

# Government Open Data Portals: A Measurement of Data Veracity Coverage



Nor A. M. Sabri, Nurul A. Emran, Norharyati Harum

**Abstract:** *Open data initiative has been adopted by many countries worldwide due to the need for establishing agile open government and knowledge-based economy. As a result, we can witness an increasing amount of government open data shared on public government's portals that become sources of rich big data. While this scenario provides data transparency and eases accessibility for public data consumers, the quality aspect, or the veracity property (as commonly coined to big data) of open data is the topic of concern. Not only poor quality data causes misleading results, the reputation of the government as an open data provider can also be negatively affected. Thus, to understand how the government's portals deal with the veracity aspect of their data, in this paper, we present the results of examining quality criteria imposed by selected government's data portals for their data contributors. In particular, we extract quality criteria from the open data policy of the government's data portals under study. The result shows that out of 108 portals, only 27% of the portals explicitly state their quality criteria in the policy, with varying coverage of quality criteria. The frequency of the identified 15 quality criteria shows the types of quality criteria that receive more (and less) attention by the open data portals based on their relative importance. We conclude with suggestions on the areas of further research and development in the government's open data.*

**Keywords :** *Data Veracity, Quality criteria, Open government data, Open data policy, Open data principles.*

## I. INTRODUCTION

Many governments worldwide are encouraged to be more transparent to their citizens by increasing the amount of government data published on the web through open data portals. As a result, government data sets that are used to be 'hidden' behind the firewalls are now can be easily accessed by anyone. Open government data (OGD) is defined as "any data held by the government that can be reused, and redistributed by anyone, for any purposes, including commercial reuse, free of charge and without any

restrictions" [1]. The establishment of Global Open Data Index<sup>1</sup> website organized by Open Knowledge Network shows that a large volume of government data of countries worldwide are freely used, modified and shared. From this website, countries are ranked based on fifteen categories of government data sets. There are several drivers for data 'openness'.

For Oman Government, OGD initiative is important to create agile open government infrastructure by 2020 [2]. OGD is also needed to transform the government into a sustainable knowledge-based economy towards the year 2020 [3]. Moreover, the internet economy has grown to be an important source to gain and provide data with the potential of boosting the whole economy, innovation, competitiveness and contribution to society. The Seoul Declaration stated that the government is considering a new way to revive the economy after the financial crisis in 2008 by opening the public sector information to generate value from the data shared where the data reused and published freely by anyone [4]. The openness concept is the preliminary instrument of Korea to prevent corruption and mistake as stated in presidential statements for Government 3.0 [5]. Furthermore, in Oman, the government's open data concept sets that the datasets in the database of government must be made accessible publicly and its reusability and be redistribution must be made free-of-charge, with no restrictions for any purposes including the commercial use [3].

The data sets shared by government agencies such as ministries, municipalities, and other authorities (such as national statistics, budgetary information, geographical data, laws and regulations, and education and transport data) are produced, maintained and updated in massive amounts. This massive volumes of data is a source of new knowledge and management resources for public and private sectors [2]. OGD initiative can encourage stronger and interconnected societies to meet the needs of citizens with innovation and prosperity. This is because the open government becomes a platform for maximizing public value. It has been reported that the most widely used sources of open data are government and government-supported institutions (called open government data), rather than social media, companies, and NGOs [5]. Besides, three factors characterize open data which are open access, compatible database format and freedom of reuse. Open access refers to the way of collecting data by everyone without discrimination, compatible database format means that the data sets are accessible in bulk and compatible with many systems.

**Revised Manuscript Received on October 30, 2019.**

\* Correspondence Author

**Nor A. M. Sabri**, Computational Intelligence Technologies (CIT), Center of Advanced Computing Technologies (CACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia (UTeM), 76100 Durian Tunggal, Melaka, Malaysia. Email: noramalinamohdsabri@gmail.com

**Nurul A. Emran**, Computational Intelligence Technologies (CIT), Center of Advanced Computing Technologies (CACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia (UTeM), 76100 Durian Tunggal, Melaka, Malaysia. Email: [nurulakmar@utem.edu.my](mailto:nurulakmar@utem.edu.my)

**Norharyati Harum**, Information Security Forensics and Computer Networking (INFORSNET), Center of Advanced Computing Technologies (CACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia (UTeM), 76100 Durian Tunggal, Melaka, Malaysia. Email: [norharyati@utem.edu.my](mailto:norharyati@utem.edu.my)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

<sup>1</sup> <https://index.okfn.org/>

Meanwhile, freedom of reuse means everyone can use, reuse, redistribute and mix the data free-of-charge without bureaucratic procedures [6].

OGD has been useful in improving many areas namely job creation and raises economic public services, government-citizens transparency, accountability and participation of the citizen in government decision making, and information sharing between the government and citizens [5]. Hiramoto (2013) stated among the objectives of using open data is to enhance transparency and confidence. By raising the transparency of government, corruption can be prevented and the trust level of citizens can be increased. Another objective is to promote participation and collaboration of public and private sectors in fostering creative innovations. This achieved by providing efficient public services and rapid response to the diversification of needs and values. To create economic stimulus in government is also the objective open data. Providing the stimulus to the entire economy can increase the efficiency of operations between national and local governments [2].

