International Journal of E-Planning Research
Volume 9 • Issue 4 • October-December 2020

A Survey of Municipal Open Data Repositories in the U.S.

Bev Wilson, University of Virginia, USA
Cong Cong, University of Illinois at Urbana-Champaign, USA

ABSTRACT

Cities in the United States are increasingly embracing open data as a means of advancing a variety of interests. Promoting transparency, facilitating public engagement, proactively managing records requests, and fostering innovation in the public and private sectors are among the commonly cited motivations for this phenomenon. While there is an extensive literature on the benefits and challenges of open government data, there are far fewer empirical studies that explore and document how these initiatives are unfolding at the local government scale. This article asks what kinds of data are being made open in U.S. cities and to what extent do open data policies and related regulatory actions matter in shaping the content and structure of public-facing repositories. The authors conclude that population size and regulatory actions exert a positive influence on the amount and variety of datasets provided through municipal open data portals. Implications for the design and governance of open government data initiatives at the local level are also discussed.

KEYWORDS

Civic Technology, Data Portals, Open Data, Open Government

1. INTRODUCTION

Many cities in United States have launched open data initiatives over the past decade (Johnson et al., 2017; Thorsby et al., 2017; Mergel et al., 2018). Open data have been defined as "data and content [that] can be freely used, modified, and shared by anyone for any purpose" (Open Knowledge Foundation, 2017) and is generally expected to be provided in a machine-readable format (Janssen et al., 2012). It has been argued that if government data is released in an open format, a variety of stakeholders and users can leverage this information for purposes that support and facilitate more inclusive and effective planning (Batty, 2013). While the arguments in favor of open government data (OGD) are well-documented (Kitchin, 2014; Wirtz & Birkmeyer, 2015), the effectiveness of these initiatives in achieving commonly cited aims of greater transparency and accountability, enhanced public engagement, and value creation (both commercial and social) remains unclear (Janssen et al., 2012; Attard et al., 2015). Prior studies have noted the need for systematic evaluation of OGD initiatives (Veljković et al., 2014), but in addition to a lack of conceptual clarity, progress toward benchmarking these efforts has been constrained by insufficient empirical evidence and a limited

DOI: 10.4018/IJEPR.2020100101

understanding of how the data, technologies, and over-arching policies shape and interact with the actors and institutions involved in using these resources to co-produce social and commercial value (Janssen et al., 2012). A comprehensive response to all these gaps and issues is beyond the scope of a single article, but this paper contributes to the existing literature by characterizing the landscape of OGD initiatives at the municipal level in the United States. We consider formal regulations governing the provision of open data and present empirical evidence that documents the content and structure of public-facing data repositories in order to answer two related research questions:

- What kinds of data are being provided through municipal open data repositories in the United States?
- Do open data policies and regulatory actions matter for the content and structure of municipal OGD repositories in the United States?

This paper begins with an overview of open data and open government initiatives that highlights common goals and motivations for these endeavors. Next, we present the results of an inventory of municipal open data repositories around the United States and discuss both similarities and differences in their contents. The influence of formal open data policies and open government regulations on the kinds of data that are provided is then considered by reviewing a smaller sample of cities from the larger inventory with contrasting regulatory contexts regarding data sharing. Much of the existing literature on open data focuses on national efforts and the leading edge of research continues to be conducted outside of the United States. As a result, this article responds to a gap in the literature with respect to the provision and governance of open government data in the United States at the local level.

Empirical studies that focus on the provision of open government data hosted on public-facing portals are few in part because of "the large number of diverse data structures that make the comparison and aggregate analysis of government data practically impossible" (Attard et al., 2015, p. 400). We present a novel approach for gathering information on the contents of public-facing open data portals, then standardize the results to compare across cities, drawing inferences about the priorities and other factors shaping municipal open government data initiatives. We hypothesize that in addition to the size and by extension, financial resources of a city, the existence of a formal legislation and executive leadership are the most important predictors of the extent open government data provision. Finally, we expect that information that is part of a mandated reporting processes (e.g., crime statistics, budgets), that is easy to share due to existing standards (e.g., GTFS) and a lack of privacy concerns, and that are of interest to an audience external to government are the most likely to be published.

2. OPEN DATA AND OPEN GOVERNMENT DATA INITIATIVES

The motivations for open government data initiatives vary but mostly center on promoting participation and fostering business and civic innovation (Huijboom & Van den Broek, 2011; Robinson et al., 2009). However, what currently draws the most attention from researchers in this arena is the barriers to achieve the stated goals of open data initiatives. Janssen et al. argues that open data does not automatically guarantee transparency and participation (2012). Several major impediments need to be addressed from streamlining data publishing process, stimulating value creation at the user's end, to promoting the transformation of data governance framework. The major cause of these barriers is believed to come from the fact that most of these OGD initiatives are supply-driven (Janssen et al., 2012), giving much attention to making datasets available while overlooking the users' need, willingness and ability to use these data. However, how to measure the use and impact of open data is still inconclusive, partly due to the scarcity of empirical evidence of the generated value (Zuiderwijk & Janssen, 2014), and partly the appropriate metrics for evaluating use and impact in a complicated system under varied local contexts (Bertot et al., 2012).

2.1. Modes of Open Data

Open data is also an umbrella term where specific examples or initiatives routinely incorporate a wide array of resources, involve diverse configurations of actors, and achieve varying degrees of openness. Prior studies have argued that the conflation of open data with open government data is problematic in that it assumes "the kinds of data that might be used to deliver on the promise of open data will be held by governments" and ignores important differences between developed and developing country contexts as overlooks the role of citizens, businesses, and other non-governmental actors (Davies & Perini, 2016). Open government data (OGD) represents an important mode or manifestation of open data within the broader open data framework, and is centered on authoritative information collected, stored, and shared by government entities acting unilaterally or in concert with other actors (e.g., nonprofit organizations, corporations). In addition to open data provided by government, the volume of volunteered geographic information (VGI) that is being produced by "citizen sensors" (Goodchild, 2007) and contributed to repositories like OpenStreetMap has rapidly increased, fueled by the proliferation of GPS-enabled mobile phone (Neis & Zielstra, 2014). A recent study by Meijer and Potjer (2018) documents the range of actors as intermediaries involved in the creation and use of citizen-generated data to inform and serve as a check on public policy and governance. Academic studies often yield data, and there are increasing calls for opening these resources to other researchers and the public (Molloy, 2011; Gewin, 2016), with independent verification and reproduction of results seen as enhancing the rigor and reliability of scientific findings. Still other types of open data are generated through public-private partnerships (Susha et al., 2017) similar to that offered by Google's Waze navigation app to cities and transit operators around the world (Waze, 2019).

In contrast to diversity in its sources and aims (e.g., government transparency, reproducibility of research), the growth of open data has fueled convergence in the tools and infrastructure used for its storage and dissemination. Public-facing data portals are increasingly common among local governments, evidenced by a crowdsourced list of portals maintained by the Open Knowledge Foundation (2019). In the United States, proprietary options like Socrata and ESRI's ArcGIS Portal and the open-source CKAN are the most commonly used platforms for the provision of open government data, which can be attributed to factors like familiarity in the case of ESRI products, ease of use for Socrata, or cost considerations with CKAN. The practical aspects of open government data have also led to establishment of data standards such as the General Transit Feed Specification (GTFS) and the General Bikeshare Feed Specification (GBFS), which facilitate the development of apps that can leverage real-time information related to these to modes of transportation (Zack, 2019). Another example is Open 311, which provides a common framework for storing and sharing data related to non-emergency calls placed to city government for services or information (Nalchigar & Fox, 2018).

