Investigating Metadata Adoptions for Open Government Data Portals in US Cities

Fanghui Xiao
*School of Computing and Information,
University of Pittsburgh, United States. fax2@pitt.edu*

Wei Jeng
*Department of Library and Information Science & Center
for Research in Econometric Theory and Applications,
National Taiwan University, Taiwan. wjeng@ntu.edu.tw*

Daqing He
*School of Computing and Information,
University of Pittsburgh, United States. dah44@pitt.edu*

ABSTRACT

Open government data (OGD) is a valuable resource for both policy transparency and government accountability. All levels of the United States government are working hard to promote open data and its portals. However, there is still a lack of studies on local-level OGD portals in the United States, particularly on the quality of metadata adopted by these portals. By examining 200 US cities, a list of 112 local-level portals is sampled and we investigate the current usages of open data platforms for building local-level OGD portals. This study further investigates and discusses the adoption and potential issues of metadata on those OGD portals. Our result findings discuss the platform distributions among US local-level OGD portals, and also highlight several critical issues associated with metadata on the portals. We anticipate the results will inspire further studies on identifying solutions to improve the metadata and enhance the usability of open government data portals.

KEYWORDS

Open Government Data (OGD), Metadata, data standard, open data platform, open data platform

INTRODUCTION

Open government data (OGD) refers to any data or information generated by public bodies, such as all government levels, and are made available to the public to access, reuse, and redistribute without copyright restrictions (Kassen, 2013; Veljković et.al., 2014; Ubaldi, 2013). OGD has proven to enhance policy transparency and government accountability, and augment both economic and social development (Assaf et al., 2015; Ubaldi, 2013; Thorsby, 2017). It not only engages the public with better knowledge and understanding of government affairs, but also helps the government with better informed improvements on their services or policies (Assaf et al., 2015 & Ubaldi, 2013). Manyika et al. (2013) also state that open data could produce $3.2 to $5.4 trillion in economic value per year across several domains of the global economy, such as the fields of education and transportation. In addition, certain types of open data, such as transportation and air quality data, could stimulate innovations for improving the quality of people's lives (Assaf et al., 2015).

An OGD portal can be defined as "*an official web-portal launched at the federal or local level aimed at making certain types of governmental datasets publicly accessible via internet in a machine-readable format* (Kassen, 2013)." To date, as the concept of openness and data science related application are flourishing, the number of OGD portals has also been increasing rapidly in the United States. OGD portals can be divided into federal, state and local (cities and counties) levels. This paper focuses on local level OGD portals, the reason being that there is little prior work analyzing local level government open data. In addition, a local level OGD is important because of its closer connection with local organizations, neighborhoods, and communities, which could have more direct impacts on a citizen's daily life and neighborhood. For instance, New York officials indicate that analyzing open data about building inspections helps improve predictions of fires (Thorsby et al., 2017).

Despite the rapid adoption of OGD portals, there are still many challenges in enabling effective and user-friendly usages of open government data, such as data quality concerns (e.g., Conradie and Choenni, 2014; Huijboom & Van den Broek, 2011; Janssen et al, 2012), lack of common standards (Huijboom & Van den Broek, 2011; Zhang et al., 2015), and understandability issues about OGD (Weerakkody et al., 2017; Bayley et al., 2013; Verburg & Neumann, 2011). According to Neumaier (2017), metadata quality issues in OGD portals have been recognized as a crucial challenge for broader adoption and also as a barrier for the overall success of open data. In fact, many existing studies indicate there are many metadata issues in OGD portals. For example, in some cases the provided metadata was scarcely checked in terms of its correctness (Kučera, Chlapek, & Nečaský, 2013), and there is a lack of descriptive metadata, causing misinterpretation of government information (Dawes, 2010). Metadata plays an important role in OGD because it allows datasets to be more searchable, discoverable, interpretable and reusable. Incomplete metadata may influence users' ability to discover and locate the datasets they need, while incorrect or meaningless values for metadata elements could mislead users with misinterpretations of the data, and even wrong usage of the

data (Neumaier, Umbrich, & Polleres, 2017). Therefore, having comprehensive metadata schemas and accurate metadata element values are crucial to the usage and overall mission of OGD portals; not to mention the fact that metadata also plays a central role in long-term preservation of data on those portals.

