Tracking metadata changes in the government open data portals

Milena Frtunić Gligorijević*, Miloš Bogdanović*, Leonid Stoimenov*
* Faculty of Electronic Engineering, University of Niš, Niš, Serbia
milena.frtunic.gligorijevic@elfak.ni.ac.rs, milos.bogdanovic@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs

***Abstract*— Due to the transparency and open government initiatives, a large amount of open government data has been published on the open data portals. In order to provide data reusability, open data portals have different search mechanisms which depend on the quality of datasets metadata. Therefore, within this paper we present the results of the analysis of metadata changes on 40 government open data portals over a two year period. One of the main dataset search resources on the open data portals are datasets categories and tags used for datasets description. Therefore, the focus of our analysis is on tracking changes in the number of datasets on the portals, and changes that consequentially occurred in categories and tags available on the portals. We present different valuable results that can be used for tracking the trend of metadata changes and growth of the metadata that can affect their findability.**

Keywords: open government data, open data portal, metadata, e-government

I. INTRODUCTION

Publishing of data in open formats on the Web has become extensive and overwhelming trend. The great idea surrounding the growth of publicly available data is the idea of Open data. The idea of Open data has a goal of making public data available to anyone to use and republish, without restrictions from copyrights, patents or other controlling mechanism [1]. Benefits were foreseen to be harvested by individuals, public and private institutions, which, in turn, additionally focused governments towards the data [2]. In its desire to become a well-informed population, civil societies are forcing public administrations to publish data and increase their transparency [3]. As a response, Open Government Data (OGD) and Open Data Portals (ODPs) are dominating the last decade in terms of making significant information publicly available and inhibiting public resources misuse.

Traditionally, Open Government Data (OGD) originates from governments [4], both in the form of recourses offered and as a philosophy of promoting transparency and making government data available to all. Open Government Data in most cases consists of documents regarding economy, education, health, transportation and other records relevant to society. Envisioned as free to be redistributed, it is expected to enhance transparency and economic growth [5]. In support to this statement, a study financed by European Commission in 2018 foresees public sector information total value growth to reach 215 billion in the next decade [6]. By representing the data from various fields, such as economy, environment, social care and agriculture, Open Government Data, as well as other parts of Open data, need efficient tools for publishing and discovering the data. Hence, Open Data Portals (ODPs) emerged as means for publishing and browsing OGD.

From the user standpoint, Open Data Portals (ODPs) are perceived similar to digital libraries. They contain catalogues and use datasets for file grouping purposes. File formats, as well as publishing policies, may vary. Datasets can be downloaded manually or programmatically - by querying these portals using their APIs [7]. Portal administrators are usually responsible for data publishing process. They organize data into categories and, more significant, assign specific metadata to datasets while organizing them. Dataset metadata is basic descriptive information in structured format regarding the published resource. Metadata is organized as key-value pairs (meta-keys and values), where the key represents the property label while the value holds numerical or textual representation. Using these metadata, ODP developers provide users with ability to create more or less advanced search queries and narrow the results of their interest.

Being able to accurately and efficiently discover the needed data is crucial for OGD consumers. For this to be possible, ODP administrators and Open data publishers use metadata to describe the data in an accurate and comprehensive manner. Thus, metadata becomes the essence of OGD discovery and usage. In contrast, having incomplete, or even worse, inaccurate metadata, hinders ODP usability, decreases the significance of the available data and prevents users from finding the data they need.

In this paper, we will present a comparative analysis of the state of the dataset's metadata and changes in terms of the growth of the number of datasets on selected portals. We will investigate meta-keys used for categorization and labeling the datasets since these offer an intuitive way for users to search for datasets. We will primarily concentrate on the analysis of this meta-keys completeness and the way datasets are usually categorized and tagged, the number of categories datasets are usually assigned to, and the number of tags that usually label one dataset.