2.2. Challenges for Municipal OGD Initiatives

In the United States, much of the foundation for open government data was laid at the federal level by the Obama Administration with the implementation of measures designed to make government more transparent, participatory, and collaborative (McDermott, 2010). Many of these provisions were adopted by local governments, including the establishment of clear standards for the collection and sharing of information on public-facing websites and portals, but there are also important differences in open government data initiatives at the city versus national level. The nature of the information collected by local governments is different, with a higher degree of granularity and as a result, more obvious opportunities to engage questions of privacy (e.g., mapping of reported crimes, identities of property owners, etc.). Lourenço (2015) offers a series of criteria for evaluating how well open government data portals advance transparency aims and uses this list to evaluate seven portals maintained by national governments. The author concludes that "the structure and organization of open government portals is not suitable to support transparency for accountability" and do not adequately engage the question of what data should be reported or made available (Lourenço, 2015, p. 331). Mergel et al. (2018) reviewed open data policies for cities in the U.S. and conducted interviews

with 15 city managers to evaluate the achievement of the stated aims of those policies and conclude that the "vast majority of U.S. cities included in our study engage in product innovations, only very few seek to intentionally develop internal process innovations" (p. 629). Each of these prior studies suggests that there is a disconnect between the range of purported benefits of open government data and existing evidence that those positive impacts are currently being realized. In addition to a clear leadership role for larger cities, the way that local governments approach the task of managing open data initiatives also varies. While most studies conclude that open government data initiatives at the local level have not yet delivered on the more ambitious goals that are typically used to frame and justify these efforts (Norris and Reddick, 2012), other research argues that it has demonstrated the potential to catalyze civic engagement and to serve as a basis for advocacy by creating new audiences and actors (Kassen, 2013; Meijer & Potjer, 2018; Wilson & Chakraborty, 2019) within the open government data ecosystem (Dawes et al., 2016).

2.3. Governance of Open Government Data

Scholars have reached consensus on the need for changes in how the public sector views and manages open data in the future. Blauer (2018) argues that cities must begin thinking about data as "an enterprise asset" and understand that open data that are shared with the public represent only one element of what should be a larger and more comprehensive data governance strategy (p. 157). While some local governments have adopted an "open by default" orientation where all information that is not subject to legal protections or is not deemed sensitive from a privacy standpoint, are collected and maintained for eventual publication, other jurisdictions have not yet fully made this cultural and managerial shift. Welch et al. (2016) hypothesize that "coercion, persuasion, technical management capacity, and technical engagement capacity" are the most important determinants of local government data sharing and use responses from a survey of 2,500 city managers in the United States to explore their effects. While the existing of formal data sharing policies were not found to be useful predictors of actual data sharing, the authors conclude that by cultivating the "human capital and technical infrastructure needed to respond to data sharing demands" (p. 400), local governments may be able to better align the needs of individual departments or agencies with the broader aims of open government data initiatives. A broader interpretation of data governance that includes not only regulatory provisions, but also engages target user groups within and outside government is also important. In order to achieve these goals, local government must play multiple role within the open government data ecosystem. For example, local government is expected to serve as a catalyst for data use and reuse, (Zuiderwijk & Janssen, 2014) and as a sophisticated data analyst and data consumer (Chui et al., 2014), in addition to its role as a data provider. Technological advancements have made co-production defined as an environment where "government treats the public not as customers but as partners" in the delivery of public services (Linders, 2013) more feasible. This paradigm shift will create "feedback loops in which the government can learn from the public" (Janssen et al., 2012, p.259) allowing cities to be more responsive to the needs of its citizens (Goldsmith & Crawford, 2014), while helping to close "the large gap between the promises of open data and what is actually realized" (Janssen et al., 2012, p. 266).

3. RELATED WORK

A number of recent studies attempt to evaluate the performance of municipal open data portals in terms of whether their content and features align with and advance key open data principles (Thorsby et al., 2017), encourage user interaction and engagement (Chatfield & Reddick, 2017; Zhu & Freeman, 2019), and maintain high-quality data products (Ozmen-Ertekin & Ozbay, 2012). Most of these studies focus on the supply side of the open government data ecosystem, analyzing the kinds of datasets provided and the functionality available to users of the portal, based on a manual evaluation of the website. Empirical studies that document these characteristics and contrast observed trends within

the stated aims and asserted benefits of open government data initiatives deepen our understanding through critique. For example, Nahon et al. (2015) evaluated OGD initiatives of 16 U.S. cities using metadata from public-facing portals maintained using a popular third-party software platform. Fields in the metadata offer insight into how cities in the sample select and categorize their datasets, and information on how often and how regularly the cities update each dataset sheds some light on specific priorities and the overall commitment to municipal open data. An earlier study by Ubaldi (2013) proposed a larger set of metrics for assessing OGD initiatives including demand (e.g., number of views and downloads per day), and re-use (e.g., number of apps developed with the data) to close the knowledge gap between data provision and its value creation potential. As application programming interfaces (APIs) become more widely available and more user-friendly, a key technical barrier to more detailed and systematic analysis of open data repositories is removed.

Methods for comparing and benchmarking open data achievements typically rely on generic assessment frameworks developed by researchers. The majority of these frameworks are developed and applied at the national level, including Open Data Barometer (Open Data Barometer, 2016), Data Openness Indicator (Bogdanović-Dinić et al., 2014), Open Data Index (Open Knowledge Foundation), and Open Data Portal requirement (Lourenço, 2015), with less focus on the specific priorities and needs of local governments. At the local level, previous studies have proposed several evaluation indices with a variety of aims and strengths. A recent study by Zhu and Freeman (2019) compared the user-interaction features of 34 municipal open government portals based on a standardized scoring framework and noted that a common limitation of cross-sectional comparisons is difficulties in understanding and accounting for difference in the context-specific experience of local users of OGD within and across cities. This point echoes earlier research that concluded understanding "why cities gravitate toward one type of OGD behavior or another would require a nuanced analysis of the socioeconomic and political context within which OGD policies are adopted, as well as an assessment of how the use of OGD by individuals, organizations, or businesses affects the city's commitment to its OGD policy" (Nahon et al., 2015, p.129). To address this gap, we extend the evaluation of OGD initiatives to include a consideration of how the type of action upon municipal OGD initiatives are based (e.g., ordinance, executive order, open data policy) may influence how these efforts unfold.

Studies that explore the determinants of where open government data initiatives are being implemented and how they are being governed have yielded inconsistent results, which invites further research. For example, population size is commonly identified as a critical factor on the openness of government data (Thorsby et al., 2017) as it is believed to be positively related to the financial resources devoted to OGD initiatives, data demand from citizens, and the technical capabilities of local government staff, yet it has lower explanatory power in studies conducted outside the U.S. (Wang et al., 2018). A recent study (Riggs et al., 2019) concluded that larger cities are more likely to have an open data repository, while only 28 percent of the full sample of cities considered (N = 602) did, perhaps underscoring the financial costs and infrastructure requirements of open data initiatives (Johnson et al., 2017). The presence of open data policy is also a factor with conceptual importance and that has been the subject of study. Chatfield and Reddick (2017) examined the effect of policy intensity with respect to promoting meaningful open data portal services in 20 large cities in Australia. The authors consider "external goals-focused policies" which tend to target enhancement of citizen engagement and creating public benefit, in contrast to low-intensity, "internal goals-focused policies" that aim at enhancing government transparency and efficiency. The study found that cities with high-intensity open data policies experienced a larger increase in the number of datasets released on their portals over a two-year period. This aligns with the common understanding that open data policies promote the implementation of OGD initiatives by providing guidance on data sharing and by building trust with the public. However, empirical studies like this typically covers a small fraction of cities engaging in open data initiatives and there is a need for research that involves larger data samples and clearer policy hierarchies in order to advance our understanding of OGD initiatives and the local level.