However, a recent study about metadata quality ranked data portals in North America at the bottom (Kubler et al., 2017). This study examined 259 open data portals across 43 countries, using Analytic Hierarchy Process method, and North American portals range between 130-220. This motivated us to look more closely at the metadata of US local government data portals. Since there are few studies on this topic, we first examine the current adoption of various open data platforms in local government data portals. This is because open data platforms often have corresponding metadata schemas, which could influence the adoption of metadata on the platforms. Furthermore, metadata schemas, as well as the naming and values of metadata elements, can sometimes be changed during the development of open data portals. This is beneficial for providing flexibility in the metadata to more tightly associate with the actual data on the portal, but it could also cause inconsistencies and difficulties for others to understand and use the metadata. Therefore, our study examines the commonly available metadata schemas such as CKAN, Socrata, and ArcGIS, and also studies their adoptions on local OGD portals. One outcome of our study is to propose suggestions for improving the information descriptions on these open data portals. In summary, the research questions of this study are:

RQ1: What is the current state of the adoption of open data portal platforms and the corresponding metadata schemas in major US cities?

RQ2: What are the existing metadata issues in US local government data portals?

LITERATURE REVIEW
Open data platforms
There are many open data platforms: some are developed by commercial companies, and others are open source projects with the support of active communities. Among the platforms developed by commercial companies, Socrata and ArcGIS are most commonly used. Socrata allows government organizations to publish their data online and make the data accessible, searchable, and usable by citizens (Socrata, 2018). ArcGIS Open Data is specialized in geographical data management, helps users collaborate on projects and access maps and apps, and has both analytic and administrative usages (ArcGIS Open Data, 2018).

In comparison, Junar, OpenDataSoft, and OpenGov are less common commercial platforms. The purpose of Junar is to assist governments, organizations or companies in publishing their data and to make the data reusable by providing tools, visualizations, maps, dashboards and APIs (Junar, 2018). OpenDataSoft is designed for business users to publish, share and re-use data, with the intention of inspiring new applications (OpenDataSoft, 2017). OpenGov provides cloud-based solutions that focus on transforming the way governments budget, measure performance and engage the public. (OpenGov, 2018).

There are two well-known open source open data platforms. Comprehensive Knowledge Archive Network (CKAN) is more commonly used. Its goal is to provide tools to ease publishing as well as to enable the finding, usage, and sharing of data (CKAN, 2018). DKAN is the other open source platform. Its goal is to provide freedom for both originations to publish structured information and for the public to use the structured data (DKAN, 2018).

All these platforms have some form of metadata schema to help describe the data stored on their platforms. Socrata, ArcGIS and CKAN operate with their own metadata schemas, whereas OpenDataSoft, Junar and DKAN all follow the DCAT standard in their metadata schema design.

Metadata in OGD
Metadata has been recognized as a significant factor in discovering, managing, preserving, and migrating data, but there is limited scholarly research on metadata within the context of open government data (OGD). Martin et al. (2017) conducted a study that evaluates the quality and usability of open data for public health research (three open data portals: federal, state, and city levels). They conclude that standardized metadata can improve the usability of open data and is likely to lead to better understanding of the data's limitations and the process of data collection. Assaf et al. (2016) introduced prevalent metadata models and harmonized them. Lisovska (2016) also compared the metadata styles that have been used for OGD, and discovered a serious issue of lack of interoperability between the metadata of open data portals. Furthermore, Lisovska claims that "if open data portals expose their metadata in different standards then searching and discovering datasets across platforms is impossible." The Australian Government Productivity Commission (2017) reports that "Data standards exist but are not widely used" and points out that much of Australia's metadata does not use clear and consistent standards. In the past two years, the issue of metadata quality has started to receive more attention. Neumaier et al. (2016) developed a generic metadata quality assessment framework to assess metadata quality for different open data portals. They propose that metadata quality issues on open data portals are a crucial problem for interoperability and may influence the overall success of open data. Kubler et.al (2017) adopted

the assessment framework to compare metadata quality among OGD portals by using the analytic hierarchy process. They found that today's organizations do not sufficiently manage their published datasets, resources and corresponding metadata.

Most of the extant work focuses on the analysis of metadata styles on a conceptual level, and, to the best of our knowledge, very little research is based on open data portals in operation to examine how their metadata is practically used. Hence, this work focuses on inspecting the metadata issues of open data portals in the US.

