This is the accepted manuscript (post peer-review) of an article which has been published in *Telematics and Informatics*. The published version can be found here: https://doi.org/10.1016/j.tele.2021.101678

The citation for the published paper is:

Lim, Theodore C. 2021. "Patterns in Environmental Priorities Revealed through Government Open Data Portals." *Telematics and Informatics* 64 (November): 101678. https://doi.org/10.1016/j.tele.2021.101678.

Patterns in environmental priorities revealed through government open data portals

Abstract

The ways in which environmental priorities are framed are varied and influenced by political forces. One technological advance--the proliferation of government open data portals (ODPs)--has the potential to improve governance through facilitating access to data. Yet it is also known that the data hosted on ODPs may simply reflect the goals and interests of multiple levels of political power. In this article, I use traditional statistical correlation and regression techniques along with newer natural language processing and machine learning algorithms to analyze the corpus of datasets hosted on government ODPs (total: 49,066) to extract patterns that relate scales of governance and political liberalism/conservatism to the priorities and meaning attached to environmental issues. I find that state-level and municipal-level ODPs host different categories of environmental datasets, with municipal-level ODPs generally hosting more datasets pertaining to services and amenities and state-level ODPs hosting more datasets pertaining to resource protection and extraction. Stronger trends were observed for the influences of political conservatism/liberalism among state-level ODPs than for municipal-level ODPs.

Keywords: open data, open government, environmental policy, local and state government

1 Introduction

Decreasing costs of data collection and storage technologies are resulting in an explosion in the amount of data available for analysis, and data about our built and natural environments are no exception. For example, NASA's Earth Observing System Data and Information system will ingest satellite-collected environmental data at a rate of >246 Petabytes/year by 2025 (https://earthdata.nasa.gov/about/eosdis-cloud-evolution). In addition to data collected for scientific research purposes, data are also collected for administrative purposes by all levels of government. Relating to the environment, governments may collect data that tracks the state of natural resources, pollution levels, and environmental services such as waste collection and water provision; they may also collect data in order to provide relevant information to citizens, such as the spatial locations of environmental hazards and amenities, and environmental conditions such as climate or meteorological data, topography, or land cover.

Increasingly, governments are opening their datasets to public access and use. Open government data refers to non-confidential, non-privacy-restricted data collected using public funding that is made freely available for anyone to download. These data are often hosted on open data portals (ODPs), which are online repositories that include features to aid dataset discoverability and visualization. These data reflect the priorities of governments, and data that are either made more or less accessible have been shown to correspond to transitions between political parties in the United States (Janz 2019). Data should therefore be examined for their roles within broader governance structures. For open government *environmental* data, I posit that this should include a need to understand what roles open government environmental datasets may play with respect to both theories of e-governance, and theories of environmental governance.

For this study, I hypothesized that the environmental datasets hosted on municipal and state-level government ODPs would reflect contexts of environmental governance, including: (1) differences in priorities across jurisdictional scales and (2) differences in priorities according to political leanings of the place. I found that there were statistical differences between the datasets hosted on municipal ODPs and those hosted on state ODPs. Municipal ODPs were more likely to include datasets related to environmental services and amenities, while state ODPs were more likely to include datasets related to resource protection and extraction. Preliminary evidence suggests that being politically liberal and having more financial resources are both correlated with collecting more data tracking environmental performance (such as the greenhouse gas emissions of a place), but that states with more financial resources but that are politically conservative are less likely to collect data tracking environmental performance.

The above findings have implications for understanding how and where data “gaps” are occurring, and how different places (for example: more/less conservative, urban/rural, etc) have different conceptualizations to “the environment” and humans’ relationship to nature. As is described in greater detail in the “Methods” section, parts of the analysis demonstrate how techniques such as natural language processing and machine learning algorithms can allow for automated tracking of trends over time. This makes this research distinct and complementary to other studies that have conducted qualitative research evaluating characteristics of ODPs as tools of e-governance.

In the next section, I provide more background information on theories from environmental governance and e-governance that motivated this study.

2 Background Literature

2.1 Politics and Multiple Levels of Environmental Governance

Environmental challenges, including: managing conflicts around environmental resources, reducing or cleaning up pollution, and the balance between conservation and extraction, are best understood as complex socio-ecological problems (Proctor and Larson 2005). Such problems usually do not have one solution, but involve a process of collaboration and negotiation between stakeholders and regulators involved, and are sometimes referred to as “wicked” socio-environmental problems (DeFries and Nagendra 2017; Rittel and Webber 1973). As a result, the identification and prioritization of an environmental problem and the definition of its boundaries and potential solutions are highly dependent on both social and environmental contexts, involving diverse actors and multiple levels of governance (e.g. Meadowcroft 2002).

How such problems are conceived of and framed is partially a function of our evolving scientific understanding, and by extension, what data we choose to collect and analyze about our environments. For example, John Snow’s 1854 cholera outbreak map, is often credited as being the first spatial urban environmental analysis, and was used as evidence of water-borne disease originating from a contaminated well. Snow’s collection and analysis of environmental data debunked the dominant scientific theory of *miasmas* or “bad smells” as the primary cause of disease. Government health workers then prioritized the efficient channelization of wastewater away from growing urban populations, even if that meant polluting downstream locations. Later, better scientific understanding of the hydrological cycle led to more scientists viewing wastewater and stormwater runoff as resources to be harnessed within cities to meet ecological and sustainability goals, instead of treating it as a waste stream to be disposed of (Melosi 2000).

The above example illustrates *how* different perspectives in time and space, with the help of data and information, can result in very different prioritizations of social and environmental outcomes. But, data itself is often contested politically in environmental decision-making. For example, in the Chesapeake Bay Watershed on the East Coast of the United States while better data and scientific understanding have enabled policymakers to concentrate on the contributions of nonpoint sources of pollution to the Bay, the same data and models have enabled resistance. Diffuse actors, such as rural landowners wary of more environmental regulation, have been able to impede implementation of environmental policies, through questioning the data being used for decision-making and requesting that more, or different data be collected (Layzer 2011; Lim 2021). Therefore, while scientific knowledge and data may be able to prescribe what *should* be done in order to remedy a particular environmental problem to government regulators, reliance on scientific knowledge or data alone fails to recognize the political powers that are very influential in environmental policy-making and policy implementation (French 2019).

In contrast to *government*, which refers to the top-down regulatory powers of state actors, *governance* is useful for forming theory around the how participation of broader groups of stakeholders-- such as non-government organizations, community groups, and individual citizens-- shapes the creation and implementation of environmental policies (e.g. Brondizio, Ostrom, and Young 2009; Cash et al. 2006). The institutional fit of organizations involved in environmental governance-- or matching the scale and scope of the governance arrangement to the scale and scope of the environmental challenge--is a determinant of successful environmental governance (Newig and Fritsch 2009). Therefore, both the variability in environmental conditions and the variability in governance scale (Moss and Newig 2010) will influence the data and information needs of environmental governance.

Developments in approaches to environmental challenges are also shaped by political and social trends, which influence data and information needs. With respect to environmental planning in the US, Daniels identified five environmental “eras,” which focused on different aspects of humans’ relationships to the environment, and therefore have conceptualized “the environment” in different ways (Daniels 2009). These five eras included: The period from the 19th century to the early 20th century, which was a response to industrialization that focused on urban parks and playgrounds as well as conservation of natural resources; 1920 - 1969, in which regional ecological planning and wilderness conservation grew; 1970 - 1981, the birth of modern environmental planning, which emphasized pollution clean-up and state-level planning; 1982 - 2008, simultaneously a “bridge” and “backlash” era, which introduced regulatory flexibility and financial incentives to environmental regulation and the increased roles of land trusts and NGOs; and finally 1992 - present, in which concepts such as sustainability, the global environment, and urban ecological planning became more prominent (Daniels, 2009). **Table 1** shows the data and information needs associated with the latter three eras. Although presented in eras, these three approaches to environmental policy and planning exist simultaneously today and have been analyzed through discourse analysis to understand the meanings attached to various environmental governance approaches (Dryzek 2009).