II. RELATED WORK

Many countries worldwide have shifted their focus away from the quantity of data that is being made publicly available. Instead, governments and agencies are focused on uprising interest regarding the quality of open data and metadata. Improving the quality of metadata is expected to enable the development of high quality applications capable of reusing available OGD. Further, metadata usage is expected to provide the ability to learn OGD insights and develop new services relying on contemporary technologies emerging from artificial intelligence. The European Commission, in its Open Data

Maturity report for 2021 [8], observes OGD in four dimensions: Policy, Impact, Portal and Quality. As reported, special attention was given to the fourth dimension (Quality), which stands for "open data quality". More precisely, this dimension is focused on ensuring the quality of both data and metadata. As stated in the report, the Commission tries to "ensure the systematic and timely harvesting of metadata as well as monitoring mechanisms that are in place to ensure high-quality publication of metadata, compliant with the DCAT-AP standard and compliant with several deployment quality requirements" [8]. A fact that stands out in this report is that the quality dimensions improved only limitedly, making it the least mature dimension of the 2021 assessment.

Since the additional recommendation regarding how metadata should be published within OGD, W3C Data on the Web Best Practices (DWBP) Working Group emphasizes that metadata should be created in a way that makes both humans and computers capable of understanding the description of a dataset [9]. At the same time, scientific community is emphasizing metadata importance for data discovery and reuse. Various researchers worldwide suggest that insufficient metadata will significantly hinder automatic dataset discovery [10], that correct interpretation and use of Open Data will not be possible unless datasets are coupled with accurate metadata [11] and that dataset discovery is irreversibly bound to metadata [12]. In response, FAIR Data Principles were suggested [13]. As envisioned, applying these principles upon OGD, public datasets would become findable (F), accessible (A), interoperable (I) and reusable (R), mostly due to metadata.

The volume of Open data is growing rapidly and this volume, as well as the variety it brings, demands for efficient access and query methods. Since ODPs are, in most cases, access points towards OGD, research community has diverted its focus towards examining Open Data portals' performance. These findings are mostly based on the quality of metadata and they report undisputable deficit in metadata; in particular, the conducted research reports datasets lacked metadata [14][15], or presented a naive version of them [16]. Whether being centralized or decentralized [15], powered by CKAN or Socrata platform [14], metadata quality needs constant monitoring - preferably using automatic evaluation platforms [3][12][17]. If automatic metadata assessment tools were to be implemented and used, the performances of ODPs would be improved through improved metadata completeness. High quality seems still to be an unachieved goal, but the benefits it is expected to bring surpass the time and the effort needed. Research community seems to be united in its demand for greater portal managers' engagement in making better use of their platforms for greater user participation.

III. ANALYSIS OF OPEN DATA PORTAL METADATA

Datasets metadata information, such as information about categories and keywords that describe datasets, has an important role in datasets findability and consequently reusability which is the key for open data to reach its full potential and value. Furthermore, the possible incompleteness of the datasets metadata accompanied by a large amount of datasets on the portal can further complicate the usage of the portal and make it even harder for users to navigate and find information of interest.

Within this paper, we analyze all available datasets and its accompanying metadata from 40 government open data portals and compare the state of these portals in July 2020 and December 2021. In this analysis, the focus is given to the open data metadata used for searching the datasets on the portal, especially meta-keys that represent categories the dataset belongs to, and keywords used for datasets description. Our goal is to investigate the state of the portals for these meta-keys, in terms of the number of datasets on the portal, the number of categories defined on the portal, and the number of tags used for datasets labeling. Further, we analyzed metadata completeness and publishing habits, in terms of the number of datasets that have the category specified, the number of categories one dataset is usually assigned to, are datasets usually described with keywords and with how many, and how this information changed over time.

