Government Information Quarterly 34 (2017) 53–61

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Understanding the content and features of open data portals in American cities

Jeffrey Thorsby, Genie N.L. Stowers *, Kristen Wolslegel, Ellie Tumbuan

Public Administration Program, School of Public Affairs and Civic Engagement, Downtown Campus, Suite 679, San Francisco State University, 835 Market Street, San Francisco, CA 94103, United States

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form 9 July 2016
Accepted 14 July 2016
Available online 3 August 2016

Keywords:
Open data
Open data portals
Open government
Civic technology
Data portals

ABSTRACT

In this paper, we present the results of research on features and content of open data portals in American cities. Five scales are developed to categorize and describe these portals: the Open Data Portal Index (ODPI), Data Content Index (DCI), a compilation of the two (Overall Index), the Number of Datasets and Number of Datasets per 100,000. Regression models explaining variation between cities on these scales indicate city population as an important influence, along with participation in a regional consortium. More variation could be explained in the number of datasets model (79.8%) than in any other model. Overall, results indicate portals are in a very early stage of development and need a great deal of work to improve user help and analysis features as well as inclusion of features to help citizens understand the data, such as more charting and analysis.

While descriptive research has been conducted on open data and open data portals at the national level, and in other countries at the national and sub-national levels, there is still a need to conduct basic empirical research upon American urban open data portals. In this article, we present not just descriptive research on these portals but also much needed results of inferential analysis of the characteristics of the first open data portals in American cities. The article begins with a description of open data, open data portals and a brief history of their beginnings. Methods for the study are then discussed and characteristics of the first portals are described, based upon examination of the first 37 urban open data sites to be posted on the Cities site of data.gov. Indices based on these features are described and regression models posited to explain the variation among cities in these models. Finally, the results of these models are described.

1. Open data and data portals

The Internet has permanently altered the process by which information is obtained on a daily basis, particularly within the realms of news consumption and business commerce. The term *Web 2.0*, coined by Tim O'Reilly (2005) to describe the second wave of innovation in information and communication technologies (ICTs), utilizes advanced participatory features such as podcasts, widgets, and Really Simple Syndication (RSS). More recently, the Internet has begun to influence how societies use ICTs within their governments. Sandoval-Almazan and Gil-Garcia (2012) point out that the definition of electronic government (e-government) has been and continues to be debated by scholars. On one end, it is narrowly defined, with its purpose being simply to enhance government services. On the other end, it is broadly defined and includes various degrees of participatory engagement aimed at increasing democratic systems and empowering citizens.

With the emergence of Web 2.0, e-government has been presented with the opportunity to take an active role in further reshaping its definition. Taewoo Nam (2012) explains that e-government now includes Government 2.0 and Open Government as the means and the ends of current e-government rationale. Specifically, "the transition of government [breaks down] into new modes in terms of goals and tools. The new aspects of e-government are not only for [the] government, but also for the public as customers and users" (p. 347).

2. The principles of open data

In 2009, a group of activists calling themselves the Open Government Working Group laid out eight principles for the use of public data. They proposed data was open when it is: 1) complete; 2) primary; 3) timely; 4) accessible; 5) machine-processable; 6) non-discriminatory; 7) non-proprietary; and, 8) license-free (Dawes, 2010). With these principles, they believed data would be available to all without the need for purchase or special software, its source and purpose would be transparent, and data could be used by anyone who wanted to access it, thereby enhancing participation and collaboration around the topic of the data.

* Corresponding author.
E-mail address: gstowers@sfsu.edu (G.N.L. Stowers).

http://dx.doi.org/10.1016/j.giq.2016.07.001
0740-624X/

Further movement towards open data was the shift of U.S. federal policy towards e-government as "open government", underscored in President Obama's Memorandum on Transparency and Open Government in 2009. This policy document defined three foundational principals: transparency, participation and collaboration (Ganapati and Reddick, 2012, p. 115; McDermott, 2010; Jaeger and Bertot, 2010).[1]

In 2010, Tim Berners-Lee (the founder of hypertext linking) extended our understanding by laying out the levels of open data with his Linked Open Data 5 Star plan (Berners-Lee, 2010). This scheme is found in Table 1. One star data is available to all, two stars mean the data is machine-readable, three stars means it is machine-readable and in a non-proprietary format. With four stars, the data has the properties of the three previous levels plus is based upon standards (RDF, or Resource Description Framework) which allow precise pointing to the data online. Five stars, or the most accessible open data level, includes everything else plus the ability to link one's data to another's data.