METHODOLOGY

We believe that a well-designed OGD portal should have a clear presentation of its functions, its metadata schema, and a procedure for publishing dataset on its site, so that any user of the portal can obtain these critical pieces of information online at the portal. Consequently, the method used in this study is to directly analyze the portal's online content for answers to our research questions, rather than obtaining such information by engaging either the users or the management teams of the portals. As a preliminary study of the portals' metadata issues, our research method is adequate. We acknowledge the need to engage both the users and the management teams of the portals, which we plan to do in the future based on the specific problems we identify through this preliminary study.

Sample selection and cleaning criteria

Thorsby et al. (2017) indicate that a city's population size has a strong correlation with the possibility of that city having a local government data portal and the size of the datasets on the portals. Therefore, our study sampled the top 200 major cities in the United States based on their population (United States Census, 2018).

Besides using population as a sampling criterion, we also developed two rules for determining whether an online website is an open data portal or not. These two rules are as follows:

Step 1: check if the website uses a common open data platform, such as Socrata or CKAN. If yes, we annotate it in the data collection table.

Step 2: if a website is not an existing, well-known open data platform, it is examined based on four of Dawes' principles of open data: complete, accessible, machine-processable and downloadable (Dawes, 2010). If the platform satisfies the four principles, it is identified as an open data platform and annotated the related information. Otherwise, the website is not recognized as an open data platform.

These two steps helped us remove certain local city websites whose data do not have descriptive information or are not down-loadable, or where the data link information is missing, since these issues are unable to satisfy the very basic criteria for open data.

Data collection procedure

To locate an open data website of a city in the top 200 US cities, we performed a Google search with the query "the city of + [city name] + open data." The reason that we took this approach rather than searching inside data.gov for the local open data portal information is because data.gov does not have a complete list of local governments' open data portals. There could be some data centers in a city with different purposes; however, in our study, we only focus on the government data portals that are generally supported or incorporated by their local governments. Through repeated Google searches for the 200 cities and applying the two open data portal identification rules, in total we found 112 online OGD portals from the selected 200 cities.

Based on each identified OGD portal, we performed deeper analysis on the platform used, the metadata schema, and metadata elements. To maintain the quality of the data annotation, two authors of this paper independently annotated all 112 OGD portals. Most data extracted did not involve subjective judgement, such as the number of datasets shown on the website. The complete annotation was conducted in February and March 2018, and all results are stored in a spreadsheet.

Table 1 shows the identified platform information for a portal, which includes the city name, portal URL, number of datasets, platform information, and metadata schema information.

Parameters	Description	Examples
City's name and state	the city's name	Pittsburgh, Pennsylvania
Binary	if the city has a portal then the value is 1, if no: 0	0 or 1
Portal's link	the link of the local-level OGD portal	http://www.wprdc.org
Number of dataset	The total number of datasets that the portal hosted as of January	349
Platforms of city	The system that the portal adopted	CKAN
The model of metadata	The metadata schema that the portal adopted	CKAN

Table 1. Example of the identified platform information

Based on the results from RQ1, we manually analyzed the top three most popular platforms for their metadata collection. Table 2 shows some examples of the metadata information we collected.

CKAN	CKAN Version	maintainer	maintainer_email	contact point phone number	license	……
WPRDC	2.4.4	data steward name	data steward email		license	
Houston	2.6.2					
Phoenix	2.5.7					
Philadelphia	2.2.4	maintainer	maintainer email	maintainer phone		
Boston	2.5.7	contact point	contact email	contact point phone number	license	

Table 2. Examples of the identified metadata information from datasets on CKAN platforms

Data analysis method

Our data analysis employed multiple methods. To answer RQ1, we performed descriptive statistics analysis on the open portal platforms and metadata schemas we collected. To answer RQ2, we decided to adopt the National Information Standards Organization (NISO) standard (Riley, 2017) for assessing the functionality of the metadata on the portals. Metadata has a wide range of functions, which include assisting the systems maintaining the content and facilitating users to find, obtain and share the information with others (Riley, 2017). Therefore, whether the metadata functioned well or not can greatly influence both the portals and users, such as if there is missing data or failure to discover a dataset.