For this research we used 10 government open data portals that use Socrata and 30 government open data portals that use CKAN platforms. Although the structure of metadata on these platforms differs, similar tags are used to describe the information that is the subject of analysis in this paper. Within ODPs that use Socrata platform, meta-key *category* was used to store information about the category the dataset belongs to. The meta-key *tags* was used to store keywords used for dataset's description. From the ODPs that use CKAN platform, some portals adapted the meta-keys to their needs; therefore not all portals use the same meta-keys for category and keywords. Within this research, from all ODPs that use CKAN platform, Canada ODP uses meta-key *keywords* to store keywords, and other portals use meta-key *tags* for this purpose. As for information about the category, 26.6% of analyzed portals use custom meta-keys such as *sector*, *theme*, *topic*, *theme-primary*, *theme-secondary*, *subject*, and *categorization* to save this information, and the rest of the ODPs use CKAN's default meta-key *groups* for this purpose.

The list of all open data portals that are part of this research is given in Table 1, along with the summary information about the total number of datasets on the portal (ND), a number of categories used on the portal (NC), and a number of different tags used for dataset description (NT) in July 2020 and December 2021. Furthermore, Table 1 holds the information about the percentage of datasets that are not assigned to any category (ND (-) C(%)), the percentage of datasets that are not described with any tags (ND (-) T(%)), and a percentage of datasets that have neither category nor tags (ND (-) CT(%)) for both years. Lastly, Table 1 presents the information about the maximum and the average number of categories one dataset belongs to on the portal, and the maximum and the average number of tags used for dataset's description.

As it can be seen from Table 1, most of the open data portals had a growth in the total number of available datasets. The only portal that had a huge decrease in the number of datasets 63.45% is Canada's ODP due to the data consolidation [18]. On an average scale for all analyzed ODPs, there was a 20.7% increase in the published datasets. However, after a deeper analysis of the sets of data in 2020 and 2021, it was concluded that the growth of the overall number of datasets was not the result of only adding a new datasets to the portal, but also