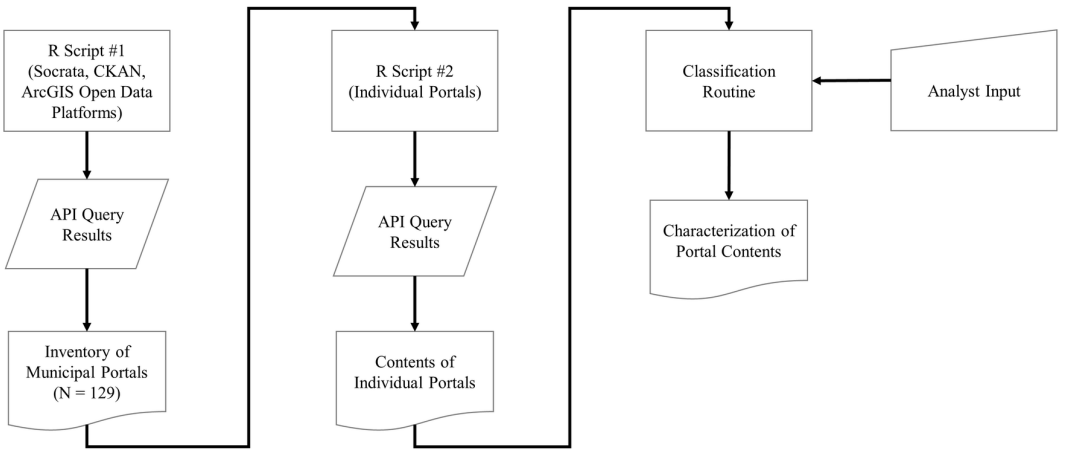
4. DATA AND METHODS

We rely primarily on data collected through APIs to characterize the contents of municipal open data portals in our sample. Information on the regulatory framework (e.g., is there an open data policy) of each public-facing data repository considered has been compiled and is used to explore the extent to which the mode of leadership matters in shaping open government data in cities.

4.1. Data Collection

In order to assess the contents of municipal open data portals, we created a series of R scripts that sent queries or requests for information to the API for the three most common open data repository platforms: Socrata, Comprehensive Knowledge Archive Network (CKAN), and ESRI's ArcGIS Open Data product (Xiao et al., 2018). As shown in Figure 1, we first collected a list of local open data portal URLs from various sources, including the Socrata Discovery API (Socrata, 2019), CKAN instances census (CKAN, 2019), ArcGIS Open Data API (ESRI, 2019), Open Knowledge Foundation's *dataportals.org* service, and the cities section of the *data.gov* repository. We adopted this approach because there is currently no single repository or centralized source with a complete list of local open data portal information (Xiao et al., 2018; Zhu & Freeman, 2019). From the initial listing of all open data portals derived from the aforementioned sources, we retained[1] those supported or initiated by municipal governments and resolved duplicated and broken portal URLs, yielding a convenience sample of 129 active municipal open data portals (Figure 2).

Figure 1. Overview of data collection and data elaboration procedures



Since 2009, over 140 local governments have adopted an official policy on open data (Sunlight Foundation, 2018) and we obtained a sample of municipal open data policy documents (Open Data Policy Hub, 2019) maintained by the Open Data Policy Hub of the Sunlight Foundation. Policies available through this library are characterized according to the legal means (e.g. ordinance, executive order) through which they were created and the date when they were passed or signed.

Table 1 presents a summary of the 129 repositories in this study, including the number of portals using the Socrata, CKAN or ArcGIS Open Data platforms; the number of datasets on the portals; whether the city has an open data policy; and the city population taken from the 2013-2017 American Community Survey (ACS) 5-year Estimates. All data were collected between August and September 2019.

Figure 2. Location of cities in the sample (N = 129)

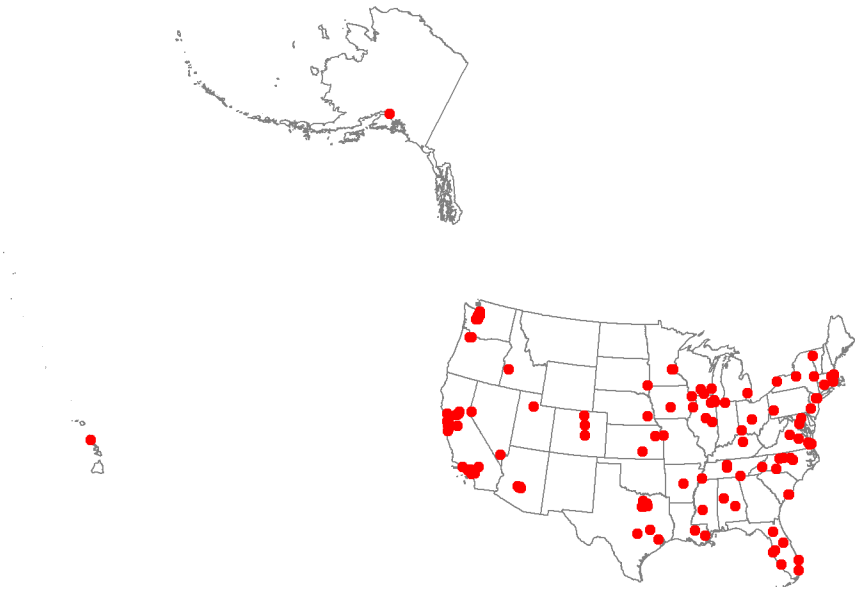


Table 1. Municipal open data portals by platform (N = 129)

Platform	Portals (#)	Data Policy (%)	Median Datasets	Median Population
CKAN	14	85.7	127	310,061
ArcGIS Open Data	43	62.8	88	230,964
Socrata	72	50.0	150	195,984
Full Sample	129	58.9	108	214,778

As shown in Table 1, the open-source CKAN platform (CKAN, 2019) comprised the smallest percentage of the overall sample, but a very high proportion of cities using this platform in the sample also had a formal open data policy in place. Municipalities where the level of commitment to open government data is high enough to support an open data policy, might also favor open-source software as part of an ethos that emphasizes digital freedom and the democratization of information (Sullivan, 2011). Because CKAN is an open-source platform, support is provided on a voluntary basis through its user community, unless a municipality contracts with a private company for paid support. Despite the lack of license fees for the software, the technical demands of deploying and maintaining CKAN may help to explain why fully proprietary alternatives like ArcGIS Open Data and Socrata are better represented in this sample.

Next, a second set of queries were sent to the platform-specific API for each of the cities in the sample to collect more detailed information about the repository contents and about specific datasets. From the results of these queries, we retrieve essential information such as the format, categories, and keywords assigned to the datasets available on the portal to support further analysis and to answer the research questions articulated in the preceding section. Some of these data elements (e.g., category) are not included or are stored under different names due to the heterogeneous metadata schemas of the three most commonly used data publishing frameworks (Neumaier et al., 2016; Kubler et al., 2017).

To address this challenge, we adapted the dataset harmonization approach of Neumaier et al. (2016) to map the specific metadata keys of Socrata and CKAN to a standardized set of names, then derived an equivalent mapping of ArcGIS metadata keys from the ArcGIS Open Data documentation as shown in Table 2.

Table 2. Reconciling metadata keys across platforms

Dataset Attribute	Socrata	CKAN	ArcGIS
Name	resource.name	title	title
Category	classification.domain_ category	groups	—
Tag	classification.domain_tags	tags	keyword
Description	resource.description	notes	description
Data format	resource.type	resource$format	distribution$format
Contributor	resource.attribution	organization.title	publisher.source
Last update	resource.updatedAt	resource$last_modified	modified

4.2. Analysis Overview

The primary goal of this study is to better understand what types of information are being provided on existing data portals in U.S. cities and how these initiatives are being governed. In addition to providing empirical data about the status of the OGD movement in U.S. cities, this study offers recommendations to help local governments design and conduct benchmarking activities. Our analysis proceeds in two steps. First, we document and analyze the current open data provision by indicators available from the metadata that differentiate the cities' activities in releasing government data. We then examine the regulatory actions regarding open data initiatives in different cities and explain the variation between cities through hypothesis tests and regression analysis. We hypothesize that cities with a high commitment to open data will consistently publish a variety of government data, engage audiences from within and outside government and facilitate data use and reuse through the design and maintenance of their public-facing portals. A series of indicators—*number*, *completeness*, *formats*, and *variety* of datasets—derived from the second round of automated API queries (Figure 1) are used to test these hypotheses in a regression modeling framework.