Later scholars and activists further refined open data definitions with three related principles: *availability and access*, *reuse and redistribution* and *universal participation* (Gurstein, 2013). Gurstein argues that data should be made freely available for no more than its reproduction cost, in its full form, and inconvenient and modifiable machine-readable formats (*availability and access*). Providing data more freely might mean using the non-proprietary CSV format, which is modifiable and widely used today on open data portals. Further, data need to be made available under terms that allow for their reuse and redistribution, including the ability to combine data with other datasets (*reuse and redistribution*). A CSV formatted dataset would enable users to download the dataset, combine with other CSV datasets and redistribute to other users interested in the same issue. Lastly, data must also be made available to the public in an equitable fashion (*universal participation*), such as a simpler format that would allow users without advanced technical skills, or more advanced software, to use the data. Machado and De Oliveira (2011) further defined open data as "the publication of data in open raw formats and ways that make it accessible and readily available to all and allow reuse, such as the creation of data mashups and applications" (p. 449).

The three laws of open government data (herein referenced as 'open data' for this paper) summarize what format it should take, and why:

1. *If it [(data)] can not be spidered or indexed, it does not exist.*
2. *If it [(data)] is not available in open machine readable format, it can not engage.*
3. *If a legal framework does not allow it [(data)] to be repurposed, it does not empower* (Eaves, 2009).

Kassen (2013) noted that the value of open data developed through releasing government data to the public in machine-readable formats because presumably both government transparency and civic engagement will increase. Such reasoning is consistent with Tolbert and Mossberger's (2006) findings that e-government initiatives have a positive impact on citizen attitudes of trust in their local government. Even more specifically, the positive effects of open data can be broadened to include economic innovation, increased accountability, increased government agency efficiency, and social and economic innovation (Kitchin, 2013). In just two cities (New York and Chicago), numerous instances have been cited of ways open data has benefitted city government or citizens themselves. New York officials point to the analysis of open data about building inspections to help improve predictions of fires; Chicago points to their use of crime data and the development of helpful applications from their open data (Goldstein and Dyson, 2013).

Despite the potential of positive changes towards accountability and transparency that open data could bring, concerns are increasingly being raised about the intended and/or unintended consequences of open data. For example, Kitchin (2013) raises concerns over neoliberalization and marketization of public services as open data is increasingly used primarily to build applications by private sector firms, further exacerbating inequality. While activists hope that open data can further citizen engagement with their governments, little evidence so far exists that this is happening.

Critics have raised other concerns, such as the important distinction between open data and the Right to Information Movement (RTI movement), "which promotes access to government information as a fundamental right" (Janssen, 2012, p. 3). Specifically, Janssen contends that if governments predominantly rely on economic innovation as a vehicle for transparency and accountability, then "open data risks the illusion of transparency and accountability, while in reality [it could be] causing information inequality and disempowerment of citizens" (p.12). The fact that there exists a digital divide, where access to information technology is not equitably distributed across populations, gives credence to this argument. Bates (2012) similarly concludes that a real danger of open data initiatives is that they would become framed in a manner that "champions the superiority of markets over social provision…potentially becoming, little more than corporate subsidy" (p. 10).

Such concerns raise a potential conflict between the goals of accountability and transparency and those of economic stimulation. Further, in this era of information scandals such as the case of Edward Snowden, who leaked information about a National Security Agency program that was spying on citizens within the United States (Risen, 2013), urgent concerns about government and transparency are apparent. Janssen (2012) argues that the type of data released through open data platforms has a direct effect on the outcome of accountability. Geographic data, postcodes, transportation data, corporate data and business data will stimulate economic innovation; whereas, budgetary data, spending expenditures, procedural data, and legal data will stimulate accountability efforts (p. 11).

Table 1
Tim Berners-Lee linked open data 5 star scheme.
Source: Berners-Lee, Tim. (2010). Is your linked open data 5 star? Retrieved on October 16, 2015, from http://www.w3.org/DesignIssues/LinkedData.html.

Stars	Interpretation
*	"Available on the web (whatever format) *but with an open license, to be Open Data*
**	"Available as machine-readable structured data (e.g., excel instead of image scan of a table)
***	"as (2) plus non-proprietary format (e.g. CSV instead of excel)
****	"All of the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
*****	"All the above, plus: Link your data to other people's data to provide context"

3. Open data evolution

In 2007, a group of open software activists met in Sebastopol, California and outlined the basic concepts of open data, defining public data as a public good that should be available to all (Chignard, 2013). President Obama, in his 2009 executive order, codified their ideas, a groundswell of support emerged around the Open Government concept, and the federal government created data.gov. Collectively called the open data movement, it has sprung up throughout the majority of states, in many cities throughout the country (Townsend, 2013) and now, across the world.