TABLE 1

SUMMARY INFORMATION OF OPEN DATA PORTAL ANALYSIS

ODP	ND		NC		NT		ND(-)C(%)		ND(-)T(%)		ND(-)CT(%)		MaxNC		AvgNC		MaxNT		AvgNT	
	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021
Africa	5782	6442	66	70	3146	3304	30.40	29.11	10.29	13.43	7.90	9.19	3	3	0.70	0.97	38	38	5.77	4.20
Alberta Gov	22879	26687	28	28	22697	25302	6.37	5.59	0.52	0.43	0.09	0.07	15	15	1.36	1.37	88	88	5.92	5.87
Aragon	2149	2485	22	22	1143	1281	0.00	0.00	0.00	0.00	0.00	0.00	1	1	1.00	1.00	40	40	6.51	6.36
Austria	28125	36769	14	14	4818	5095	0.14	0.11	0.16	0.14	0.13	0.11	14	14	1.04	1.03	28	29	6.26	6.73
Brazil	9393	10631	22	20	5972	7308	98.06	98.86	4.61	6.64	4.60	6.63	7	3	0.03	0.01	138	120	3.11	2.99
Brisbane	345	1210	16	16	482	592	0.00	0.00	0.00	0.00	0.00	0.00	8	6	2.27	2.11	22	20	8.23	9.96
Buenos Aires	349	422	11	12	753	360	0.00	0.47	0.57	67.77	0.00	0.47	4	4	1.18	1.25	29	25	4.47	1.62
Canada	85546	31270	19	19	17225	27551	0.00	0.00	0.00	0.00	0.00	0.00	18	19	3.41	2.03	187	231	4.40	7.34
Chile	3915	4069	23	23	1829	1886	62.43	65.86	40.95	41.09	40.66	40.82	20	20	0.88	0.81	29	29	1.68	1.76
Colorado	1997	2090	47	46	1872	1963	3.10	4.78	10.87	13.30	1.80	3.83	1	1	0.97	0.95	84	84	6.56	6.53
Dutch	14847	16294	95	94	5380	6311	0.00	0.00	51.15	52.77	0.00	0.00	10	10	1.16	1.22	50	50	2.00	2.01
Germany	35960	56032	13	13	28352	37584	6.84	6.09	8.63	5.89	1.63	1.30	12	13	2.01	1.55	718	720	10.44	12.03
Ireland	10151	12949	14	14	5070	11778	0.05	0.05	5.63	5.84	0.03	0.04	1	1	1.00	1.00	183	183	6.68	5.49
Italy	34548	50014	13	13	10054	14904	26.93	39.55	23.76	24.06	7.00	11.59	7	7	0.89	0.73	133	133	2.88	3.52
London	879	1018	18	18	1453	1489	2.16	2.46	23.09	25.05	1.48	1.47	12	12	1.58	1.62	61	61	5.49	4.97
Maryland	5125	2990	61	53	2399	2391	2.89	11.00	26.99	20.70	1.95	9.83	1	1	0.97	0.89	45	63	4.80	7.39
Melbourne	216	221	7	7	551	557	0.00	0.00	0.00	0.00	0.00	0.00	1	1	1.00	1.00	78	78	6.76	7.04
Mexico	8985	9444	12	12	9821	9891	38.59	38.45	3.83	3.23	3.06	2.29	1	1	0.61	0.62	172	172	5.70	5.51
New Jersey	501	524	11	11	280	351	0.60	1.91	17.56	16.60	0.00	0.95	1	1	0.99	0.98	19	30	2.67	2.93
New South Wales	10611	16065	14	14	3270	7804	95.32	69.82	15.47	10.66	14.84	9.67	2	2	0.05	0.30	72	25	1.65	3.11
New York State	4811	5267	20	16	4928	5249	3.06	2.73	17.56	18.00	1.48	0.74	1	1	0.97	0.97	51	51	4.68	4.64
New Zealand	15493	31009	35	38	6576	9450	55.91	57.53	25.10	50.86	0.61	24.47	7	7	0.69	0.62	112	84	2.60	2.16
Niagara	310	300	17	17	1052	1071	20.65	22.33	32.90	33.33	4.84	5.00	7	7	1.05	1.04	52	52	4.45	4.56
Oregon	2205	2189	9	9	839	869	25.85	28.00	37.64	39.56	24.44	25.99	1	1	0.74	0.72	18	18	2.42	2.48
Pennsylvania	383	637	20	24	582	794	1.31	3.45	31.07	29.20	1.04	2.35	1	1	0.99	0.97	21	24	4.52	5.39
Queensland Gov	2793	3189	31	31	1975	2147	9.17	18.78	15.04	14.27	1.25	2.23	4	6	1.07	0.96	21	29	3.43	3.43
Romania	1897	2526	10	10	1013	1102	96.42	97.19	34.53	29.22	32.74	27.83	3	3	0.04	0.03	28	28	2.32	2.23
Slovakia	2256	2754	0	0	859	1056	100.00	100.00	22.03	19.72	22.03	19.72	0	0	0.00	0.00	13	13	1.97	2.11
SouthAustralia	1632	1670	14	14	2252	2387	9.68	11.56	3.31	2.46	0.92	0.72	6	6	1.02	1.01	53	53	4.59	4.21
Surrey	380	395	12	12	757	781	11.84	15.44	0.26	0.76	0.00	0.51	3	3	1.00	0.96	32	32	4.93	4.98
Swiss	7547	6650	24	24	4179	5315	4.07	4.98	26.42	35.56	3.95	4.84	24	24	1.46	1.58	478	478	10.80	8.89
Texas	5027	5550	34	31	2350	3127	28.25	36.00	38.97	40.86	21.88	25.32	1	1	0.72	0.64	46	59	3.26	3.76
UK	54909	51600	25	27	11102	12568	38.17	45.90	74.51	67.74	12.70	13.68	11	11	0.87	0.75	269	269	1.39	1.75
Uruguay	2281	2336	22	22	571	692	3.16	2.78	6.01	4.49	2.41	1.63	3	4	1.07	1.08	15	15	2.27	2.36
USA	212090	337304	20	7	60114	304160	89.05	94.85	70.23	2.22	66.47	1.50	7	2	0.11	0.05	1458	32719	2.02	17.30
Vermont	221	283	12	10	114	113	9.95	25.80	19.46	45.23	7.24	23.67	1	1	0.90	0.74	12	12	2.33	1.28
Victoria	4936	5442	15	15	3113	3239	0.83	0.64	1.32	0.06	0.00	0.00	1	1	0.99	0.99	22	25	4.90	4.47
Washington	2915	2485	15	18	1090	998	28.10	11.71	27.75	14.61	22.71	8.53	1	1	0.72	0.88	84	57	3.44	3.93
WesternAustralia	1923	2421	59	58	3020	3145	57.31	63.49	10.24	8.30	7.80	6.28	9	8	0.85	0.67	26	26	7.19	5.64
Western Pennsylvania	333	332	14	15	1067	1115	24.62	34.04	1.50	1.51	0.30	0.30	2	2	0.76	0.66	34	34	6.51	6.79