Table 1. Data and information needs associated with modern eras of environmental protection (adapted from Mazmanian and Kraft 2009)

	1970 - 1981: Birth of Modern Environmental Planning	**1982 - 2008: Backlash to regulation, bridge to sustainability**	**1992 - Present: Toward Sustainable Communities**
Goals	Regulating for environmental protection	Efficiency-based regulatory reform and flexibility	Emphasis on long-term societal and natural needs, urban ecological planning
Data and Information Needs	- firm-level emissions - waste stream contents and tracking - human health effects - environmental compliance	- costs of environmental harms and benefits of reduced pollution and economic development - environmental accounting - toxics release inventory and “right-to-know” programs	- sustainability criteria and indicators - region/ community/ global interaction effects - material and energy inventories/accounting - computer modeling of human-natural systems interactions

In addition to recognizing the importance of state-level governance within the US’ federal system, we must also recognize the role of municipal-level governance in environmental issues and cities’ political motivations for adopting sustainability goals. Municipal governments are increasingly leading efforts to adapt to and mitigate global climate change (Bulkeley and Betsill

2013). Many cities, where sustainability efforts are particularly salient with urban residents, have their own sustainability policies and plans (Saha and Paterson 2008). There are several theories that explain why sustainability is a particular interest at the municipal level, including that: urban areas tend to be more politically liberal (supporting “liberal policies” such as government intervention and provision of public services, progressive tax regimes, etc), which is associated with higher values placed on natural resource protection and environmentalism (Blomquist 1991; Clark and Allen 2004; Drummond 2010; May and Koski 2007), and in higher levels of personal-efficacy, the belief that individual actions can make a difference in environmental outcomes (Dunlap, Xiao, and McCright 2001; Lester 1995; D. Mazmanian and Sabatier 2016); that urban areas may use sustainability efforts as an indicator of quality of life in order to compete against other urban areas to retain and attract residents (Lubell, Feiock, and Handy 2009); and, that sustainability policies can help cities save money, through reducing waste of materials and energy (Portney 2009).

2.2 Open Government Data

Government open data portals (ODPs) are online platforms that serve as data repositories where anyone can download non-confidential, non-privacy-restricted data collected using public funding (Attard et al. 2015). Government ODPs are increasingly recognized as a key part of the “smart city” identity, promising to improve transparency, accountability, democratic participation, and efficiency in governance (Safarov, Meijer, and Grimmelikhuijsen 2017; Barns 2018; Townsend 2013; Yadav et al. 2017). In the United States, President Obama’s 2009 *Memorandum on Transparency and Open Government* promoted the idea of e-government and data transparency and spurred efforts to create standards for government ODP metadata and architectures in order to promote these ideals. Canada’s federal government has been developing its *Action Plan on Open Government* since 2012 (Gill and Corbett 2017). Indeed, government transparency is an important motivation for opening governmental datasets that require systematic planning of ODPs to ensure that this goal is met (Lnenicka and Nikiforova 2021).

However, there are still many questions about whether opening government data actually results in improvements to governance. There are numerous examples of powerful actors simply wielding their power through the seemingly “objectivity” of data and data analysis. For example, in an example of why it is important not only to consider the transparency of the data itself, but also the power context in which data are released, Gurstein describes how opening of land titles in Bangalore, India were used by privileged classes to exploit gaps in land titles, leverage legal resources, and grab land from the poor (Gurstein 2011). Government data coverage is also highly uneven at multiple scales and has raised concern about further deepening the “digital divide,” or creating new ones (Graham et al. 2014). The opening of government data has been shown to be dependent on: the presence of local “champions” of ODPs (Chatfield and Reddick 2018); perceptions of government “giving up” control (Janssen, Charalabidis, and Zuiderwijk 2012); admnistrative barriers, such as a lack of key policies or economic incentive (Barry and Bannister 2014); perceived risk to public servants (Wirtz et al. 2016); and the legislative and organizational characteristics of governance (Safarov 2019). A comprehensive review of open government and roles of ODPs identifies a disconnect between the ideals of open government data's capacity to increase transparency and participation and what actually occurs in practice, and points out how more empirical research is necessary to examine the theoretical relationships between opening government data and governance transformation (Tai 2021).

Previous research has explored the relationships between the features of government ODPs with the characteristics of the local population and government structure, finding a correlation between city population and the number of datasets, the different types of portal content, and the portal features and content. This is because higher populations are associated with having more economic/financial resources to implement open data initiatives (Thorsby et al. 2017). Others have carried out descriptive, qualitative, and case analyses, pointing out that ODPs have different functions, acting as data repositories, showcases, city service score cards, and data marketplaces (Kassen 2013; Lourenço 2015; Kubler et al. 2018; Barber and MacLellan 2019; Gessa and Sancha 2019; Wilson and Cong 2021; Nikiforova and McBride 2021); and that open data programs are actually part of an "ecosystems" of data, information, and technological activities, policies, and services (Dawes, Vidiasova, and Parkhimovich 2016; Goldstein and Dyson 2013). Yet, the connections between level of governance, political leaning, and a systematic evaluation of the kinds of data that are made available is lacking.

A closer and more comprehensive look at environmental datasets on municipal and state ODPs is an opportunity to learn more about the governance and political context influences on open data initiatives. In this study, I attempt to explore this opportunity across as many ODPs as possible, distilling insights from as many data sets hosted on these ODPs as possible.

2.3 Hypotheses

Given the above findings from the literature, the following patterns were hypothesized to be present in open government environmental datasets:

- Compared to politically conservative places, liberal places host more datasets related to environmental sustainability (environmental protection and performance, such as tracking greenhouse gas emissions)
- Compared to states, municipalities will seek to showcase "livability" of their communities by highlighting environmental services and amenities and would be more likely to track sustainability metrics, such as greenhouse gas emissions
- Places with more financial resources (higher populations) will host more environmental data tracking environmental performance and environmental protection

Others have addressed roles that policy could play in shaping the availability of data, the application of standards to spatial government data, and the interoperability between schools and contexts of environmental data (e.g. Mulcahy and Clarke 1994). However, the relationships between political contexts and what information is made available deserves more attention. This paper examines how patterns of place are associated with different conceptualizations and aspects of the environment. It therefore contributes to understanding of (1) how government

ODPs represent another dimension of political environmental discourse; and (2) how open data reflect the particularities of place and why they have resisted externally prescribed standardization.

3 Methods

3.1 Overview of approach

To test the above hypotheses, this analysis relies on the systematic collection of all datasets (hereafter referred to as “data objects”) present on municipal and state government ODPs and analyzes patterns from their titles and descriptions. Two complementary approaches are used. In the first approach, each data object is coded according to *a priori*, manually assigned environmental themes that correspond to different conceptualization of the environment. Once coded for the presence of these themes, the data objects are analyzed through statistical correlation and regression. A potential shortcoming of assignment of *a priori* codes however, is that it relies on the correct identification of codes, and may impose subjective structure/meaning on the titles and descriptions. Therefore a second approach, which relies on natural language processing and machine learning classification was also tested. In this approach, no *a priori* codes were used. Instead, machine learning algorithms were used to identify groups of words that tended to be predictive of what kinds of ODPs from which the data objects originated. This approach requires the analyst to then provide interpretation and meaning to the identified terms. If both approaches yielded complementary results, this was considered to strengthen the evidence either supporting or rejecting the above hypotheses. **Figure 1** gives a graphical overview of the main differences of between the two complementary approaches used in this study.

