

Article

Using OpenGovB Transparency Indicator to Evaluate National Open Government Data

Petar Milić ^{1,*} , Nataša Veljković ² and Leonid Stoimenov ²

¹ Faculty of Technical Sciences, University of Priština—Kosovska Mitrovica, Kneza Miloša 7, 38227 Kosovska Mitrovica, Serbia

² Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia; natasa.veljkovic@elfak.ni.ac.rs (N.V.); leonid.stoimenov@elfak.ni.ac.rs (L.S.)

* Correspondence: petar.milic@pr.ac.rs; Tel.: +381-654-627-132

Abstract: Transparency evaluation in open government is a process of measuring the extent of transparency against a predefined set of indicators. In this paper, we address the existing initiatives regarding data and government transparency evaluation as two separate indicators and present the analysis of their advantages and drawbacks. Based on that analysis, we extend a part of the OpenGovB benchmark related to transparency in open government. What is unique about this benchmark is that it utilizes metadata of data published on the open government data portals to calculate the majority of indicators related to data transparency indicators. For the government transparency indicator evaluation, the benchmark utilizes some of the well-known transparency indicators. The article shows concrete results obtained from the application of the defined transparency evaluation model on 22 open data portals, thus demonstrating the possibilities of its application as well as the gains regarding generated results. The proposed model bridges the gap between available methodologies for evaluating transparency based on collaboration and participation and methodologies for evaluating transparency based on open data.

Keywords: Open Government benchmark; transparency evaluation; data transparency; government transparency



Citation: Milić, P.; Veljković, N.; Stoimenov, L. Using OpenGovB Transparency Indicator to Evaluate National Open Government Data. *Sustainability* **2022**, *14*, 1407. <https://doi.org/10.3390/su14031407>

Academic Editors: Hans Jochen Scholl, Luis Felipe Luna-Reyes and Manuel Pedro Rodríguez Bolívar

Received: 3 December 2021

Accepted: 24 January 2022

Published: 26 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Open government fulfills society's demand for responsible and responsive government and emphasizes the role of transparency as a determinant of government performance. Thanks to the global initiative and rapid development of the e-government idea, the concept of e-government has been raised to a higher, more sophisticated level of open government [1,2]. This new e-government model involves data openness, transparency, and participation as the main openness features, and is driven by modern technologies and created concepts that enable greater online collaboration [3]. The concept of open data has had the most impact on shaping the open government and directing its growth towards data and not services [4]. Data openness represents an essential precondition for building transparency, and it is promoted around the world as part of the open government initiatives. Data openness is not solely focused on making information obtainable, but also on ensuring that data are well-known, easily accessible, and open to all. Publishing data is the first step, while the second step is the provision of data in a way that creates opportunities for users to go beyond passive recipients [5].

Transparency cannot be separated from open data, as the principle of open data is the main prerequisite for building transparency. A transparent government is one that allows free access to open data and guarantees the authenticity of publishing agencies, data integrity, accuracy, and quality, providing users with understandable datasets in reusable formats. Harrison et al. [6] state that transparency cannot be achieved by mere downloading of data sets. The data must be useful and allow users to create more value. Jaeger

and Bertot [7] emphasize that the issue of transparency does not have only short-term considerations, regarding information availability to all, but also long-term considerations referring to information usability by all. They further elaborate that such demands require establishing tasks related to achieving information usability and accessibility, promoting government and information and technology literacy, making appropriate and accurate content and services available, meeting user expectations, promoting trust, and encouraging lifelong usage. The accomplishment of such defined tasks requires the definition of comprehensive descriptive information [8], as a form of transparent government information, and its deliverance to the public through online services.

Evaluation of open government as a concept consisting of evaluation of data openness, transparency, participation, and collaboration has not been addressed in the academic literature in the manner we propose with the OpenGovB [2]. Furthermore, we could not find a benchmark that evaluates transparency in the open government context, in a way that allows automatic calculation of given parameters solely based on metadata of published datasets. Automating evaluation would make the process of open data analysis much faster, and it would contribute to the objectivity of results compared to analysis in which data are gathered and analyzed manually. Another motive behind this evaluation is to provide a comparison of the proposed evaluation framework regarding the existing standards, or globally accepted definitions for values being measured, with a lack of standards. The proposed framework for transparency evaluation in open government is designed to address transparency by using government and data transparency. Combining these indicators, we have gained a unique transparency indicator for open government, which will be explained more thoroughly in the next sections of the paper. We will explain the theoretical background of the evaluation model and provide results from the analysis of transparency on 22 open data portals.

2. Background

Transparency is a noun that represents the condition of being transparent. However, if we observe the adjective, it means that the associated object may be interpreted differently, depending on the object itself. This preliminary remark is necessary since in this paper we will look at data transparency and government transparency and embrace a particular approach accordingly. Keeping this in mind, in the following paragraphs we will provide a revision of the literature on transparency made up of these two important factors.

In literature, data transparency is often not separated from the concept of data openness and is defined accordingly [9,10]. Tauberer [10] points out three principles of openness: accessibility, authenticity, and accuracy. The accessibility principle complies with openness demands which promote online availability of data and free access without discrimination or the need to agree to a license. Authenticity relates to users' trust in published data, as well as data relevance. This principle refers to the authenticity of data sources in terms of their reliability and reputation. Accuracy refers to data precision and represents one aspect of data quality. Seen through these principles, transparency could be defined as a measure of openness, data sources' authenticity, and data accuracy and integrity. However, there is more to transparency. Veljković et al. see data transparency as not only authenticity of data and data sources, but also as data understandability and reusability [11]. Understandability denotes clarity of data—the principle that each user should comprehend the information contained—and reusability refers to the possibility of using the same data in a different manner.

The United Nations Global E-Government Survey on the quality of government services and products for the UN Member States addresses e-government readiness and e-participation [12]. An e-participation index addresses quality and information usefulness and services aiming at engaging citizens in public policy through information and communication technologies. It is related to data transparency since it focuses on data quality and usability, as well as the aspect of data openness towards citizens. The Accenture E-Government Leadership Survey shows the government performance by measuring service

maturity and delivery maturity, which reflects different delivery aspects of government [13], but it could not be related to the data transparency feature. The e-government benchmark conducted by [14], popularly named eGovBE, has initially been focused on reviewing twenty basic public services and e-government progress in terms of core indicators: sophistication, availability of services, user orientation, and national portal [15,16]. In 2009, the model's measuring criteria and core indicators were enhanced by replacing national portal and user-centricity with eProcurement and user experience indicators, making the benchmark more oriented towards outcomes and impact. In the light of these new features, data transparency could be seen in user-centricity as an attempt to make citizens more involved with governmental procedures.