The CIOs of Chicago, Los Angeles, Boston, New York, Seattle, Washington, D.C. and San Francisco, working informally as the Group of Seven in 2009, focused on the need for a common open interface for

[1] McDermott (2010) explains that the Open Government Initiative is legal based on pre-existing laws including, but not limited to, The Paperwork Reduction Act of 1980, The E-Government Act of 2001 and 2002, The Freedom of Information Act and its amendments and the Open Government Act, 2007 (p. 404–411).

applications like 311[2] data (Douglas, 2010). Other cities like London, Palo Alto, Boston and others started developing portals, often with the help and support of organizations like Code for America and local non-profits wanting to use data to create new apps and solve community problems. Code for America created a Fellows program, recruiting coders to go to participating cities and develop innovative projects in civic technology. "Hackathons" were held where data experts created computer applications from city data. Nonprofits, cities, and the federal government held competitions for the best applications for solutions to government problems using government data. Community groups like Open Government Chicago, Smart Chicago Collaborative, and OpenOakland, formed and worked with cities (to various extents) on their data. (Goldstein and Dyson, 2013; Kassen, 2013).

Cities created new positions called Chief Data Officers (CDOs) to develop policies and organize their data. A new era of civic technology activism seemed to have arrived, with city staff, computer programmers, academics and activists working together on managing city data. Chicago highlights an entire series of new applications created from their open data, including How's Business, Plow Tracker (tracking snow plow work and cleared streets), Chicago Councilmatic (highlighting legislative business of the City Council), and Look at Cook (government budget visualization app). New technology startups were created (such as OpenGov), basing their products on the data made newly available to them from government. The new field of data journalism developed from the older field of computer-assisted reporting (CAR), also taking advantage of open data and the new software tools available to create easy to understand visualizations for their publications. This phenomenon can be seen in the work of the New York Times Graphics Department (Goldstein and Dyson, 2013) or projects like ProPublica's Surgeon Scorecard, with journalists acting as moderators for a database on local medical professionals.

Many cities were already familiar with success stories like Baltimore's CitiStat (begun with crime data https://data.baltimorecity.gov/) and CompStat (Austin https://www.austintexas.gov/page/compstat-reports/ San Francisco http://sanfranciscopolice.org/compstat) in many local police departments. Cities became further invigorated by stories from cities like New York, which showed great success with popular restaurant inspection data and used city data to solve problems in forecasting house fires. (Goldstein and Dyson, 2013; Helbig, Cresswell, Burke and Luna-Reyes, 2012). To date, the movement continues to grow.

In attempts to understand what was happening in this movement, scholars hypothesized models of open government data portals and their growth (Kalampokis, Tambouris and Tarabanis, 2011; Sieber and Johnson, 2015). Kalampokis et al. (2011) based their model on e-government growth models. In their scheme, there are four levels: the aggregation of government data, the integration of government data across all agencies, Stage 3 with the integration of government data with non-government data, and the last stage, the integration of government data with non-government formal and social data. Lee and Kwak (2011) also took a maturity approach that incorporated social media engagement, suggesting four stages: increasing data transparency, improving open participation, enhancing open collaboration and realizing ubiquitous engagement. Sieber and Johnson (2015) took a different approach, with a set of comparative models:

- "Data over the wall." Government publishes open data in a one-way flow directly to citizens and businesses in the community.
- "Code exchange: Government as open data activist." Under this model, government agencies take a more proactive role in encouraging users to develop new products and in turn, receives the benefits of applications developed from its data.
- "Civic issue tracker." With this model, there is more of a two-way street with data and information flowing from government to citizens and then back, as citizens use the data to track problems in the community and then report them back to the government, as a type of crowdsourcing of fixable problems.
- "Participatory open data: open data as open government." This last model describes a fully two directional exchange, with data coming from the government and citizens, alike, in a more fully co-production approach.

[2] Like 411 is designated in the United States for information and 911, for emergencies, 311 is designated for service calls. Many cities have now set up 311 service call centers with services being requested just via phone calls but have now expanded to web and texting requests. This 311 data is available now on several open data portals in cities.

4. Related work

There has yet been little systematic and empirical research on the subject of open data portals at the American sub-national level. However, research on other components of open data portals has been conducted. Kalampokis, Tambouris and Tarabanis' (2011) and Sieber and Johnson (2015) have provided initial theoretical work.