There are some issues raised on OGD portals involving metadata, search ability, discoverability, and usability of the data [7]. Furthermore, open data, in general, has been reported to introduce many challenges due to the diverse data features such as size, structure, and encoding [8]. The trend that the shared datasets are published in an ad-hoc manner raises data consumers' concern on its quality. Nevertheless, the existence of OGD policy is a sign of governments' initiative in ensuring the quality of data published in their web portals, where a set of quality criteria (also called as OGD principles) are stated explicitly in the document, along with other information such as the scope and objectives of OGD.

### A. OGD Policy

OGD policy is endorsed by a government as a guideline for the public and also OGD contributors in understanding the objectives, scopes and the criteria (or principles) of OGD in ensuring the quality of the data shared in government's web portals.

For example, OGD policy for 29 OECD (Organisation for Economic Co-operation and Development) countries<sup>2</sup> states four main objectives the policy as 1) to create economic value, 2) to deliver more efficient public services, 3) to increase government openness, and 4) to increase government transparency (OECD Publishing, 2016).

It has been explicitly stated in British Columbia's (a province of Canada) OGD policy that, among the key objectives of its policy are to make government data more accessible to the public, to respond to citizens needs and expectations, to engender a participatory environment in which citizens are engaged with their government, communities and public policy issues, and to enhance trust between citizens and the government (British Columbia CIO Office, 2011).

The non-OECD country like Oman uses the policy as the basis towards the creation of innovative and value-added services for public participation (Sultanate of Oman, n.d.). Moreover, high-quality OGD may encourage its citizen to foster

creativity in developing positive economic and social benefits through innovative applications. Besides, through the policy, Oman made its commitment to open government explicit, where a practical policy framework is designed for the release of high-value datasets to the public.

For New South Wales (Australia), OGD policy aims to assist the government to release more data for business and research conducted by community and industries. Similar to OECD countries, its policy sets to achieve better public services through OGD [9].

Open data portals serve as big data sources that supports data-driven discoveries and software application developments. Thus, open data is characterized with the volume, variety, velocity and veracity properties of big data. The veracity aspect, has been recognized as an important aspect for big data utilization along with the other three properties, but yet, limited attempts can be seen to measure it [10].

While knowing the objectives of OGD stated in the policy is helpful in understanding governments' expectation in realizing open government initiative, in this paper, we are more concern on how the policy can be used to guide data contributors to share quality data that meets the veracity aspect. Quality criteria in OGD policy which are also called as OGD principles represent a government's commitment in fulfilling the veracity property. By setting clear and precise quality criteria, the public's trust in using OGD for related applications (such as developing innovative software applications) can be built.

In the next section, we will see the quality criteria stated in OGD policy of several governments.

### B. OGD Quality Criteria

Tauberer (2014) characterized OGD with four main qualities namely "Accessible", "Accurate", "Analyzable", and "Authentic" (Tauberer, 2014). By consolidating white papers, online documents, blogs on open data, and some online policies, Tauberer breaks these qualities into 14 quality criteria as shown in Table I.

If we consult OGD policies that are available from several governments' open data we can see overlaps between Tauberer's quality criteria and the quality criteria stated in those policies. Nevertheless, for better coverage, the non-overlapping quality criteria are also considered in this paper. The 31 governments' policy (at country and state levels) that are examined to extract OGD quality criteria can be referred in Table II. We can observe from OGD quality criteria presented so far, there are cases where different terms are used to refer the same quality. Therefore, term filtering and combination is conducted before quality coverage can be measured against the government portals under study.

In the next section, similar works that study quality criteria for OGD will be presented. Section III

<sup>2</sup> <http://www.oecd.org/about/membersandpartners/>

**Table I: Quality Criteria of OGD from [11]**

Quality Criteria	Description
Online and free	Data are available on the Internet for free.
Primary	Data are collected at the source, with the finest possible level of granularity, not in aggregate or modified forms
Timely	Data are made available as quickly as necessary to preserve the value of the data.
Accessible	Data are available to the widest range of users for the widest range of purposes
Analysable (Machine-readable)	Data are in a machine-processable format that allows the public to carry out its analyses of raw data, rather than relying on a government's analysis.
Non-proprietary	Data are available in a format over which no entity has exclusive control.
Non-discrimination	Data are available to anyone, with no requirement of registration
License-free	Dissemination of the data is not limited by intellectual property law such as copyright, patents, or trademarks, contractual terms, or other arbitrary restrictions.
Permanent	Data should be made available at a stable Internet location indefinitely.
Safe file formats	Published data should be in formats that do not include executable content.
Provenance and trust	Published data should be digitally signed or include attestation of publication/creation date, authenticity, and integrity.
Public input	The public is allowed to inform data on demand according to their needs.
Public review	The presence of contact person to respond for feedback from the public regarding compliance to OGD quality criteria.
Interagency coordination	Data has interoperability property that enables data sharing among government agencies.

covers the methodology used to measure coverage of quality criteria and their relative importance, followed by Section IV for results and discussion. Finally, Section V concludes this paper.