ND - total number of datasets
NC - total number of categories
NT - total number of tags
ND(-)C(%) - percentage of datasets without category
ND(-)T(%) - percentage of datasets without tags
ND(-)CT(%) - percentage of datasets without category and tags
MaxNC - maximum number of categories per dataset
AvgNC - average number of categories per dataset
MaxNT - maximum number of tags per dataset
AvgNT - average number of tags per dataset

of removing some of the existing ones. After comparing datasets identification numbers in 2020 and 2021, we concluded that on 23 portals more than 90% of datasets from 2020 analysis remained on the portal in 2021. On the other hand, on 8 open data portals no more than 50% of datasets were part of the available set in both 2020 and 2021, making the overall percentage of 80.1% of datasets that were available in both data retrievals.

The number of the available categories per ODP remained the same in most of the portals, with only one portal without having available categories. However, some portals like USA open data portal had a decrease in the number of categories. Overall, there is an average growth of 28.48% in the number of tags on open data portals. Some portals, like USA, Ireland, and South New Wales ODP, recorded significantly more tags at the end of 2021 than in 2020. Buenos Aires ODP is the only portal with a significant drop in the number of tags (52.19%).

However, this increase in the overall number of tags did not affect the number of datasets without accompanying tags, or datasets without both accompanying tags and categories. For both situations, only a few ODPs, like USA ODP, had a significant reduction of these numbers. The ODPs that already had a good percentage of tagged datasets maintained that trend. However, almost half of the portals had an increase in the number of datasets without tags, having the overall percentage of untagged datasets rise on average from 18.5% to 19.14%. On the other hand, the average percentage of datasets without both categories assigned and tags defined, has decreased from 8% in 2020 to 7.34% in 2021.

Similar results were obtained for the number of uncategorized datasets, where more than half of the portals recorded an increase in the number of uncategorized datasets. Only portals that already had a good percentage of datasets with assigned categories maintained good numbers. As it can be seen, of the 40 portals analyzed, only 6 ODPs had all datasets categorized in 2020. In 2021, of those six portals, five kept that trend, while one had 2 datasets uncategorized. Several ODPs had a very poor percentage of data categorization in 2020 with more than 85% of datasets being uncategorized and on most of these portals the numbers remained similar in 2021. Overall the percentage of uncategorized datasets for all portals within this research in 2020 was 24.78% and in 2021 it grew to 26.28%.

Depending on the open data portal, one dataset may belong to one or more categories. Within our research, we have 14 ODPs where datasets belong to maximum of one category. From those 14 portals, 10 use Socrata platform and 4 use CKAN. On other open data portals, it can be noted that there are datasets that belong to multiple categories, in some cases to all categories defined on the portal. However, those situations are rare and one dataset usually belongs to a few available categories. The number of tags used to describe one dataset depends on the portal and the dataset itself, but it is common to use multiple tags to label the dataset. As it can be seen in Table 1, there are cases where a lot of tags were used for a description of a particular dataset, but those situations are not that common. Therefore, the average number of tags used to describe one dataset in 2020 was 4.55 tags, and in 2021 4.99. This increase is mostly due to the significant changes on the USA open data portal whereas on the other portals percentages remained similar. Visualization of the average number of tags and the average number of categories per dataset on all analyzed open data portals is presented in Figure 1 where clustered columns represent the average number of tags per dataset, and dots represent the average number of categories per dataset in 2020 and 2021.