These benchmarks were applied to earlier e-government models, which is why data transparency is not their concrete area of focus. Considering that an official open government benchmark still does not exist, we can come to the conclusion that there is no unique evaluation model of data transparency as part of open government in general. This claim is confirmed by the systematic research of Matheus and Janssen, who suggest a comprehensive model of determinants that will foster transparency enabled by open government data [17]. These data are exploited through open data portals, which, according to Lnenicka and Nikiforova [18], play the role of an interface that creates transparency. Release of government data on open data portals influences the processes that go toward contribution to transparency where corruption, wastage, and inefficiency take place the most. However, there are some initiatives, amongst which Osimo's is the most referenced one. Osimo proposes a set of basic public data to measure data reusability and transparency [19]. For the evaluation of basic public data, he recommends a five-level scale, the levels of which gradually change from no data to sheer data availability and reusability. A similar proposal for assessing the availability of open-linked data comes from Sir Tim Berners-Lee who has defined a 5-star model for assessing data reusability based on formats of published data [20]. As data availability is one of the key features of data transparency, this model could also be considered for building a framework for transparency evaluation in open government.

Based on addressed benchmarks and initiatives, we have identified that data transparency is also affected by data quality in the form of data accuracy, data accessibility and availability, data reusability, and data openness. These features are used to gain insight into the e-government openness, but have not been analyzed as part of one transparency evaluation framework. Our intention is to use all these features jointly in our open data transparency evaluation, since their importance is clearly significant, considering their presence in evaluation models applied in earlier stages of e-government. Keeping that in mind, we will utilize OpenGovB transparency indicator and extend it to include identified quality components.

Government transparency evaluation is a process of measuring the extent of transparency against a predefined set of indicators. As transparency is also an e-government feature, existing e-government benchmarks need to be considered to obtain a comprehensive view on assessment methods that will unveil the role of transparency in open government. There are a variety of studies on the impact of transparency in the government domain [21–24], but none of them consider quantitative indicators that will relate transparency to the potential usage of government information by stakeholders.

When it comes to creating indicators of government transparency, we should start from its definition. Government transparency is mostly defined as a measure of citizens' insight into business, processes, and operations of the government [25]. It is about disclosure, providing citizens with a window to find and view information, process flows, issues, events, projects, policies, and other matters of their interest. Meijer et al. [26] distinguish three types of government transparency: active, passive, and forced release of information. Having this in mind, transparency is a complex measure that must encompass all aspects of government interaction with the public. However, transparency is often simplified to a single domain: the provision of information of public importance. For this

simplified view on government transparency, there are some straightforward and generally accepted indicators.

If we look at the government transparency through enabling data openness, then we can talk about two groups of indicators; first, one relies on generally accepted rules of openness, while others are bound to legal regulations [27,28]. Generally accepted rules of data openness depend on the country itself, its history and traditions, political preferences of the nation, tradition, culture, etc. The methodologies that are based on these rules depend on the views of researchers and can be freer in the choice of indicators. On the other hand, legally binding regulations are defined by the law in the field and, therefore, methodologies that accompany these regulations have predictable indicators. One of the most easily measurable and comparable elements of government transparency is the legislation on freedom of information (FOI) [24,29,30]. Methodologies for assessing FOI are mainly based on expert assessment of whether there is an appropriate legal framework that enables access to information of public importance. This kind of transparency is often called passive transparency since dissemination occurs upon user request and is not actively released by authorities. What governments often do is publish information of public importance, which is proactive disclosure of information, and these published data are considered by some authors as a proxy for measuring active transparency [31]. Nevertheless, González-Gallego and Nieto-Torrejón [32] reported that the availability of government data to the relevant stakeholders, of which the most important are the citizens, leads to them having more trust in public institutions, as well as the creation of trustworthy governments [33]. In particular, this is achieved by promoting open government data usefulness among them, thus making governments more accountable and trustable.

The Centre for Law and Democracy (CLD), a non-profit corporation from Canada, created a methodology for comparative assessment of a legal framework for the right to information (RTI). They define 61 indicators divided into seven groups: the right of access, scope, requesting procedures, exceptions and refusals, appeals, sanctions and protections, and promotional measures. Each indicator is scored between 0–2 points, and the total number of points that can be achieved is 150. This methodology was launched in 2013, and since it was oriented towards the legal framework only and not towards the implementation of the framework, some of the most developed countries in Europe had a very low ranking. For example, Scotland has an FOI law originating from 1766 and is ranked 40 (92 out of 150 points), even though they had a long history of implementation. On the other hand, Serbia, with its law on FOI from 2003 and very weak implementation of this law, is ranked as first on the list (135 out of 150 points). If we exclude these negative effects of the methodology, we can say that strong law is of great importance to government transparency since it supports government openness and facilitates access to information of public importance.

Other methodologies for measuring government transparency are aimed at certain transparency aspects, such as budget transparency. The International Budget Partnership (IBP) organization ranks countries based on a questionnaire with 125 questions related to the transparency of government budget [34].

We can also observe the absence of corruption as one of the government transparency aspects. The Corruption Perceptions Index (CPI) or Control of Corruption (CC) are the two most famous indexes of corruption [35]. CPI takes the value from the range (0, 100) where 0 indicates high corruption and 100 denotes the absence of corruption in the government administration. CC indicator takes values from the range (−2.5, +2.5), whereas the upper and lower range limits are related to low and high corruption respectively.

As can be noted, government transparency is a multifaceted concept that is often measured by looking at only some of the facets, such as corruption or FOI law. Moreover, as there is a lot of effort coming from non-governmental organizations and academic institutions to measure different aspects of government transparency, they will be included in our transparency calculation. According to the above, we can conclude that RTI and CPI indicators represent a valuable source for measuring government transparency. These

two will be used for the utilization and extension of the OpenGovB transparency indicator, which will be further evaluated as a constituent of transparency in open government.

3. OpenGovB Transparency Indicator

OpenGovB is a model for assessing the extent of a government's openness in accordance with well-defined and globally embraced openness principles [2]. The benchmark addresses openness of government through four main principles: open data, transparency, participation, and collaboration, and is unique in this approach. In essence, it serves to determine the extent of fulfillment of the main goals of open government. The focus of this paper is on the transparency evaluation using the transparency indicators given in OpenGovB and the usage demonstration on open data portals.