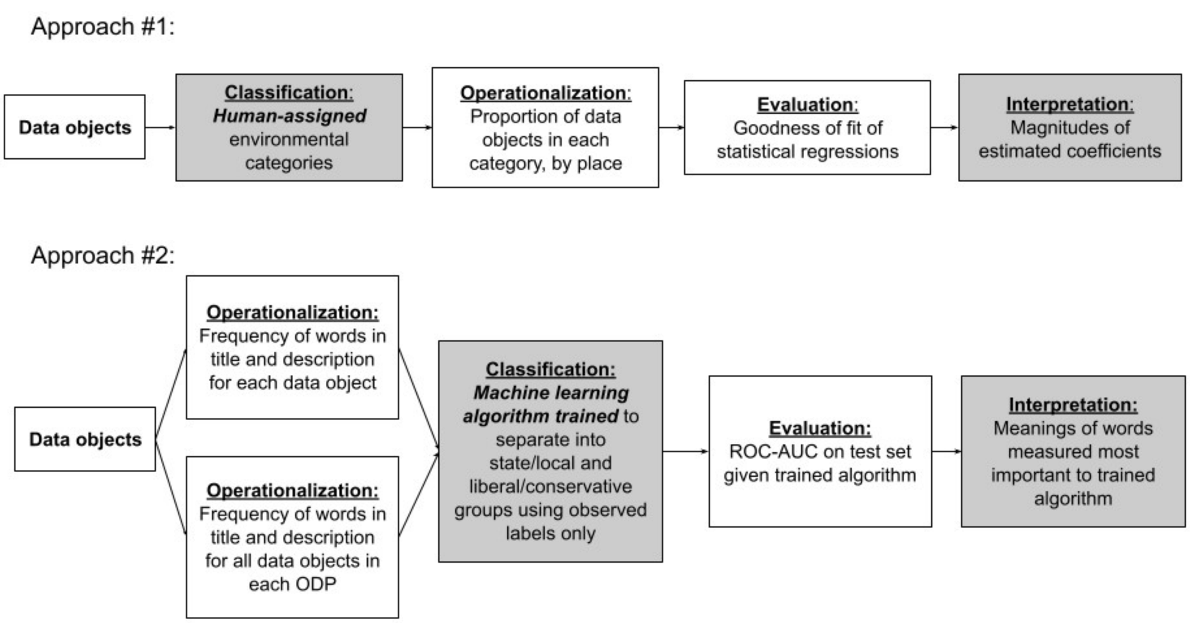

Figure 1. Graphical overview of two complementary approaches used in this study.

3.2 Description of data and correlation/regression analysis

As of 2019-04-24, 17 of 50 US states had ODPs that met federal open data standards; of the 100 most populous municipalities in the US, 64 municipalities had ODPs that met federal open data standards. From these ODPs, a total of 49,066 data objects were collected, of which 28,309 came from the 64 municipal-level ODPs and 20,757 came from the 17 state-level ODPs. The locations of the municipalities and states whose ODPs were included in the analysis are shown in **Figure 2** and are available with more detail in the **Supplemental Information.**

Figure 2. Locations of ODP sources of data objects included in this study. Black filled circles (64) are locations of municipalities with ODPs included in the analysis, and size of the circle indicates realtive population size of the municipality. Unfilled circles show locations of top 100 most populous municipalities that did not have federal open data standards-compliant ODPs. Darker shaded gray sates (17) have ODPs included in the analysis. Note: the state of Hawaii and teh municipality of Honolulu are not shown int he figure, but both had ODPs used in the analysis (See **Supplemental Information** for more detailed information).

Based on titles and descriptions, I then selected only those data objects that could be considered "environmental data" based on how environments have been historically managed through environmental policy, from the literature review above. The environmental data objects were then further labeled with the categories in **Table 2**, which break down the ways the environment may be conceptualized through data. These categories were established before the analysis based on previous knowledge about US federal/state environmental policy and municipal environmental planning. A total of 6,522 environmental data objects were categorized.

To ensure reliability, two coders, a research assistant and I, used **Table 2** to code a random sample of 1,000 data objects. After confirming that >95% consistency between the two coders could be reached, I categorized all 6,522 data objects.

Table 2. Environmental Categories used to classify types of environmental data objects

Environmental Category	**Examples**
Environmental conditions	topography, contours, land cover/land use, parking lots, meteorology (air temperatures, climate, rainfall), geology, soil properties, streams, rivers, ponds
Environmental hazards	flooding, earthquakes, hurricanes, sources of pollution, fire/wildfire risk, dumps/landfills, wastewater treatment, toxics, superfund sites
Environmental resources	hunting/fishing, extraction, agriculture
Environmental services	infrastructure (water/wastewater, hydrants), trash/recycling, snow removal, permits, potable water quality, 311
Environmental amenities	parks, playgrounds, community gardens, green infrastructure, tree inventories, water features
Environmental protection	conservation areas, wildlife species, endangered species, habitats, restoration
Environmental programs	voluntary programs (stormwater, tree planting), grants, rebates, clean-ups, park/nature programming, green streets
Environmental equity	Environmental justice, including social vulnerability
Environmental performance	energy efficiency/consumption, alternative fuel vehicles, charging stations, emissions estimates, LEED buildings, climate mitigation, air quality, surface/groundwater quality, recycling rates

It should be noted that sometimes, a data object could fit within multiple categories. When this occurred, the category appearing lower on on **Table 2** was used. This is because the order presented in the table roughly reflects the “eras” of approaches to environmental governance presented in the introduction. Environmental hazards and resources data are descriptive of conditions; environmental services and amenities data reflect conventional sanitation, infrastructure services, and beautification and recreational opportunities; and environmental protections, programs, equity, and performance data reflect increasingly normative values of humans’ relationship with, and responsibility to, the environment.

After categorization, each data object was joined to two variables reflecting the state or municipality from which it came: population and a scale of political liberalism/conservatism. As in Thorsby (2017), population was used as a proxy for the economic resources available to the municipality/state. The measure of political liberalism/conservatism was developed by Tausanovitch and Warsaw (http://www.americanideologyproject.com/) (2013; 2014), based on a comprehensive survey of the types of policies enacted in each state/municipality. While voting trends may reveal individual-level political conservatism or liberalism, I used the Tausanovitch and Warsaw because these indexes were created based on comprehensive surveys of types of legislation passed at the municipal and state government levels, explicitly to determine whether government policies, including environmental policies, adopted in places correlated with measures of individual residents' political liberalism/conservatism. Since governments determine what data will be released on ODPs, these correlations could reflect both the role of political ideology on opening government data, and on environmental policy. For the municipalities included in the analysis, the Tausanovitch and Warsaw index ranged on a continuous scale from -1.0 (San Francisco, CA) to 0.3 (Virginia Beach, VA), where negative values indicate political liberalism and positive values indicate more political conservatism.

Pearson's r correlations were calculated between the number of total datasets on each ODP, the number of environmental datasets, the populations of the state/municipality, and the measure of political conservatism for the state/municipality. Linear regression analysis was also used to control for multiple, and potentially interactive effects on the types of environmental data hosted on ODPs: level of governance, population (which may be understood as a proxy for economic resources), and political liberalism/conservatism. The dependent variable in the linear regressions was the proportion of environmental datasets in a given environmental category, for each of the 81 places, with the full specification shown in **Equation 1**.

Equation 1

$$y_i = \beta_0 muni_i + \beta_1 state_i + \beta_3 muni_i * liberal_i + \beta_4 state_i * liberal_i + \beta_5 muni_i * population_i + \text{¿}$$
$$\beta_6 state_i * population_i + e_i$$

Where,

- y_i is the proportion of data objects categorized as the given environmental category for ODP_i
- $muni_i$ and $state_i$ are binary variables indicating whether ODP_i is from a state or municipal ODP
- $liberal_i$ is a binary variable indicating whether the place is more liberal than the mean value according to the Tausanovitch and Warsaw index and the ODPs included in the analysis
- $population_i$ is the total population in municipality or state *i*
- e_i is the error term for ODP_i

3.3 Detecting statistically significant word patterns through "separability" of data objects

To supplement the above analysis based on my *a priori* environmental categories, I prepared two corpus-based datasets composed of the text used to describe each data objects (titles and descriptions) to examine content of data objects based on the textual content in their metadata. The first dataset preserved the data objects as the unit of observation (each "row" in the dataset corresponded to one data object, total of 6,522 observations). The second dataset aggregated the corpuses of all data objects by ODP (each "row" in the dataset corresponded to a unique ODP , total of 81 observations). Since the text analysis that these corpuses would be used for is based on word frequencies as a proxy for importance, several data cleaning steps were added to remove frequently used English words (such as: "the," "a," and "and"), standardize the treatment of plurals and tenses, and to remove phrases specific to ODPs that were not descriptive of the data objects themselves. If left in, these non-meaningful words would skew the frequency analysis. An example of an ODP-specific phrase that needed to be removed was "Trouble downloading or have questions about this City dataset? Visit the [OpenDataPhilly Discussion Group](http://www.phila.gov/data/discuss/)," which was included in the description of every Philadelphia data object. This short phrase repeated in every data object would have influenced the frequency-based analyses, but offers no insight into the environmental meaning of the data objects, and therefore needed to be removed. All pre-processing of text data was done with tools built using the Python natural language package *nltk* (http://www.nltk).