The number of available datasets is an important indicator because the amount of data published on the portal in many cases reflects the extent the city is willing and able to share information. This is especially true for more mature open government data initiatives (i.e., those that have been in operation for several years), where enough time has passed to allow resources to be allocated and for technical hurdles to be overcome. We evaluate the completeness of the data provided by assessing the provision of attributes that enable users to discover individual datasets and understand what is being provided (e.g., categories, tags) as well as attributes that are necessary for data use and reuse (e.g., description, spatial reference, contact information). The format in which open data are provided is consequential in that many formats presuppose a specific software in order to access and use the underlying information. However, there is more than one way to define what open means in the context of a data format. For example, open data are by definition provided in a machine-readable format, which includes but is not limited to, *xls, csv, rdf, wms, kml, shp, geojson* and *xml* (Neumaier et al., 2016). Pushing beyond the machine-readable criterion, the 5-star data format schema (5-Star Open Data, 2015) also considers to what extent data are discoverable and linked on the web. Given that open data portals exhibit some degree of standardization by virtue of the dominance of the Socrata, CKAN, and ESRI platforms,

those considered in our sample are performing well in providing downloadable and machine-readable data to users. As a result, we focus more on the variety of data formats available on the repositories that include both machine-readable data for civic-minded developers to build products and cultivate values around them, and visualization, data stories (narrative), and mapping that remove technical demands for non-technical users. The breadth of data made available to the public is measured by counting the formal categories specified on the portal (Nahon et al., 2015), but proxies like this cannot address the question of whether the information provided by the government is actually the kind of information that people need (Open Data Barometer, 2016). We standardize the formal data categories in the sample based on their semantic features as a means of characterizing the most common domains and issues addressed by municipal open data offerings. While we adapted our procedures from Thorsby et al. (2017) and Nahon et al. (2015) who performed similar re-categorization in their research, the question of how to cluster data categories across platforms and which keywords align with which grouping largely depends on the analysts' judgement. Our approach began by using the *stringdist* packages in R (Van der Loo, 2014) to form coarse groupings, then we manually reviewed and edited the results to yield 18 data categories to serve as a consistent taxonomy of municipal open data—*Government*; *Public Safety*; *City Services*; *Planning and Regulation*; *Transportation*; *Finance*; *Community*; *Education*; *Geographic Information*; *Infrastructure*; *Environment*; *Health; Economy; Housing; Recreation; Business; Arts, Culture and History;* and *Other*. We further hypothesize that the existence of formal legislation and executive leadership are important indicators of the extent and nature of open government data provision at the local level. While the adoption of an open data policy and other non-binding resolutions have symbolic value, we expect that these gestures are less likely to provide the detailed guidance, allocate necessary funds, and establish institutional infrastructure necessary to sustain municipal open data initiatives. In order to explore this relationship, we aggregate the over 30 types of regulatory actions related to open government data provision in our sample to six categories, with higher levels indicating a more formal and enforceable commitment—*Laws and Ordinances*; *Executive Orders/Directives*; *Memorandums*; *Resolutions*; *Policies and Procedures*; *Other Policy Statements*.

5. RESULTS

The central research questions focus on characterizing the contents and structure of public-facing open data repositories maintained by 129 U.S. cities in the study sample. The results of our analyses presented here are organized by the *number*, *completeness*, *formats*, and *variety* indicators described in the preceding section.

5.1. Number of Datasets

Using the approach outlined in Figure 1, we obtained data on 36,973 datasets across all 129 public-facing portals considered. The number of available datasets varies greatly ranging from five (Mesquite, TX) to 3,642 (Kansas City, MO), with a median of 108, a mean of 286, and standard deviation of 543. Most of the cities in the sample (89 percent) publish less than 500 datasets on their portals, which suggests that a few outliers provide a much larger number of datasets. Table 3 summarizes the average number of datasets by city size, region, and whether the city has an open data policy.

The number of datasets provided has a moderate positive correlation with city population ($r = 0.509$, $n = 129$, $p < 0.001$) and a significant difference ($t = 3.354$, $p = 0.001$) in the average number of datasets for cities with open data policies (mean = 402) and cities without an open data policy (mean = 127). There are also notable regional differences, with a small number of cities located in the Northeast publishing the largest amount of open data, on average.

Table 3. Number of datasets by population, region, and policy adoption (N=129)

	Number of Cities	Mean Datasets
Population		
< 100,000	25	96
< 450,000	77	176
>= 450,000	27	778
Region		
Midwest	22	370
Northeast	14	478
South	46	290
West	47	187
Open Data Policy		
No Policy	54	127
With Policy	75	402

5.2. Completeness of Datasets

Categories help users to discover potentially useful information by narrowing their search to a subset of the total datasets available on the repository that have been assigned to one or more a small number of substantive groups. Tags are a more flexible way to document and describe the contents of a dataset and as a result, tend to be more numerous and specific. Despite the efforts of dominant platforms like Socrata to standardize and simplify the hosting and distribution of open data, we found that not all datasets on the repositories studied had been assigned to categories or tagged with keywords (Table 4). We also found key differences in

Table 4. Dataset completeness by population, region, and policy adoption (N=129)

	Number of Cities	Categorized Datasets (%)	Tagged Datasets (%)	Datasets With Description (%)	Datasets With Category, Tag, Description (%)
Population					
< 100,000	25	84.2	70.8	78.5	59.9
< 450,000	77	83.1	71.3	76.3	57.5
>= 450,000	27	82.6	67.0	86.0	60.8
Region					
Midwest	22	83.2	76.1	83.2	66.5
Northeast	14	82.9	75.7	89.0	63.5
South	46	81.6	66.6	75.6	55.3
West	47	84.8	69.7	76.8	56.8
Open Data Policy					
No Policy	54	81.1	64.9	70.2	52.8
With Policy	75	84.7	74.2	84.9	62.9

the use of categories and tags across platforms. For example, datasets hosted on ArcGIS Open Data portals only have tags (keywords) assigned, but those who build and maintain these sites customize the list of categories encountered by users at the entry page to navigate by using certain tags as query keywords. Socrata and CKAN separate the use of the categories and tags by limiting the number of categories assigned to each dataset to one but allowing multiple tags and it should be noted that neither of these is a mandatory field. Surprisingly, descriptions of the dataset are also not a required component for hosting data on either of the three platforms, but providing guidance on how to understand, access, and use the data provided is essential to involving citizens and realizing many of the broader aims of open government data initiatives (Ruijer et al., 2017; Zuiderwijk et al., 2018).

On average, 60 percent of datasets are provided with categories, tags, and descriptions. This is clearly an area that should be targeted for improvement given that providing discoverable and documented data is among the fundamental tenets of OGD (Sunlight Foundation, 2014) and is important for making the data useful (Open Government Working Group, 2013). The completeness of tags is significantly lower than that of categories or descriptions, possibly due to less priority given to tagging the datasets. However, there is no statistically significant difference across cities with differing populations, those located in different regions, or with varying open data policy frameworks with respect to how comprehensively cities populate the three repository attributes. This is an interesting finding because it suggests that the level of documentation the data receive before publication is more likely driven by judgement rather than by a systematic data management process or by the needs of external users.