Online data are organized into various data categories; however, OpenGovB examines only nine categories: Finance and Economy, Environment, Health, Energy, Education, Transportation, Infrastructure, Employment, and Population [4], to establish a standard evaluation measurement model. This source is used for assessing BDS, DOI, and T indicators. Additionally, BDS indicates the presence of a predefined set of high-value open data categories, which is necessary for the automatic evaluation of data transparency. User involvement is considered to be a valuable indicator of government transparency, and it is included in the benchmark model in order to express the willingness and readiness of the government to accept and utilize users' perspectives and points of view. It only influences the evaluation of user participation indicators in the proposed model. In OpenGovB, transparency is viewed as the average function of all types of transparency, data, and government, which are equally important for the constitution of the transparent government.

3.1. Transparency Evaluation Model

Based on research on transparency and its meaning in the context of open government, we define transparency in open government through two separate indicators: data transparency and government transparency. Data transparency is measured through authenticity, understandability, and reusability of data available on the government's open data portal. These three measures tell us whether the government publishes necessary information about data sources, whether their formats are reusable, and if they are properly described, which are prerequisites for automatic processing. Furthermore, they are also used for the calculation of e-government openness. Government transparency utilizes existing transparency indicators, primarily RTI and CPI. The final value of transparency is the average value of the two indicators as given in Equation (1).

$$T = \frac{DT + GT}{2} \quad (1)$$

3.2. Data Transparency (DT) Indicator

Keeping in mind the fact that an open data portal publishes large amounts of datasets organized in different data categories, using the complete data collection may have large time and cost impacts. To calculate data transparency, it is necessary to choose a relevant subset of datasets for each category as a sample for the calculation process. Data transparency evaluation is thus a two-phase process where the first phase relates to choosing a relevant subset of datasets, while the other implies formulas' application to key dataset features.

Generating results from sample data, rather than from the complete population, is a statistical challenge that we are confronted with. When we have a reason to believe that the sampled and the unsampled data have essentially similar or equivalent characteristics, we can make conclusions about the entire population based on an incomplete subset of the population. To ensure that the sample is representative not only of itself but also of the unsampled group, it is necessary to consider the level of confidence, the margin of error, and the expected accuracy [36]. The confidence level usually takes a value of 95% or 99%, which corresponds to the standardized value of the confidence interval of ± 1.96 and ± 2.58 .

The confidence interval is the interval within which we expect the mean of the population to be, while confidence level denotes the probability that an unknown population mean value is within the interval. We have chosen a 95% confidence level for which Z takes the value of 1.96 in the calculation, according to the table of standard normal curve areas values. The margin of error allows for a deviation of the sample size results in comparison to population results. In our calculations, the margin of error is valued at 10%. Accuracy denotes a percentage of sampled data that would truly satisfy the required characteristics. As there is no trustworthy way for a reliable prediction of such percentage, we have used the 50% value. For choosing the sample size of the population, we are using the following equations [37]:

$$n_0 = \frac{Z^2 \times p \times (1 - p)}{e^2} \quad (2)$$

$$n = \frac{n_0 \times N}{n_0 + (N - 1)} \quad (3)$$

Equation (2) determines the sample size using the confidence level (Z), the margin of error (e), and expected accuracy (p), while Equation (3) represents a correction of calculated sample size according to the true size of data category (N).

After choosing a sample size for each data category, the data transparency indicator is calculated as an average of authenticity (A), understandability (U), and reusability (R) for datasets contained in the sample (see Equation (4)). We equally weighted all three components as our conducted research on transparency definition and its meaning showed that they are all equally important for building transparency.

$$DT = \frac{A + U + R}{3} \quad (4)$$

Authenticity is a measure of trust in data publishers and the accuracy of data itself, and it could be seen as a complex feature comprised of two parameters: Data Sources (DS) and Data Accuracy and Integrity (DAI). Considering that DAI represents a summary value based on data characteristics from a sample subset and that, by nature, it is a measure of trust in the precision and credibility of published data, while DS is a global view on data providers by which we can obtain insight into the reliability of data providers, we have decided that DS should constitute 40%, and DAI 60%, of the Authenticity indicator (Equation (5)).

$$A = 0.4 \times DS + 0.6 \times DAI \quad (5)$$

Data sources are governmental and non-governmental authorities, institutions, and agencies that provide raw data for publishing on an open data portal. For users to utilize data safely and without prejudice and trust in published data integrity, it is necessary for publishers to be well-known and have a good reputation. Most of the responsibility for this requirement is on governmental authorities and we define it through the maintenance of a list of data sources available on the data portal ($F1$), the possibility of reviewing datasets published by a particular data source ($F2$), and the existence of data source description ($F3$). Users are another important source of feedback through grading data sources ($F4$) based on their experience with data published by the source. Information provided by the government ($F1$, $F2$, and $F3$) makes up 80% of the DS value, while user-provided feedback is involved with 20%. Features marked as $F1$ and $F2$ are considered more influential on building DS parameters, as they represent high-frequency user requests, while features $F3$ and $F4$ are considered to have less influence. $F1$, $F2$, and $F3$ are scored 1 if satisfied and 0 if unsatisfied, while $F4$ presents an average grade scaled to a range (0, 1). The formula for DS calculation is given in Equation (6).

$$DS = 0.3 \times F1 + 0.3 \times F2 + 0.2 \times F3 + 0.2 \times F4 \quad (6)$$

DS takes its values from the range (0, 1) where 0 signifies that none of the features are present on open data portals, while 1 marks maximum data sources' credibility.

Data Accuracy and Integrity is a parameter that represents a measure of trust in the precision and credibility of published data. To provide a meaningful and reliable evaluation method for calculating such delicate characteristics, three assessment features need to be addressed: government grading (E1), user grading (E2), and quality certification (E3). Government grading relates to government-provided feedback on data accuracy and integration for each published dataset. The establishment of the evaluation procedure is left to the government itself, while the final feature value is calculated as an average grade of all grades in a sample data subset and scaled to range (0, 1). User grading represents numerable user feedback on published data, usually obtained via online scoring of data. Government and user grading are very important since they represent an opinion, an experience, and a point of view, which is why they are equally scored and together represent 70% of the DAI parameter. The third feature is quality certification, and it simply acknowledges the existence of a certifying document for a dataset that represents an electronic proof of data accuracy and integrity of contained information and is provided by the data publisher. The final value of the DAI feature is calculated according to Equation (7).

$$DAI = 0.35 \times E1 + 0.35 \times E2 + 0.3 \times E3 \quad (7)$$

DAI takes its values from the range (0, 1) where 0 denotes complete data inaccuracy, while 1 signifies maximal satisfaction of all three features.

