

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Do open government data (OGD) portals show signs of knowledge management (KM) practices?

Chokki, Abiola Paterne; Alexopoulos, Charalampos; Matheus, Ricardo; Saxena, Stuti; Frénay, Benoît; Vanderose, Benoît

Published in:
Technology Analysis & Strategic Management

DOI:
[10.1080/09537325.2023.2280512](https://doi.org/10.1080/09537325.2023.2280512)

Publication date:
2023

Document Version
Peer reviewed version

[Link to publication](#)

Citation for pulished version (HARVARD):
Chokki, AP, Alexopoulos, C, Matheus, R, Saxena, S, Frénay, B & Vanderose, B 2023, 'Do open government data (OGD) portals show signs of knowledge management (KM) practices? An empirical investigation', *Technology Analysis & Strategic Management*. <https://doi.org/10.1080/09537325.2023.2280512>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Do Open Government Data (OGD) portals show signs of Knowledge Management (KM) practices? : An empirical investigation

Abstract

Open Government Data (OGD) is a build-up of the data accumulated in the government organizations pertaining to the structural and functional dimensions and it is imperative for OGD to be high-value for facilitating value creation and innovation. The present study purports to provide a launchpad to the aforementioned truism by advancing the concept of Open Government Data Capital (OGDC) resting on the principles of Knowledge Management (KM) given that the high-value OGD can result only with the engagement of the concerned administrative agencies in knowledge sharing for being made accessible for wider use via dedicated web portals. To drive home the arguments, an empirical investigation is conducted with four top-notch countries, viz., Canada, Australia, New Zealand and the United States, in terms of the quantitative evaluation of their OGD portals' quality and inferences are drawn as to how OGDC may be furthered with the provision and maintenance of high-value datasets. Thus, it is shown that the Australian OGD portal is qualitatively robust and leads in terms of OGDC which may be beefed up with more integration of the KM practices in terms of the inter-governmental agencies' coordination and the other countries are lagging behind in terms of the quality parameters. While little attention has been paid towards the integration of KM in the OGD realm, the present study is a step forward in this direction.

Keywords:

Open Government Data, OGD, Open Government Data Capital, OGDC, Quality, Knowledge Management

1. Introduction

Open Government Data (OGD) pertains to the provision of high-value datasets pertaining to the structural and functional facets of administrative bodies via dedicated web portals for facilitating the re-use of the same by a range of stakeholders to further value creation and innovation (de Juana-Espinosa & Lujan-Mora, 2019; Jetzek, Avital and Bjorn-Andersen,

2014; Peled, 2011; Wirtz, Weyerer & Rosch, 2018). OGD is considered as datasets but it needs to be appreciated that without the linking of datasets, meaningful interpretation cannot be done (Kalampokis, Tambouris & Tarabanis, 2011; Muñoz, Bolívar & Arellano, 2022; Shadbolt et al., 2012) and this implies that knowledge creation-the resultant OGD repositories- happens and this OGD assumes form as an Open Government Data Capital (OGDC) which is further processed into valuable outputs thereby leading to value creation and innovation. OGDC happens when there is a high-value OGD sourced from different sources such that inter-governmental agencies coordinate with one another in this process of knowledge sharing, and, hence, knowledge creation. Thus, in the present study, OGDC is defined as: "the holistic OGD which gets accumulated in the public web repositories as 'knowledge warehouse' and is transmuted into valuable information for value creation and innovation". Knowledge Management (KM) includes "knowledge acquisition, encoding, storage, transfer, application and sharing" (Zhao et al., 2020: 372). In the context of OGD, knowledge sharing assumes significance on account of the involvement of different government departments and even other stakeholders where the citizens become the contributors to the OGD repositories-case in point being the users themselves who engage in knowledge exchange with the OGD providers and this goes a long way in OGD quality improvisation besides serving as a feedback-and-control apparatus (Ruijter & Meijer, 2020). The chief impediment towards OGDC is linked with the quality of OGD per se (Sadiq & Indulska, 2017). Therefore, knowledge handling strategies assume criticality with the technical infrastructure in place and the requisite personnel acknowledging their responsibilities (Lai et al., 2020; Rhee & Choi, 2017) in administrative organizations so that datasets are provided in complete and accurate formats via the dedicated web portals.

But for touching base on the manner in which knowledge sharing happens via linked OGD (Davies & Edwards, 2012; Geiger & von Lucke, 2012), there has been scant academic attention on the fact that organizational factors play a significant role in knowledge sharing in the public sector (Welch, Feeney & Park, 2016; Zhang, Dawes & Sarkis, 2005) but the same has not been appreciated in the domain of OGD: the present study seeks to plug this gap in two ways: 1. It provides a brief regarding the need for knowledge management (KM) vis-a-vis OGD, and 2. It provides an empirical investigation regarding the knowledge management of four out of top-10 countries ranked across OGD initiatives given that without quality

maintenance, knowledge management cannot happen. The research question addressed by the study is: "To what extent is KM furthered across the OGD portals in countries faring well in the OGD initiatives' standards?" As a theoretical contribution to the extant KM literature wherein the impetus upon knowledge sharing practices has been forwarded to result in increased knowledge corpus creation via value derivation and innovation, the present study hinges itself in line with the assertion that "knowledge management representation systems have been created and continue evolving in order to link different data" (Charalabidis, Alexopoulos & Loukis, 2016: 48), the present study addresses the call made by Charalabidis and his colleagues that further research is warranted in the "integrated knowledge base" domain vis-a-vis OGD.

The rest of the paper is structured as follows: following a brief on the implications of OGD for knowledge management and the creation of OGDC (Section 2), the research methodology is detailed (Section 3) with the results (Section 4) and the discussion of the findings (Section 5) and conclusion is provided (Section 6) with a rounding off with the indications for future research veering around OGDC (Section 7).