The first corpus dataset (unit of observation: one data object) was used in an analysis to test the extent to which the words used in the data objects themselves could be used to separate different "types" of ODP sources from each other: (1) state/municipal level, and (2) politically liberal/conservative. For example, if the words in the title and description of an environmental dataset could reliably predict whether it originated from a state or municipal ODP, this would be evidence that the text descriptions of the datasets between state and municipal ODPs were generally different (descriptions can be used to separate the two groups). Words that typically appear on state and municipal ODPs could then be compared to examine what kinds of environmental data tended to be collected and hosted by state versus municipal governments. Similarly, if the words in the titles and descriptions of the data objects were used to predict whether the originating source was a conservative or liberal place, this would be evidence that liberal places generally use different words in their environmental data than conservative places. The binary classification of conservative (1) or liberal (0) was made by reclassifying Tausananovitch and Warsaw's measures of political conservatism for states and cities above and below the mean values.

The second corpus dataset (unit of observation: one ODP) was used in an analysis to see whether the whole corpus of all data objects' text could be used to predict its liberalism/conservatism. This analysis tests whether the whole corpus of environmental text from many data objects is associated with political liberalism/conservatism, even if individual data objects are not found to be associated with political liberalism/conservatism. This might happen for example, if a distinguishing feature of conservative places is that certain words are used in higher proportions across all datasets.

For both corpus-based analyses, the binary outcome variable in the first analysis, city (1) or state (0); or conservative (1) or liberal (0) was predicted by a random forest classifier algorithm where the predictors were a matrix of count frequencies of the words in the corpus (the process of representing the data in this way is called “vectorization”), normalized by their frequency across all datasets (this is referred to as the inverse document frequency) (Baeza-Yates and Ribeiro-Neto, 2011). The random forest classifier works as a bagged decision tree algorithm, which is well-suited for handling the impact of presence/absence and frequency of influential words on the classification (in this case, of city or state-originated data or liberal or conservative city-origination). A range of hyper-parameters for the random forest classifier were tested, including the size and depth of the trees. The performance of the classifiers fit using each hyper-parameter set trained on a training set (90% of the data) was evaluated on the test set (10% of the data) to reduce the effects of overfitting on the training data, which would lead to artificially inflated performance metrics. Because of the imbalanced ratio of municipal and state-hosted datasets (local data was about 60% of the total), classification accuracy is not an appropriate metric for comparing performance and model training. Instead, recall and precision metrics were examined separately, as well as the combined, Receiver Operator Curve-Area Under the Curve (ROC-AUC) metric, which is insensitive to imbalanced data sets to evaluate performance.

4 Results

4.1 Correlation and Regression Analyses

Table 3 Shows a summary of the frequencies of environmental data products in each environmental category.

Table 3. Frequencies of environmental data products in each environmental category

Environmental Category	Count
Amenities	840
Conditions	794
Equity	7
Hazards	958
Performance	929
Programs	271
Protection	850
Resources	883

Services	990

The category with the fewest data objects was for equity. Examples of this category included: San Francisco's "Community Resiliency Indicator System" and "Flood Health Vulnerability" data objects, Pittsburgh's "Allegheny County Environmental Justice Areas," Austin's "Equity & Livability - % of residents within ½ mile of a park" and "City Park Acres per 1,000 Population" data objects, New York State's "Office of Environmental Justice (OEJ) Grants Awards," and Idaho's "Environmental Justice Map" data object. Although other data objects (such as those that dealt with environmental hazards and amenities) could certainly be analyzed to draw insights about environmental equity, only the seven data objects listed above explicitly acknowledged environmental equity issues in their titles or descriptions.

Figure 3 shows a comparison between municipal and state ODPs of the proportions of data objects in each environmental category. The largest discrepancies between municipality and state ODPs were observed between the "amenities" and "services" categories, in which a larger proportion of municipal data objects were classified, and the "performance," "programs, " "protection," and "resources" categories, in which a larger proportion of state data objects were classified.

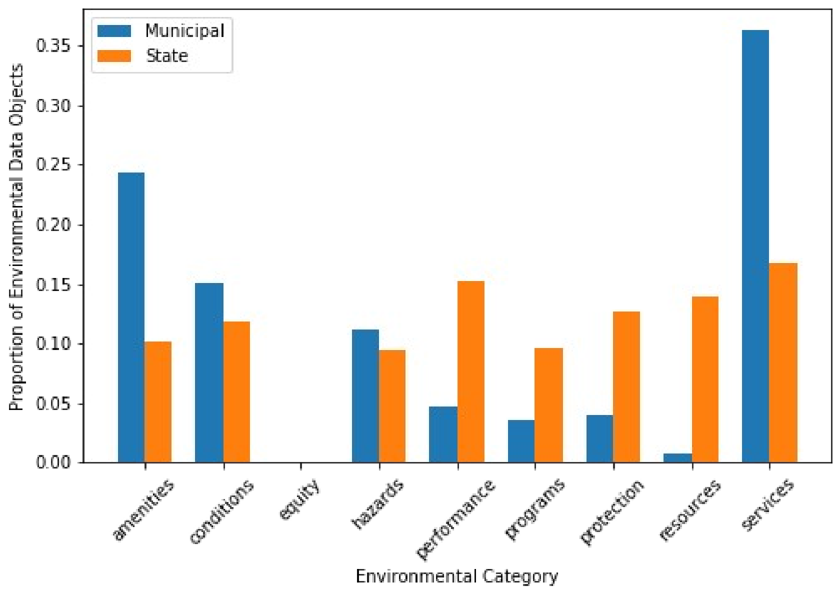

Figure 3 Comparison of proportions of each environmental category among municipal and state ODPs.

Among cities, I found that the total number of data objects is positively correlated with population size (rho = 0.45, p = 0.0002), which is in agreement with previous findings (Thorsby et al., 2017). This is important because throughout this analysis, population size is used as a proxy for the financial resources of the place. The number of environmental data objects is slightly more strongly correlated with population size than the number of datasets in general (rho = 0.47, p = 6.6 e-5). It was also found that the number of datasets was weakly correlated with the measure of political conservatism (rho = -0.29, p = 0.02), with more liberal cities having more environmental datasets than conservative cities. The number of environmental datasets is more strongly correlated with political ideology (rho = -0.35, p = 0.005). **Figure 4** shows a scatterplot showing the relationship between the number of environmental datasets and the log of the population in each city. The figure also demonstrates that larger cities also tend to be more liberal (rho = -0.38, p = 0.002).

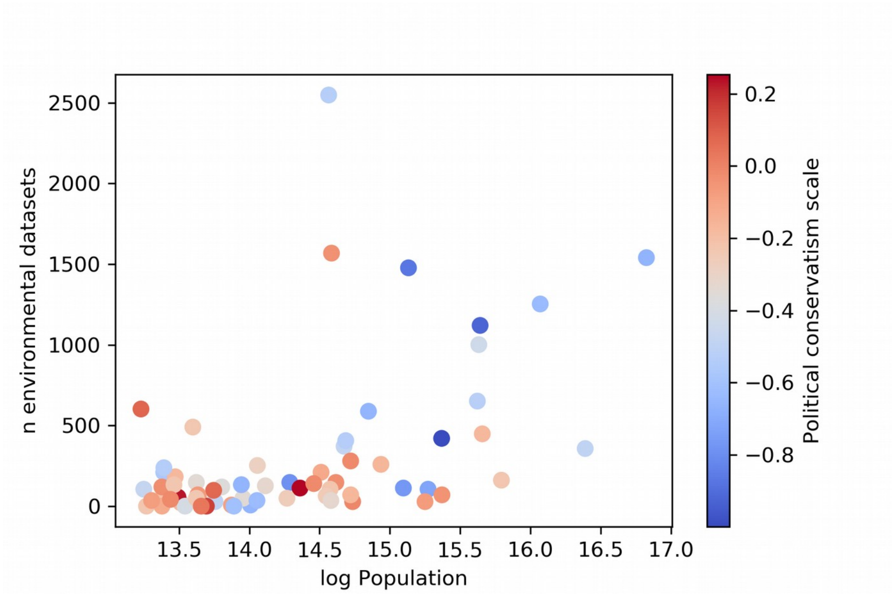

Figure 4. [COLOR] Relationship between municipality population and the number of environmental data objects hosted on the ODP. Colormap indicates measure of political conservatism, with red being more politically conservative, and blue being more politically liberal.