5.3. Data Formats

A defining characteristic of open data is that they are provided in a machine-readable format, which allows users to perform their own analyses and visualizations without additional and often time-consuming processing (Neumaier et al., 2016). There are also notable differences in how the three most common open data platforms store and categorize data based on their format. CKAN lists all downloadable resources in various formats for each dataset, ranging from conventional (e.g., .pdf. .xls), to proprietary (e.g., .gdb), to proprietary but open (e.g., .shp, .zip), to strictly non-proprietary formats (e.g., .csv, .json, .xml). Socrata integrates many of these basic data formats to create new data formats unique to the platform with names like “dataset”, “map”, “story”, and “file”, in addition to offer mechanisms for export and download of many resources. The Socrata “dataset” format is in most cases machine-processable tabular data, the “file” and “document” formats represent conventional formats (e.g., .pdf. .xls), while “map”, “chart”, “story” and “datalens” are dynamic visualizations or data analytics provided directly on the data portal. The often-cited assertion that open government data initiatives can facilitate public engagement and democratic governance (Evans & Campos, 2013; Attard et al., 2015), providing data in a variety of formats reflects an awareness of and sensitivity to the needs of different audiences and user groups. For example, geospatial data formats like geoJSON and shapefiles cater to the needs of users with the technical proficiency and software necessary to create maps, while platform-based visualizations allow non-technologists to explore spatial relationships immediately and without venturing beyond their web browsers.

In order to more easily compare contents of data repositories in our sample and to better understand the diversity of data formats provided by each city, we calculate an entropy measure that quantifies this variety. Our entropy index is based on the Shannon Diversity Index (Shannon, 1948) which has been used to measure mixing of land uses (Cervero & Kockelman, 1997; Soria-Lara et al., 2016), among other planning applications:

$$Entropy = -\sum_{i}^{n} p_i ln\left(p_i\right) \tag{1}$$

In (1), p_i is the proportion of datasets provided in each format and n is the number of unique formats that we found on each data repository. The value of entropy index varies between 0 for total concentration on one data format, and $ln\left(n\right)$ for total dispersion across the format represented on the portal. Prior studies have also used a relative entropy measure (Sarvestani et al., 2011) that normalizes the absolute entropy value to a scale from 0 to 1 by the number of variables. However, we want to account for the influence of the number of data formats in use so that in Table 5, cities with higher entropy values present both a broader array of data formats and a more even distribution across those data formats.

Table 5. Data format entropy by population, region, and policy adoption (Socrata and CKAN platforms, N = 86)

	Cities (#)	Data Formats (#)	Data Format Entropy
Population			
< 100,000	18	5	1.23
< 450,000	47	7	1.34
>= 450,000	21	10	1.33
Region			
Midwest	13	8	1.34
Northeast	12	11	1.53
South	29	6	1.22
West	32	7	1.30
Open Data Policy			
No Policy	38	5	1.14
With Policy	48	9	1.40

For this portion of the analysis, we only considered cities using the Socrata and CKAN platforms because data products available on the ESRI ArcGIS Open Data platform are more specialized and are stored in a uniform geodatabase format with no variation among repositories that use it. The mean and standard deviation of entropy index value across 86 cities in this constrained sample are 1.31 and 0.46, with the highest value (2.36) observed in Philadelphia, PA.

5.4. Data Variety

A total of 356 unique dataset categories were identified in the raw data compiled by querying the Socrata and CKAN platform APIs. We again excluded data repositories built on ESRI's ArcGIS Open Data because there is no attribute equivalent to category associated with each dataset (Xiao et al., 2018). In our constrained sample of 86 cities, the total number of dataset categories used within a given repository ranges from one (multiple cities) to 61 (Kansas City, MO) with six as the median number of categories. Most cities in the sample organize their datasets into four to eight categories, while about 26 percent of the cities present more than ten categories. We first organized the initial 356 unique dataset categories into 18 subject groups. Next, we calculated the frequency of their appearance, the number of datasets that fell within each group, and the relative size of each group measured as a percentage of all datasets provided across all 86 data repositories considered (Table 6).

The four largest subject groups (*Government*, *Public Safety*, *City Services*, and *Finance*) encompass over one-half of the total datasets available within our sample and within these four

Table 6. Descriptive statistics for subject groups (Socrata and CKAN platforms, N = 86)

Subject Groups	Example Categories	Data Portals (#)	Datasets (#)	Percent
Government	Government, Governance, City Administration, Enforcement	74	3980	17.29
Public Safety	Public Safety, Police, Crime, Safety	66	3488	15.15
City Services	311, City Services, Open Data, Utilities, Public Works	67	2544	11.05
Finance	Budget and Finance, Expenditures	56	1718	7.46
Planning and Regulation	Planning, Land Use, Zoning Code, Property, Quality of Life	33	1574	6.84
Community	Community, Social Services	38	1467	6.37
Transportation	Transportation, Vision Zero, Parking	40	1325	5.75
Education	Education, Education/Youth	18	1104	4.79
Geographic Information	Map, Geodata, Boundaries, GIS	25	1058	4.60
Housing	Housing and Development	15	769	3.34
Others		60	736	3.20
Environment	Environment, Energy, Healthy and Sustainable Environment	32	637	2.77
Economy	Economy, Economy and Workforce	39	593	2.58
Health	Health, Public Health	20	565	2.45
Infrastructure	Infrastructure	23	545	2.37
Recreation	Parks and Recreation	22	453	1.97
Business	Business	18	259	1.12
Arts, Culture, and History	Arts and Culture, Historic Preservation	15	209	0.91
Total	—	—	23,024	100

subject groups, information related to government administration (e.g., budgeting and performance metrics), crime and safety, and city services (e.g., 311 service requests, building permits, power and water supply, and recycling) were the most likely to be provided.

We again focus on the diversity of categories offered by each city and use the entropy measure (1) to quantify subject group varieties. Here, the p_i in (1) is the proportion of datasets within each subject group and n is the number of subject groups present on a city's data portal. A higher entropy index value indicates greater diversity in the categories of datasets available on public-facing portals that cover a wider array of local government domains and that potentially cater to the needs and interests of more a varied user community. The mean and standard deviation of this entropy value in our constrained sample are 1.30 and 0.64, with the highest value (2.40) observed in Philadelphia, PA. Multiple cities in the sample have entropy index values close to zero, and in these cases the datasets they have provided can be represented using one or two subject groups (Table 7). Cities with formal open data policies show significantly higher entropy values ($t = 3.946$, $p < 0.001$) and number of subject groups ($t = 4.611$, $p < 0.001$) than those without open data-related regulations.

Table 7. Data variety entropy by population, region, and policy adoption (Socrata and CKAN platforms, N = 86)

	Cities (#)	Subject Groups (#)	Variety Entropy
Population			
< 100,000	18	5	1.12
< 450,000	47	6	1.25
>= 450,000	21	8	1.55
Region			
Midwest	13	6	1.34
Northeast	12	7	1.54
South	29	8	1.19
West	32	6	1.28
Open Data Policy			
No Policy	38	4	1.01
With Policy	48	7	1.52