2. Literature review

Fundamentally, "data can be seen as the lowest level of abstraction from where information and then knowledge are derived" (Ubaldi, 2013: 5) and as far as the OGD initiatives are concerned, it is important that "co-development of knowledge" happens by the "encouraging external inputs and new sources of knowledge" (Ubaldi, 2013: 14). Besides formulating a long-drawn strategy and capable leadership, a sustainable OGD initiative mandates that its OGD initiative is well-entrenched alongside the "necessary coordinations with other agencies" (Solar, Concha & Meijueiro, 2012: 214). Inter-departmental coordination and integration is significant for ensuring that data is shared across the departments for furthering knowledge sharing (Dawes, 2012; Sanderson, Banks, Deakin & Udagawa, 2015; Zhao et al., 2022) and making it available on the dedicated portals for public re-use. This coordination is impeded on account of several reasons, for instance, the perceived risk and hierarchical organization, bureaucratic and autocratic decision-making culture (Wirtz, Piehler, Thomas & Daiser, 2016). In the empirical study conducted on the Taiwan government agencies' OGD initiatives, it was shown that governments must engage in data exchange for the success of

the OGD initiatives given that they stand to gain in terms of the potential advantages of the OGD initiatives (Wang & Lo, 2016). Thus, it has been attested that OGD initiatives involve inter-departmental collaboration along with the support from the senior management executives and the persistence of the personnel at the middle levels of management and the analysts (Krishnamurthy & Awazu, 2016). This knowledge sharing culture among the organizations would result in OGDC: the repository created with the sharing of datasets pertaining to the administrative agencies which is replete in all aspects. OGDC is facilitative of open innovation via OGD by a range of stakeholders. Open innovation is furthered by the data-driven knowledge management and knowledge sharing endeavors (Chaston, 2012; Del Vechhio et al., 2018). OGD engagement is instrumental in value derivation for entrepreneurial pursuits (Kitsios & Kamariotou, 2023). Likewise, Open Social Innovation (OSI) involving the multitudinous stakeholders' engagement is linked with the extent to which the open platforms permit openness, accountability, resource availability and involvement (Fortunato et al., 2017; Gegenhuber et al., 2023). Also, given the possibilities of engaging in open innovation via open data, users engage in deriving nuances from visualization and statistical analyses (Park & Gil-Garcia, 2022).

However, OGDC can happen only when the OGD is high-value (Nikiforova, 2021) and complete, accurate, updated, available in user-friendly formats like CSV, XLS that do not require any specialized software. The motivation to conceptualize OGDC was that in line with the extant research which is replete with instances where the OGD quality is a cause of concern (Grimmelikhuijsen & Feeney, 2017; Martin, De Rosario & Perez, 2015; Matheus, Janssen & Maheshwari, 2020; Yang, Lo & Shiang, 2015), it becomes important for the government agencies to adopt a strategy for refurbishing the OGD initiative. For instance, in the case of the Czech Republic, it was deduced that the data catalogs are of low quality and that the OGD is scattered here and there and such haphazard implementation of the OGD initiative results in problems for the users in terms of data discovery and re-use (Kucera, Chlapek & Necasky, 2013). Six parameters for knowledge sharing quality pertain to the completeness, reliability, timeliness, relevance, ease of understanding and accuracy (Chiu, Hsu & Wang, 2006) and these dimensions constitute the sine qua non of OGD thereby clinching the main argument advanced in the present study. Thus, it is imperative to assess the quality of the OGD that is provided via the dedicated portals to understand if OGDC can

be furthered with value creation and innovation by the stakeholders or not- the following section (Section 4) provides an initiation of this assertion.

3. Research methodology

Figure 1 summarizes the different phases of the research methodology.

3.1. Selection of country portals and determination of number of datasets to be analyzed per portal

To select the country portals to evaluate in this study, we rely on the Open Data Barometer (ODB), a well-acclaimed benchmark for OGD initiatives across countries in terms of their trajectories and progress over the years (World Wide Web Foundation, 2018). The top 10 country portals ranked in the 2018 ODB report (Canada (CA), United Kingdom (GB), Australia (AU), France (FR), South Korea (KR), Mexico (MX), Japan (JP), New Zealand (NZ), United States (US), and Germany (DE)) are initially selected for further analysis. However, due to certain research considerations, we end up selecting only 4 (CA, AU, NZ, US) of these 10 top country portals. The research considerations are: (a) the interface, resources (i.e., the file containing the dataset) and metadata (i.e., the information about the dataset) in the portals should be in English, as most authors are not familiar with other languages, (b) the portals should provide an Application Programming Interface (API) to facilitate access to their datasets and metadata, (c) The resources on portals should be available in Comma Separated Values (CSV) format and well structured (e.g., no combination of metadata of columns and data in one single CSV file or existence of several empty rows in the CSV file as in the case of the GB portal) to facilitate automatic evaluation of the datasets.

Since there are thousands of CSV datasets for each portal and there is a restriction on the maximum number of data to be retrieved during a period on the portals, we decide to use a sample size of CSV datasets from each portal for the evaluation instead of the full CSV datasets. The formula below (Equation 1) was used to determine the number of sample datasets (N') to collect in order to obtain statistically significant results from the initial number of CSV datasets (N) (Alexander *et al.*, 2017). The sample size for each portal was estimated at 600 datasets based on the following settings: $p = 0.5$, $e = 4\%$, and $z = 1.96$ (this value is obtained by considering a confidence interval of 95%). Table 1 presents the selected

country portals with information on their access link, platform used to develop the portal, ODB rank (as of 2018), number of total CSV datasets (as of 20/10/2022).

$$N' = \frac{\frac{z^2 * p (1-p)}{e^2}}{1 + \left(\frac{z^2 * p (1-p)}{e^2 N} \right)} \text{Equation 1}$$

Where p is the population proportion, N = total number of CSV datasets in the portal, e = margin of error, z = z-score.

Table 1: List of the country portals retained.

Country	ODB Rank (2018)	Portal Access Link	Platform	Total Number of CSV datasets (N) As of 20/10/2022	Sample Size (N')
Canada (CA)	1	https://open.canada.ca	CKAN	12,143	600
Australia (AU)	3	https://data.gov.au	MAGDA	11,117	
New Zealand (NZ)	8	https://data.govt.nz	CKAN	16,126	
United States (US)	9	https://www.data.gov	CKAN	33,955	

3.2. Benchmarking framework

To assess the data quality of the retained portals, we have used the data quality framework proposed in (Vetrò *et al.*, 2016). Since our goal in this study is to automate the data quality of the portals without the need to recruit participants, which is quite difficult (especially finding a significant number of open data experts for the evaluation) and time-consuming, we therefore determine which (sub-) dimensions of the selected framework need to be taken into account. Some inclusion and exclusion criteria are therefore proposed to achieve our goal. These are: (a) including all (sub-) dimensions that can be automated, i.e., that do not require the intervention of the participants in order to quantify them, and (b) excluding all (sub-) dimensions that rely on information not available on the evaluated portals. Applying these criteria, we retained 4 main dimensions (traceability, completeness, compliance, and accuracy). For example, the dimensions “currentness” and “expiration” were excluded because in most of the retained portals there is no record of previous versions of the datasets (resources), nor is there any information on the metadata of the previous updates of the datasets, the frequency of updates, and a complete list of updates dates (except the last update

date). Table 2¹ lists all dimensions and sub-dimensions of the (Vetrò *et al.*, 2016) framework with information on their description, formula or reason for exclusion when applicable (see gray rows for more detail about the excluded (sub-) dimensions). Once we identify the sub-dimensions to be automated, we then implement them in a python application (source code available at https://github.com/stutisaxenaogd/KM_Table-2). The implemented application is subdivided in four main steps:

- Random selection of 600 datasets from each portal. For each dataset, we download the associated CSV resources and metadata. Since there is a restriction on the number of data to be retrieved during a period on the portals, we decide to download a maximum of 10 resources per dataset and also limit the size of each resource to be downloaded to 200MB to prevent this error.
- Calculation for each dataset of the sub-dimensions: “Track of creation”, “Track of updates”, and “eGMS Compliance”. For each dataset, we rely on their metadata to calculate the values of the mentioned sub-dimensions.
- Calculation for each resource of the sub-dimensions: “Percentage of complete cells”, “Percentage of complete rows”, and “Percentage of accurate cells”. For the first two sub-dimensions, their calculations were straightforward. However, for the last sub-dimension “Percentage of accurate cells”, for each resource, we first detect the datatype of each column by enhancing the `csv_detective`² package (a python package that relies on regular expressions and column names to detect the datatype) in terms of regular expressions and supported datatypes. Then, we calculate the number of cells that have correct values according to the previously detected datatypes. The list of all 20 datatypes considered so far is as follows: address, country, country code, latitude, longitude, geo shape, geo point, boolean, email, integer, float, image, phone, sex, url, color, date, datetime (iso & rfc), hour & minute, year.
- Calculation for each dataset of the sub-dimensions: “Percentage of complete cells”, “Percentage of complete rows”, and “Percentage of accurate cells”. For each dataset, the values of these sub-dimensions are represented by the means of the values of the sub-dimensions of the related resources calculated in the previous step.

¹Online supplementary material: https://github.com/stutisaxenaogd/KM_Table-2

² <https://github.com/etalab/csv-detective>

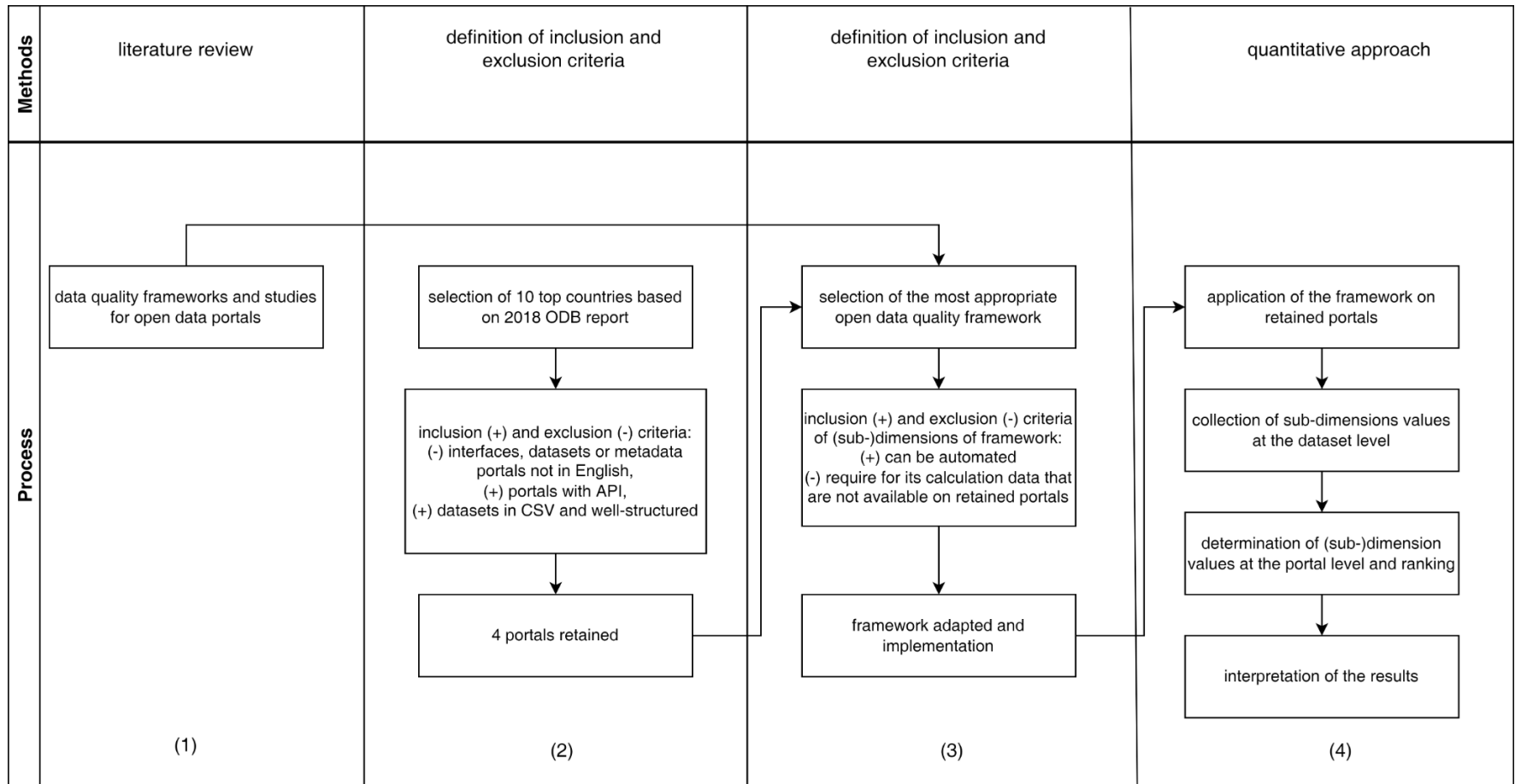


Figure 1: Research methodology.

3.3. Data collection

Once the implementation of the automated framework is done, we then run the application on each of the 4 retained portals to collect 600 datasets for each of them. The data collections were performed on a Lenovo ThinkPad with 8 GB 2133 MHz RAM and 2.29 GHz Intel Core i5 CPU, running Windows 8.1. The runtime to collect the datasets for each portal was between 5h and 8h because for some datasets, the application needs to collect multiple resources (up to 10 resources) and some resources can be large files (up to 200MB). For each portal, the application generates at the end of the run, a CSV file that includes information about the collected datasets (e.g., identifier, title, description, publisher, creation date, etc.) and the values of the sub-dimensions: “Percentage of complete cells”, “Percentage of complete rows”, and “Percentage of accurate cells”. Other sub-dimensions such as “Track of creation” and “eGMS compliance” are calculated using the excel formula since the information needed for their calculation has already been collected. The remaining sub-dimensions “Track of updates” and “Five star Open Data” are set to 0.25 and 0.6, respectively, because for each dataset only the last update is provided (“Track of updates”) and the format considered in this study is only CSV (“Five star Open Data”).