Table 4 shows the results of Pearson's correlations and t-tests of means between the proportion of data objects in each environmental category and the liberalism/conservatism of the state or municipality. For Pearson's correlations, the Tausanovitch and Warsaw index (a continuous variable between -1 and 1) was used; for the t-test of means, conservative and liberal groups were formed based on the Tausanovitch and Warsaw index's mean value to compare means. Of all the correlations and t-tests tested, only four were statistically significant. Based on the Pearson's correlation, more liberal municipalities tended to have a larger proportion of environmental equity data objects on their ODPs and more liberal states tended to have a greater proportion of environmental performance data objects on their ODPs; based on the t-test of means, more liberal municipalities tended to have a greater proportion of environmental performance data objects and more conservative states tended to have a greater proportion of environmental services data objects.

Table 4. Relationships between prevalence of data objects in environmental categories and

political conservatism/liberalism of municipality or state

	Pearson Correlation						T-Test of Means					
	Municipality ODP			**State ODP**			**Municipality ODP**			**State ODP**		
Environmental Category	**rho**	**p value**		**rho**	**p value**		**t statistic**	**p value**		**t statistic**	**p value**	
Amenities	-0.005	0.970		0.101	0.699		0.546	0.587		0.361	0.723	
Conditions	0.051	0.704		0.360	0.155		0.333	0.741		1.001	0.333	
Equity	**-0.321**	**0.013**	*	0.285	0.268		-1.693	0.096		0.635	0.535	
Hazards	-0.209	0.113		-0.012	0.964		-1.454	0.151		-0.981	0.342	
Performance	-0.213	0.105		**-0.547**	**0.023**	*	**-2.266**	**0.027**	*	-1.710	0.108	
Programs	-0.101	0.445		-0.285	0.268		-0.556	0.581		-1.498	0.155	
Protection	0.212	0.107		0.136	0.604		1.527	0.132		0.003	0.997	
Resources	0.105	0.430		0.141	0.588		0.471	0.640		0.830	0.419	
Services	0.104	0.432		0.224	0.387		0.466	0.643		**2.019**	**0.062**	*

* denotes statistical significance at the alpha = 0.10 level. ** denotes statistical significance at the alpha = 0.01 level

The linear regressions allow for the effects of multiple variables to be controlled for, which is useful in teasing apart whether the correlations observed above may be interpreted as "causal" or simply correlative. In particular, places with higher populations (and therefore more economic resources to implement new technologies, environmental policies and programs and collect and distribute data) tend to be correlated with political liberalism. Controlling for population in regression can help to tease apart this conflation of effect. **Table 5** shows the regression model performance and parameter estimates for all environmental categories.

Table 5. Regression model performance and parameter estimates for all environmental categories

		muni			**state**			**muni*liberal**	
Environmental Category	**Adj R2**	**param est**	**p value**		**param est**	**p value**		**param est**	**p value**
Resources	**0.477**	0.010	0.436		**0.082**	**0.011**	*	-0.005	0.789
Performance	**0.324**	0.004	0.845		**0.137**	**0.003**	**	0.026	0.312
Protection	0.100	0.059	0.003		**0.102**	**0.036**	*	-0.032	0.246

Programs	0.097	0.024	0.179		0.008	0.860		0.003	0.900	
Services	0.035	**0.389**	**0.000**	**	-0.181	0.191		-0.029	0.707	
Amenities	0.032	**0.265**	**0.000**	**	-0.211	0.039		-0.025	0.660	
Hazards	-0.021	**0.089**	**0.006**	**	-0.002	0.982		0.074	0.101	
Equity	-0.022	0.000	0.982		0.003	0.326		**0.002**	**0.099**	*
Conditions	-0.022	0.161	0.000		0.063	0.461		-0.014	0.770	
		state*liberal			**muni*population**			**state*population**		
Environmental Category (contd)		**param est**	**p value**		**param est**	**p value**		**param est**	**p value**	
Resources		-0.031	0.394		-3.20E-12	0.999		**8.94E-09**	**0.012**	*
Performance		**0.090**	**0.092**	*	**1.15E-08**	**0.004**	**	**-1.74E-08**	**0.001**	**
Protection		0.023	0.682		-1.46E-09	0.722		-2.58E-09	0.629	
Programs		**0.107**	**0.040**	*	3.68E-09	0.327		-1.98E-09	0.686	
Services		-0.107	0.508		-4.44E-09	0.706		7.85E-09	0.609	
Amenities		0.019	0.875		-3.76E-09	0.662		1.08E-08	0.338	
Hazards		-0.027	0.768		-4.30E-09	0.523		2.30E-09	0.794	
Equity		-0.004	0.189		-1.20E-11	0.956		-5.08E-11	0.859	
Conditions		-0.069	0.491		-1.18E-09	0.871		-7.92E-09	0.406	

* denotes statistical significance at the alpha = 0.10 level. ** denotes statistical significance at the alpha = 0.01 level

Of the nine regression models (one for each of the environmental categories), only two were able to capture over 30% of the variation in the dependent variable (the proportion of data objects categorized in the environmental category): Environmental Resources, and Environmental Performance. Controlling for all other factors, state ODPs were more likely to host data objects relating to environmental resources, and states with larger populations had a higher proportion of data objects relating to environmental resources.

Controlling for all other factors, state ODPs had a larger proportion of environmental performance data objects. Politically liberal states had an even greater proportion of environmental performance data, and this was estimated to be an *opposite* effect to the effect of state populations, which was negative. In contrast, the population effect for municipalities was positive: the larger the population, the greater the proportion of data objects classified as tracking environmental performance.

4.2 Analysis of data object and ODP corpuses

4.2.1 Differences between words describing municipal and state environmental data objects

The trained classifier correctly sorted 1,621 out of 1,773 data objects in the test set (91%) into whether they originated from municipal or state ODPs (true negatives: 1,030, false positives: 62, false negatives: 90, true positives: 591). This corresponds to a recall of 87%, precision of 91%, and accuracy of 91%. The ROC-AUC for the test set was 0.91, considered to be very good performance.

Having confirmed that the descriptions of each data set could indeed be used to satisfactorily separate local from state data sets, I then used a chi-squared test to measure the association between each word in the descriptions to the city or state labels. Words with highest chi-squared statistics, and that were present in the data descriptions of at least five different ODPs are presented in descending order of influence in **Table 6**. Many of the words most closely associated with local environmental data sets involved amenities and services (e.g.: “park,” “recreation”, “tree,” “sewer,” “sanitary,” “[garbage] collection,”etc), while many of those most closely associated with state environmental data involved natural resources-- both extractive and recreational, pollution monitoring, and toxics (e.g.: “drinking [water],” “school [testing],” “juvenile,” and “lead,” which all have to do with drinking water monitoring, “production,” “coal,” “oil,” “fish,” “pollutant,” “monitor,” and “toxic”). These meanings were confirmed by searching for these keywords in the original data objects, since taken out of context they could have had different meanings.