5.5. Regulatory Context

While there are some similarities in the way that OGD initiatives have evolved in the U.S. at the national, state, and local levels, cities and other local governments enjoy considerable flexibility in how they structure and manage these efforts. This is due in large part to the federalist system of government and the tradition of states delegating significant authority to local governments to govern themselves. Only 75 of the 129 U.S. cities in our sample have formalized their commitment to opening government data through legislation, executive orders, official resolutions, or policies. Further, nearly two-thirds (48) of those cities that have taken these steps use the Socrata and CKAN platforms, while the remainder (27) have opted for ESRI's ArcGIS Open Data platform. Cities in our sample have followed different paths with respect to developing the governing of their open data initiatives. For example, New York City passed an open data law in 2012 that directed all city departments to publish all data within one year using open standards (New York City Council, 2012). In contrast, some of the earliest open data policy adopters including Chicago, IL, Philadelphia, PA and Los Angeles, CA launched their open data portals through executive orders rather than legislative action and built momentum under strong senior leadership. For example, the 2012 executive order issued in Chicago appointed dedicated staff to lead the development of open data policies and required all appropriate datasets and associated metadata to be made available online "to the extent practicable" (City of Chicago, 2012). Formal resolutions have also been used to direct the regular and proactive sharing of government data with citizens and other portal users. The City of Oakland started its open data efforts much earlier, but its 2013 open data resolution[2] formalized those efforts (City of Oakland, 2013). Open data regulations can take many forms including administrative and internal policies, as in the case of Houston, TX, Sacramento, CA, and Durham, NC. Different types of regulatory actions vary in gravity (e.g., the binding power of legislation is stronger than plans and procedures), intensity (e.g., executive orders oftentimes require certain tasks to be completed within a maximum timeframe), and enforcement guidance (e.g., contents detailing the scope, budget, implementation, and monitoring enable sustained enforcement of open government initiatives), which may contribute to not only variation in data-sharing practices, but ultimately in the content provided via public-facing data repositories. In addition to these types of regulatory actions, we also consider the cumulative years elapsed since the first open data-related policy was adopted in the city as an important aspect of its open data regulatory context. Discussion and provision of open data policies over longer periods of

time should allow for feedback and contributions from both inside and outside of local government to influence what is provided on open data portals.

As shown in Table 8, there appears to be a relationship between the nature of regulatory actions taken in support of open government data and which subject groups are the most dominant in terms of representation on the public-facing data repository. Datasets about *Government and Public Safety* appealing during the early development stage of OGD initiatives (i.e., in operation for less than two years) or that are regulated by non-binding policies, perhaps because these datasets are easy to gather from standard administrative reporting routines (e.g., crime statistics, building permits, 311 calls) or that are already available in standard formats (e.g., GTFS). Geospatial data—including city boundaries, zoning maps, and property maps—are also easy to share for these cities, due in part to the growing popularity of digital archives and geospatial analysis software. Cities with more mature OGD initiatives have included datasets from a growing variety of sources and contributors including public schools, libraries, utility providers, and regional planning organizations.

Table 8. Number, completeness, data format and data variety by regulatory context

	All Platforms With Policy, N = 75			Socrata and CKAN Platforms With Policy, N = 48					
	Cities (#)	Datasets (#)	Datasets With Category/Tag, Description (%)	Cities (#)	Format Entropy	Variety Entropy	Main Subject Groups		
Policy Maturity									
>5 Years	18	710	61.5	11	1.47	1.55	Government	Education	Planning and Regulation
3-5 Years	39	359	59.2	29	1.37	1.53	Public Safety	City Services	Government
<=2 Years	18	186	72.4	8	1.40	1.45	Government	Public Safety	Geographic Information
Regulatory Action									
Laws and Ordinances	12	748	72.1	8	1.43	1.75	Government	Education	City Services
Executive Orders/Directives	25	510	61.4	17	1.49	1.45	Planning and Regulation	Public Safety	Finance
Resolutions	21	303	62.3	12	1.42	1.58	Public Safety	City Services	Government
Policies and Procedures	17	120	59.5	11	1.22	1.39	Public Safety	Geographic Information	Government

5.6. Regression and ANOVA

In order to more rigorously explore the observed differences in the provision of municipal open data, we estimate linear regression and analysis of variance (ANOVA) models to statistically assess the influence of conceptually important factors related to the open government data initiatives. We first tested the relationship between the number of datasets provided and four independent variables—population, region, platform and the existence of open data policy—using the full sample and ordinary least squares (OLS) regression. As shown in Table 9, the adjusted R squared value suggests that almost one-third of the variation in the volume of data provided is explained by these four factors, with population exerting a significant and positive influence upon the number of dataset available. This supports previous findings that population size is the most important consideration in predicting the number of datasets available on local government repositories (Thorsby et al., 2017). These results also corroborate our hypothesis that the public declaration of an open data policy spurs departments

Table 9. OLS regression results (Full sample, N = 129)

	Est.	Std. Error	Beta Weight
Intercept	3.93	0.38 ***	0.00
Population (100,000)	0.39	0.09 ***	0.33
Region (Midwest Reference Category)			
Northeast	-0.03	0.34	-0.01
South	-0.47	0.25	-0.19
West	-0.49	0.25	-0.20
Platform (CKAN Reference Category)			
ESRI	0.22	0.32	0.09
Socrata	0.86	0.30 **	0.36
Open Data Policy	0.64	0.20 **	0.26
Adjusted R^2	0.29		
Maximum VIF	1.28		
Shapiro-Wilk	0.98 (p = 0.37)		
Breusch-Pagan	13.50 (p = 0.06)		

Note: The dependent variable is the natural log of total datasets available.
*p < 0.05. **p < 0.01. ***p < 0.001.

within local government to more regularly and proactively share their data with citizens and other portal users.

Next, we used ANOVA to examine the influence of the same independent variables on the remaining indicators—*completeness*, *formats*, and *variety*—using a constrained sample of 86 cities due to the aforementioned lack of consistency between ESRI's ArcGIS Open Data platform and the schemas used by the other two platforms. Here, we use the number of data formats and the number of subject groups to represent *formats* and *variety*, respectively. As shown in Table 10, population measured here using three ranges (i.e., less than 100,000; between 100,000 and 450,000, and more than 450,000 residents) matters for all of the dependent variables considered. This is consistent with prior research that have noted the fiscal requirements of open data for local governments (Janssen et al., 2012; Johnson et al., 2017), but also reflects the reality that larger cities have not only greater financial resources, but are also more likely to have the technical capacity and external stakeholder engagement needed to sustain these initiatives (Wang & Feeney, 2016). Significant regional differences were found only with respect to the number of data formats represented with both the South and West regions scoring lower than Northeast region, based on Tukey post-hoc tests (Tukey, 1949).

Table 10. ANOVA results (Socrata and CKAN platforms, N = 86)

	Number	Completeness	Formats	Variety
Population Range	(2, 78) = 12.80 ***	(2, 78) = 0.95	(2, 78) = 7.46 **	(2, 78) = 6.45 **
Region	(3, 78) = 1.97	(3, 78) = 1.42	(3, 78) = 4.82 **	(3, 78) = 1.37
Platform	(1, 78) = 5.78 *	(1, 78) = 0.18	(1, 78) = 23.35 ***	(1, 78) = 1.36
Open Data Policy	(1, 78) = 4.56 *	(1, 78) = 0.15	(1, 78) = 0.73	(1, 78) = 7.15 **

*p < 0.05. **p < 0.01. ***p < 0.001.

The platform used mattered for the number of datasets provided as well as the variety of data formats represented with Socrata-based data portals offering more datasets but in a narrower range of formats than CKAN-based portals. Further, a formal commitment to open data in the form of an official policy was associated with a higher volume of data as well as greater variety in the substance of the dataset provided (Table 10). This finding suggests that at the very least, there is symbolic value in adopting an official policy that may enhance the data publishing process and increase the potential for value-creation through open government data.

Finally, we focus on the 48 cities with open data policies that also use Socrata or CKAN platforms to further explore the influence of policy maturity and regulatory instruments on *number, completeness, formats,* and *variety* using the ANOVA approach described above. As shown in Table 11, there are significant differences by region, platform, and type of regulatory action in the number of datasets provided. Cities located in the Northeast region of the United States, those using the Socrata platform, and supported by a formal Law or Ordinance were associated with higher dataset provision rates. This further supports our hypothesis that strong leadership and explicit support for open data by local government translates into more datasets and greater diversity of content on an open data platform.