All dimensions and sub-dimensions of the framework are weighted equally to facilitate comparison given the significance of all the dimensions and sub-dimensions. Thus, we calculate the value of each dimension for each dataset by computing an average across the sub-dimensions of the dimension. We also calculate the data quality for each dataset by calculating an average across the dimensions which would further an understanding of the extent to which the data are high-value. Once the (sub-) dimensions and data qualities are calculated for each dataset, we then calculate for each (sub-) dimension and data quality the means and the standard deviations (SD) in order to have an overview of these (sub-) dimensions and data quality at the portal level.

3.4. Interpretation of results

The results obtained in the previous section are then interpreted at three levels: (1) sub-dimension level, (2) dimension level and (3) portal level. For the sub-dimension level, we rely on the means and standard deviations calculated for each sub-dimension for each portal to compare the performance of the portals. For the dimension level, we use the means and

standard deviations calculated for each dimension for each portal to compare portal performance. For the portal level, we look at the means and standard deviations calculated for data quality for each portal to compare the performance of the portals.

For each level, we also performed a T-test³ and used the p-values generated by the T-test to determine whether there is a significant difference between the values of two portals. If the p-value ≤ 0.05 , then we conclude that there is a statistically significant difference. Otherwise, there is not enough evidence to state that there is a statistically significant difference.

4. Results

4.1. Results by sub-dimensions

Figure 2 shows the mean values by sub-dimension for the evaluated country portals. In order to confirm the higher score of one country portal compared to another portal, some statistical tests are performed and the results are summarized in Table 3. Regarding the sub-dimension “track creation”, all portals obtain an excellent score. These results show that all portals have provided the information about the data source and the creation date for each of their datasets. For the sub-dimension “track of updates”, all the portals have the same score and do not perform well. These scores can be justified by the fact that all portals do not provide the required information about their previous updates of their datasets, except the last modified date which is provided. For the sub-dimension “percentage of complete cells”, the best result is obtained by the AU portal, followed by the NZ and US portals which have approximately the same results. The worst result is obtained by the CA portal. As for the sub-dimension “percentage of complete rows”, the worst result is obtained by the CA portal. On the other hand, the best result is performed by the AU portal, followed by the US portal and then by the NZ portal. These results show that the retained portals published some datasets which have empty cells and consequently impact on the percentage of complete rows. Regarding the sub-dimension “eGMS compliance”, the CA portal performed the best for this sub-dimension followed by the AU and NZ portals. The US portal has the lowest result. The overall good results can be justified by the fact that most portals provide the required information (e.g., title, description, and publisher) for the metadata of their published datasets. Since our study focuses on CSV resources as we want to automate the calculation of

³<https://rb.gy/cdw5so>

data quality, we set the sub-dimension “five star open data” to 0.6 (3 stars) for all evaluated portals. For the sub-dimension “percentage of accurate cells”, the best result is obtained by the NZ portal, followed by the AU, CA, and US portals, which have approximately the same results. These excellent results show that most of the portals are filling the cells of their datasets with the appropriate datatypes.

In addition to these separate evaluations, we also average each sub-dimension across all portals to determine which sub-dimension performed best or worst. Based on Figure 3 (see gray bars), we note that the sub-dimension “track of creation” is the metric where most of the portals performed well, followed by the sub-dimensions “eGMS compliance”, “percentage of complete cells”, “percentage of accurate cells”, “five star open data”, “percentage of complete rows” respectively. The worst result was performed by the sub-dimension “track of updates”.

4.2. Results by dimensions

Figure 4 presents the mean values by dimension for the selected country portals. In addition, we perform some statistical tests whose results are summarized in Table 4 to confirm whether one country portal performs better than another for each dimension. For the dimension “traceability”, all the portals have the same results since their results for the sub-dimensions “track of creation” and “track of updates” were also the same. As for the dimension “completeness”, the best result is performed by the AU portal followed by the US portal and the NZ portal respectively. The CA portal has the worst result for this dimension. For the dimension “compliance”, the worst result is obtained by the US portal. The best result is obtained by the CA portal followed by the AU and NZ portals which have the same results. These results show that most of the portals are trying to follow the standards in terms of data publication, especially in providing metadata information, but there is still room for improvement to achieve better results. For the dimension “accuracy”, most of the portals perform very well however the NZ portal performed better than the three other portals which have the same results.

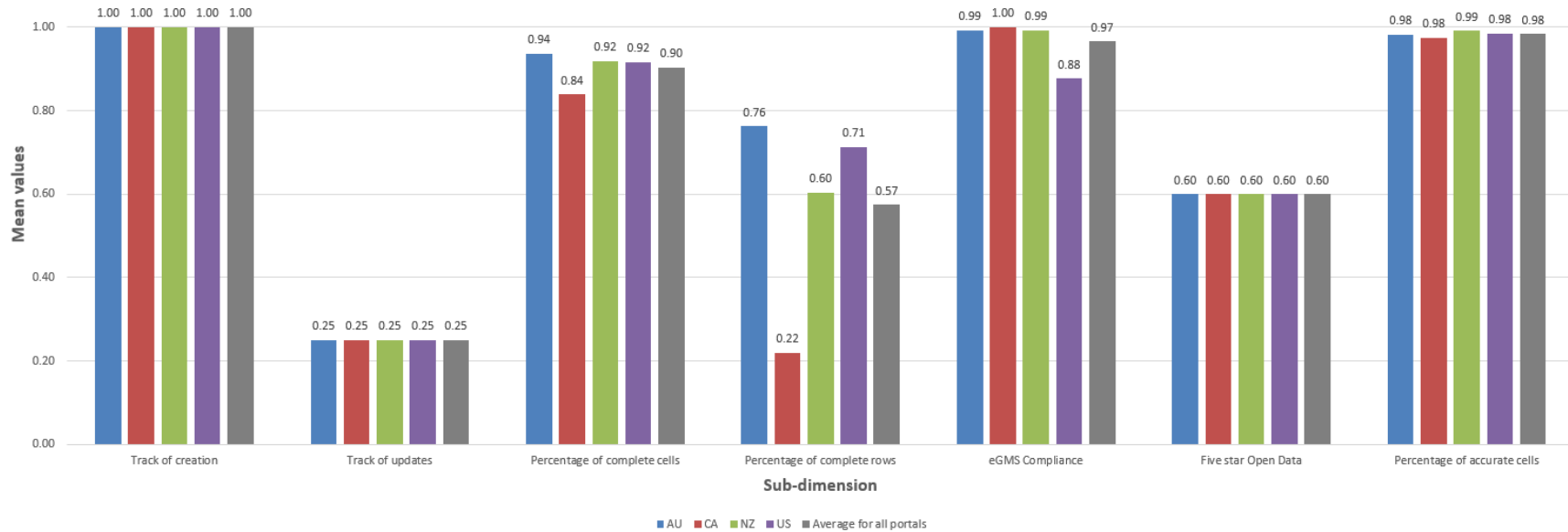


Figure 2: Mean values by sub-dimension for country portals.