Case studies have been developed which have delved deeply into the workings of portals at other levels and other countries (Alexopoulos, Spiliotopoulou and Charalabidis, 2013; Goldstein and Dyson, 2013; Lathrop and Ruma, 2010). In a case study of Greece, Alexopoulos, Spiliotopoulou and Charalabidis (2013) provide descriptive analysis of the functional, semantic and technical features of open data portals of various levels of government. They examined the types of open data licenses used, types of searches available, and types of platforms, metadata and data format found. They did not incorporate the subject categories of the data or results of analysis of the open data.

With a comparative approach, Ganapati and Reddick (2012) conducted a survey of state CIOs that focused on state level portals.

At the international level, Petychakis, Vasileiou, Georgis, Mouzakitis and Psarras (2014) examined EU countries, focusing descriptive analysis of functional and technical perspectives. They compared attributes like the technical standards utilized by various national open data portals, types of searches available, numbers of languages used and categories of data (i.e., law enforcement, budget). In 2014, they concluded that EU countries had made steady progress in improving the quality of open data portals but there is still a great deal of variability among countries.

Other, practitioner research, such as from the PIRG Education Foundation and the Open Knowledge Foundation, has also provided descriptive information on open data and portals. A 2013 study (Davis, Baxandall and Perannunzi, 2013) for the PIRG Education Foundation examined open data portals but focused only on fiscal transparency. They reported that 17 of the largest 30 cities had analyzed budgetary data down to what they term the "checkbook" level of detail. However, this report was limited because it only focused on those 30 cities and budgetary data. The authors define two levels of budget transparency—1.0 and 2.0. Transparency 1.0 provides only limited access to what they term "scattered" data about public expenditures, intended for determined residents to find data and tools normally available only to informed insiders. Transparency 2.0, on the other hand, has sites that are encompassing (user-friendly with comprehensive data and tools), one-stop and searchable and downloadable. The researchers then applied grades of A to F for the 30 cities. Their scale utilized financial data only.

The Open Knowledge Foundation, in cooperation with Code for America and the Sunlight Foundation (2015), supports an ongoing, crowd-sourced project begun in 2014 where users can contribute information about open data portals. The focus is entirely on the subject content of the databases available on various portals. They then publish pages comparing different cities on types of data disclosures (i.e., asset disclosure or budget). Most interesting is the weighting scheme they have developed, which includes features like whether the data is machine-readable, provided on a timely basis, and measures the quality

and user-friendliness of the data. However, the Open Knowledge Foundation cities project focuses upon on the data itself, not on the open data portal, as is the purpose of this project. In this project, only existing American city open data portals that include machine readable data are included; another feature used in this project as opposed to the crowd-sourced project is the content of the data.

Open data indices have also been developed by various researchers. At the international level, indices of open data have advanced more significantly, including by the Open Knowledge Foundation. Their Global Open Data Index is a percentage index that incorporates components like whether the data is available online, is publicly available, in digital form, in machine-readable format, openly licensed, and provided on a timely basis (Open Knowledge Foundation, 2015). Most of these features are not relevant for our research project, as by definition the data in American urban open data portals is publicly available online and is machine-readable in digital format plus openly licensed.

The same is true of the Web Foundation's Global Open Index (http://index.okfn.org/) (Web Foundation, 2016a; Web Foundation., 2016b). This Index includes questions like, does the data exist? Is the data in digital form, publicly available, available for free? Is it online and is it machine-readable and openly licensed? These international efforts are important but for this research project; the Index is focused too closely on the features of national systems or on these basic questions. For this research project on American urban open data portals, these features are all assumed to be included.

Beyond case studies and the work by advocacy groups and some work at the international level, systematic research at the American sub-national level is still lacking. Kassen (2013) notes, "in practice it is not yet clear how the potential of the open data concept can be realized at the local level [as a political phenomenon] as there has been no analysis of current projects thus far" (p. 2). A recent workshop hosted by scholars of the International Development Research Centre, the World Wide Web Foundation, and Berkman Center at Harvard University discussed the potential outcomes of open data. The group found that "[t]here is a widespread lack of empirical evidence underlying the implementation of Open Data initiative… that can guide better practice and policy formulation" (International Development Research Centre; World Wide Web Foundation; Berkman Center at Harvard University, 2012).

At least one study in this neglected area, the necessary empirical work on open data portals in American cities, has been initiated. Karen Mossberger and Kuang-Ting Tai (2014) presented research in progress on the topic of cities and open data. Their research examined the proportion of cities with open data portals among those larger than 250,000 in population, a cross-sectional approach similar to the project reported in this article.

The research in this paper, providing descriptive findings on open data portals in American cities builds on this earlier research (Mossberger and Tai, 2014; Goldstein and Dyson, 2013; Lathrop and Ruma, 2010) then goes further by providing the first inferential analysis of urban portals and the factors that influence their adoption. Descriptive results of features in portals are provided, building upon those seen in Davis, Baxandall and Perannunzi (2013) and the Open Knowledge Foundation (2015). Finally, a model for predicting the features and content of open data portals is provided, using new indices.