## II. RELATED WORKS

One of the aims of open government movement is to gain trust and confidence from the public against their government. While this aim seemingly applies to the local citizens, OGD can benefit data consumers worldwide as the data are published in an online web portal that is accessible by anyone with an Internet connection. Information on how a government monitor data contributors' compliance against quality criteria of its OGD policy can be a useful instrument to gain the public's trust. Nevertheless, this may require governments' willingness to disclosing their quality measures, which might be confidential (or debatable). Alternatively, we can measure the extent of quality criteria covered by a particular government, by examining OGD policy published in the government's web portals. Even though knowing the 'coverage' of quality criteria cannot determine actual quality compliance, this measure can be used as a sign of the government's level of commitment and awareness on quality assurance. In addition, it is also important to learn which quality criteria that receive more (or less) emphasis by the governments which can be used as the basis for future improvement.

Quality assessment is usually performed to evaluate compliance against quality standards [12], where its importance has been reported in ensuring data trustworthiness [13]. There are several methods of quality assessment proposed to date involving the quality of data published on the web.

Hartig and Zhao in 2009 proposed assessing the quality of data on the web by using data provenance [13]. Their work focuses on timeliness quality where missing provenance information is addressed by associating certainty values with calculated quality values. In 2012, Mendes, Mühleisen and C. Bizer proposed a framework called as Sieve for quality assessment for linked data where four concepts which are assessment metrics, data quality indicator, scoring function and aggregate metric [14]. Completeness, conciseness, and consistency are the quality criteria considered in the framework, where data from English and Portuguese versions of DBpedia are integrated. In 2016, AlRushaid and Saudagar reported the result of measuring openness level of Saudi Arabia, Taiwan, United Kingdom, Denmark, Colombia, and Finland [15]. To measure openness, a scoring model called as Global Open Data Index is used, where the model is enhanced with more survey questions that cover all areas of OGD maturity model ([16] that is used as a benchmark. A total of 13 quality criteria are considered in this study, where experts' judgment is used to determine the weight of quality criteria.

The final openness score is calculated by aggregating the scores of individual quality criteria for the six countries, which also indicate the extent of quality criteria coverage.

The proposals presented so far provide useful insight into the types of measures used in assessing the quality and the quality criteria of concerned. Nevertheless, these proposals are limited in terms of examining quality criteria' relative importance among the governments. In addition, for openness measure, in particular, the government understudy is limited to six governments only.

**Table II. Government Portals of OGD**

	COUNTRY/ STATE/ CITY	URLS
1.	Australia	<a href="https://data.gov.au/">https://data.gov.au/</a>
	New South Wales	<a href="https://data.nsw.gov.au/">https://data.nsw.gov.au/</a>
2.	Brunei	<a href="https://www.data.gov.bn/">https://www.data.gov.bn/</a>
3.	Canada	<a href="https://open.canada.ca/en">https://open.canada.ca/en</a>
	British Columbia	<a href="https://data.gov.bc.ca/">https://data.gov.bc.ca/</a>
	Montreal	<a href="http://donnees.ville.montreal.qc.ca/">http://donnees.ville.montreal.qc.ca/</a>
	New Brunswick	<a href="https://open.canada.ca/data/en/dataset/38f09431-f26b-5afb-811e-fed0a912ff7d">https://open.canada.ca/data/en/dataset/38f09431-f26b-5afb-811e-fed0a912ff7d</a>
4.	Denmark	<a href="http://www.opendata.dk/">www.opendata.dk/</a>
5.	France	<a href="http://www.data.gouv.fr/fr/">http://www.data.gouv.fr/fr/</a>
6.	Germany	<a href="https://www.govdata.de/">https://www.govdata.de/</a>
7.	India	<a href="https://data.gov.in/">https://data.gov.in/</a>
8.	Indonesia	<a href="https://data.go.id/">https://data.go.id/</a>
9.	Italy	<a href="https://www.dati.gov.it/">https://www.dati.gov.it/</a>
10.	Japan	<a href="https://open-data.pref.shizuoka.jp/">https://open-data.pref.shizuoka.jp/</a>
11.	Malaysia	<a href="http://www.data.gov.my/">http://www.data.gov.my/</a>
12.	New Zealand	<a href="https://www.data.govt.nz/">https://www.data.govt.nz/</a>
13.	Norway	<a href="http://data.norge.no/data">http://data.norge.no/data</a>
14.	Oman	<a href="http://www.oman.om/wps/portal/index/opendata">http://www.oman.om/wps/portal/index/opendata</a>
15.	Philippines	<a href="http://data.gov.ph/">http://data.gov.ph/</a>
16.	Russia	<a href="https://opengovdata.ru/">https://opengovdata.ru/</a>
17.	Singapore	<a href="https://data.gov.sg/">https://data.gov.sg/</a>
18.	South Africa	<a href="http://data.gov.za/terms-of-use.html">http://data.gov.za/terms-of-use.html</a>
19.	South Korea	<a href="http://data.seoul.go.kr/">http://data.seoul.go.kr/</a>
20.	Spain	<a href="http://datos.gob.es/en/documentacion">http://datos.gob.es/en/documentacion</a>