Table 3: Pairwise country portal comparisons by sub-dimension.

Sub-dimension	Pairwise country portal comparisons					
	AU-CA	AU-NZ	AU-US	CA-NZ	CA-US	NZ-US
Track of creation						
Track of updates						
Percentage of complete cells	(AU > CA)*	(AU > NZ)*	(AU > US)*	(NZ > CA)*	(US > CA)*	
Percentage of complete rows	(AU > CA)*	(AU > NZ)*	(AU > US)*	(NZ > CA)*	(US > CA)*	(US > NZ)*
eGMS Compliance	(CA > AU)*		(AU > US)*	(CA > NZ)*	(CA > US)*	(NZ > US)*
Five star Open Data						
Percentage of accurate cells		(NZ > AU)*		(NZ > CA)*		(NZ > US)*

Only statistically significant comparisons are shown. Star represents significance level: * $p \leq 0.05$.

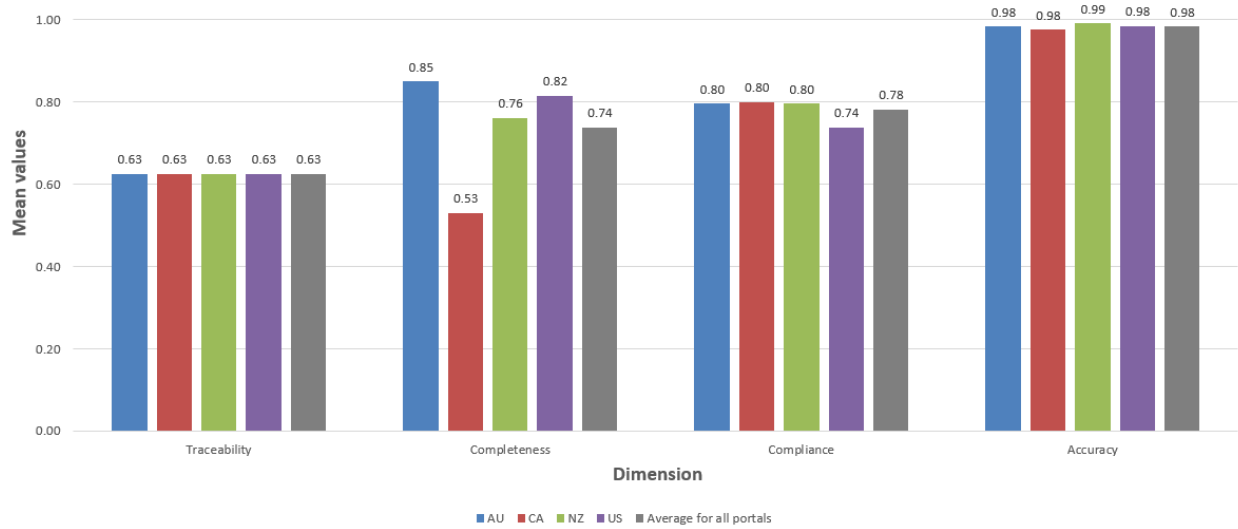


Figure 3: Mean values by dimension for country portals.

Table 4: Pairwise country portal comparisons by dimension.

Dimension	Pairwise country portal comparisons					
	AU-CA	AU-NZ	AU-US	CA-NZ	CA-US	NZ-US
Traceability						
Completeness	(AU > CA)*	(AU > NZ)*	(AU > US)*	(NZ > CA)*	(US > CA)*	(US > NZ)*
Compliance	(CA > AU)*		(AU > US)*	(CA > NZ)*	(CA > US)*	(NZ > US)*
Accuracy		(NZ > AU)*		(NZ > CA)*		(NZ > US)*

Only statistically significant comparisons are shown. Star represents significance level: $*p \leq 0.05$

Similar to the sub-dimension, we also average each dimension across all portals to determine which dimension performed best or worst. As shown in Figure 4 (see gray bars), the dimension “traceability” has the lowest value, followed by the dimensions “completeness” and “compliance”. The dimension “accuracy” is the one for which most portals performed best.

4.3. Results by portals

Based on the results of the dimensions for each portal, we then average the values of the dimensions to calculate the data quality for each portal to determine which portal performs better. Referring to Figure 4 and Table 5, the AU portal has the highest data quality, followed by the NZ and US portals which have the same scores. The CA portal has the lowest data quality. These results can be justified by the fact that with reference to the previous section, the AU portal performed best on most dimensions compared to the other country portals.

Furthermore, these results show that there are still some improvements to be done to portals to ensure their reuse. First, portals should provide details about their updates by implementing, for example, a versioning module in their portals. Second, portals should fill in all cells of their datasets to improve completeness, for example by using a simple Excel file to check for empty cells and fill them in before publication. Third, they should provide information about the metadata of their datasets and especially categorize each dataset to make it easier for reusers to search and navigate, because in the US portal (which has the lowest result on this dimension) the main issue was on this metric. Fourth, the portals should try to fill the cells of each column with the corresponding datatype and format to improve accuracy.

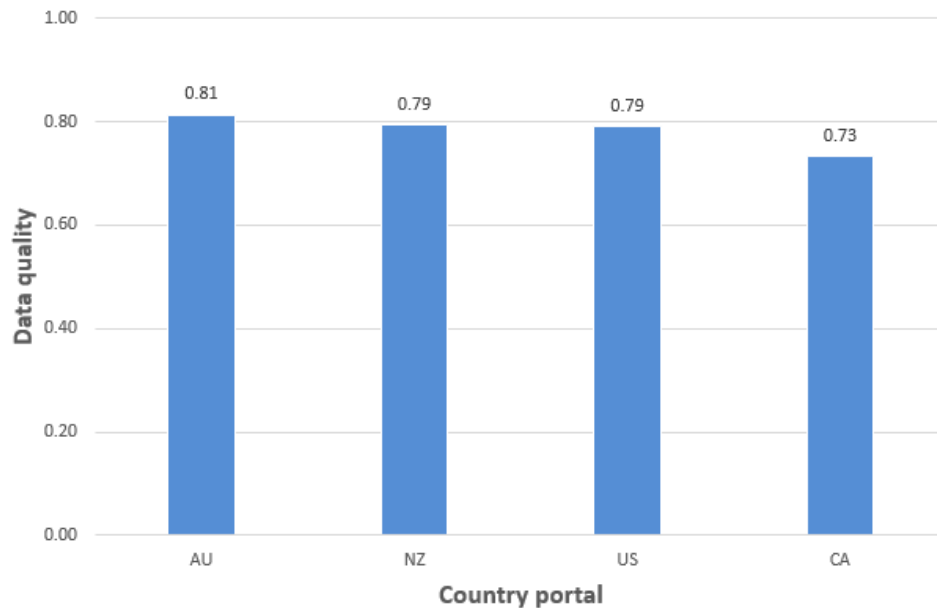


Figure 4: Data quality by country portal.