5. Methodology and data collection

There are now several possible types of benchmarks and measures for open data adoption (Susha, Zuiderwijk, Janssen, and Gronlund, 2015; Veljkovic et al., 2014). The systems reviewed by Susha, et al. (World Bank-Open Data Readiness Assessment, the Open Data Barometer of the Open Data Institute and World Wide Web Foundation, the Open Data Index of the Open Knowledge Index, the PSI Scorecard from the ePSI Platform and Open Data Economy from Capgemini Consulting) were primarily based on the maturity and readiness of nation-state sites. The system proposed by Veljkovic et al. is based on features from the definition of open government (participation, collaboration, data and government transparency, and open data). The Global Open Data Index focuses on features that are a given in American open data portals. Other systems are based on survey research or interviews with site operators.

As seen above, there is clearly a need to conduct basic descriptive empirical research upon U.S. open data portals in American cities. Thus, this study focuses on examining the actual open data portals themselves—what features they contain and the kinds of datasets they are hosting. The research reported here goes beyond the kind of basic information provided in the Global Open Data Index and other indices, and then provides inferential analysis utilizing the descriptive findings.

The research methodology utilized here was a cross-sectional analysis of the open data portals of American cities. The cities chosen were all those listed by the United States government premier data portal (data.gov) in their cities section (http://www.data.gov/open-gov/). The list from data.gov contained 37 cities (counties were not included in this study). Of these, Atlanta was not included in the study because it required an account and password to enter and therefore, was not considered "open." Other listings of possible cities, such as DataPortals.org (supported by the Open Knowledge Foundation) were available but, data.gov was considered to be the definitive site. All of the open data portals offered by American cities listed in data.gov and so, in our study provide publicly available data for free, in digital format, online and openly licensed. Providing data with these features is simply how these projects are all structured.

Each site on the list was examined numerous times so members of our four-person research team could become fully familiar with the features available on each; each reviewer made notes on features they found on various portals as they explored. After this initial investigation, a draft list was developed that delineated the specific portal features to be studied, using the various definitions of critical elements of open data (machine-readable, accessible) (Gurstein, 2013, Berners-Lee, 2010). Local experts working in open data nonprofits curated and further revised the list since they are working in the trenches of the open data movement. Their input led to an adjustment of the features list to one that is incorporated into two different indices. Given the approach of this article, providing initial descriptive data and analysis of open data in cities, the indices in this research were developed from an applied perspective, as opposed to the more theoretical approach proposed by Veljkovic, Bogdanovic-Dinic, and Stoimenov (2014).

This research examines the features of the portals as well as the content area of the datasets provided. The first empirically-based index developed for this study, the Open Government Data Portal Index (OGDPI), focused upon portal features and ranged in scale from 1 to 100 (the initial points sum to 125 but were weighted so that the range was limited to 100, to ensure comparability with the other indices). Table 2 offers a list of the open data portal features included in this index, as well as the weights assigned to each (developed based upon discussions with open data experts).

The features incorporated in the OGDPI were those seen across multiple portals as necessary for both citizens and information technology professionals; these features were categorized into content, help, policy, and results. Content incorporated the number of datasets per 100,000 population, how the data appeared on the portal, whether the files were machine-readable (Berners-Lee, 2010), included metadata about the files, and whether users could manipulate or analyze the data online. Help features included whether users could search if there were tutorials or lists of help topics available, and whether clear contact information was available. Policy features included whether or not an open data policy or terms of use statement was present. The results category incorporated whether or not an app showcase or other information about results achieved from posting the data was present, whether analytics was posted, and whether APIs or another information for developers to use to create applications with the data was available.

Table 2
Features of open data portal index.

