested simplifying the presentation of weight setup at the dimension and metric level. As for new features, they proposed adding a dimension to measure the correctness of the (meta)data description in comparison to its title and content, as well as the accuracy of the column descriptions and cell values to ensure that the dataset reflects real-world data accurately. One participant also recommended adding an option to export the (meta)data quality results in PDF or Excel format. Another participant suggested including more details on certain metrics, such as presenting the specific columns that make the rows incomplete in the data for the "Percentage of complete rows" metric.

## Conclusion

The aim of this paper is to identify quality dimensions for assessing metadata and data separately in the context of OGD, integrate them into a tool that accommodates user preferences, and assess its effectiveness. To accomplish this, we conducted a literature review to identify quality dimensions, metrics, and initial user requirements for the prototype. We then incorporated this information into a conversational agent and evaluated its ease of use and usefulness in assisting users to evaluate the quality of a dataset through a user test that involved 14 participants. The user test results show that users found the prototype easy to use and useful to assess the quality of data and metadata, to highlight the errors related to a dataset, and to take into account their preferences.

This research makes theoretical contributions in several aspects. Firstly, it builds upon prior studies (Data.europa.eu 2020; Reiche et al. 2014; Vetrò et al. 2016) to propose a comprehensive list of dimensions and metrics (sub-dimensions) necessary to evaluate the quality of metadata and data content separately (Table 2). Secondly, unlike previous studies, this study explicitly specifies which dimensions are applicable to metadata and data content and provides corresponding metrics to be considered for each dimension. This categorization enables infomediaries to be aware of the relevant dimensions and metrics when evaluating metadata or data content. Additionally, it provides an overview of the dimensions and allows for a more in-depth understanding of the errors in the data. Thirdly, through existing tools, we have identified a set of features required to develop a (meta)data quality assessment tool. This list can serve as a reference for infomediaries or researchers to compare or analyze (meta)data quality assessment tools.

One significant limitation of this study is the potential lack of representativeness of the evaluation sample. While the number of participants may be considered small, previous studies (Faulkner 2003; Nielsen 2000) suggest that a minimum of five participants for usability tests is a good starting point, and Lallemand and Gronier (2018) indicate that around 15 users is enough to study general tendency. To improve representativeness, we suggest exploring alternative communication channels or conducting on-site evaluations at universities or with open data users. Another limitation of the study is the small number of dimensions/metrics selected on the data level. However, this decision was made deliberately, as only objective dimensions/metrics were chosen. This was to ensure that their measurements could not differ from one user to another, unlike subjective measures such as column understandability, which can vary depending on the user's perception. In future work, we plan to (re)define (existent) additional dimensions/metrics that are relevant to the (meta)data and explore ways to automate them through machine learning methods, and also assess whether the quality assessment computed by the prototype is consistent with expert judgment. Second, we will assess the impact of a conversational agent's presence on an OGD portal by utilizing the Technology Acceptance Model (Davis 1989) to measure citizens' intent to use it. Third, we plan to conduct an experimental study to identify which types of infomediaries (developers, data scientists, businesses, or citizens) are most likely to find the conversational agent helpful.

## REFERENCES


Attard, J., Orlandi, F., Scerri, S., and Auer, S. 2015. 'A Systematic Review of Open Government Data Initiatives', *Government Information Quarterly* (32:4), Elsevier Ltd, pp. 399–418.

Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. 2009. 'Methodologies for Data Quality Assessment and Improvement', *ACM Computing Surveys* (41:3).

Berends, J., Carrara, W., Engbers, W., and Vollers, H. 2020. 'Re-Using Open Data'.

Bhandari, S., Ranjan, N., Kim, Y. C., Park, J. Do, Hwang, K. Il, Kim, W. H., Hong, Y. S., and Kim, H. 2021. 'An Automatic Data Completeness Check Framework for Open Government Data', *Applied Sciences* (11:19), pp. 1–17.

Brooke, J. 1986. 'SUS, A Quick and Dirty Usability Scale', *Usability Evaluation In Industry*.

Budde, R., Kautz, K., Kuhlenkamp, K., and Züllighoven, H. 1992. 'What Is Prototyping?', *Information Technology & People* (6:2/3), pp. 89–95.

Cantador, I., Viejo-Tardío, J., Cortés-Cediel, M. E., and Rodríguez-Bolívar, M. P. 2021. 'A Chatbot for Searching and Exploring Open Data: Implementation and Evaluation in E-Government', in *DG.O2021*, pp. 168–179.

Chokki, A. P., Alexopoulos, C., Saxena, S., Frénay, B., Vanderose, B., and Ali, M. 2022. 'Metadata Quality Matters in Open Government Data (OGD) Evaluation! An Empirical Investigation of OGD Portals of the GCC Constituents', *Transforming Government: People, Process and Policy*, pp. 1–21.

Chu, P. Y., and Tseng, H. L. 2016. 'A Theoretical Framework for Evaluating Government Open Data Platform', in *EGOSE*, pp. 135–142.

Crusoe, J., Simonofski, A., Clarinval, A., and Gebka, E. 2019. 'The Impact of Impediments on Open Government Data Use: Insights from Users', *in RCIS19,* pp. 1–12.

Data.europa.eu. 2020. 'Metadata Quality Assessment Methodology'. (https://data.europa.eu/mqa/methodology, accessed January 9, 2023).

Davis, F. D. 1989. 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *Management Information Systems* (13:3), pp. 319–339.