Policy readiness is one of the criteria measured by Open Data Barometer (ODB) organization where fifteen kinds of data that belongs to selected governments countries are evaluated. [17]. These countries are included based on their commitments to adopt the International Open Data Charter Principles<sup>3</sup> or the G20 Anti-Corruption Open Data Principles. The countries are measured (based on survey results) on their readiness to share data that fulfill ten open data principles (quality criteria) as follows:

1. Completeness

<sup>3</sup> List of open data principles in Open Data Charter: <https://opendatacharter.net/principles/>

2. Primacy
3. Timeliness
4. Ease of Physical and Electronic Access
5. Machine readability
6. Non-discrimination
7. Use of Commonly Owned Standards
8. Licensing
9. Permanence
10. Usage Costs

Policy readiness in the evaluation conducted by ODB organization focuses on the extent of a well-defined open data policy and/or strategy in a country. It also indicates the consistency of open data management and publication. Nevertheless, it does not measure the coverage of quality criteria stated in the policy. In the next section, the methodology used to measure OGD quality criteria relative importance and their coverage will be presented.

### III. METHODOLOGY

Fig.1 shows the stages of the methodology adopted in examining OGD quality criteria in this study. The first stage is the data extraction stage. In this stage, OGD portals are identified. Initially, there are 108 OGD portals identified from the U.S open data portal<sup>4</sup>. These government portals are organized into the country, state and city levels. Open data policies are downloaded from these portals for offline quality criteria extraction. In this stage, we discovered that 31 portals (of 25 countries) publish their OGD policy online as shown in Table II. Then, we extract quality criteria from the policies. In total there are 36 raw quality criteria extracted. After filtering synonyms and related terms, the total quality criteria under study is reduced to 15, as shown in the Appendix.

The next stage is to measure the frequency of quality criteria that appear in the policies. A qualitative software called as QSR NVivo10 is used in this step. To begin the analysis, the software needs the set of government portals as the input. Another input needed is the set of quality criteria identified in the previous step, where these quality criteria are defined as 'nodes' in the software. In this study, 15 nodes are used and assigned to their corresponding portals. Each node will be assigned to at least one portal. Each portal will have a set of nodes assigned to it.

The frequency scores are computed in percentage by the software. For example, if the frequency score for a quality criteria is 100% it means that particular criteria is stated in all countries' policy.

In the final stage, we measure the coverage of quality criteria of the government portals under study. A simple coverage measure (similar in (Emran, 2015)) is used. The following are descriptions of the notations.

Suppose that:

- P is a set of OGD portals under the measure, where n is the total number of portals,
- P<sub>i</sub> is the i<sup>th</sup> portal in P,
- QC<sub>i</sub> is the set of quality criteria of P<sub>i</sub>
- UQC is the union of quality requirement sets of n portals,

<sup>4</sup> <http://www.data.gov/open-gov>



Then, coverage of  $P_i = |QC_i| \cap |UQC| / |UQC|$ .

An average value is taken for countries that have multiple open data portals such as Canada, United States and Australia.

We use the report from ODB as a benchmark to evaluate the results. In particular, assessment report in 2016 of Policy Readiness is taken as it covers almost all countries included in this study.

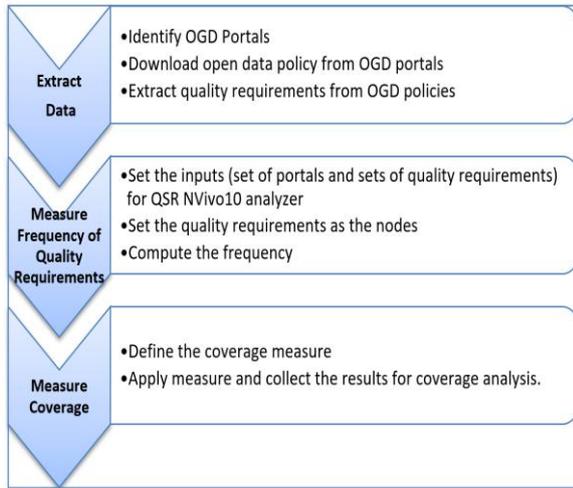


Fig. 1: Stages of OGD Quality Criteria Measurement

IV. RESULTS AND DISCUSSION

The results of the study can be divided into two parts: The frequency (the number of times) the quality criteria can be found in the countries’ open data policy, and the coverage of quality criteria by countries.

Fig. 2 shows the frequency results. X axis shows the frequency of the quality criteria that appears in Government open data policy in percentage, while Y axis represents the quality criteria under measure. These quality criteria overlaps with all quality criteria considered by ODB organization as stated in Section 2 except for “Usage Cost” criteria. The quality criteria that can be found in the policy but not covered by ODB are “Public”, “Reusable”, Manageable” “Trusted” “Protected” and “Non-proprietary”. This indicates that all countries under measure are in line with the main quality criteria set by ODB organization. With additional quality criteria added in the policy, these countries introduce additional requirements that need to be followed by data contributors.