The understandability feature reflects the existence of comprehensible descriptions for each dataset category, as well as contained raw data. This could be accomplished by publishing textual descriptions of data categories with detailed explanations of contained data types, as well as descriptions of each contained dataset. Understandability is a complex measure comprised of two parameters: Data Categories Description (DCD) and Data Sets Description (DSD), where the first one represents descriptions of data categories, while the second one denotes the existence of a description for each dataset. Understandability is calculated based on DCD and DSD parameters as given in Equation (8). DCD parameter constitutes 40% of the final Understandability value, while stronger influence is given to the DSD feature due to its direct relation to data. The Understandability indicator ranges from 0 to 1, where 1 implies a complete availability of both data categories and datasets descriptions, and 0 implies the lack of descriptions.

$$U = 0.4 \times DCD + 0.6 \times DSD \quad (8)$$

DCD parameter is evaluated against three features: textual description (D1) as existence of textual data description for data category, tags (D2) as the existence of searchable tags associated with data category, and linked information (D3) in a form of links towards additional information regarding data category. All three are considered equally important and are thus equally scored. Each category is separately scored against all three features as $(0.33 \times D1 + 0.33 \times D2 + 0.33 \times D3)$ with a final value from range (0, 1). DCD is then calculated as an average value of all categories' values. Value 0 of the DCD parameter means that not even one data category satisfies any of the three features, while value 1 signifies that all dataset categories have adequate descriptions, tags, and linked information. DSD is calculated against the same three features as DCD, but for each dataset from a chosen sample subset. Each dataset is valued from range (0, 1), following the same equation applied to data categories in the DCD parameter. The final DSD value represents an average value of all datasets. DSD values range from 0 to 1 where 0 implies a total lack of data descriptions, tags, and linked information, while 1 signifies that all datasets are properly described.

Data Reusability (R) is a parameter that refers to providing data in open formats so that a user can search, index, and download data via common tools without any prior knowledge of data structures. We have defined an assessment scale, inspired by Sir Tim Berners-Lee's 5-star model for measuring openness of linked open data [20], by adapting it

to a 4-level model that meets the requirements of estimating data reusability in an open government context.

The calculation of the R parameter is performed against each dataset in the chosen sample subset by determining its reusability level. A dataset is scored according to the level it belongs to and then scaled to the range (0, 1). Reusability's final value is an average value of the dataset's reusability. It ranges from 0 to 1, where 0 represents the lowest data availability level and denotes that no dataset is available online, while the value of 3 represents that all datasets are not only available but also maximally reusable in the open data context.

Keeping in mind the complexity of sub-indicators, we present the final equation for the data transparency indicator as the average data transparency for all categories of open data:

$$DT = \text{AVG}\left(\frac{0.4 \times DS + 0.6 \times DAI + 0.4 \times DCD + 0.6 \times DSD + R}{3}\right) \quad (9)$$

Data transparency values are in range (0, 1) or (0–100%) and spread over five transparency levels, as illustrated in Figure 1. Each level represents a range of min-max values of authenticity, understandability, and reusability with equal participation. To explain this further, let us look at an example for level 2. When all three sub-indicators (A, U, R) have values ranging from 0.036 to 0.759, then, looking at Equation (3), data transparency is in the range of 36–75.9%. The lowest level in the data transparency scale, named cradle, represents either a complete lack of transparency or the beginnings of the transparency initiative. As a government advances in opening and enriching open data, it will rise higher on a transparency scale through basic, average, and transparency levels up until the final (high transparency) level.

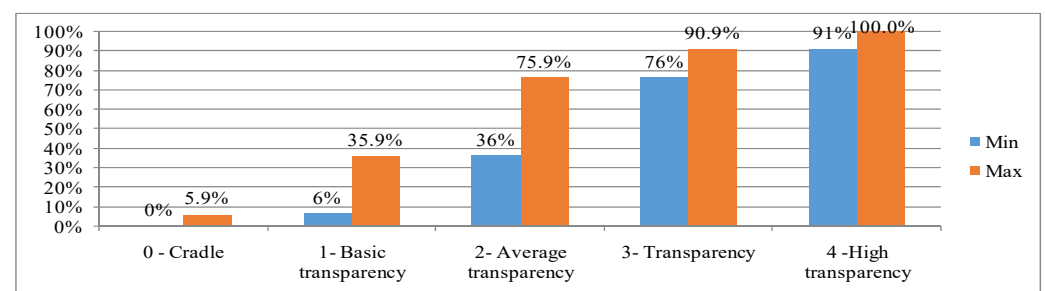


Figure 1. Data transparency (DT) scale range.

3.3. Government Transparency (GT) Indicator

For government transparency, we adopted two mentioned indicators, RTI and CPI. We used both for comparison purposes. It is interesting that RTI acts as a static measure since it only changes when a new FOI law is enacted, while CPI is more dynamic and changes each year. Since GT and DT are included in the overall transparency indicator value with an equal percentage, we had to scale RTI and CPI to range (0, 1) since transparency final value should be in that range. Using RTI we calculated GT as presented in Equation (10), while the CPI-dependent calculation is given in Equation (11).

$$GT = \frac{RTI}{150} \quad (10)$$

$$GT = \frac{CPI}{100} \quad (11)$$

4. Use Case: Transparency Evaluation

In the beginning, open data initiatives were heterogeneous in nature, open licenses differed between initiatives, and there were still heterogeneous formats and a lack of metadata [38,39]. Datasets are often published in a format defined by the system from

which the data originate but without the syntax, semantics, and context, i.e., without the crucial elements which make data more usable [40]. Openly publishing datasets should include syntax (structure), semantics (understanding), and context (metadata) to enable data reusability. Luckily, more recently open-source data portals have begun to offer a standardized system for data publishing, viewing, and retrieval. Comprehensive Active Knowledge Network (CKAN) developed an open-source software framework built on open standards which ensures that published data are compatible with other such portals. CKAN uses Data Catalog Vocabulary (DCAT), which is a standard vocabulary for describing datasets in data catalogues. It also exposes API for developers, to allow retrieving datasets and their metadata properties based on DCAT.