NISO provides introductory guidance to users about metadata. NISO has been recognized as an organization that develops and maintains well-known technical standards related to information objects in the LIS field, e.g., Z39.50. In its publication "Understand Metadata," NISO discusses six essential attributes of metadata designed for different uses. These attributes are:

- *Discovery* means that metadata can aid users in looking for necessary information by allowing users to search or browse. Elements could include title, author, subject.
- *Display* shows that metadata can assist users in identifying or understanding a resource. Example properties could be title or publication data.
- *Interoperability* points out that to effectively exchange data between systems, high quality metadata could help effectively match materials and improve interoperability. Examples are file type and license.
- *Digital-object management* refers to the fact that metadata can provide information that points out the digital material or appropriate version to satisfy user needs, which supports digital-object management. The elements, for example, cover creation data/time and license terms.
- *Preservation* means that metadata plays a significant role in long-term preservation. For example, technical metadata and preservation metadata help verify content integrity, especially after a migration. Examples are the elements of file type and checksum.
- *Navigation* means that elements or attributes of metadata (such as tags or groups) can navigate users toward relevant information from one page or section to another, which helps users more quickly obtain the information they seek.

Consequently, our examination of issues in the metadata of US OGD portals consists of the six aspects above.

RESULT

Current state of the adoption of open data platforms and metadata schema

As shown in Table 3, among the 112 portals we collected, the top three adopted platforms are Socrata (38.4%), ArcGIS Open Data (tie, 38.4%), and CKAN (13.4%). This finding is similar to Kubler et al. (2017)'s study, where Socrata is the most common platform in North America; we would like to note that ArcGIS open data is also commonly used in US OGD portals.

Platforms	Socrata	ArcGIS Open Data	CKAN	homegrown	Others	Total
N	43	43	15	5	6	112
%	38.4%	38.4%	13.4%	4.5%	5.4%	100%

Note: The Other category includes Junar (n=2); dkan (n=2); OPEN GOV (n=1); and OpenDataSoft (n=1)

Table3. The distribution of the platforms adopted by US OGD portals (N=112)

As for metadata adoption, we found the distribution of metadata schemas is mostly consistent with the platform a portal adopts (see Table 4). The metadata schemas on each of the top three platforms (Socrata, ArcGIS Open Data, and CKAN) are developed as part of the platform. Among the remaining less common platforms, OpenDataSoft and many homegrown platforms have designed their own metadata schemas. Some platforms still decide to follow standard metadata schemas. For example, Junar adopts the RDF metadata standard as presented in Dublin Core and DCAT (Lisowska, 2016), and DKAN builds its metadata schema by selecting elements from DCAT and Project Open Data (DKAN Docs, 2017).

Metadata schema	Mapping to platform	N
Socrata	Socrata	42
ArcGIS Open Data	ArcGIS Open Data	43
CKAN	CKAN (n=15); homegrown (n=1)	16
homegrown	homegrown (n=4); Socrata (n=1)	5
Dublin Core and DCAT	Junar	2
DCAT & Project open data schema	DKAN	2
OpenDataSoft	OpenDataSoft	1
No metadata	OpenGov	1

Table 4. The cross-tabulation of platforms and corresponding metadata schema.

However, there are several exceptions. For example, Denver's OGD portal uses a homegrown platform, but adopts CKAN's metadata schema. Through investigation, we found that this is because Denver simply borrowed the metadata schema from its state's OGD portal, which uses CKAN as the platform. This is why there is one homegrown platform that uses CKAN metadata schema in Table 4. Another exception is San Francisco's OGD portal (DataSF), which uses the Socrata platform but with its own metadata schema. This is why the Socrata platform has one instance with a homegrown metadata schema. DataSF claims that their open data policy requires a metadata standard to assist users in understanding data, so they developed the DataSF metadata standard (DataSF, 2018).

Some portals totally abandon metadata in their platform. For example, Santa Clarita employs the OpenGov platform for its portal, but does not integrate any metadata information for its data. Instead, it provides various visualizations, including percentage graphs and line graphs, to provide means for describing its data on the portal.

Existing metadata issues in the US local OGD portals

To answer RQ2, as stated, we employed NISO metadata standards as the criteria and examined metadata based on six aspects: *discovery, display navigation, interoperability, digital object management, and preservation.* Our examination concentrates on

analyzing the top three metadata schemas used by US OGD portals: Socrata, ArcGIS and CKAN. In the rest of this section, we will present the metadata issues according to these six aspects.