IV. Conclusion

In order for open data to reach its full potential, it needs to be discoverable, easy to access and reuse. The quality of datasets metadata has a direct effect to datasets discoverability and reusability. Therefore, in many countries, the focus has recently shifted from the quantity

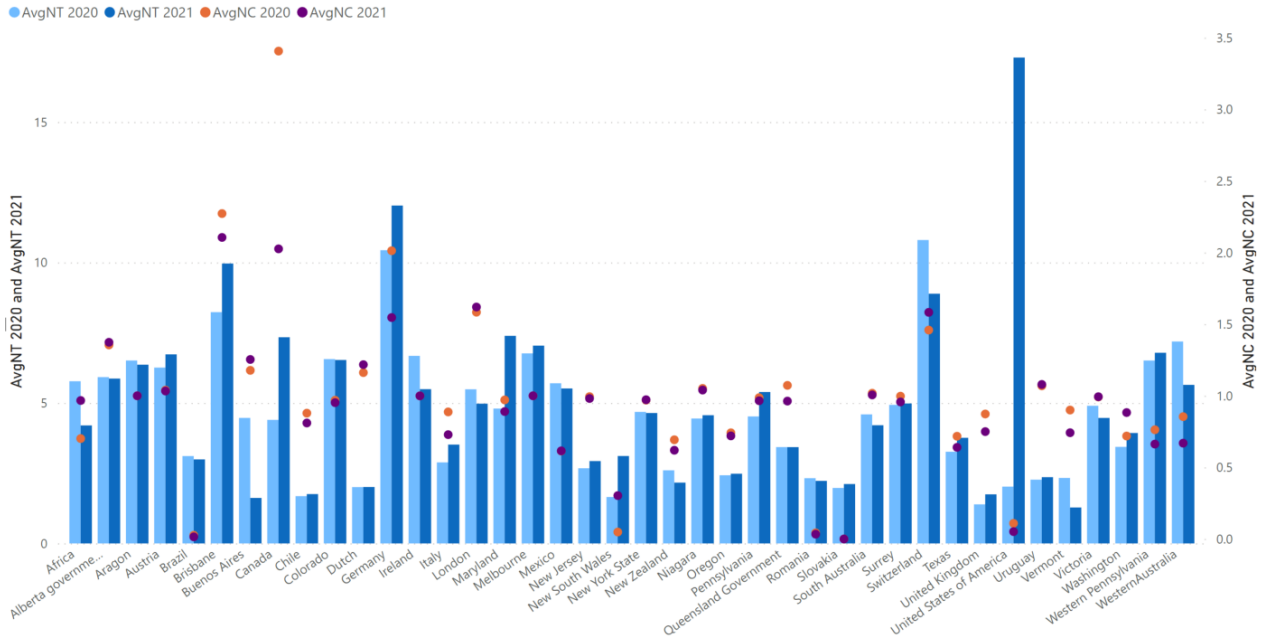

Figure 1. Average number of tags and average number of categories per dataset per open data portal in 2020 and 2021

of published data to the quality of both published data and its metadata. Within this paper our aim was to investigate the quality of datasets metadata that is usually used for filtering datasets on the selected portals in 2020 and 2021, and to compare the results. Searching for datasets by meta-keys that represent categories and tags are recognized as intuitive search modes for open data portals' users. Therefore, our analysis was focused on investigating completeness of these meta-keys and analyzing publisher's habits in terms of categorizing and tagging datasets.

From the presented analysis, it can be noticed that the number of datasets on open data portals has an increasing trend, but also that the meta-keys that represent categories and tags on the portals, often do not have defined values, which can drastically affect the quality of data queries. Only a few open data portals have made progress in the analyzed period in terms of reducing the percentage of uncategorized and untagged data. Most portals have maintained similar numbers. Furthermore, it was noticed that some of the portals had significant modifications in the organization of datasets on the portals, having changed the list of available categories and reorganized datasets. This indicates that these open data portals are still changing and developing. However, from the presented analysis, it can be concluded that the data publishing habits did not change much in this period of time. Furthermore, metadata quality in terms of metadata incompleteness regarding categories and tags also remained the same.