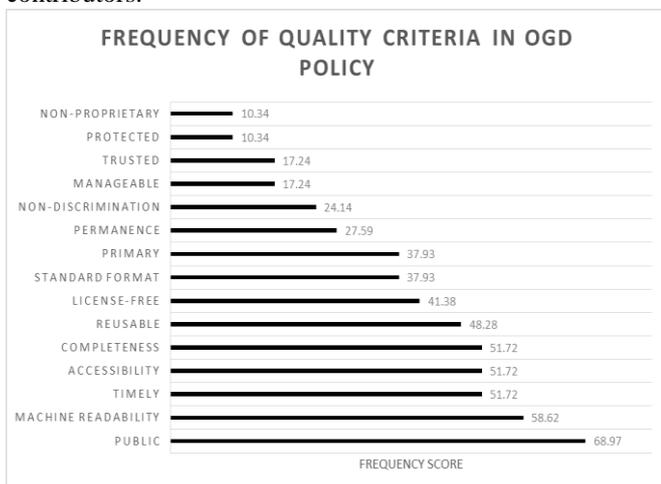


Fig. 2: Relative Importance of OGD Quality Criteria

The result shows that, the most common quality criteria that can be found in the policy is “Public” with percentage value of 68.97%, followed by “Machine Readability” (58.62%). “Timely”, “Accessibility”, and “Completeness” criteria have the same score of 51.72%. Quality criteria that receives the least attention are “Protected” and “Non-proprietary” with 10.34%. The low score for these two quality criteria can be explained by the absence of these criteria in the ODB quality list. Nevertheless, even though “Public” does not exist in the ODB quality list, it becomes a popular criteria as this quality criteria is by definition the inherent characteristic of all data sets contributed by the open government countries audited by ODB. All countries that are evaluated by ODB must comply the “Open” criteria set by the organization [17]. On average, the score of the frequency is 37, with standard deviation of 18.4, and coefficient variation of 0.497. With coefficient variation value less than 1.0, the result indicates that the frequency score for the quality criteria tend to be close to the average score and thus the variation of the score is low. The results presented in Fig. 2 signifies the relative importance of quality criteria of OGD as stated in the policy. Nevertheless, the question regarding on what extent these quality criteria are taken seriously by the OGD countries under study (by manifesting them in their policy) cannot be answered by the results just presented. Therefore, the coverage score as shown in Fig. 3 is used to answer the question.

Based on the results, we can see that the highest coverage is 60% and the lowest is 6.67%. On average, the countries under study include 37.71% (about 6) out of 15 data quality criteria in their policy. This is considerably low as less than half of the quality criteria are covered. With coefficient variation of 0.42, the result indicates the consistency of the scores across countries from the mean value.

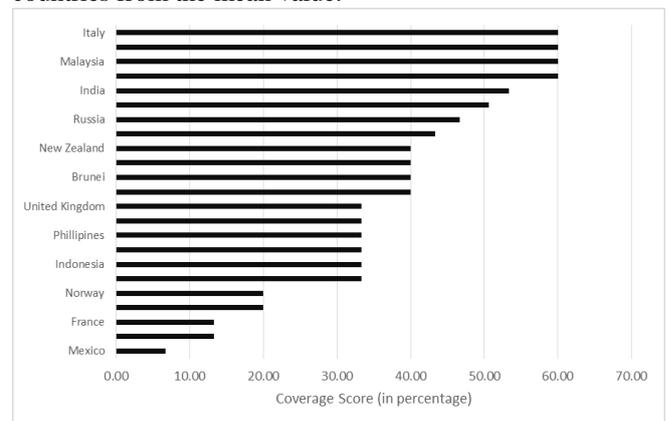


Fig. 3: OGD Policy Coverage by OGD Countries

If we further analyze the result by including the countries’ policy readiness score in ODB, we can see the distribution of the countries based on the policy coverage score and the policy readiness score (published by ODB) in four quadrants as shown in Fig. 4.

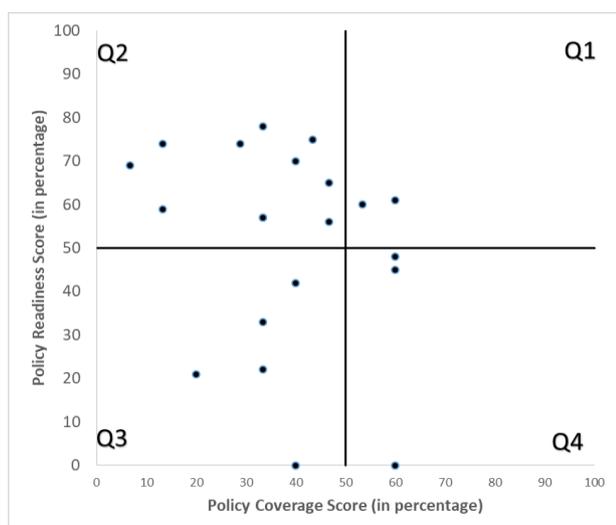
The upper quadrant (Q1 and Q2) represents high policy readiness scores (with 50% and above scores) and the lower quadrants (Q3 and Q4) represent low policy readiness scores (with 50% and lower scores).