For the practical evaluation of the transparency indicator, we have chosen 22 open government data portals that publish their open data using the CKAN platform. These portals were selected based on the software platform they use for the back-end since the CKAN offers the ability to query the portal's data using a defined API. All the calculations for the data transparency (DT) feature are done based on the API call results, i.e., open data JSON representation. For the government transparency (GT) we took two indexes: RTI and CPI. Table 1 presents results for the Transparency indicator (T) of the OpenGovB benchmark. In the table, we presented results for the indicators: data transparency, government transparency using RTI, government transparency using CPI, and final transparency when using RTI or CPI. Since RTI and CPI indexes have different maximal values, 150 and 100 respectively, to have comparable results we used scaled values in the calculation. As can be seen from Table 1, the first five countries have the best results when scoring by T-RTI. Finland also has an excellent score for T-CPI and DT values just above the countries' average. Italy and Russia are best in DT with a score of 0.39 but, as can be observed in Figure 1, they are only at the average data transparency level. Serbia has the best score in GT-RTI, which resulted in a better overall rank, but looking at the data transparency only it is very low: two times lower than the DT average.

Table 1. Transparency indicator (T) results.

Country	Portal	DT	GT-RTI	GT-CPI	T-RTI	T-CPI
Brazil	http://dados.gov.br/api	0.33	108	40	0.5	0.4
Finland	http://www.hri.fi/api	0.28	105	89	0.5	0.6
Iceland	http://opingogn.is/api	0.35	92	78	0.5	0.6
Russia	http://hubofdata.ru/api	0.39	98	29	0.5	0.3
Serbia	http://rs.ckan.net/api	0.12	135	42	0.5	0.3
Australia	https://data.qld.gov.au	0.33	83	79	0.4	0.6
Czech Republic	http://cz.ckan.net/api	0.4	72	55	0.4	0.5
Hungary	http://opendata.hu/api	0.31	87	48	0.4	0.4
Ireland	http://data.gov.ie/api	0.36	64	73	0.4	0.5
Italy	http://www.opendatahub.it/api	0.39	57	47	0.4	0.4
Slovakia	http://data.gov.sk/api	0.38	70	51	0.4	0.4
Sweden	http://oppnaddata.se/en/api	0.17	92	88	0.4	0.5
Switzerland	http://opendata.admin.ch/api	0.38	77	86	0.4	0.6
United States	https://catalog.data.gov/api	0.19	89	74	0.4	0.5
Austria	http://www.data.gv.at/api	0.36	37	75	0.3	0.6
Belgium	http://portal.openbelgium.be/api	0.27	59	77	0.3	0.5
Canada	http://data.gc.ca/data/api	0.1	79	82	0.3	0.5
Germany	https://www.govdata.de/ckan/api	0.22	52	81	0.3	0.5
Greece	http://ckan.okfn.gr/api	0.2	65	44	0.3	0.3
Romania	http://data.gov.ro/api	0.1	83	48	0.3	0.3
United Kingdom	http://data.gov.uk/api	0	99	81	0.3	0.4
Africa	http://africaopendata.org/api	0.37	0	0	0.19	0.19

To demonstrate the process of the data transparency calculation (Figure 2), we will use the Australian open data portal as an example. It has a high value for data transparency, and it also has high rankings in the CPI and RTI indexes.

Step 1: Identify categories from the Basic Data Set.

On the Australian open data portal all BDS categories were present and therefore we applied data transparency calculation for each category separately.

Step 2: Calculate the sample size of datasets to be evaluated for each category.

For the identified categories, we calculated the total number of datasets in category (N) and a sample size (n) for each category (Table 2).

Step 3: Query random datasets from each category to the number of the sample size and calculate data transparency sub-indicators.

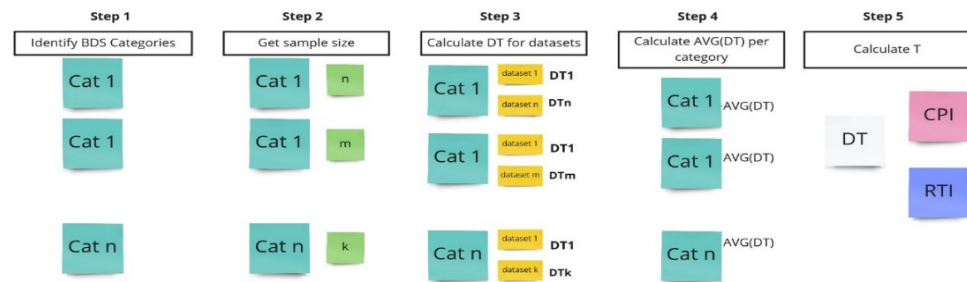


Figure 2. Steps for calculating the transparency indicator.

Table 2. Sample size calculation.

BDS Category	Tag	N	n
Finance and Economy	<i>economy</i>	9	8
	<i>finance</i>	46	28
Environment	<i>climate</i>	6	6
	<i>environment</i>	82	37
Health	<i>health</i>	44	27
Energy	<i>energy</i>	21	16
Education	<i>education</i>	76	36
Transport	<i>transportation</i>	3	3
	<i>transport</i>	75	36
Infrastructure	<i>infrastructure</i>	21	16
Employment	<i>employment</i>	17	14
Population	<i>population</i>	15	12

Authenticity sub-indicator is not dependent on individual datasets and therefore it can be calculated based on API calls that return data on organizations that publish data on open data portals. For calculating Authenticity, we used the API call: http://data.gov.au/api/3/action/organization_list which returns a list of all organizations that publish data on the Australian open data portal. There were exactly 288 organizations in total and for each one we examined required subcomponents of Authenticity. The first four sub-indicators are related to data sources (F1–F4) and the other three (E1–E3) on the data themselves. F2 and F4 describe the possibility for users to rate data sources portal and can provide feedback on the assumed data, but this information could not be obtained through the CKAN API calls, and thus the sub-indicators received score 0. F1 sub-indicator relates to the data sources list and it received score 1 since there was a list of organizations available on the open data portal. F3 sub-indicator, relating to the existence of the meta-data description of the data source, receives a value 1 if there is a value for the [description] and [is_organization] meta-tags; otherwise, the obtained value is 0. This is evaluated for each data-publishing organization separately and then a final F3 value is taken as the average.