REFERENCES

[1] S.R. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, "DBpedia: A Nucleus for a Web of Open Data," *The Semantic Web*, Lecture Notes in Computer Science, pp. 722-735, 2007.

[2] N. Veljković, S. Bogdanović-Dinić, L. Stoimenov, "eGovernment openness index," *Proceedings of the 11th European Conference on eGovernment*, Ljubljana, pp. 571–577, 2011.

[3] S. Neumaier, J. Umbrich, A. Polleres, "Automated quality assessment of metadata across open data portals," *Journal of Data and Information quality*, vol. 8, no.1, pp. 2:1-2:29, 2016.

[4] Davies T, Walker SB, Rubinstein M et al. The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and IDRC, 2019. DOI: 10.5281/zenodo.2668475. Published by African Minds and the IDRC.

[5] Berends J, Carrara W and Radu C. The economic benefits of open data. Analytical report 9, 2017.

[6] Barbero M, Bartz K, Linz F et al. Study to support the review of directive 2003/98. EC on the re-use of public sector information 2018; DOI: 10.2759/ 373622.

[7] Sasse T, Smith A, Broad E et al. Recommendations for open data portals: from setup to sustainability, 2017.

[8] European Commission, Directorate-General of Communications Networks, Content and Technology, OPEN DATA MATURITY REPORT 2021, Last update 02-12-2021, Written by Daphne van Hesteren and Laura van Knippenberg

[9] W3C Data on the Web Best Practices (DWBP) Working Group, Data on the Web Best Practices, W3C Recommendation 31 January 2017, Eds. Bernadette Farias Lóscio, Caroline Burle, and Newton Calegari

[10] Braunschweig K, Eberius J, Thiele M et al. The State of Open Data Limits of Current Open Data Platforms, 2012.

[11] Zuiderwijk A, Volten C, Kroesen M et al. Motivation perspectives on opening up municipality data: Does municipality size matter? Information 2018; 9: 267. DOI: https://doi.org/10.3390/info9110267.

[12] Reiche KJ and Höfig E. Implementation of metadata quality metrics and application on public government data. In 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops. pp. 236–241. DOI: 10.1109/COMPSACW.2013. 32.

[13] Wilkinson MD, Dumontier M, Aalbersberg IJ et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 2016; 3. DOI: https://doi.org/10.1038/sdata.2016.18.

[14] Máchová R and Lnenicka M. Evaluating the Quality of Open Data Portals on the National Level. Journal of theoretical and applied electronic commerce research 2017; 12: 21 – 41. DOI: http://dx.doi.org/10.4067/S0718-18762017000100003.

[15] Vetrò A, Canova L, Torchiano M et al. Open data quality measurement framework: Definition and application to open government data. Government Information Quarterly 2016; 33(2): 325 – 337. DOI: https://doi.org/10.1016/j.giq.2016.02.001.

[16] Oliveira MIS, de Oliveira HR, Oliveira LA et al. Open Government Data Portals Analysis: The Brazilian Case. In Proceedings of the 17th International Digital Government Research Conference on Digital Government Research. dg.o '16, New York, NY, USA: ACM. ISBN 978-1-4503-4339-8, pp. 415–424. DOI: 10.1145/ 2912160.2912163.

[17] Kubler S, Robert J, Neumaier S et al. Comparison of metadata quality in open data portals using the analytic hierarchy process. Government Information Quarterly 2018; 35(1): 13 – 29. DOI: https://doi.org/10.1016/j.giq.2017.11.003.

[18] Ross Thompson, Growing by Shrinking - Consolidating Data on the Open Government Portal, retrieved from: https://open.canada.ca/en/blog/growing-shrinking-consolidating-data-open-government-portal, retrieved on: February 2022.