Discovery. Socrata and CKAN provide reasonable metadata information for users to discover data, but ArcGIS only has limited support for this aspect. Socrata and CKAN include attributes of title, categories or group of datasets, and tags, which provide basic functions to assist users in searching and browsing data. On the other hand, the default metadata in ArcGIS includes the elements of title and tags, but there is no attribute in the group level, i.e., categories or group, that could prevent users to find data from browsing categories.

Display. The displays of Socrata, CKAN and ArcGIS are clear. As shown in Figure 1, Socrata displays the metadata under the dataset and highlights important information, such as updated time and downloads times. This arrangement clearly presents the metadata information. CKAN presents the metadata elements closer to the relevant dataset, which highlights the association clearly (see Figure 2). Compared to Socrata and CKAN, metadata elements displayed on an ArcGIS interface are less efficient. When we reach a dataset page (see the screenshot in Fig 3), "View Metadata" must be clicked in order to see the elements in the dataset (Fig 4).
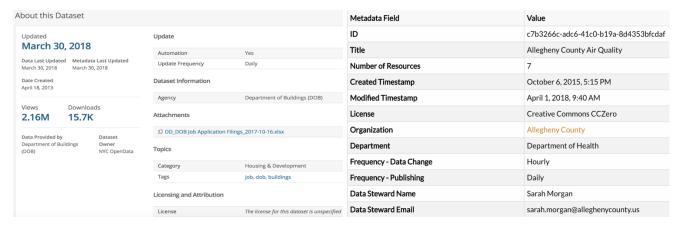

Fig.1. Socrata metadata page (from New York City OpenData) **Fig.2.** CKAN metadata page (from Western PA Reg. Data Ctr)
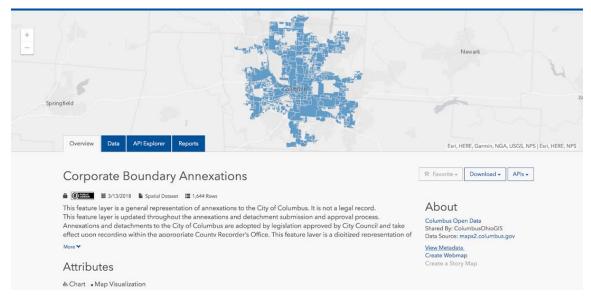

Fig.3. ArcGIS Metadata page1 (from the city of Columbus open data)
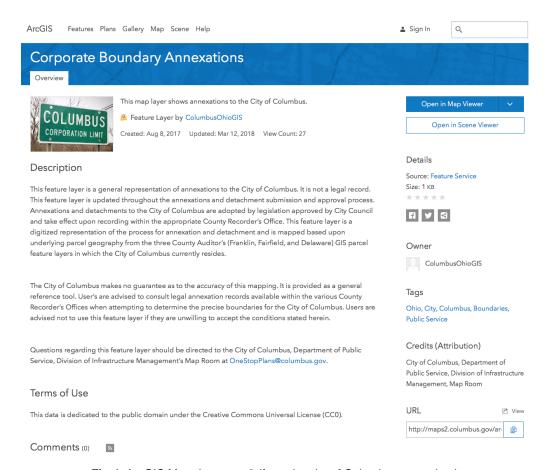
Fig.4. ArcGIS Metadata page2 (from the city of Columbus open data)

Navigation. Overall, all three metadata standards provide clear directions for users to look for relevant information and to navigate between datasets.

Interoperability. The three platforms provide very basic metadata elements for interoperability, and all of them have unique drawbacks. First, in order to effectively exchange content between systems, technical metadata and preservation metadata are essential, such as file format, file size and created or modified time. However, only CKAN has the "format" element in the resource level of the metadata schema. Socrata and ArcGIS do not include "format" as a metadata element, and only show it when users are trying to download a dataset. Also, among the three platforms' metadata schemas, only ArcGIS has the element of "file size" in their metadata information. Second, the metadata about data ownership is also crucial for interoperability. ArcGIS' metadata does not show license or right field, which is a major issue when using the data.