Table 5: Data quality pairwise country portal comparisons.

	Pairwise country portal comparisons					
	AU-CA	AU-NZ	AU-US	CA-NZ	CA-US	NZ-US
Data quality	(AU > CA)*	(AU > NZ)*	(AU > US)*	(NZ > CA)*	(US > CA)*	

Only statistically significant comparisons are shown. Star represents significance level: * $p \leq 0.05$.

5. Discussion

Findings from the present study vis-a-vis the qualitatively advanced OGD portals of Australia are in line with the previous research which attests to the early adoption of OGD practices under the aegis of the leading "policy entrepreneurs", i.e. the lead departments and agencies of the government (Chatfield & Reddick, 2018). Even the usability assessment of the OGD portal of Australia among other countries showed Australia scoring the highest in terms of the three parameters, viz., data specification, feedback and requests (Machova, Hub & Lnenicka, 2018)-the research acknowledged the success of Australian OGD initiative in terms of the state-of-the-art data explorer and visualization tools. Australian OGD portal is regularly updated and extensive (Power et al., 2015). Thus, Australia has succeeded to a great deal in making a transition from being secretive towards openness and transparency (Hardy & Maurushat, 2017; Henninger, 2018).

In the context of New Zealand, the relatively low performance of the OGD portal in terms of the quality given the variability of OGD file formats-as deduced in the present study-may be attributed to the fact that there are limited resources for cross-agency coordination and lack of interest in the government agencies to overly invest in ensuring the institutionalization of the OGD initiatives (Oh, 2013). The OGD portals of New Zealand are lacking in terms of data adequacy, accessibility, interoperability and infrastructure apart from intergovernmental coordination and user capabilities (Stas NZ, 2021). Similar conclusions were also deduced in another study wherein the complexity and heterogeneity of OGD led to inefficacious OGD publishing procedures (Schindler, Dionisio & Kingham, 2018). Such bottlenecks dampen the OGDC processes, to admit the least.

In line with the findings from the present study, data quality of US portal is inadequate and this is owing to the inefficacious support from the key organizational personnel within and without (Krishnamurthy & Awazu, 2016) and this serves as a major challenge in the development of OGDC. Conclusions similar to the present study were also found for the US OGD portals wherein it was shown that only whilst data-mostly incomplete- have been published via the portals, they are still "level one" OGD, i.e., in PDFs and HTML formats and thus, they are not machine-processable (Yi, 2019)-implying that the data quality is

inappropriate (Dawes, 2012). Complementing these findings with the previous research deductions that the OGD initiatives are “unevenly” developed in the US (Nugroho, Zuiderwijk, Janssen & de Jong, 2015), it is apparent that OGDC institutionalization is far-fetched.

In the case of Canada, findings from the present study attest previous research findings that besides the fact that the Canadian national authorities are rampant upon furthering their OGD initiatives with relatively better quality, the municipal authorities are lagging behind and there is lack of coordination between the government agencies vis-a-vis data sharing (Roy, 2014) and both of these facets are impediments towards an OGDC culture. Furthermore, the OGD initiatives are not being universally adopted by all the government agencies: case in point may be counted the Research Councils in the higher education sector which are not furthering OGD policies, let alone the publishing and dissemination dimensions (Lasthiotakis, Kretz & Sa, 2015). Likewise, in another case study, it was deduced how the Immigration, Citizenship and Refugees Canada (IRCC) did not engage completely in its OGD initiative's refurbishment or furthered citizen engagement, for that matter (Gintova, 2019). Such lacuna may be attributed to the lack of attention and policy directives from the end of the government agencies coupled with limited user engagement with the OGD platforms (Longo, 2017). Concomitantly, other challenges linked with the OGD initiatives of Canada are associated with the changes in public policies, changes in the names of the administrative departments, the tendencies of removal or alteration of the content of the OGD portals (Paterson, 2018) and even modulating the transparency commitments which is suggestive of the "closed", unaccountable and insular tendencies of the government (Clarke, Lindquist & Roy, 2017). Thus, it is clear that steps need to be taken for OGDC to happen.

6. Conclusion

Two of the key challenges pertaining to OGD initiatives relate to digital asset management and archiving and preservation and both these dimensions are linked with KM practices. Taking cue from this, the present study sought to present a concept of OGDC besides providing an empirical grounding with four top-notch countries across OGD quality standards wherein it was deduced that quality of the datasets is important for the success of OGD initiatives and quality can be maintained only via robust KM practices among the

administrative agencies. It needs to be appreciated that establishing a data governance infrastructure with facilitating foundations for knowledge sharing among the administrative agencies is not an easy task, however, short-term and long-term goals need to be chalked out for ensuring that sustainable and collaborative KM practices are being adopted by the government departments for furthering the cause of OGDC.

To appreciate the nuances of the research objectives, an empirical assessment was done of four of the top ten countries' OGD portals in terms of their quality metrics given that quality is a factor of the KM practices being adopted by the government agencies. Findings from the study show that whilst Australia is in the forefront in terms of the quality parameters, and, hence, testifies its commitment towards developing an OGDC culture, other countries are laggards and steps need to be taken for quality improvisation through the interventions aimed at furthering inter-departmental coordination and support from the senior management executives. By proposing a concept of OGDC which rests its edifice on the inter-agency collaboration with a specific focus on the quality maintenance of OGD, the present study contributed to the twin-pronged challenges identified in the OGD initiatives' efficacy, viz., implementation barriers and the barriers to use by the stakeholders. Furthermore, the study contributed to the knowledge corpus of OGD-KM dyad with its conceptual formulation, viz. OGDC, followed by the empirical validation.

Thus, the present study leaves some recommendations for the policy-makers, practitioners and society: first, a strategic vision needs to be chalked out by the governments horizontally and vertically with a point device aim of maintaining quality standards of the OGD portals via dialogue and discussion; second, and, in line with the first, a comprehensive plan document needs to be in place for chalking out the manner in which the OGDC culture may be imbibed by the governments; third, training and development needs of the concerned personnel as also the users should be met and this mandates a strong commitment towards OGD initiatives wherein budgetary allocations need to be made; and finally, myriad societal stakeholders' involvement is necessary for ensuring that the KM practices are furthered with the involvement of all those involved in data publishing and data consumption.