To get the score for E1 and E2 (Table 3), it was not possible to use existing data on the portal, since the information about user grading and government grading could not be obtained; therefore, they received value 0. On the other hand, the E3 sub-indicator can be viewed at the organizational level so that if an organization has certification on data

quality it will mean that published data are also certified. To score E3 on the organizational level we examined values for meta-data tag [state] on value “active”, [approval_status] on value “approved” and if [is_organization] is set to “true”. If all three conditions are met, E3 received score 1 for that organization. In the future, upon the availability of the possibility to query and obtain information on data certification, this calculation will be changed so that it is applied to the data inside each data category within the organization.

Table 3. Authenticity calculation (A).

DS	F1	Existence of data sources' list	1
	F2	Grading data sources by users	0
	F3	[description] && [is_organization]	AVG(F3_Organization ₁₋₂₈₈)
	F4	Grading data by users	0
DS_Total		0.48	
DAI	E1	Government grading	0
	E2	User grading	0
	E3	[state]==active && [is_organization]==true && [approval_status]==approved	AVG(E3_Organization ₁₋₂₈₈)
	DAI_Total		0.3
A	$(DS_Total + AVG(DAI_Total_Organization_{1-N}))/2$		

In Table 4 we presented a calculation for the Understandability indicator in the light of the two sub-indicators, DCD—related to categories and DSD—related to datasets. Understandability of data sets is analyzed by looking for defined properties of random datasets within the sample, while data categories' understandability analyzes properties of categories. If [description] property has value, D1 is scored 1, otherwise 0. For D2 we were looking for the value of [tags] property, and D2 has scored 1 or 0 accordingly. Finally, for D3 both [extras] and [links] properties were examined and, if both had values, D3 received score 1, otherwise score 0. Calculating values for DCD and DSD uses scoring of D1–D3 as described. For the Australian portal, the understandability of data categories received a value of 0 since there were no values for D1–D3 sub-indicators. Understandability of datasets is an average value of understandability for all analyzed data, and it has the value of 0.2.

Table 4. Understandability calculation (U).

DCD	D1	[description]
	D2	[tags]
	D3	[extras] && [extras]=>[links]
	DCD_Total	AVG(DCD_Category ₁₋₉)
DSD	D1	[description]
	D2	[tags]
	D3	[extras] && [extras]=>[links]
	DSD_Total	AVG(DSD_Category ₁₋₉)
U	$U = 0.4 \times DCD_Total + 0.6 \times DSD_Total = 0 + 0.6 \times 0.33 = 0.2$	

For example, the Climate BDS category has the results for DSD as given in Table 5. To obtain these results, we executed API call <https://data.qld.gov.au/api/action/tag/climate> to find out that there are six datasets tagged with the word climate, and then we requested each dataset and checked their properties [description], [tags], [extras], and [extras[links]]. We obtained a score of 0.33 for this category.

Table 5. DCD results for climate category.

Tag	Data Set	D1	D2	D3	DSD
Climate	acorn-sat	0	1	0	0.33
	air-conditioner-location-running-hours-data	0	1	0	0.33
	australian-solar-energy-information-system-aseis-v3-0	0	1	0	0.33
	economics-of-australian-soil-conditions-199697-factor-most-limiting-yield-aciditysodicitysalinity	0	1	0	0.33
	waite-arboretum-spatial-data	0	1	0	0.33
	waite-arboretum-catalogue	0	1	0	0.33
	DSD Climate			0.33	

Reusability is calculated on the dataset level. For the sample size of each dataset category, we chose random datasets. As datasets can have many resources, we randomly picked their sample and checked the formats. Each format was scored as given in Table 6, and the average level was determined for the dataset.

Table 6. Reusability calculation (R).

R	level 0	No Format
	level 1	machine-readable structure semantic
	level 2	structure && non-proprietary
	level 3	semantic && non-proprietary
	R_Category	AVG(R_Category_resources)
R	AVG(R_Category₁₋₉) = 0.4	

Data resources in the Australian portal are made available in a variety of formats (Figure 3). Approximately 63% of the analyzed resources are in XLS and XLSX formats, and a large number of resources are available in other machine-readable formats, most commonly CSV, XML, KML, and JSON. Only a small part (up to 4%) of resources are published in non-reusable formats such as PDF, plain text, or images, which is a direct barrier for further use.

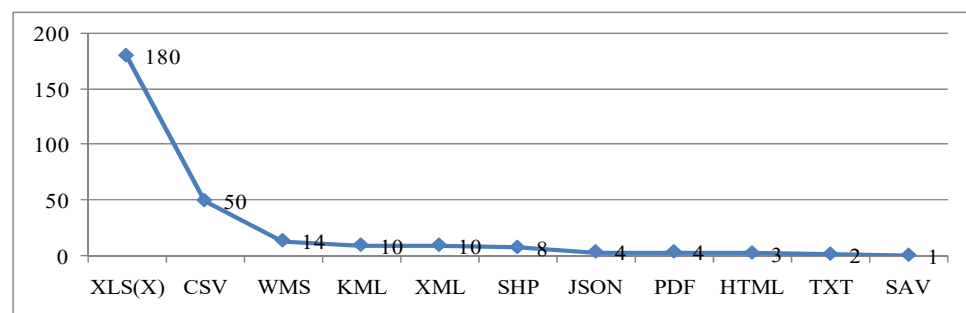


Figure 3. Formats of resources in sampled datasets.

Step 4: Calculate average data transparency for each category and for the portal.

For the Australian open data portal, we obtained the results for data understandability and reusability for each data category and the authenticity indicator of the portal. Then we could calculate data transparency for each data category, as presented in Table 7, and DT for the portal by applying Equation (9).

Table 7. Data transparency (DT) calculation.

Category	<i>U</i>	<i>R</i>	<i>A</i>	<i>DT</i>		
				0.56		
				0.79		
				0.8		
				0.77		
<i>economy</i>	0.20	1.11	0.37	0.67		
				0.84		
				0.58		
				0.64		
				0.77	<i>RTI</i>	<i>CPI</i>
<i>environment</i>	0.20	1.39	0.37	0.79		
<i>health</i>	0.20	1.41	0.37	0.8		
<i>energy</i>	0.20	1.34	0.37	0.77		
<i>education</i>	0.20	1.14	0.37	0.67		
		1.48				
<i>transportation</i>	0.20	1.35	0.37	0.84		
		0.97				
<i>infrastructure</i>	0.20	0.97	0.37	0.58		
<i>employment</i>	0.20	1.08	0.37	0.64		
<i>population</i>	0.20	1.33	0.37	0.77		
<i>AVG</i>	0.2	0.42	0.37	0.33	83/150 = 0.56	79/100 = 0.79
Transparency-RTI					0.4	
Transparency-CPI					0.6	

Step 5: Use the existing transparency index (RTI or CPI) and calculate the Transparency indicator.