Table 6. Words most closely associated with local and state environmental data sets

Predictive Rank	City	State
1	park	drinking
2	recreation	school
3	tree	production
4	building	juvenile
5	street	lead
6	sewer	oil
7	trash	abundance

8	city	coal
9	district	summer
10	capture	wild
11	community	river
12	collection	gas
13	frequency	winter
14	sustainable	fall
15	sensor	habitat
16	storm	host
17	court	fish
18	purpose	summarize
19	property	well
20	illegal	earthquake
21	schedule	toxic
22	beach	pollutant
23	plan	spring
24	sanitary	monitor
25	dumping	harvest

4.2.2 Differences between words describing data objects from conservative and liberal places

For municipalities, the trained classifier performed only moderately well in sorting whether data objects originated from more conservative or more liberal municipalities (ROC-AUC metric of 0.80). For states, the trained classifier exhibited very good performance in sorting whether data objects originated from more conservative or more liberal states (ROC-AUC metric of 0.92). Again, the chi-squared test to measure the association between each word in the descriptions and the classification of liberal or conservative was ranked from highest to lowest. **Table 7** shows the top 25 words associated with liberal and conservative cities and states' environmental data. Among the municipalities' data objects, the two words most highly predictive of a liberal city were: "recreation" and "sensors." Energy data was also more likely to come from a liberal city than a conservative city. In contrast, conservative municipalities' datasets were more strongly predicted with many words associated with ensuring environmental services such as “[public] works,” “solid [waste],” “garbage,” “trash,” and “waste” collection. Conservative states were much more likely to include extraction-related data words, including “production,” “oil,” “gas,” “coal,” “field,” and “mining.”

Table 7. Words most closely associated with liberal and conservative local and state environmental data sets

	Cities		States	
Predictive Rank	**Liberal**	**Conservative**	**Liberal**	**Conservative**
1	recreation	daily	school	production
2	sensor	miss	lead	lid
3	automate	solid	wild	oil
4	require	works	program	gas
5	field	dumping	summer	coal
6	energy	brush	fall	division
7	court	east	department	summarize
8	track	refresh	bay	ground
9	indicator	designate	indicator	pollutant
10	quality	class	mile	well
11	capture	waste	terrestrial	abundance
12	action	illegal	acquisition	field
13	beach	valve	station	mining
14	temperature	trash	creek	emission
15	frequency	garbage	climate	wildlife
16	performance	hour	performance	deer
17	measurement	central	preservation	reserve
18	environment	trail	energy	geological
19	release	flood	ecology	pollution

20	sample	creation	electricity	farm
21	acre	parcel	record	chemistry
22	ward	zoning	trend	fish
23	bridge	bike	riparian	toxic
24	weather	disposal	elevation	chemical
25	earth	cleanup	cost	natural

5 Discussion

The results of this study reveal some broad patterns in the types of environmental data available municipal and state-level ODPs in the US and have some correlations to political liberalism/conservatism. Below, I present a discussion of the findings related to the three hypotheses posed in the Introduction.

5.1 Liberal places host more datasets related to environmental sustainability

If "sustainability" is operationalized through environmental data related to tracking environmental performance and protecting resources so that they are available for future generations, then this hypothesis was supported in several ways. First, more liberal municipalities and states were both associated with a higher proportion of environmental data in the "performance" category. Data objects in this category included Washington DC's "Private Building Energy and Water Performance Benchmarks," San Francisco's "Community-wide Greenhouse Gas Inventory," and Philadelphia's annual "Litter Index" datasets. From the states, it included Oregon's "Greenhouse Gas Reporting" updates, Maryland's annual "Ozone Exceedance Days," and the annual updates of amounts of recycled wastewater used in Hawaii (all more liberal than the mean according to Tausanovith and Warsaw's index). The regression analysis also provided some support of this hypothesis, but only for states. The estimated coefficient for the interaction between the binary variable for *state* and the binary variable for *liberal*, indicated that when controlling for population (a proxy for economic resources of the state), more liberal states were more likely to have a higher proportion of their environmental data dedicated to tracking performance.

The analysis of text in the corpuses of data object descriptions moderately supported these findings for municipal ODPs, and more strongly supported these findings for state ODPs. Words such as "performance," and "measurement," were among the most associated with liberal cities, while words describing services such as "[public] works," "solid [waste]," "garbage," "trash," and "waste" collection were most associated with conservative cities. Among states, words used to describe tracking of pollutants were among the most associated with liberal states, while words associated with extractive industries and natural resources were most associated with conservative states.

5.2 Municipalities will seek to showcase “livability” of their communities by highlighting environmental services and amenities

This hypothesis was clearly supported by the high proportions of environmental data objects in the “amenities” and “services” categories on municipal ODPs compared to state ODPs, shown in **Figure 2.** However, it was not true that municipalities hosted more environmental performance data than states. Equity-related environmental data appeared exclusively on municipality ODPs, and was highly correlated with higher population, more liberal municipalities (though, overall very few equity-related environmental data were coded). Among municipalities, more liberal places were more likely to host performance data.

The analysis of the text in the corpuses of municipal ODPs and state ODPs also highlighted municipalities’ increased focus on concepts related to “livability.” Words that were most predictive of a municipal ODP data object included: “park,” “recreation,” “community,” and “sustainable.”

5.3 Places with higher populations will host more environmental data tracking environmental performance and environmental protection

Through the regression analysis, I could control for multiple factors to better understand the influence of financial resources (total population), given political liberalism/conservatism and governance level. The model fit for environmental protection did not reach a sufficiently high goodness-of-fit to warrant interpretation of its estimated coefficients, but the model of environmental performance was able to account for more than 30% of the variation in the proportion of data objects related to environmental performance. I found that when controlling for political liberalism/conservatism, states with higher populations actually had fewer data objects relating to environmental performance. This suggests that it is not merely financial resources that determine how much is devoted to tracking environmental performance, but the political alignment inclination to do so. Among municipal-level ODPs however, higher populations (more financial resources) were associated with more environmental performance tracking data, while the effect of political liberalism/conservatism was not statistically significant.

One reason that political liberalism/conservatism may not have been as significant in determining the proportion of environmental performance data in cities might be because in general, cities tend to be more liberal than states according to Tausanovitch and Warsaw’s index. Cities that were included in this study, those that have large populations and have implemented ODPs that meet federal guidelines for standards, also tend to be especially liberal-leaning, and therefore resulted in less variation present in this variable compared to the state level.

5.4 Limitations

There are several limitations to this research. First, while I found evidence of content differences between liberal/conservative cities and states, this analysis does not address the causal reason. The differences in coverage could be because of resource constraints (although I attempted to use population as a proxy for financial resources, this is a crude proxy), values and priorities attached to the environment and environmental policy, or differences in underlying environmental conditions experienced in different places. Second, parts of this analysis relied heavily on word frequencies in data object titles and descriptions as a proxy for the importance of concepts being expressed by that word. In order to scale qualitative analysis of meaning to thousands of data objects hosted on dozens of ODPs, necessitated reducing the meaning of words in this way and many important environmental themes captured by particular data objects may not always have had high word frequencies. In my observations working with the data, high word frequencies tended to be associated with datasets that were updated frequently, had been collected over a long period of time, or were associated under a specific program collecting many datasets. These are plausible indicators of importance. However, one could also say that with the decreasing cost of storing data, collecting many datasets or slicing them in different ways to host on an ODP may not actually be an indicator of the importance of the topic-- just the "cheapness" of storing data. Lastly, there are several limitations with the political conservatism/liberalism data used in the analysis. Political conservatism/liberalism was operationalized using Tausanovitch and Warsaw's index for state and municipal government policy preferences, which were created in 2013 and 2014, while the titles and descriptions of data products from the ODPs were collected in 2019. Therefore, there could be inaccuracies associated with states or municipalities that have become more conservative or liberal since 2013/2014.

6 Conclusions

This research offers several contributions. First, while others have conducted comparative, descriptive, and index-based research on the number of datasets and features of ODPs, this study is the first to use quantitative and automated text processing techniques on all the available metadata of US city and state ODPs to extract patterns in the content themes in the datasets themselves. As was shown in this research, examining the content themes provides better understanding of the variation in the kinds of environmental datasets that are being collected in different communities. Second, this research is able to identify where data "gaps" may be occurring and how conceptualizations of "the environment" may be diverging. The presence of sustainability performance data was more likely to have been collected in larger, more liberal municipalities, while the presence of environmental sanitation services data was more associated with smaller, more conservative municipalities, for example. This highlights where additional political action or data collection efforts are needed. The study also showed large differences between the environmental concerns at the municipality and state scales, with cities more focused on environmental amenities and services, and states more focused on resources and risks. This could have implications both for tracking gaps in urban-rural differences in human relationships to the environment, but also for how such gaps are perpetuated through the collection and availability of data.