Feature	Points given for each feature
Content	
Datasets per 100,000 population	1 pt. for 1–25 datasets per 100k, 2 for 26 to 50 per 100k, 3 for 51 to 75, 4 for 76 to 100 per 100k, 5 for more than 100 per 100k population
Organization of datasets by categories	5 if yes, 0 if no
Are users able to manipulate data in datasets online	5 if yes, 0 if no
Number of categories into which datasets are organized	1 for 1 to 4, 2 for 5 to 8, 3 for 9 to 12, 4 for 13 to 16, 5 for over 16 IF logically organized
Are users able to create visualizations online with the data	5 if yes, 0 if no
Are the data files in machine readable formats	20 if yes, 0 if no
Are users able to suggest new datasets	5 if yes, 0 if no
Is metadata available to define and explain data	5 if yes, 0 if no
Help features	
Can users search through datasets	5 if yes, 0 if no
Can users search through help topics	5 if yes, 0 if no
Are tutorials available	5 if yes, 0 if no
Is a list of help topics available	5 if yes, 0 if no
Is clear contact information available for users if they need help	10 if yes, 0 if no
Policy	
Is there data or an open data policy	5 if yes, 0 if no
Is there terms of use statement	5 if yes, 0 if no
Results	
Is there an app showcase available	5 points for an app showcase, 3 for linkage to some collection of apps
Is there a call for action or invitation for citizens to use the data	5 if yes, 0 if no
Is other information about results of datasets available	5 if yes, 0 if no
Are analytics available	5 if yes, 0 if no
Are APIs and other information for developers available	5 if yes, 0 if no

Table 3
Content area features for dataset content index.

Content areas
311 data
Animals
Arts and Cultural Resources
Budgets, finances, expenditures
Salaries (employee)
Businesses
Demographics
Education/schools
Elections
Health
Housing
Library
Parking
Parks/environment
Permits
Planning & development
Public safety (police, fire stations, emergency management)
Crime
Procurement
Real estate
Restaurants, food inspections
Traffic, transportation
Utilities
Zoning
Parking
Other

Also, data was collected on the platform used for the open data portal; however, this was not included in the OGDPI index.

Table 3 describes the content areas incorporated into the second index, the Dataset Content Index (DCI). This index was developed and tested by the research team's initial examinations of the open data portals in the study. Each content area received one point if databases were available in that area on a given portal; the number was then multiplied by 4 to have a comparable index ranging from 1 to 100. Unlike the content areas highlighted by the Open Knowledge Foundation City Census, these categories are broader, incorporating the possibility of a variety of different types of datasets in each category.

These indices were used to test hypotheses developed in two basic ways, through 1. grounded research, incorporating input from practitioners in the field and examination of documents and artifacts such as open data portals themselves; and, 2. suggestions from earlier research. The six hypotheses developed for this study utilized both approaches in suggesting various characteristics of cities that could relate to the development of their open data portals.

H1. The content and sophistication (i.e., types of features) of open data portals will increase with the population size of cities, as larger cities have more resources and therefore, more ability to develop innovations (as shown by many, including Moon and Norris (2005) for e-government adoptions). The size of each city will be measured using the 2013 population of each city.

H2. The content and number of features will increase with the level of education of the city; in particular, we hypothesize that university cities or others with high tech industries would be more likely to have more features in their open data portals. The level of education will be measured by the percent bachelor's degrees in each city; its importance was suggested by examination of open data portals in American cities. For example, portals in university towns, even if they were small (like Austin), seemed to be more innovative and cities located near to high-tech areas, even if they were small (Palo Alto), seemed to show high numbers of features and sophistication.

H3. We hypothesize that the type of government (city-manager or mayor-council) will have an impact on the number of datasets or features of open data portals, resulting from the more effective and up to date professionalism of city managers in these systems. City government is measured by a binomial variable with each case either a 1 or a 2 for each type of government. In studies of American cities, the inclusion of the government type is necessary due to previous findings in many studies that professionalism of city manager governments predicts city outcomes.

H4. The degree of innovation in cities is expected to be an important feature in efforts to develop open data portals (Moon and Norris, 2005); some of the original so-called "G7" cities were not the largest cities, but they were more innovative, technologically or otherwise. Based upon grounded research and observation, we hypothesize that more innovative cities would be more likely to have more sophisticated open data portals. Code for America's Fellows Program selects fellows each year who have expertise in coding or another facet of civic technology and sends them to a participating city, also awarded through competition, to spend a year working on projects will be more innovative. Cities participating in this process are likely to be more innovative and interested in civic technology, as they pay to be involved in what are specifically civic technology projects. Due to the technological nature of this innovation, this variable is coded with a one if the city had ever been awarded a Code for America Fellow and a 0 if it had not.

Table 4
Descriptive statistics for indices and datasets.

	Number of datasets	Number of datasets per 100,000 residents	Open government data portal index	Dataset content index	Overall index
Mean	176.7	102.9	59.9	60.6	117.5
Median	94	56.7	58.4	56	112.4
Mode	18	3.8	58.4	56	98.4
Std. dev	259.0	333.4	17.1	23.1	32.9
Minimum	8	1.8	12.8	16	52.8
Maximum	1292	1983.5	89.6	96	182.4

H5. Based upon observation from grounded research approaches, the age of each portal was expected to make a difference, as we observed that the earliest open data portals had progressed further than newer sites. Thus, we hypothesize that the number of datasets and features on the portals would increase with age and degree of organizational development. The number of months since the creation of the portal measures the age of the portal.