7. Further research pointers

With the conceptualization of the OGDC, several indicators emerge that might be addressed in future academic pursuits. For one, it may be pertinent to understand how the commonality of OGDC across the administrative agencies of different countries might lend cues for the “best practices” to be adhered to by all. Second, issues like governmental nature in terms of democracies and non-democracies as also the constitutional and statutory foundations might be of interest for those keen on unraveling the OGDC dimensions vis-a-vis legal informatics. Likewise, an understanding of the Knowledge Sharing Behaviors (KSBs) at the individual levels, i.e., the employee per se, might be of interest for the academicians interested in the behavioral public administration domains. Finally, though not conclusively, further research is merited to appreciate the relevance and implications of OGDC in the newer digital government formats like Gov 4.0 and beyond or the tension between Smart Cities vis-a-vis the conventional ones.

References

- Alexander, H., Illowsky, B. & Dean, S. (2017). *Introductory Business Statistics*. Samurai Media Limited.
- Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2016). A taxonomy of open government data research areas and topics. *Journal of Organizational Computing and Electronic Commerce*, 26(1/2), 41-63. <http://dx.doi.org/10.1080/10919392.2015.1124720>.
- Chaston, I. (2012). Knowledge management and open innovation in United Kingdom local authorities. *International Journal of Public Administration*, 35(4), 248-258. <https://doi.org/10.1080/01900692.2011.651293>.
- Chatfield, A.T., & Reddick, C.G. (2018). The role of policy entrepreneurs in open government data policy innovation diffusion: An analysis of Australian Federal and State Governments. *Government Information Quarterly*, 35(1), 123-134. <https://doi.org/10.1016/j.giq.2017.10.004>.
- Chiu, C.M., Hsu, M.H., & Wang, E.T.G. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems*, 42(3), 1872-1888. <https://doi.org/10.1016/j.dss.2006.04.001>.
- Clarke, A., Lindquist, E. A., & Roy, J. (2017). Understanding governance in the digital era: An agenda for public administration research in Canada. *Canadian Public Administration*,

60(4), 457-475. <https://doi.org/10.1111/capa.12246>.

Davies, T., & Edwards, D. (2012). Emerging implications of open and linked data for knowledge sharing in development. *IDS Bulletin*, 43(5), Institute of Development Studies. Blackwell: Oxford (UK).

Dawes, S. S. (2012). A realistic look at open data. Center for Technology in Government. Working Paper. https://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf.

Del Vecchio, P., Di Minin, A., Messeni Petruzzelli, A., Panniello, U., & Pirri, S. (2018). Big data for open innovation in SMEs and big corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management*, 18, 6-22.

Fortunato, A., Gorgoglione, M., Petruzzelli, A.M., & Panniello, U. (2017). Leveraging big data for sustainaing open innovation: The case of social TV. *Information Systems Management*, 34(3), 238-249. <https://doi.org/10.1080/10580530.2017.1330000>.

Gegenhuber, T., Mair, J., Luhrsen, R., & Thater, L. (2023). Orchestrating distributed data governance in open social innovation. *Information and Organization*, 33(1), 100453. <https://doi.org/10.1016/j.infoandorg.2023.100453>.

Gintova, M. (2019). Use of social media in Canadian public administration: Opportunities and barriers. *Canadian Public Administration*, 62(1), 7-26. <https://doi.org/10.1111/capa.12298>.

Grimmelikhuijsen, S.G. & Feeney, M.K. (2017). Developing and testing an integrative framework for open government adoption in local governments. *Public Administration Review*, 77(4), 579-590. <https://doi.org/10.1111/puar.12689>.

Hardy, K., & Maurushat, A. (2017). Opening up government data for Big Data analysis and public benefit. *Computer Law & Security Review*, 33(1), 30-37. <https://doi.org/10.1016/j.clsr.2016.11.003>.

Henninger, M. (2018). Reforms to counter a culture of secrecy: Open government in Australia. *Government Information Quarterly*, 35(3), 398-407. <https://doi.org/10.1016/j.giq.2018.03.003>.

Jetzek, T., Avital, M. and Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), 100-120. <https://doi.org/10.4067/S0718-18762014000200008>.

Kalampokis, E., Tambouris, E. & Tarabanis, K. (2011). A classification scheme for open government data: Towards linking decentralised data. *International Journal of Web*

- Engineering and Technology*, 6(3), 266-285. <https://doi.org/10.1504/IJWET.2011.040725>.
- Kitsios, F., & Kamariotou, M. (2023). Digital innovation and entrepreneurship through open data-based platforms: Critical success factors for hackathons. *Heliyon*, 9(4), e14868. <https://doi.org/10.1016/j.heliyon.2023.e14868>.
- Krishnamurthy, R., & Awazu, Y. (2016). Liberating data for public value: The case of Data.gov. *International Journal of Information Management*, 36, 668-672. <http://dx.doi.org/10.1016/j.ijinfomgt.2016.03.002>.
- Kucera, J., Chlapek, D., & Necasky, M. (2013). Open government data catalogs: Current approaches and quality perspective. In: Kö, A., Leitner, C., Leitold, H., Prosser, A. (eds) *Technology-Enabled Innovation for Democracy, Government and Governance. EGOVIS/EDem 2013. Lecture Notes in Computer Science*, 8061. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40160-2_13.
- Lai, K.K., Chen, H.C., Chang, Y.H., Kumar, V., & Bhatt, P.C. (2020). A structured mpa approach to explore technological core competence, knowledge flow, and technology development through work patentometrics. *Journal of Knowledge Management*, 25(2), 402-432.
- Lasthiotakis, H., Kretz, A., & Sa, C. (2015). Open science strategies in research policies: A comparative exploration of Canada, the US and the UK. *Policy Futures in Education*, 13(8), 968-989. <https://doi.org/10.1177/1478210315579983>.
- Longo, J. (2017). The evolution of citizen and stakeholder engagement in Canada, from Spicer to #Hashtags. *Canadian Public Administration*, 60(4), 517-537. <https://doi.org/10.1111/capa.12229>.
- Machova, R., Hub, M., & Lnenicka, M. (2018). Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. *Aslib Journal of Information Management*, 70 (3), 252-268, <https://doi.org/10.1108/AJIM-02-2018-0026>.
- Martin, A. S., Rosario, A. H. D., & Pérez, M. D. C. C. (2016). An International Analysis of the Quality of Open Government Data Portals. *Social Science Computer Review*, 34(3), 298-311. <https://doi.org/10.1177/0894439315585734>.
- Matheus, R., Janssen, M. and Maheshwari, D. (2020), "Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities",

Government Information Quarterly, Vol. 37 No. 3, Article 101284.

<https://doi.org/10.1016/j.giq.2018.01.006>.