Table 11. ANOVA results (Socrata and CKAN platforms with policy, N = 48)

	Number	Completeness	Formats	Variety
Population Range	(2, 36) = 2.48	(2, 36) = 0.22	(2, 36) = 1.15	(2, 36) = 1.05
Region	(3, 36) = 3.07 *	(3, 36) = 1.28	(3, 36) = 1.91	(3, 36) = 0.93
Platform	(1, 36) = 11.52 **	(1, 36) = 0.40	(1, 36) = 7.58 **	(1, 36) = 0.003
Policy Age (yrs)	(2, 36) = 3.02	(2, 36) = 1.59	(2, 36) = 0.58	(2, 36) = 0.49
Regulatory Action	(3, 36) = 2.92 *	(3, 36) = 0.66	(3, 36) = 0.57	(3, 36) = 2.40

*p < 0.05. **p < 0.01. ***p < 0.001.

There were also significant differences in the variety of data formats used on public-facing repositories with CKAN-based portals containing a wider range of data formats, on average.

6. CONCLUSION

This article presents the results of a multi-faceted analysis that characterizes the landscape of municipal open data in the United States. In addition to documenting the amount of data being provided, we have also investigated how cities organize, document (e.g., metadata), store, and distribute data through the innovative use of API queries to collect information from 129 repositories hosted on each of the three platforms. The findings of this study increase our understanding of the content and structure of public-facing data repositories and provides more empirical evidence to inform future research on open government data initiatives and their impacts. Our approach makes it possible to make comparisons across platforms and to monitor the evolution of individual cities' open data provision over time in a systematic and even-handed way. The indicators we employed are tractable without sacrificing the nuance necessary to serve as one source of feedback for the governance of municipal open government data initiatives and offer a generalizable framework for local governments to engage in benchmarking activities. We also examined the role of regulatory context in facilitating open government data provision at the municipal level, which has been less studied than state or national open data initiatives (Thorsby et al., 2017). We find that population size and the existence of regulatory actions that advance open government values exert a positive influence on the amount and variety of datasets provided through municipal open data portals. In addition, we conclude that

when local governments establish legally binding obligations of data-sharing, they are more likely to realize expected policy outcomes.

Our findings have important implications for OGD provision and research. First, open data provision requires strong and explicit support from government leaders, consistent with prior research arguing that barriers to "true openness" of open government data mainly exist in the willingness to share data and resistance borne from the risks of privacy violation and legal liability (Zuiderwijk & Janssen, 2014). Rising political pressure that emphasizes an obligation to open government data makes it easier for all actors to participate and helps to ensure that necessary financial and human resources are devoted to the OGD initiatives. Second, more attention should be paid to the ongoing refinement of data publishing processes and provisions should be incorporated into broader OGD regulations to mandate review and improvement as the needs and technical capacity of users evolves. The current dominance of three data repository platforms makes it easier for cities to learn from each other, but the lack of response in data completeness and variety to changes in the predictors considered here as well as the entropy measures used in our analysis, indicates the risk of cities simply mimicking their peers in response to the pressure of publicizing their data (Zuiderwijk & Janssen, 2014). For this reason, it is important for decision-makers to consider the foundational aim of knowledge co-production in open data governance. Although little information is available at this stage on who is using these data and for what purposes, providing mechanisms for users to suggest that a dataset be opened, holding hackathons and data literacy programs, and pursuing closer relationships with nonprofits are potential strategies for cultivating demand for data resources.

Our study also has several limitations that should be noted. First, we focus on the nature of regulation rather than the policy content, which may not fully represent the variation in the goals, scope, and values that different policies advocate within the local context nor the specific provisions these policies contain. Whether these policies are effectively implemented is also beyond the scope of this study. Questions regarding the characteristics of open data decision-making context and policy enforcement may be further investigated through case studies and interviews with stakeholders both within and outside local government. Second, we emphasize certain attributes of open data and constrain our sample in response to the availability of reliable and comparable information accessible across platforms. For example, checking the completeness of contact information attached to each dataset is helpful to see whether the published data invite feedback and suggestions from users, but it can be filled with a name of an agency or a default input (e.g., GIS Admin) on some platforms rather than a meaningful contact person on others, which limits the potential for meaningful comparison. There are other challenges in interpreting the metadata fields from data repositories that are described in further detail by Xiao et al. (2018). Lastly, our quantitative analysis may not fully reflect the underlying causes of the variation of data provision and it is likely that there are other factors—especially those internal to local government—that could be explored by future studies.

Several promising areas that are ripe for further research emerge from the work presented here, including a need to better understand the impact of OGD initiatives on government accountability and public engagement. One of the consequences of providing open data is that it can be accessed anonymously and as a result, there is little information on who is using these resources and for what purposes unless users make an effort to communicate. Cultivating a closer relationship with current and potential consumers of open government data advances the goal of co-production while also increasing the likelihood of realizing a return on the investment cities are making in open data. Finally, we would like to recommend a more consistent approach to the design and management of open government data repositories. Establishing shared data standards would make interoperability and the integration of data from a variety of sources less challenging. In addition, longitudinal analysis using API queries routinely executed over time may reveal more interesting insights into the commitment of cities to providing data that respond to the needs and priorities of residents and other users external to local government.

REFERENCES

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. doi:10.1016/j.giq.2015.07.006

Batty, M. (2013). Big Data, Smart Cities and City Planning. *Dialogues in Human Geography*, *3*(3), 274–279. doi:10.1177/2043820613513390 PMID:29472982

Bertot, J. C., McDermott, P., & Smith, T. (2012, January). Measurement of open government: Metrics and process. In *2012 45th Hawaii International Conference on System Sciences* (pp. 2491-2499). IEEE.

Blauer, B. (2018). Building the data city of the future. *The Annals of the American Academy of Political and Social Science*, *675*(1), 151–165. doi:10.1177/0002716217746359

Bogdanović-Dinić, S., Veljković, N., & Stoimenov, L. (2014). How open are public government data? An assessment of seven open data portals. In *Measuring E-government efficiency* (pp. 25–44). New York, NY: Springer. doi:10.1007/978-1-4614-9982-4_3

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D, Transport and Environment*, *2*(3), 199–219. doi:10.1016/S1361-9209(97)00009-6

Chatfield, A. T., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly*, *34*(2), 231–243. doi:10.1016/j.giq.2017.02.004

Chui, M., Manyika, J., & Kuiken, S. V. (2014). What executives should know about open data. *The McKinsey Quarterly*, (January): 2014.

City of Chicago. (2012). *Open data policy*. Retrieved from https://www.chicago.gov/city/en/narr/foia/open_data_executiveorder.html

City of Oakland. (2013). *Open data policy*. Retrieved from https://localwiki.org/oakland/Open_Data_Policy

CKAN. (2019, October 25). *Instances*. Retrieved from https://ckan.org/about/instances/

Data.gov. (2019, October 25). *Open Data in the United States*. Retrieved from https://www.data.gov/open-gov/

Davies, T., & Perini, F. (2016). Researching the emerging impacts of open data: Revisiting the ODCC conceptual framework. *Journal of Community Informatics*, *12*(2), 148–178.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. doi:10.1016/j.giq.2016.01.003

ESRI. (2019, October 25). *ArcGIS Open Data*. Retrieved from https://akharris.github.io/arc-swag/#/

Evans, A. M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management*, *32*(1), 172–185. doi:10.1002/pam.21651

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, *529*(7584), 117–119. doi:10.1038/nj7584-117a PMID:26744755

Goldsmith, S., & Crawford, S. (2014). *The responsive city: Engaging communities through data-smart governance*. San Francisco, CA: Jossey-Bass.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. doi:10.1007/s10708-007-9111-y

Gurstein, M. (2013). Should "open government data" be a product or a service (and why does it matter?). *Gurstein's Community Informatics*. Retrieved October 22, 2019, from https://gurstein.wordpress.com/2013/02/03/is-open-government-data-a-product-or-a-service-and-why-does-it-matter/

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

Johnson, P. A., Sieber, R., Scassa, T., Stephens, M., & Robinson, P. (2017). The cost(s) of geospatial open data. *Transactions in GIS*, *21*(3), 434–445. doi:10.1111/tgis.12283