Australia has scored 83 for RTI and 79 for CPI, using the values for DT sub-indicators and applying Equation (1), we receive 0.4 and 0.6 as transparency indicator values respectively.

The steps-based explanation serves as a guide for anyone who wants to try to calculate transparency. We automated the calculation process by developing a web-based application for calculating OpenGovB indicators, including transparency. The calculation is initiated upon entering the URL of the open data portal and represents the current indicator value that depends on the randomly chosen datasets within the sample.

What have we learned from assessing data transparency of 22 open government portals? Who can benefit from this study, and how do they benefit?

Currently, all portals have their weaknesses, which mainly originate in poor data quality. Reasons for producing poor data quality are varied, starting from insufficient budget, lack of internal manual resources, lack of relevant technology, inadequate data strategy, human error, lack of internal communication between departments, or inadequate management support. By improving data quality, bars would raise for all transparency sub-indicators. By adding missing dataset descriptions, category descriptions, and responsible organizations that maintain data, and by linking relevant datasets, tagging categories, and datasets, governments would significantly improve the score for the Understandability indicator. Exposing the list of trusted data sources and enabling open-source modules that would allow user grading of data sources and datasets could improve Authenticity indicator scores. To gain better results for the Reusability indicator, readability of resources should be improved, meaning that open data formats need to prevail over proprietary data formats. That would be the prerequisite for reusing the data, building new datasets, services, and solutions.

The transparency evaluation results should be considered by governments to improve their own open data policies, and in so doing improve government transparency and users' trust. The results are also a good starting point for the researchers who want to build new transparency models or suggest future improvements for the current ones. It is

necessary to monitor improvement on certain model features and scale weights accordingly as governments progress to a more mature transparency model.

5. Conclusions

Transparency in open government is a relatively new concept that has arisen from the previous e-government models and initiatives. It has proven to be one of the most significant features of open government and a concept every government strives towards. The idea of building a transparent government in the context of openness towards citizens has quickly spread from Obama's office to governments around the globe. Today, a great number of countries emphasize transparency as a major milestone in their smart-city strategies and action plans. Transparency in open government makes steps forward towards the usage of data in a meaningful way. When there are no obstacles for data transparency, there are no obstacles for data consumption. A smart city vision can gain value from monitoring transparency and our transparency evaluation model can be the right tool for this.

The model for evaluating transparency that was proposed here offers an assessment method for determining the level of progression based on the most critical open data concerns: data accuracy and integrity, data quality, data sources' credibility, data clarity, and reusability, which we consider a significant advantage of such an approach. Numerous sources in the literature that already dealt with benchmarking of open government data support this claim [2,27,41,42]. They all directly relate to open data, making the model completely applicable for open government. The other part of the model is government transparency, for which we decided to include some of the existing indicators. At the same time, this could present a drawback, since the model addresses only some of the transparency features, while flexibility and comparison of transparency are overshadowed by different features. This we see as the challenge that we need to overcome in the following period to strengthen the model even more and make it more trustworthy. We applied the model on 22 open data portals that are built on the CKAN platform to demonstrate the assessment possibilities, and in the future we plan to enable support for other data platforms as well. This is especially important in order to obtain comprehensive overview of open data portals powered by other platforms (DKAN, Udata, Socrata, etc.). Increasing the number of surveyed portals will also contribute to the validity of this study.

Future research aspects should be focused on the development of transparency indicators related to back-office experience and efforts. Governmental representatives need to be actively involved in evaluating transparency, as they could provide valuable information regarding legislative background, initiatives, and strategies, which are all important parts of a transparent government. Although the Authenticity indicator does involve government feedback in the evaluation process, the model needs to be further enriched with back domain information that would increase the administration's participation in the final assessment. This could be achieved through thoroughly designed questionnaires that would be filled in by government representatives. Consequently, this will provide an adequate environment for the comparison of those indicators with our benchmark model and its indicators from which unambiguous conclusions can be drawn about how the problem of transparency in open government should be observed. Comparative analysis of qualitative and quantitative indicators of transparency in the government domain also can contribute to the future establishment of a standard evaluation model for transparency in open government.