Muñoz, L. A., Bolívar, M. P. R., & Arellano, C. L. V. (2022). Factors in the adoption of open government initiatives in Spanish local governments. *Government information quarterly*, 39(4), 101743.

Nikiforova, A. (2021). Towards enrichment of the open government data: A stakeholder-centered determination of high-value data sets for Latvia. *14th International Conference on Theory and Practice of Electronic Governance*, 367-372.

<https://doi.org/10.1145/3494193.3494243>.

Nugroho, R.P., Zuiderwijk, A., Janssen, M., & de Jong, M. (2015). A comparison of national open data policies: lessons learned. *Transforming Government: People, Process and Policy*, 9(3), 286-308. <http://dx.doi.org/10.1108/TG-03-2014-0008>.

Oh, J. (2013). Literature review-Open Data in New Zealand.

https://www.stat.auckland.ac.nz/~joh024/LitReviews/LitReview_NZOpenData.pdf.

Park, S., & Gil-Garcia, J.R. (2022). Open data innovation: Visualizations and process redesign as a way to bridge the transparency-accountability gap. *Government Information Quarterly*, 39(1), 101456. <https://doi.org/10.1016/j.giq.2020.101456>.

Paterson, S. (2018). The government information landscape in Canada. *Working Paper*.

<http://hdl.handle.net/2429/66480>.

Peled, A. (2011). When transparency and collaboration collide: The USA open data program. *Journal of the American Society for Information Science and Technology*, 62(11), 2085-2094. <https://doi.org/10.1002/asi.21622>.

Power, R., Robinson, B., Rudd, L., Reeson, A. (2015). Scenario Planning Case Studies Using Open Government Data. In: Denzer, R., Argent, R.M., Schimak, G., Hřebíček, J. (eds) *Environmental Software Systems. Infrastructures, Services and Applications. ISESS 2015. IFIP Advances in Information and Communication Technology*, 448. Springer, Cham. https://doi.org/10.1007/978-3-319-15994-2_20.

Rhee, Y.W., & Choi, J.N. (2017). Knowledge management behavior and individual creativity: Goal orientations as antecedents and in-group social status as moderating contingency. *Journal of Organizational Behavior*, 38, 813-832. <https://doi.org/10.1002/job.2168>.

Roy, J. (2014). Open data and open governance in Canada: A critical examination of new

- opportunities and old tensions. *Future Internet*, 6, 414-432.
<https://doi.org/10.3390/fi6030414>.
- Ruijter, E. & Meijer, A. (2020). Open government data as an innovation process: Lessons from a living lab experiment. *Public Performance & Management Review*, 43(3), 613-635,
<https://doi.org/10.1080/15309576.2019.1568884>.
- Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, 37, 150-154. <http://dx.doi.org/10.1016/j.ijinfomgt.2017.01.003>.
- Sanderson, P., Banks, D., Deakin, S., & Udagawa, C. (2015). Encouraging inter-regulator data sharing: the perceptions of regulators, Cambridge Centre for Housing and Planning Research,
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/416626/15-213-data-sharing-external-report.pdf.
- Schindler, M., Dionisio, R. & Kingham, S. (2018). A multi-level perspective on a spatial data ecosystem: Needs and challenges among urban planning stakeholders in New Zealand. *International Journal of Spatial Data Infrastructure Research*, 13, 223-252.
<https://ijsdir.sadl.kuleuven.be/index.php/ijsdir/article/view/480>.
- Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W. & Schraefel, M.C. (2012). Linked open government data: lessons from Data.gov.uk. *IEEE Intelligent Systems*, 27 (3), 16-24. <https://doi.org/10.1109/MIS.2012.23>.
- Solar, M., Concha, G. & Meijueiro, L. (2012). A model to assess open government data in public agencies. In: Scholl, H.J., Janssen, M., Wimmer, M.A., Moe, C.E., Flak, L.S. (eds) *Electronic Government. EGOV 2012. Lecture Notes in Computer Science*, 7443. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33489-4_18.
- Stats NZ (2021). COVID-19 lessons learnt: Recommendations for improving the resilience of New Zealand's government data system. <https://data.govt.nz/assets/Uploads/Covid-19-lessons-learnt-full-report-Mar-2021.pdf>.
- Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, 22, OECD Publishing.
<http://dx.doi.org/10.1787/5k46bj4f03s7-en>.
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C.O., Iemma, R. & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open

- Government Data. *Government Information Quarterly*, 33(2), 325–337.
<https://doi.org/10.1016/j.giq.2016.02.001>.
- Wang, H.J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly*, 33(1), 80-88. <https://doi.org/10.1016/j.giq.2015.11.004>.
- Welch, E.W., Feeney, M.K., & Park, C.H. (2016). Determinants of data sharing in U.S. city governments. *Government Information Quarterly*, 33, 393-403.
<http://dx.doi.org/10.1016/j.giq.2016.07.002>.
- World Wide Web Foundation. (2018), *Open Data Barometer: Leaders Edition*, available at:
<https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf>.
- Wirtz, B.W., Piehler, R., Thomas, M.J., & Daiser, P. (2016). Resistance of public personnel to open government: A cognitive theory view of implementation barriers towards open government data. *Public Management Review*, 18(9), 1335-1364.
<https://doi.org/10.1080/14719037.2015.1103889>.
- Wirtz, B.W., Weyerer, J.C., & Rosch, M. (2018). Citizen and open government: An empirical analysis of antecedents of open government data. *International Journal of Public Administration*, 41(4), 308-320. <https://doi.org/10.1080/01900692.2016.1263659>.
- Yang, T.M., Lo, J., & Shiang, J. (2015). To open or not to open? Determinants of open government data. *Journal of Information Science*, 41(5), 596-612.
<https://doi.org/10.1177/01655515155586715>.
- Yi, M. (2019). Exploring the quality of government open data: Comparison study of the UK, the USA and Korea. *The Electronic Library*, 37(1), 35-48. <https://doi.org/10.1108/EL-06-2018-0124>.
- Zhang, J., Dawes, S. S., & Sarkis, J. (2005). Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. *Journal of Enterprise Information Management*, 18(5), 548–567. <https://doi.org/10.1108/17410390510624007>.
- Zhao, S., Jiang, Y., Peng, X., & Hong, J. (2020). Knowledge sharing direction and innovation performance in organizations. Do absorptive capacity and individual creativity matter? *European Journal of Innovation Management*, 24(2), 371-394.
<https://doi.org/10.1108/EJIM-09-2019-0244>.
- Zhao, L., Zhang, L., Sun, J., & He, P. (2022). Can public participation constraints promote green technological innovation of Chinese enterprises? The moderating role of government

environmental regulatory enforcement. *Technological Forecasting and Social Change*, 174, 121198. <https://doi.org/10.1016/j.techfore.2021.121198>.