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

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

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:10.1016/j.giq.2017.11.003

Linders, D. (2013). Towards open development: Leveraging open data to improve the planning and coordination of international aid. *Government Information Quarterly*, *30*(4), 426–434. doi:10.1016/j.giq.2013.04.001

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, *32*(3), 323–332. doi:10.1016/j.giq.2015.05.006

McDermott, P. (2010). Building open government. *Government Information Quarterly*, *27*(4), 401–413. doi:10.1016/j.giq.2010.07.002

Meijer, A., & Potjer, S. (2018). Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly*, *35*(4), 613–621. doi:10.1016/j.giq.2018.10.004

Mergel, I., Kleibrink, A., & Sörvik, J. (2018). Open data outcomes: US cities between product and process innovation. *Government Information Quarterly*, *35*(4), 622–632. doi:10.1016/j.giq.2018.09.004

Molloy, J. C. (2011). The open knowledge foundation: Open data means better science. *PLoS Biology*, *9*(12), e1001195. doi:10.1371/journal.pbio.1001195 PMID:22162946

Nahon, K., Peled, A., & Shkabatur, J. (2015). OGD heartbeat: Cities' commitment to open data. *eJournal of eDemocracy and Open Government*, *7*(2), 116–136.

Nalchigar, S., & Fox, M. (2019). Achieving interoperability of smart city data: An analysis of 311 data. *Journal of Smart Cities*, *3*(1), 1.

Neis, P., & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet*, *6*(1), 76–106. doi:10.3390/fi6010076

Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality*, *8*(1), 2. doi:10.1145/2964909

Norris, D. F., & Reddick, C. G. (2013). Local e-government in the United States: Transformation or incremental change? *Public Administration Review*, *73*(1), 165–175. doi:10.1111/j.1540-6210.2012.02647.x

Open Data Barometer. (2016). *The Open Data Barometer*. Retrieved from https://opendatabarometer.org

Open Data Policy Hub. (2019, October 25). *Open Data Policy Collection*. Retrieved from https://opendatapolicyhub.sunlightfoundation.com/collection/alpha/

Open Knowledge Foundation. (2017). *The open definition*. Retrieved December 3, 2018 from https://opendefinition.org

Open Knowledge Foundation. (2019). *Data portals: A comprehensive list of open data portals from around the world*. Retrieved September 13, 2019 http://dataportals.org

Ozmen-Ertekin, D., & Ozbay, K. (2012). Dynamic data maintenance for quality data, quality research. *International Journal of Information Management*, *32*(3), 282–293. doi:10.1016/j.ijinfomgt.2011.11.003

Riggs, W., Steins, C., & Shukla, S. G. (2019). *City Planning Technology, 2019 Benchmarking Study*. Retrieved September 13, 2019 from https://www.planetizen.com/features/102851-city-planning-technology-2019-benchmarking-study

Robinson, D., Yu, H., Zeller, W. P., & Felten, E. W. (2009). Government data and the invisible hand. *Yale Journal of Law & Technology*, *11*, 159–175.

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, *34*(1), 45–52. doi:10.1016/j.giq.2017.01.001

Sarvestani, M. S., Ibrahim, A. L., & Kanaroglou, P. (2011). Three decades of urban growth in the city of Shiraz, Iran: A remote sensing and geographic information systems application. *Cities (London, England)*, *28*(4), 320–329. doi:10.1016/j.cities.2011.03.002

Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423, 623–656.

Socrata. (2019, October 25). *Discovery API*. Retrieved from: https://socratadiscovery.docs.apiary.io/#

Soria-Lara, J. A., Aguilera-Benavente, F., & Arranz-López, A. (2016). Integrating land use and transport practice through spatial metrics. *Transportation Research Part A, Policy and Practice*, *91*, 330–345.

Sullivan, J. L. (2011). Free, open source software advocacy as a social justice movement: The expansion of F/OSS movement discourse in the 21st century. *Journal of Information Technology & Politics*, *8*(3), 223–239. doi:10.1080/19331681.2011.592080

Sunlight Foundation. (2014). *Open data resources*. Retrieved from https://sunlightfoundation.com/our-work/open-cities/projects-resources

Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as "bazaars"? *Transforming Government: People*, *Process and Policy*, *11*(1), 157–172.

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. doi:10.1016/j.giq.2016.07.001

Tukey, J. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*(2), 99–114. doi:10.2307/3001913 PMID:18151955

Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives*. OECD Working Papers on Public Governance, No. 22. Paris, France: OECD Publishing.

Van der Loo, M. P. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*(1), 111–122. doi:10.32614/RJ-2014-011

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

Wang, D., Chen, C., & Richards, D. (2018). A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments. *Government Information Quarterly*, *35*(4), 644–656. doi:10.1016/j.giq.2018.10.006

Waze. (2019). *Waze for Cities Factsheet*. Retrieved September 13, 2019 from https://web-assets.waze.com/partners/ccp/Waze_for_Cities-TP_v01.pdf

Welch, E. W., Feeney, M. K., & Park, C. H. (2016). Determinants of data sharing in US city governments. *Government Information Quarterly*, *33*(3), 393–403. doi:10.1016/j.giq.2016.07.002

Wilson, B., & Chakraborty, A. (2019). Planning smart(er) cities: The promise of civic technology. *Journal of Urban Technology*, *26*(4), 29–51. doi:10.1080/10630732.2019.1631097

Wirtz, B. W., & Birkmeyer, S. (2015). Open government: Origin, development, and conceptual perspectives. *International Journal of Public Administration*, *38*(5), 381–396. doi:10.1080/01900692.2014.942735

Xiao, F., Jeng, W., & He, D. (2018). Investigating metadata adoption for open government data portals in US cities. *Proceedings of the Association for Information Science and Technology*, *55*(1), 573–582. doi:10.1002/pra2.2018.14505501062

Zack, R. (2019). A Data Standard for New Mobility. Institute of Transportation Engineers. *ITE Journal*, *89*(6), 26–28.

Zhu, X., & Freeman, M. A. (2019). An evaluation of US municipal open data portals: A user interaction framework. *Journal of the Association for Information Science and Technology*, *70*(1), 27–37. doi:10.1002/asi.24081

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. doi:10.1016/j.giq.2013.04.003

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. Information Polity. *The International Journal of Government & Democracy in the Information Age*, *19*(1/2), 17–33.

Zuiderwijk, A., Shinde, R., & Janssen, M. (2018). (in press). Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results? *International Review of Administrative Sciences*.

ENDNOTES

1 Open data portals maintained by state governments, nonprofit organizations, or federal government agencies were removed at this stage.

2 A local government resolution does not carry the enforcement authority and certainty of an ordinance or law, but represents a formal position taken by elected officials on a particular topic.

Bev Wilson is an Associate Professor of Urban and Environmental Planning at the University of Virginia's School of Architecture. He is currently conducting research that highlights strategies for open data initiatives to more effectively achieve their stated goals of transparency, efficiency, and responsiveness and is also identifying ways for urban planners to more fully participate in the deployment and governance of technology in the city. The use of tools like GIS, mobile phone apps, and video to inform better planning and decision-making is a key thread connecting his research and teaching. He is also serving as Vice-Chair of the American Planning Association's Technology Division.

Cong Cong is a Ph.D. student in Urban and Regional Planning at the University of Illinois at Urbana-Champaign (UIUC), where she also received her Master's in Urban Planning and Civil Engineering. She studies and researches ways that data and technology can be used to address critical urban planning issues. She works on spatial modeling in Land Use Evolution and Impact Assessment Model (LEAM) lab at UIUC and has worked as a Data Science Research Fellow at the Biocomplexity institute at the University of Virginia. She was a transportation modeler following her Bachelor's education in Transportation Engineering at Tongji University in China.