Digital item/object management. Socrata and CKAN provide adequate metadata support for digital object management, but ArcGIS cannot. Socrata and CKAN provide the information of "License", "Group or Categories", "Version" and "Maintainer" or "Department" to point out the datasets that could match users' needs. These two metadata schemas cover some elements that can support the digital item management, but the lack of file size information would influence long-term management. On the other hand, ArcGIS' metadata does not have "License", "Version" or even "Group" elements, crucial information for managing and delivering the appropriate content or necessary version to users. ArcGIS metadata is not helpful to digital object management.

Preservation. Socrata and CKAN have basic preservation metadata, but ArcGIS' default preservation metadata is limited. As a whole, Socrata and CKAN include administrative or descriptive metadata: title, resource, created date and license terms. New versions of CKAN also provide a group of resource metadata elements, such as package id and position, which enable long-term preservation and data management/discovery. These elements can assist in managing digital information. However, the default ArcGIS metadata has no resource, file types or license elements.

Finally, OGD's metadata schema is constantly expanding in order to cope with the demands of representing new types of data. For example, the latest version of the CKAN metadata schema added geospatial related fields. In addition, Socrata added social

media related metadata fields, such as comments, community rating, and your rating. These are not considered metadata elements in the traditional view of metadata but are added in order to improve social participation and engagement, and to augment the interactions between OGD and the public (Ubaldi, 2013).

DISCUSSION

Through the reported investigation on 112 US local OGD portals, we have the following observations.

Adoption of commercial platforms. Despite open source solutions, commercial platforms have occupied the majority of local-level OGD portals. Specifically, our results show that Socrata and ArcGIS, the two most common commercial platforms, make up more than three quarters (76.8%) of all the OGD portals we sampled. A commonly cited issue associated with projects based on open source platforms is the extra requirements of element tailoring and customization, which demand extra effort and labor with technical capabilities. We suspect this issue has triggered more local governments to hire commercial companies to work on their OGD portals. However, this deserves further investigation to obtain the exact explanation. We plan to conduct further interviews to develop deeper understanding.

The lack of investigation on a specific platform. We also found that lack of adequate studies on certain important commercial systems could reduce transparency in OGD portals. For example, there are many studies in the literature on Socrata, CKAN and even OpenDataSoft (Kubler et al., 2017; Neumaier, Umbrich & Polleres, 2016; Tygel et al., 2015), but few articles mention ArcGIS. Considering ArcGIS supports one-third of ODG portals in our samples, the unknown issues inside ArcGIS could significantly influence further development of open government data. In addition, because most of the data currently in OGD portals are numerical or geographical data, and many portals provide geographic visualizations to assist users with interpreting data or using data, it would be advantageous to pay more attention to the ArcGIS platform or other GIS tools so that the usage of open data can be improved.

The metadata quality issue. Overall, based on NISO's metadata analysis, all three platforms (Socrata, CKAN and ArcGIS) have metadata that meet the basic criteria. However, there are certain issues, such as ArcGIS' metadata not showing license or right fields and Socrata excluding geographic metadata. Another interesting observation is that our analysis shows that the quality of metadata, in terms of completeness and interoperability, is better in CKAN than in the other two. Combining this result with Kubler et al. (2017)'s results, where they found that CKAN is widely used in Europe and Central Asia, and Socrata is mostly adopted in North America, we hypothesize that the adopted platform might be one of the reasons why the metadata quality of open data in North America is ranked lower than other countries.

The issue of metadata variance. Flexibility in existing open data platforms, particularly their metadata schema customization, enables rapid development of OGD portals, but it may harm interoperability and cross-platform data discovery. For example, CKAN does not have adequate documentation to explain the meanings of the elements in its metadata schema, nor are there checking mechanisms to ensure that the values entered into the metadata are consistent with the intentions of the platform designers. Consequently, there are many misinterpretations of metadata elements in CKAN platforms. For instance, the element "Source" in the Boston portal has the value of "Internal", and it is not clear whether it refers to the interior of the publisher or the interior of the portal. In contrast, Jersey City's OGD portal assigns "Source" with the value of "Municipal Council of the city of Jersey City", which refers to the name of the source, whereas Denton city's "Source" is the URL of the data, which is the address of the dataset. When the value of the same elements in metadata adopt different data types and meanings, not only will users be confused, but the software that automatically processes the metadata will also encounter problems, thus harming interoperability and data discovery. The same problem could be caused by less crucial metadata inconsistency problems. For example, the standard element name "maintainer" in the CKAN platform is called "data steward name" at the Western Pennsylvania Regional Data Center (WPRDC) and "contact point" at the Boston's OGD portal. More importantly, inconsistency also exists across different platforms. For example, the meaning "topic" has three different names: "groups" in CKAN, "theme" in Project Open Data, and "category" in Socrata. Consequently, a balanced tradeoff between the flexibility to develop a portal and interoperability across many portals should be obtained. Also, documentation that explains the metadata schema in each portal could play an important role in helping users understand the data.