H6. One other observation from grounded research was that some states had created regional consortia to pool resources between communities, allowing small communities to support open data portals. Therefore, participation in a regional consortium would positively impact the number of datasets and features in an open data portal. Consortium membership coded as a one for yes or 0 for no measured whether a city was a member of a regional consortium.

6. Findings

Table 4 presents basic descriptive data on each of the indices as well as the total number of datasets and the number of datasets per 100,000 population. The average number of datasets across cities is 176.7, although there are clearly outliers because the median number is 94, suggesting that a few cities have much larger numbers of datasets. Controlled by the number of residents, the average number of datasets per 100,000 residents is 102.9 but again, there are a few cities with much higher numbers since the median number of datasets per 100,000 population is only 56.7. In both cases, the standard deviations are quite large, indicating a great deal of variations between cities. This level of variation is verified by the minimum number datasets (8 and 1.8 for cities per 100,000 population) and maximum numbers of datasets of 1292 and 1983.5 per 100,000 residents.

The mean score on the Open Government Data Portal Index (OGDPI) is 59.9 with medians and modes very close. The mean score for the Dataset Content Index (DCI) is 60.6, also with medians and modes quite close. The minimum OGDPI is only 12.8 while the maximum is 89.6, also suggesting significant variation between cities in how their portal and their data is presented. This variation is also found for the content of their data; the minimum number of content areas (DCI) is 16 while the maximum is 96. The Overall Index (OI) has a minimum score of 52.8, a maximum or 182.4 and a mean of 117.5.

Table 5
Rankings of cities by each of three indices: open data portal, dataset content, and overall index.

Cities ranked by open data portal index		Cities ranked by dataset content index		Cities ranked by overall index	
San Francisco	89.6	Austin	96	Austin	182.4
Austin	86.4	Los Angeles	96	San Francisco	177.6
Baltimore	83.2	New York	96	New York	169.6
Seattle	78.4	Chicago	92	Baltimore	167.2
Honolulu	77.6	Houston	92	Seattle	162.4
New York	73.6	Denver, Colorado	88	Los Angeles	161.6
Madison	71.2	San Francisco	88	Chicago	152
Raleigh	71.2	Baltimore	84	Raleigh	143.2
Burlington	69.6	Seattle	84	Kansas City	142.4
Providence	68	Belleville, Illinois	76	Belleville, Illinois	134.4

Table 6
Frequency of content features in open data portals.

	% of sites with each feature
Are the datasets organized into categories?	91.7%
Are the datasets in machine-readable formats?	91.7%
Can users manipulate data from the datasets online?	69.4%
Can users create charts online with the data?	69.4%
Can users suggest new datasets?	66.7%

The results of ranking by each of the three indices indicates that San Francisco, Austin, Baltimore, New York, Seattle, Chicago, Honolulu and Houston appear in the top five of each of these three indices. The remainder of the rankings, shown in Table 5, suggest that large cities, university towns, and cities surrounded by high-tech industry are likely to have more open data portal features and to post a higher variety of content on their portals. These suggestions will be tested and results shown later in the article.

Table 6 displays how cities in the study organize and present data in their portals. Practically all cities (91.7%) present their datasets in categories so that users can better locate data in their interest areas. Even more important, 91.7% present their data in some machine-readable format; this is critical since offering data in a machine-readable format is not only part of the basic definition of open data, it is also essential to most citizens being able to utilize the data effectively.

Only a bit more than two-thirds of the cities offer features on their portals so that users can either manipulate the data online (69.4%) or create visualizations (also 69.4%). Two-thirds (66.7%) feature a way in which users can suggest new datasets.

Help features are also essential for ensuring a user-friendly experience to potential data users (Table 7). Most (86.1%) of the cities provide a way for users to search through the datasets. However, only about a third provide other ways of helping users: providing a list of help topics (36.1%), providing clear contact information for where users can get help (33.3%), allowing users to search through help topics (30.6%), or providing tutorials for users (30.6).

Since open data portals are a new phenomenon for cities, establishing an appropriate policy framework around them should be developed. However, as Table 8 indicates, only 50% of these cities are posting data or open data policies on their portals. More cities exhibit terms of use policies, however (61.1%).

Table 7
Frequency of help features in open data portals.

	% of sites with each feature
Is it possible to search through the datasets?	86.1%
Is there a list of help topics available?	36.1%
Is there clear contact information available on where to get help?	33.3%
Can the user search through help topics?	30.6%
Are tutorials available?	30.6%

Table 8
Frequency of policy features in open data portals.