The left-side quadrants (Q2 and Q3) represent low policy coverage scores (with 50% and lower scores). The right-side quadrants (Q1 and Q4) are for high policy coverage scorers (with 50% and above scores). Ideally, we would desire Q1 quadrant as it represents high score for both policy readiness and policy coverage.

Based on the result, we only have two countries in Q1 which are Italy and India. Most of the countries belong to Q2 quadrants which are highly prepared in their open data policy but, with low policy coverage. Two countries with missing policy readiness scores are Oman and Brunei as these countries are not yet included in ODB assessment. Nevertheless, Oman shows high coverage score with 60% (see Q4). Our concern is for countries in Q3 quadrant that show low scores in both criteria, which are Brunei, Indonesia, Ukraine, Tunisia and South Africa.

validate government open data's veracity is an open problem. Thus, the future work can be addressed in dual contexts: the government as open data provider and the public as OGD consumers.



**Fig.4: The Quadrant of Scores of Open Government Countries**

### V. CONCLUSION

In conclusion, this paper has revealed the veracity aspect for government open data portals by examining open data policy quality criteria provided by the portals. Even though the number of portals that provide the policy is small (27% out of 108 portals), we managed to extract 15 quality criteria from 36 raw criteria. We learned from the result that, while ensuring the quality of data shared on the portals is important, little attention is given in explicitly providing quality criteria in the portal for data contributors. Popular quality attributes such as *accuracy*, *trustworthiness*, and *completeness* surprisingly received a low score in the frequency analysis for open data. Instead, more emphasis is given to the 'technical' aspects of the data, represented by *machine readability*, *public*, *timeliness*, *formats* and *reusable*. The results also indicate that, even though some portals are at a higher ranking of quality criteria coverage, information about how these criteria are validated and measured is missing from the policy. In addition, we can learn from the result that most countries are ready in preparing their policy. Nevertheless, the coverage of important quality criteria is considerably low. Therefore, the coverage of the policy is an aspect that can be improved by the open governments. Given a set of quality criteria stipulated in the policy, the question of how to measure and

APPENDIX

QUALITY CRITERIA IN OGD POLICIES OF 31 GOVERNMENTS

Quality Criteria	Government	Description
1 Machine Readability	British Columbia [19]	Machine Processable: data is structured to be manipulated and automated processed through a computer.
	Brunei <sup>5</sup>	Machine-readable: The example of formats are XLS and CSV
	Canada <sup>6</sup>	Should be stored in widely used formats of file such as CSV and XML
	Germany <sup>7</sup>	Machine readability: Computer can further processed the data formats. Formats can be read are .txt, .csv, .json, .xml, and .rss.
	Indonesia [20]	Machine-readable: the data available not always readable by machine and available in different formats.
	Montreal <sup>8</sup>	Machine readability: automate process and aid of IT process
	Oman [21]	Machine process-able: available in machine-friendly formats and widely used file formats.
	Tunisia <sup>9</sup>	Machine-readable: to be analyzed effectively and structured
	British Columbia	Public: publicly used, adapted and distributed.
2 Public (Open by default/ Available/ Perceivable )	Brunei	Public: readily available, made public for discovery. No need to register to access published datasets and can be freely used.
	New South Wales [9]	Subject to public input: share and integrate the data at the national level
	United States <sup>10</sup>	Public: openness as permitted by law and subject to the privacy, security, and confidentiality.
	Australia [22]	Open by default: the openness of data with the prerogative of data release
	New Brunswick [23]	Open data by default: available proactively and reuse without barriers
	Brunei	Available: accurate and fully described data to easy understand meaning and limitation of data and quickly available
	Tunisia	Available online: to lodge many users to use the data
	Michigan	Perceivable: present the user interface components and information to users and able to perceive the presented information
	Canada	Timely: release quickly after collected and give priority to time-sensitive data
3 Timely (Priority)	Germany	Timeliness of data: publish the data once collected
	Montreal	Timeliness: publish and update the data as promptly as possible
	New Brunswick	Timely: delay between reference period and date on data becomes available
	Oman	Timely: real-time information updates
	United States	Timely: quickly available
4 Standards Formats	British Columbia	Formats: comply with open data physical datasets format standard
	Canada	Commonly owned standards: no need to use specific software and published in open format
	Montreal	Open Standards: accessible and readable data so that no need to purchase software or licensing fee.
	Germany	Usable: data recognized internationally
	Australia [22]	Reuse: promote analysis and reuse of data
5 Reusable (Relevance)	Brunei	Reuse: acknowledge the source with metadata included
	New South Wales	Reusable: available under an open license with no restrictions
	United States	Accessibility: barrier-free, easy search and easy to find
6 Accessibility (Discoverable /Visibility /Robust)	Germany	Accessible: available with minimum time delay
	Norway <sup>11</sup>	Accessible: available in convenient and modifiable
	United States	Accessible: the ability to download and locate content
	Oman	Accessible: enable easy discovery, the data should be made public
	Brunei	Discoverable: through catalogs and search engines, clearly indicates the licenses, use conditions, and data access.
	Norway	Appropriate persistent of data, unique and resolvable
	Australia [22]	Easily discoverable: easily discoverable and accessible
	Michigan	Robust: interpret the content reliably by user agents, so that user can access the content as technologies advance and remain accessible
	Canada	Completeness: complete as possible and reflect the collected information
	Montreal	Completeness: complete as possible with described metadata
	Germany	Complete: Metadata should be included as well, along with formulas and explanations to help users understand the scope of
7 Completeness (Described/Interpretable/ Documented/Understandable)	Oman	