Deployment issues around metadata customization. In addition to interoperability, both metadata misinterpretation and inconsistency can affect data discovery. According to our findings, there are many issues in the development of OGD portals, particularly their metadata. However, the portals based on Socrata and ArcGIS do not have element misinterpretation problems at the same scale. This is likely because these two platforms belong to commercial companies, and there are dedicated professional teams to help the cities develop their open data portals and their metadata. This practice maintains consistency and knowledge across different local-level OGD portals on these two platforms. If such knowledge and experience can be openly shared in a best practice knowledge base among the whole community, it is likely that open government data would be greatly benefited. We conclude that professional development teams, as well as knowledge retainment and transfer, are important in open data portal development.

CONCLUSION

Open government data portals in the United States play an increasingly important role in government transparency and citizen engagement. Therefore, this study aims to examine local level open data portals in the US with two research focuses: 1) the adoption of open data platforms among the most populated US cities, and 2) the investigation of metadata schemas and elements in these portals. Our results show that commercial open data platforms such as Socrata and ArcGIS occupy the majority of the local level open data portals, although many portals also use open source platforms like CKAN. We further identified several issues regarding the development of metadata in these portals, including relatively limited metadata elements in ArcGIS platforms, and element misinterpretation and inconsistency in the portals using CKAN. These results and follow-up discussions contribute to the overall effort of improving the quality of metadata and the usability of open government data.

We acknowledge the limitations of this study. The sampled portals were obtained from the top 200 most populous U.S. cities; 74 out of the 112 cities we studied are ranked among the top 100. We anticipate that the number of cities with data portals will decrease drastically with the increase of population because a city's annual budget is related to the city's size. As a result, although we believe that our samples are representative enough for the purposes of this study, the identified metadata problems on the portals are far from being exhaustive.

In the future, we will expand upon our findings by including more US cities. Furthermore, as previously mentioned, we will further investigate metadata variance in the same platform, such as CKAN, as well as advance our study into data portals' decision-making strategies for adopting a particular platform, which can triangulate with our current study results. In addition, the interview method may be an appropriate method to deeply investigate metadata issues in OGD platforms from the angles of management and users, such as the intended use of portals, the issues facing portals and user's challenges regarding metadata.

ACKNOWLEDGEMENT

REFERENCES

Assaf, A., Troncy, R., & Senart, A. (2015). HDL-Towards a Harmonized Dataset Model for Open Data Portals. In *USEWOD-PROFILES@ ESWC* (pp. 62-74).

ArcGIS Open Data. (2018, March 6). Connecting Citizens and Businesses to Your Authoritative Data. Retrieved from http://www.esri.com/software/arcgis/arcgisonline/arcgis-open-data

ArcGIS for Desktop (2018, January 23). ArcGIS Metadata format. Retrieved from

http://desktop.arcgis.com/en/arcmap/10.3/manage-data/metadata/the-arcgis-metadata-format.htm

Australian Government Productivity Commission (2017). Data Availability and Use. Retrieved from:

http://www.pc.gov.au/inquiries/completed/data-access/issues/data-access-issues.pdf

Bayley, K. B., Belnap, T., Savitz, L., Masica, A. L., Shah, N., & Fleming, N. S. (2013). Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Medical care*, 51, S80-S86.

CKAN. (2018, March 6). Retrieved from https://ckan.org/,

CKAN. (2018, March 6). Metadata. Retrieved from https://ckan.org/portfolio/metadata/

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly, 31*(1), S10-S17. doi:10.1016/j.giq.2014.01.003

Dawes, S. S., & Helbig, N. (2010). Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In Wimmer M.A., Chappelet JL., Janssen M., Scholl H.J. (Eds). Lecture Notes in Computer: Vol. 6228. Science Electronic Government (pp. 50-60). doi: https://doi.org/10.1007/978-3-642-14799-9_5

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, *27*(4), 377-383.