Author Contributions: Conceptualization, N.V. and P.M.; methodology, N.V.; validation, L.S.; formal analysis, P.M.; investigation, P.M. and N.V.; writing—original draft preparation, N.V. and P.M.; writing—review and editing, N.V. and P.M.; supervision, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting reported results can be found on the links provided in Table 1 of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- McDermott, P. Building Open Government. *Gov. Inf. Q.* **2010**, *27*, 401–413. [CrossRef]
- Veljković, N.; Bogdanović-Dinić, S.; Stoimenov, L. Benchmarking open government: An open data perspective. *Gov. Inf. Q.* **2014**, *31*, 278–290. [CrossRef]
- Obama, B. *Memorandum for the Heads of Executive Departments and Agencies: Open Government Directive*; White House: Washington, DC, USA, 2009.
- Veljković, N.; Bogdanović-Dinić, S.; Stoimenov, L. Municipal Open Data Catalogues. In Proceedings of the CEDEM, Krems au Donau, Austria, 5–6 May 2011; pp. 195–207.
- Wirtz, B.W.; Weyerer, J.C.; Rösch, M. Open government and citizen participation: An empirical analysis of citizen expectancy towards open government data. *Int. Rev. Adm. Sci.* **2019**, *85*, 566–586. [CrossRef]
- Harrison, T.M.; Guerrero, S.; Burke, G.B.; Cook, M.; Cresswell, A.; Helbig, N.; Hrdinová, J.; Pardo, T. Open Government and E-Government: Democratic Challenges from a Public Value Perspective. In Proceedings of the 12th Annual International Conference on Digital Government Research: Digital Government Innovation in Challenging Times, College Park, MD, USA, 12–15 June 2011; pp. 245–253.
- Jaeger, P.; Bertot, J.C. Transparency and technological change: Ensuring equal and sustained public access to government information. *Gov. Inf. Q.* **2010**, *27*, 371–376. [CrossRef]
- Milić, P.; Veljković, N.; Stoimenov, L. Comparative analysis of metadata models on e-government open data platforms. *IEEE Trans. Emerg. Top. Comput.* **2021**, *9*, 119–130. [CrossRef]
- Stiglitz, J. *Transparency in Government. The Right to Tell, WBI Development Studies*; The World Bank: Washington, DC, USA, 2002.
- Tauberer, J. Open Government Data. 2012. Available online: <https://opengovdata.io> (accessed on 1 December 2021).
- Veljković, N.; Bogdanović-Dinić, S.; Stoimenov, L. eGovernment Openness Index. In Proceedings of the ECEG, Ljubljana, Slovenia, 16–17 June 2011; pp. 571–577.
- United Nations. *The Millennium Development Goals Report*; United Nations: New York, NY, USA, 2011.
- Hunter, D.; Jupp, V. E-Government Leadership. Rhetoric vs. Reality—Closing the Gap. 2001. Available online: www.epractice.eu/files/media/media_846.pdf (accessed on 1 December 2021).
- Capgemini. *Summary Report: Web-Based Survey on Electronic Public Services*; European Commission: Paris, France, 2001.
- Capgemini. *Method Paper: Preparing the 9th Benchmark Measurement*; European Commission: Paris, France, 2010.
- Capgemini. *The User Challenge: Benchmarking the Supply of Online Public Services*; European Commission: Paris, France, 2007.
- Matheus, R.; Janssen, M. A systematic literature study to unravel transparency enabled by open government data: The window theory. *Public Perform. Manag. Rev.* **2020**, *43*, 503–534. [CrossRef]
- Lnenicka, M.; Nikiforova, A. Transparency-by-design: What is the role of open data portals? *Telemat. Inform.* **2021**, *61*, 101605. [CrossRef]
- Osimo, D. Benchmarking e-government in the Web 2.0 era: What to measure and how. *Eur. J. Epractice* **2008**, *2*, 33–43.
- Berners-Lee, T. Linked Data. 2010. Available online: <http://www.w3.org/DesignIssues/LinkedData.html> (accessed on 1 December 2021).
- Meijer, A. Government transparency in historical perspective: From the ancient regime to open data in the Netherlands. *Int. J. Public Adm.* **2015**, *38*, 189–199. [CrossRef]
- Hood, C.; Heald, D. *Transparency: The Key to Better Governance*; Oxford University Press for The British Academy: Oxford, UK, 2006; p. 135.
- Erkkilä, T. *Government Transparency: Impacts and Unintended Consequences*; Palgrave Macmillan: London, UK, 2012.
- LoMonte, F.D. Casualties of a pandemic: Truth, trust and transparency. *J. Civ. Inf.* **2020**, *2*, iii–viii. [CrossRef]
- Cucciniello, M.; Porumbescu, G.A.; Grimmelikhuisen, S. 25 Years of Transparency Research: Evidence and Future Directions. *Public Adm. Rev.* **2017**, *77*, 32–44. [CrossRef]
- Meijer, A.; Curtin, D.; Hillebrandt, M. Open Government: Connecting Vision and Voice. *Int. Rev. Adm. Sci.* **2012**, *78*, 10–29. [CrossRef]
- Lourenço, R.P. An analysis of open government portals: A perspective of transparency for accountability. *Gov. Inf. Q.* **2015**, *32*, 323–332. [CrossRef]
- Sayogo, D.S.; Pardo, T.A.; Cook, M. A framework for benchmarking open government data efforts. In Proceedings of the 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 1896–1905.
- Calland, R.; Bentley, K. The impact and effectiveness of transparency and accountability initiatives: Freedom of information. *Dev. Policy Rev.* **2013**, *31*, 69–87. [CrossRef]
- Noveck, B.S. Rights-based and tech-driven: Open data, freedom of information, and the future of government transparency. *Yale Hum. Rts. Dev. LJ* **2017**, *19*, 1.

31. Mabillard, V.; Pasquier, M. Transparency and Trust in Government (2007–2014): A Comparative Study. *NISPAcee J. Public Adm. Policy* **2016**, *9*, 69–92. [[CrossRef](#)]
32. González-Gallego, N.; Nieto-Torrejón, L. Government Data Openness and Coverage. How do They Affect Trust in European Countries? *J. Data Inf. Sci.* **2021**, *6*, 139–153. [[CrossRef](#)]
33. Robinson, S.C. Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technol. Soc.* **2020**, *63*, 101421. [[CrossRef](#)]
34. Ramkumar, V.; de Renzio, P. *Improving Budget Transparency and Accountability in Aid Dependent Countries: How Can Donors Help?* IBP Budget Brief No. 7; International Budget Partnership: Washington, DC, USA, 2009.
35. Rohwer, A. Measuring Corruption: A Comparison between the Transparency International's Corruption Perceptions Index and the World Bank's Worldwide Governance Indicators. *CESifo DICE Rep.* **2009**, *7*, 42–52.
36. NIST/SEMATECH e-Handbook of Statistical Methods. 2020. Available online: <https://www.itl.nist.gov/div898/handbook/pri/section3/pri33a.htm> (accessed on 1 December 2021).
37. Martin, S.; Foulonneau, M.; Turki, S.; Ihadjadene, M. Risk Analysis to Overcome Barriers to Open Data. *Electron. J. e-Gov.* **2014**, *11*, 348–359.
38. Underhill, L.; Bradfield, D. *Introstat*; Juta and Company Ltd.: Cape Town, South Africa, 2011.
39. Petychakis, M.; Vasileiou, O.; Georgis, C.; Mouzakis, S.; Psarras, J. A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. *J. Theor. Appl. Electron. Commer. Res.* **2014**, *9*, 34–47. [[CrossRef](#)]
40. Bountouri, L.P.; Christos, C.; Soulikias, V.; Stratis, M. Metadata interoperability in public sector information. *J. Inf. Sci.* **2009**, *35*, 204–231.
41. Douglas, S.; Meijer, A. Transparency and public value—Analyzing the transparency practices and value creation of public utilities. *Int. J. Public Adm.* **2016**, *39*, 940–951. [[CrossRef](#)]
42. Vetrò, A.; Canova, L.; Torchiano, M.; Minotas, C.O.; Iemma, R.; Morando, F. Open data quality measurement framework: Definition and application to Open Government Data. *Gov. Inf. Q.* **2016**, *33*, 325–337. [[CrossRef](#)]