<sup>5</sup> <https://www.data.gov.bn/Pages/Principles.aspx>

<sup>6</sup> <https://open.canada.ca/en/open-data-principles#toc95>

<sup>7</sup> <http://www.kas.de/wf/de/71.15334/>

<sup>8</sup> <http://donnees.ville.montreal.qc.ca/portail/city-of-montreal->

<sup>9</sup> <http://www.data.gov.tn/24->

<sup>10</sup> <https://project-open-data.cio.gov/policy-memo/>

<sup>11</sup> <http://www.bfe-inf.org/info/data-principles>

## Government Open Data Policy: A Measurement of Data Quality Coverage

		United States	information available and examine each data item at the greatest possible level of detail.
		United States	Complete: publish in primary forms with the finest level of granularity as permitted by law
		New Brunswick	Described: fully described to understand the weakness, strength, limitation and security criteria of data
		New Brunswick	Accurate: correctly described the phenomena is designed to measure in terms of statistical estimates error
		Oman	Interpretable: the supplementary information and metadata are available
		Michigan	Documented: documentation of data sets, meaning, formats and provide sufficient information.
		Norway	Understandable: to understand the information and operation of the user interface. The content must be not beyond the user understanding
8	License-free (Licensing /Usage Costs)	Canada	Understandable: interoperable to allow users to use. Reusable and require proper contextual and metadata including provenance, quality, and indicators
		Germany	Licensing: release under open government license and agreement
		Germany	Usage costs: release data free of charge
		Montreal	Licensing: available to the public without restrictions or imposition of conditions and barriers free.
		New South Wales	Cost of use: the fee collections can exclude the use of data
		Oman	Licensing: not subject to a third party, patents, copyrights, commercial confidentiality or trademarks
		Tunisia	Usage costs: no charge to users
9	Primary (Raw)	Canada	Free: free to access
		Germany	License-free: no restrictions on dissemination, maximal openness without restrictions and not subject to the law
		New Brunswick	Open license: permission to use and reuse to all users
		New South Wales	Free of charge: can access by everyone
10	Non Discrimination	Oman	Primacy: come from the primary source include the original information
		Germany	Primacy source: publish with the highest degree of fineness and not in modified formats
		Montreal	Primacy: detailed as possible and publish in forms of collected at source
		Canada	Primacy: released as collected at the source with high granularity level
		Oman	Primary: primary source data and not modified
		Montreal	Non-discrimination: user can access the data all the time without justifying their actions
		Canada	Non-discrimination: freely accessible for all users with no justification and identification
		Oman	Non-discrimination: who can access the data anytime without providing justification and identification
11	Manageable	Michigan	Non-discriminatory: datasets available to anyone anytime, no need to identify themselves
		New South Wales	Operable: components of user interface and navigation must be operable
		Norway	Well managed: data is managed with accordance with legislative and legal including data security and protection, intellectual property and business confidentiality
		United States <sup>2</sup>	Manageable and Protected: sustainable, trustworthy repository and data management policies and plans
12	Trusted (Authoritative/Approved)	New South Wales	Managed Post-Release: designated contact to assist data use and respond to complaints
13	Permanence / Permanent (Durability/Coherent)	Canada	Trusted: Alert the users to the quality and limitations to ensure the trustworthy and authoritative
		Montreal	Permanence: capable to be found over time. The data should remain online with appropriate tracking version and archive over time
		New Brunswick	Permanence: the data remain accessible permanently online or in archives at their original publishing
		Germany	Coherent: the data can be brought together with other data within the analytic framework and overtime
14	Protected	Australia [22]	Durability: the data posted should be online documented efficiently and available for a long time
15	Non-proprietary	Oman	Protected: the collected data is suitable for use to be released and should be protected, access restricted or prohibited
			Non-proprietary: the data can be access without the need for software licensed or no control from an entity

## ACKNOWLEDGMENT

The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for supporting this research.