DKAN. (2018, March 6). Retrieved from https://getdkan.org/

DKAN Docs. (2018, March 6). Advanced Metadata Features. Retrieved from http://dkan.readthedocs.io/en/latest/components/dataset/datasetfeatures.html

DataSF. (2018, March 6). Standard: Metadata. Retrieved from https://datasf.org/resources/metadata-standard/

Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice*, *12*(1), 4-16.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management, 29*(4), 258-268. doi:10.1080/10580530.2012.716740

Junar. (2018, March 6). Open data that is easy to re-use and interpret by many different audiences. Retrieved from http://www.junar.com/

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508-513. doi:10.1016/j.giq.2013.05.012

Kapoor, K., Weerakkody, V., & Sivarajah, U. (2015, October). Open data platforms and their usability: Proposing a framework for evaluating citizen intentions. In Janssen M. et al. (Eds), Open and Big Data Management and Innovation: Vol. 9373. Lecture Notes in Computer Science (pp. 261-271). doi: https://doi.org/10.1007/978-3-319-25013-7_21

Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2017). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1), 13-29.

doi: *https://doi.org/10.1016/j.giq.2017.11.003*

Kučera J., Chlapek D., Nečaský M. (2013) Open Government Data Catalogs: Current Approaches and Quality Perspective. In Kő A., Leitner C., Leitold H., Prosser A. (eds). *Lecture Notes in Computer Science*: Vol 8061. Technology-Enabled Innovation for Democracy, Government and Governance (pp.152-166). doi: https://doi.org/10.1007/978-3-642-40160-2_13

Lisowska. (2016). Metadata for the open data portals. Retrieved from:

http://juds.joinedupdata.org/wp-content/uploads/2016/12/JUDS-DP6-Metadata-for-the-open-data-portals.pdf

Manyika, J., Chui, M., Farrell, D., Kuiken, V.S., Groves, P., & Doshi, E. A. (2013). Open data: Unlocking innovation and preformation with liquid information. Mc Kinsey Report (2013). Retrieved from

http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

Martin, E. G., Law, J., Ran, W., Helbig, N., & Birkhead, G. S. (2016). Evaluating the quality and usability of open data for public health research: A systematic review of data offerings on 3 open data platforms. *Journal of Public Health Management and Practice, 1.* doi:10.1097/PHH.0000000000000388

Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated quality assessment of metadata across open data portals. Journal of Data and Information Quality (JDIQ), 8(1), 1-29. doi:10.1145/2964909

Osagie, E., Waqar, M., Adebayo, S., Stasiewicz, A., Porwol, L., & Ojo, A. (2017, June). Usability Evaluation of an Open Data Platform. In Proceedings of the 18th Annual International Conference on Digital Government Research (pp. 495-504). ACM.

OpenDataSoft, (2018, March 6). Retrieved from https://www.opendatasoft.com/

OpenGov. (2018, March 6). Retrieved from https://opengov.com/

Riley Jenn. (2017). Understanding Metadata: What is Metadata, and What is it For?: A Primer. *National Information Standards Organization.* Retrieved from: https://www.niso.org/publications/understanding-metadata-2017

Scorata. (2018, March 6). Retrieved from https://socrata.com/

Scorata. (2018, March 6). Best Practices for Metadata Management. Retrieved from https://support.socrata.com/hc/en-us/articles/115008609707-Best-Practices-for-Metadata-Management

Thorsby, J., Stowers, G. N., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, *34*(1), 53-61.

Tygel, A., Auer, S., Debattista, J., Orlandi, F., & Campos, M. L. M. (2016, February). Towards cleaning-up open data portals: A metadata reconciliation approach. In *Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on* (pp. 71-78). IEEE.

United States Census Bureau. (2018, February 10). American FactFinder. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, (22), 0_1.

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), 278-290. doi:10.1016/j.giq.2013.10.011

Verburg, P. H., Neumann, K., & Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology, 17*(2), 974-989. doi:10.1111/j.1365-2486.2010.02307.x

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: An empirical view from the citizen's perspective. *Information Systems Frontiers*, 19(2), 285. doi:10.1007/s10796-016-9679-1

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.