Duval, E. 2001. 'Metadata Standards, What, Who & Why', *Journal of Universal Computer Science* (7:7), pp. 591–601.

Even, A., and Shankaranarayanan, G. 2007. 'Utility-Driven Assessment of Data Quality', *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* (38:2), pp. 75–93.

Faulkner, L. 2003. 'Beyond the Five-User Assumption: Benefits of Increased Sample Sizes in Usability Testing', *Behavior Research Methods, Instruments, & Computers 35*, pp. 379–383.

Grimm, P. 2010. 'Pretesting a Questionnaire', *Wiley International Encyclopedia of Marketing*, John Wiley & Sons, Ltd.

Keyner, S., Savenkov, V., and Vakulenko, S. 2019. 'Open Data Chatbot', in *ESWC 2019*, pp. 111–115.

Kubler, S., Robert, J., Neumaier, S., Umbrich, J., and Le Traon, Y. 2018. 'Comparison of Metadata Quality in Open Data Portals Using the Analytic Hierarchy Process', *GIQ* (35:1), Elsevier, pp. 13–29.

Kubler, S., Robert, J., Traon, Y. Le, Umbrich, J., and Neumaier, S. 2016. 'Open Data Portal Quality Comparison Using AHP', in *Dg.o '16*, pp. 397–407.

Lallemand, C., and Gronier, G. 2018. 'Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs'. Editions Eyrolles.

Lewis, J. R., and Sauro, J. 2018. 'Item benchmarks for the system usability scale'. Journal of Usability Studies (13:3).

Li, X. T., Zhai, J., Zheng, G. F., and Yuan, C. F. 2018. 'Quality Assessment for Open Government Data in China', in *ICIME*, pp. 110–114.

Milic, P., Veljkovic, N., and Stoimenov, L. 2021. 'Comparative Analysis of Metadata Models on E-Government Open Data Platforms', *IEEE Transactions on Emerging Topics in Computing* (9:1), pp. 119–130.

Moreau, B., Terpolilli, N., Serrano-Alvarado, P. (2019). 'A Semi-Automatic Tool for Linked Data Integration'. in ISWC2019.

Neumaier, S., Umbrich, J., and Polleres, A. 2016. 'Automated Quality Assessment of Metadata across Open Data Portals', *Journal of Data and Information Quality* (8:1).

Nielsen, J. 2000. 'Why You Only Need to Test with 5 Users', *Nielsen Norman Group*. (https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/, accessed February 7, 2023).

Nogueras-Iso, J., Lacasta, J., Urena-Camara, M. A., and Ariza-Lopez, F. J. 2021. 'Quality of Metadata in Open Data Portals', *IEEE Access* (9), pp. 60364–60382.

Nuseibeh, R. (2018). What is a Chatbot? (https://chatbotsmagazine.com/what-is-a-chatbot-6dfff005bb34, accessed April 19, 2023)

Porreca, S., Leotta, F., Mecella, M., Vassos, S., and Catarci, T. 2018. 'Accessing Government Open Data Through Chatbots', in *ICWE 2017*, pp. 156–165.

Raca, V., Velinov, G., Dzalev, S., and Kon-Popovska, M. 2022. 'A Framework for Evaluation and Improvement of Open Government Data Quality: Application to the Western Balkans National Open Data Portals', *SAGE Open* (12:2).

Reiche, K. J., Hofig, E., and Schieferdecker, I. 2014. 'Assessment and Visualization of Metadata Quality for Open Government Data', in *CeDEM14*, pp. 225–346.

Reiche, K. J., and Hofig, E. 2013. 'Implementation of Metadata Quality Metrics and Application on Public Government Data', in *IEEE 37th Annual Computer Software and Applications Conference Workshops*, pp. 236–241.

Rubin, J., and Chisnell, D. 2008. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*, Wiley Publishing, Inc.

Umbrich, J., Neumaier, S., and Polleres, A. 2015. 'Quality Assessment and Evolution of Open Data Portals', in *3rd International Conference on Future Internet of Things and Cloud*, pp. 404–411.

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., and Morando, F. 2016. 'Open Data Quality Measurement Framework: Definition and Application to Open Government Data', *Government Information Quarterly* (33:2), Elsevier Inc., pp. 325–337.

Wang, D., Richards, D., Bilgin, A. A., and Chen, C. 2023. 'Implementation of a Conversational Virtual Assistant for Open Government Data Portal: Effects on Citizens', *Journal of Information Science*, pp. 1–21.

Wang, R. Y., and Strong, D. M. 1996. 'Beyond Accuracy: What Data Quality Means to Data Consumers', *Journal of Management Information Systems* (12:4), pp. 5–33.

Webster, J., and Watson, R. T. 2002. 'Analyzing the Past to Prepare for the Future: Writing a Literature Review', *MIS Quarterly* (26:2), xiii–xxiii.

Wenige, L., Stadler, C., Martin, M., Figura, R., Sauter, R., and Frank, C. W. 2021. 'Open Data and the Status Quo -- A Fine-Grained Evaluation Framework for Open Data Quality and an Analysis of Open Data Portals in Germany', *ACM Journal of Data and Information Quality* (1:1).

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., and Auer, S. 2012. 'Quality Assessment for Linked Data: A Survey', *Semantic Web* (1), pp. 1–31.

Zhang, H., and Xiao, J. 2020. 'Quality Assessment Framework for Open Government Data: Meta-Synthesis of Qualitative Research, 2009-2019', *Electronic Library* (38:2), pp. 209–222.