## REFERENCES

1. D. Dietrich, "The context: e-Government, Open Government, Open Data." [Online]. Available: <http://workspace.unpan.org/sites/Internet/Documents/UNPAN95212.pdf>. [Accessed: 24-Apr-2018].
2. K. Hiramoto, "e-Government and Open Government Data in Japan," 2013. [Online]. Available: [https://cio.go.jp/assets/JAPAN\\_OGD.pdf](https://cio.go.jp/assets/JAPAN_OGD.pdf). [Accessed: 24-Apr-2018].
3. Sultanate of Oman, "e-Portal: The Public Authority for Consumer Protection." [Online]. Available: <https://pacp.gov.om/Open-Government-Data-Policy.aspx>. [Accessed: 24-Apr-2018].
4. W. Ritter, "Open Data in Asia: An Overview of Open Data Policies and Practices in 13 Countries," 2014. [Online]. Available: <https://www.slideshare.net/klainfo/open-dataasia092014>. [Accessed: 24-Apr-2018].
5. The World Bank Group, "Open Data for Sustainable Development: Policy Note ICT01," 2015. [Online]. Available: <http://pubdocs.worldbank.org/en/999161440616941994/Open-Data-for-Sustainable-Development.pdf>. [Accessed: 24-Apr-2018].
6. K. Izdebski, "Transparency and Open Data Principles: Why They Are Important and How They Increase Public Participation and Tackle Corruption," U. S. Department of State, 2015. [Online]. Available: <https://transparencee.org/wp-content/uploads/2015/12/open-data-principles-by-krzysztof-izdebski.pdf>. [Accessed: 24-Apr-2018].
7. S. Neumaier, J. Umbrich, and A. Polleres, "Automated Quality Assessment of Metadata across Open Data Portals," J. Data Inf. Qual., vol. 8, no. 1, pp. 1–29, 2016.
8. H. M. Kienle, "Open Data: Reverse Engineering and Maintenance Perspective," 2010.
9. State of New South Wales, "NSW Government Open Data Policy," pp. 1–6, 2016.
10. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the {EM} algorithm (with discussion)," J. R. Stat. Soc., vol. 39, pp. 1–38, 1977.
11. J. Tauberer, The Principles and Practices of Open Government Data, Second. Lulu Press, 2014.
12. L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," Data Sci. J., vol. 14, no. 0, p. 2, 2015.
13. O. Hartig and J. Zhao, "Using Web Data Provenance for Quality Assessment," in Proceedings of the First International Workshop on Semantic Web in Provenance Management, 2009, pp. 29–34.
14. P. N. Mendes, H. Mühleisen, and C. Bizer, "Sieve: Linked Data Quality Assessment and Fusion," in Proceedings of the 2012 Joint EDBT/ICDT Workshops, 2012, pp. 116–123.
15. M. W. AlRushaid and A. K. J. Saudagar, "Measuring the Data Openness for the Open Data in Saudi Arabia-Government - A Case Study," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 12, pp. 113–122, 2016.
16. G. Lee and Y. H. Kwak, "Open government implementation model: a stage model for achieving increased public engagement," in Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, 2011, pp. 254–261.
17. Open Data Barometer Organization, "Open Data Barometer - Leaders Edition," 2017.
18. N. A. Emran, "Data completeness measures," in Advances in Intelligent Systems and Computing, 2015, vol. 355, pp. 117–130.
19. British Columbia CIO Office, "Open Information and Open Data Policy," 2011. [Online]. Available: <https://www2.gov.bc.ca/assets/gov/british-columbians-our-government/s/services-policies-for-government/information-management-technology/information-privacy/resources/policies-guidelines/open-information-open-data-policy.pdf>.
20. OECD Publishing, "Open Government Data Review of Mexico: Data Reuse for Public Sector Impact and Innovation," 2016.
21. Government of Oman, "Open Government Data Policy," Oman eGovernment, 2016.
22. Government of Western Australia: Department of Premier and Cabinet, "The Western Australian Whole of Government Open Data Policy Western Australian Whole of Government Open Data Policy Version 1.1," 2015.

23. Corporate Information Management and Services City of Toronto, "Open Data Policy," 2011. [Online]. Available: [https://www.toronto.ca/wp-content/uploads/2017/11/969b-open\\_data\\_policy.pdf](https://www.toronto.ca/wp-content/uploads/2017/11/969b-open_data_policy.pdf).

## AUTHORS PROFILE



soft computing techniques in database management.

**Nor A. M. Sabri** is a PhD student at the Universiti Teknikal Malaysia Melaka. She graduated from Universiti Tun Hussein Onn in Information Systems at undergraduate level, and received Master in ICT in 2015 from the Universiti Teknikal Malaysia Melaka. Her research interest is in data mining and the application of



certification and her teaching mainly covers database-related subjects at undergraduate and postgraduate levels. Her current research interests include data quality and the application of artificial intelligence and machine learning in database management.

**Nurul A. Emran** holds Ph.D. degree in Computer Science from the University of Manchester, UK in 2011. In 2004, she joined the department of Software Engineering, Universiti Teknikal Malaysia Melaka, as a Lecturer. She holds Oracle Certified Professional (OCP)



Centre research group. From 2009 - 2011 she worked as research assistant for Glocal Center of Excellent (GCOE) program sponsored by Mombusho, Japan. From 2005-2009, she worked under Panasonic Japan in publishing patent for mobile standardization organization, 3rd Generation Partnership Project. Dr. Norharyati is now a Senior Lecturer in System and Computer Communication department and her major are related to signal processing, information theory and wireless communications.

**Norharyati Harum** received PhD Engineering in Smart Media Communication Engineering in 2012 from Keio University, Japan. From 2005-2009, she worked with Research and Development department at Panasonic Japan under Next Generation Mobile Communication