Supplemental Information

Supplemental Table. Municipality and state ODPs included in the analysis

Municipalities

Municipality	State	Metro Population	Number of data sets
New York City	NY	20,300,000	2,429
Los Angeles	CA	13,131,431	550
Chicago	IL	9,533,040	1,311
Dallas	TX	7,233,323	978
Houston	TX	6,313,158	247
Washington	DC	6,216,589	3,363
Miami	FL	6,158,824	495
Philadelphia	PA	6,096,120	529
Phoenix	AZ	4,737,270	51
San Francisco	CA	4,727,357	1,154
Detroit	MI	4,292,060	205
Riverside	CA	4,200,000	53
Seattle	WA	3,733,580	849
Minneapolis	MN	3,600,618	131
Tampa	FL	3,068,511	95
Baltimore	MD	2,808,175	2,594
Orlando	FL	2,509,454	38
Charlotte	NC	2,474,314	144
San Antonio	TX	2,473,974	275
Portland	OR	2,389,228	217
Pittsburgh	PA	2,360,867	310
Las Vegas	NV	2,227,053	497

Kansas City	MO	2,159,159	3,574
Sacramento	CA	2,149,127	95
Cincinnati	OH	2,137,406	237
Austin	TX	2,115,827	2,118
Columbus	OH	2,078,725	79
Indianapolis	IN	2,004,230	485
Nashville	TN	1,903,045	168
Virginia Beach	VA	1,725,246	52
Providence	RI	1,604,291	222
Milwaukee	WI	1,572,245	273
Memphis	TN	1,348,260	165
Raleigh	NC	1,273,985	150
Richmond	VA	1,270,158	88
Hartford	CT	1,213,123	198
Birmingham	AL	1,149,807	99
Buffalo	NY	1,136,670	177
Louisville	KY	1,293,953	0
Rochester	NY	1,077,848	326
Grand Rapids	MI	1,059,113	46
Honolulu	HI	988,650	236
Worcester	MA	942,475	19
Omaha	NE	933,316	78
Oxnard	CA	854,223	27
Allentown	PA	840,550	40
Baton Rogue	LA	834,159	198
Columbia	SC	825,033	20
North Port	FL	804,690	1,150

Charleston	SC	825,033	80
Greensboro	NC	761,184	45
Little Rock	AR	738,344	48
Colorado Springs	CO	723,878	46
Boise City	ID	709,845	54
Akron	OH	703,505	43
Lakeland	FL	686,483	28
Syracuse	NY	654,841	71
Madison	WI	654,230	141
Deltona-Daytona Beach-Ormond Beach	FL	649,202	19
Wichita	KS	645,628	69
Augusta	GA	600,151	20
Jackson	MS	578,715	39
Durham-Chapel Hill	NC	567,428	288
Chattanooga	TN	556,426	483

States

State		**State Population**	**Number of data sets**
Texas		29,360,759	1,364
New York		19,336,776	1,102
Pennsylvania		12,783,254	223
New Jersey		8,882,371	444
Virginia		8,590,563	152
Washington		7,693,612	2,170
Tennessee		6,886,834	98
Missouri		6,151,548	277

Maryland		6,055,802	1,635
Oregon		4,241,507	2,334
Utah		3,249,879	5,975
Iowa		3,163,561	1,102
Mississippi		2,966,786	20
Idaho		1,826,913	1,944
Hawaii		1,407,006	1,456
Delaware		986,809	258
Vermont		623,347	203

References

Attard, Judie, Fabrizio Orlandi, Simon Scerri, and Sören Auer. 2015. “A Systematic Review of Open Government Data Initiatives.” *Government Information Quarterly* 32 (4): 399–418. https://doi.org/10.1016/j.giq.2015.07.006.

Barber, Kathryn, and Duncan MacLellan. 2019. “Examining Open Data at the Urban Level: An Exploration of ‘Wellbeing Toronto.’” *Journal of Urban Technology* 26 (1): 107–21. https://doi.org/10.1080/10630732.2018.1558573.

Barns, Sarah. 2018. “Smart Cities and Urban Data Platforms: Designing Interfaces for Smart Governance.” *City, Culture and Society*, Innovation and identity in next generation smart cities, 12 (March): 5–12. https://doi.org/10.1016/j.ccs.2017.09.006.

Barry, Emily, and Frank Bannister. 2014. “Barriers to Open Data Release: A View from the Top.” *Information Polity* 19 (1,2): 129–52. https://doi.org/10.3233/IP-140327.

Blomquist, William. 1991. “Exploring State Differences in Groundwater Policy Adoptions, 1980-1989,” 16.

Brondizio, Eduardo S., Elinor Ostrom, and Oran R. Young. 2009. “Connectivity and the Governance of Multilevel Social-Ecological Systems: The Role of Social Capital.” *Annual Review of Environment and Resources* 34 (1): 253–78. https://doi.org/10.1146/annurev.environ.020708.100707.

Bulkeley, Harriet, and Michele M. Betsill. 2013. “Revisiting the Urban Politics of Climate Change.” *Environmental Politics* 22 (1): 136–54. https://doi.org/10.1080/09644016.2013.755797.

Cash, David, W. Neil Adger, Fikret Berkes, Po Garden, Louis Lebel, Per Olsson, Lowell Pritchard, and Oran Young. 2006. “Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World.” *Ecology and Society* 11 (2). https://doi.org/10.5751/ES-01759-110208.

Chatfield, Akemi Takeoka, and Christopher G. Reddick. 2018. “The Role of Policy Entrepreneurs in Open Government Data Policy Innovation Diffusion: An Analysis of Australian Federal and State Governments.” *Government Information Quarterly*, Internet Plus Government: Advancement of Networking Technology and Evolution of the Public Sector, 35 (1): 123–34. https://doi.org/10.1016/j.giq.2017.10.004.

Clark, Brad T., and David W. Allen. 2004. “Political Economy and the Adoption of Everyday Environmental Policies in the American States, 1997: An Exploratory Analysis.” *The Social Science Journal* 41 (4): 525–42. https://doi.org/10.1016/j.soscij.2004.08.013.

Daniels, Thomas L. 2009. “A Trail Across Time: American Environmental Planning From City Beautiful to Sustainability.” *Journal of the American Planning Association* 75 (2): 178–92. https://doi.org/10.1080/01944360902748206.

Dawes, Sharon S., Lyudmila Vidiasova, and Olga Parkhimovich. 2016. “Planning and Designing Open Government Data Programs: An Ecosystem Approach.” *Government Information Quarterly* 33 (1): 15–27. https://doi.org/10.1016/j.giq.2016.01.003.

DeFries, Ruth, and Harini Nagendra. 2017. “Ecosystem Management as a Wicked Problem.” *Science* 356 (6335): 265–70. https://doi.org/10.1126/science.aal1950.

Drummond, William J. 2010. “Statehouse Versus Greenhouse -- Have State-Level Climate Action Planners and Policy Entrepreneurs Reduced Greenhouse Gas Emissions?” *Journal of the American Planning Association* 76 (4): 413. https://doi.org/10.1080/01944363.2010.499537.

Dunlap, R.E., C. Xiao, and A.M. McCright. 2001. “Politics and Environment in America: Partisan and Ideological Cleavages in Public Support for Environmentalism.” *Environmental Politics* 10 (4): 23–48. https://doi.org/10.1080/714000580.

French, Richard D. 2019. “Is It Time to Give up on Evidence-Based Policy? Four Answers.” *Policy & Politics* 47 (1): 151–68. https://doi.org/10.1332/030557318X15333033508220.
Gessa, Ana, and Pilar Sancha. 2019. “Environmental Open Data in Urban Platforms: An Approach to the Big Data Life Cycle.” *Journal of Urban Technology* 0 (0): 1–19. https://doi.org/10.1080/10630732.2019.1656934.
Gill, Mark, and Jon Corbett. 2017. “Downscaling: Understanding the Influence of Open Data Initiatives in Smaller and Mid-Sized Cities in British Columbia, Canada.” *The Canadian Geographer / Le Géographe Canadien* 61 (3): 346–59. https://doi.org/10.1111/cag.12372.
Goldstein, Brett, and Lauren Dyson, eds. 2013. *Beyond Transparency: Open Data and the Future of Civic Innovation*. San Francisco: Code for America Press.
Graham, Mark, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. 2014. “Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty.” *Annals of the Association of American Geographers* 104 (4): 746–64. https://doi.org/10.1080/00045608.2014.910087.
Gurstein, Michael B. 2011. “Open Data: Empowering the Empowered or Effective Data Use for Everyone?” *First Monday* 16 (2). https://doi.org/10.5210/fm.v16i2.3316.
Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. “Benefits, Adoption Barriers and Myths of Open Data and Open Government.” *Information Systems Management* 29 (4): 258–68. https://doi.org/10.1080/10580530.2012.716740.
Janz, Margaret M. 2019. “The Data Refuge Project for Protecting Federal Data in the United States,” 7.
Kassen, Maxat. 2013. “A Promising Phenomenon of Open Data: A Case Study of the Chicago Open Data Project.” *Government Information Quarterly* 30 (4): 508–13. https://doi.org/10.1016/j.giq.2013.05.012.
Kubler, Sylvain, Jérérmy Robert, Sebastian Neumaier, Jürgen Umbrich, and Yves Le Traon. 2018. “Comparison of Metadata Quality in Open Data Portals Using the Analytic Hierarchy Process.” *Government Information Quarterly*, Internet Plus Government: Advancement of Networking Technology and Evolution of the Public Sector, 35 (1): 13–29. https://doi.org/10.1016/j.giq.2017.11.003.
Layzer, Judith A. 2011. *The Environmental Case: Translating Values Into Policy*. 3rd ed. Washington, D.C: CQ Press.
Lester, James P. 1995. *Environmental Politics and Policy: Theories and Evidence*. Duke University Press.
Lim, Theodore C. 2021. “Model Emulators and Complexity Management at the Environmental Science-Action Interface.” *Environmental Modelling & Software* 135 (January): 104928. https://doi.org/10.1016/j.envsoft.2020.104928.
Lnenicka, Martin, and Anastasija Nikiforova. 2021. “Transparency-by-Design: What Is the Role of Open Data Portals?” *Telematics and Informatics* 61 (August): 101605. https://doi.org/10.1016/j.tele.2021.101605.
Lourenço, Rui Pedro. 2015. “An Analysis of Open Government Portals: A Perspective of Transparency for Accountability.” *Government Information Quarterly* 32 (3): 323–32. https://doi.org/10.1016/j.giq.2015.05.006.
Lubell, Mark, Richard Feiock, and Susan Handy. 2009. “City Adoption of Environmentally Sustainable Policies in California’s Central Valley.” *Journal of the American Planning Association* 75 (3): 293–308. https://doi.org/10.1080/01944360902952295.
May, Peter J., and Chris Koski. 2007. “State Environmental Policies: Analyzing Green Building Mandates.” *Review of Policy Research* 24 (1): 49–65. https://doi.org/10.1111/j.1541-1338.2007.00267.x.

Mazmanian, Daniel A., and Michael E. Kraft, eds. 2009. *Toward Sustainable Communities: Transition and Transformations in Environmental Policy*. Second edition edition. Cambridge, MA: The MIT Press.

Mazmanian, Daniel, and Paul Sabatier. 2016. “Liberalism, Environmentalism, and Partisanship in Public Policy-Making: The California Coastal Commissions.” *Environment and Behavior*, July. https://doi.org/10.1177/0013916581133007.

Meadowcroft, James. 2002. “Politics and Scale: Some Implications for Environmental Governance.” *Landscape and Urban Planning*, Scaling and Environmental Understanding, 61 (2): 169–79. https://doi.org/10.1016/S0169-2046(02)00111-1.

Melosi, Martin V. 2000. *The Sanitary City: Urban Infrastructure in America from Colonial Times to the Present*. Johns Hopkins University Press.

Moss, Timothy, and Jens Newig. 2010. “Multilevel Water Governance and Problems of Scale: Setting the Stage for a Broader Debate.” *Environmental Management* 46: 1–6. https://doi.org/10.1007/s00267-010-9531-1.

Mulcahy, Karen A., and Keith C. Clarke. 1994. “Government Digital Cartographic Data Policy and Environmental Research Needs.” *Computers, Environment and Urban Systems*, Special Issue: GIS/LIS ’94 Central Europe, 18 (2): 95–101. https://doi.org/10.1016/0198-9715(94)90004-3.

Newig, Jens, and Oliver Fritsch. 2009. “Environmental Governance: Participatory, Multi-Level – and Effective?” *Environmental Policy and Governance* 19 (3): 197–214. https://doi.org/10.1002/eet.509.

Nikiforova, Anastasija, and Keegan McBride. 2021. “Open Government Data Portal Usability: A User-Centred Usability Analysis of 41 Open Government Data Portals.” *Telematics and Informatics* 58 (May): 101539. https://doi.org/10.1016/j.tele.2020.101539.

Portney, Kent E. 2009. “Sustainability in American Cities.” In *Toward Sustainable Communities: Transition and Transformations in Environmental Policy*, edited by Daniel A. Mazmanian and Michael E. Kraft, second edition edition. Cambridge, MA: The MIT Press.

Proctor, James D., and Brendon M. H. Larson. 2005. “Ecology, Complexity, and Metaphor.” *BioScience* 55 (12): 1065–68. https://doi.org/10.1641/0006-3568(2005)055[1065:ECAM]2.0.CO;2.

Rittel, Horst W. J., and Melvin M. Webber. 1973. “Dilemmas in a General Theory of Planning.” *Policy Sciences* 4 (2): 155–69. https://doi.org/10.1007/BF01405730.

Safarov, Igbal. 2019. “Institutional Dimensions of Open Government Data Implementation: Evidence from the Netherlands, Sweden, and the UK.” *Public Performance & Management Review* 42 (2): 305–28. https://doi.org/10.1080/15309576.2018.1438296.

Safarov, Igbal, Albert Meijer, and Stephan Grimmelikhuijsen. 2017. “Utilization of Open Government Data: A Systematic Literature Review of Types, Conditions, Effects and Users.” *Information Polity* 22 (1): 1–24. https://doi.org/10.3233/IP-160012.

Saha, Devashree, and Robert G. Paterson. 2008. “Local Government Efforts to Promote the ‘Three Es’ of Sustainable Development.” *Journal of Planning Education and Research* 28 (1): 21–37. https://doi.org/10.1177/0739456X08321803.

Tai, Kuang-Ting. 2021. “Open Government Research over a Decade: A Systematic Review.” *Government Information Quarterly* 38 (2): 101566. https://doi.org/10.1016/j.giq.2021.101566.

Tausanovitch, Chris, and Christopher Warshaw. 2013. “Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities.” *The Journal of Politics* 75 (2): 330–42. https://doi.org/10.1017/S0022381613000042.

———. 2014. “Representation in Municipal Government.” *American Political Science Review* 108 (3): 605–41. https://doi.org/10.1017/S0003055414000318.

Thorsby, Jeffrey, Genie N.L. Stowers, Kristen Wolslegel, and Ellie Tumbuan. 2017. “Understanding the Content and Features of Open Data Portals in American Cities.” *Government Information Quarterly* 34 (1): 53–61. https://doi.org/10.1016/j.giq.2016.07.001.

Townsend, Anthony M. 2013. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. 1 edition. New York, NY: W. W. Norton & Company.

Wilson, Bev, and Cong Cong. 2021. “Beyond the Supply Side: Use and Impact of Municipal Open Data in the U.S.” *Telematics and Informatics* 58 (May): 101526. https://doi.org/10.1016/j.tele.2020.101526.

Wirtz, Bernd W., Robert Piehler, Marc-Julian Thomas, and Peter Daiser. 2016. “Resistance of Public Personnel to Open Government: A Cognitive Theory View of Implementation Barriers towards Open Government Data.” *Public Management Review* 18 (9): 1335–64. https://doi.org/10.1080/14719037.2015.1103889.

Yadav, Piyush, Souleiman Hasan, Adegboyega Ojo, and Edward Curry. 2017. “THE ROLE OF OPEN DATA IN DRIVING SUSTAINABLE MOBILITY IN NINE SMART CITIES.” In *Proceedings of the 25th European Conference on Information Systems (ECIS)*, 1248–63. Guimarães, Portugal.