	% of sites with each feature
Are there terms of use policy available?	61.1%
Is there data or open data policy available?	50.0%

Table 9
Frequency of technical features in open data portals.

	% of sites with each feature
Is information for developers available?	58.3%
What platform is used to host the open data portal?	
Socrata	61.1%
Other	36.1%

Table 10
Frequency of outcomes features in open data portals.

	% of sites with each feature
Is there a call for action or invitation for citizens to use the data?	58.3%
Are results showing analysis of the data available?	25.0%
Is there attention shown to apps developed from the portal's open data?	33.3%
Is there a showcase available of apps made from the open data?	22.2%
Are there linkages to apps made from the open data portal?	11.1%
Is analytics available?	19.4%

Technical information about the data is necessary so that businesses and citizens can effectively use the data, but only 58.3% of the cities post information for app developers on their sites (Table 9). Moreover, what is very clear about these early local government data portals is that one vendor, Socrata, is powering most of their efforts (61.1%). The ubiquity of this vendor becomes significant, as Socrata's basic but "clunky" data analysis tools, such as the ability to use the data in basic graphics or select subsets of data for analysis or download, are often all that is available in open data portals and could be limiting activities online.

The actual value of open government data lies in how it will ultimately be applied and used (Table 10). There are three general areas in which government data can prove useful: 1.) to improve government functioning; 2.) to enhance citizen knowledge about government and its functions; or, 3.) to enhance local economic development and business by allowing businesses to create and sell interactive applications (apps) from the data. However, cities are not yet focusing on the first two of these outcomes (Table 9) but are mainly providing data to local businesses to build applications, through informal means. While 58.3% of the cities have some call to action, thus encouraging its use by citizens, there are relatively few actual results exhibited on the site. Only 25% of the sites present any analysis of their data, and only 19.4% of the sites offer analytics reported by the site managers (see Table 10).

Finally, only one-third (33.3%) of the portals present opportunities for showcasing applications created from the data (22.2% for a more formal showcase and 11.1% just links to any applications from the data).

From the regression analysis of the five different dependent variables (number of datasets, the number of datasets per 100,000 population and the three indices—OGDPI, the DCI and the Overall Index) several patterns are visible (see Table 11). First, the type of government (H3) had no influence on any of the dependent variables, nor did the degree of civic innovation (H4), the age of the portal (H5) or the level of education in the city (H2).

The inclusion of the city in a regional consortium (H6) had a significant and negative influence upon the OGDPI, indicating that cities not in a consortium (i.e., larger cities) were more likely to have a greater number of features in their open data portals.

Population size (H1) had the strongest impact of any of the hypothesized factors, as it often does for many technological and innovative features in urban areas. It had an extremely significant impact on the number of datasets and strong and significant impact on the DCI (types of content available online) and the Overall Index.

Overall, some support for the first hypothesis (H1: population size) and limited support was found for H6 (regional consortium). However, no support was found for the other hypotheses (H2: the education of the city, H3: the type of government, H4: the degree of civic innovation, or H5: the age of the portal).

That population size was so important to predicting the number of datasets (explaining 79.8% of the variation in the number of datasets (OGDPI), confirms the strength of population. The only two other statistically significant models were the DCI and the Overall Index. However, the level of variation explained was considerably less (24.9 and 22.9%, respectively).

7. Discussion and conclusions

Results from this study on the features of open data portals in American cities found city population size (level of resources) was a critical factor on all three of the dependent variables: the number of datasets listed in the portal, the different types of dataset content and the Overall Index (portal features and content types together). Thus, it is no surprise that the size of a city, a proxy for the availability of resources, is necessary here as it has been in predicting many other facets of technology and innovation.

This research provides critical descriptive information and initial analysis of the patterns and activities of open data portals in American cities. The impact of having one vendor dominate the market shows in the number of cities organizing their data in categories in the same way, in having machine-readable data, in enabling users to be able to analyze and chart their data online, in having search facilities readily available, and in providing needed API and other data for potential developers. This also results in many open data portals having the same "look and feel" (Zuiderwijk and Janssen, 2014). There is much more variability across cities for other features. For instance, very few sites incorporate a mechanism for users to see what results, if any, the data is having on the community.

Table 11
Regression results.

	Number of datasets	Datasets per 100,000 population	Open data portal index	Dataset content index	Overall index
Type of government	−0.041	−0.255	0.154	−0.028	0.063
Code for america fellow	0.16	−0.181	0.136	0.313	0.291
Age of portal	0.19	0.204	0.085	0.107	0.12
Regional consortium	−0.014	0.018	−0.449**	0.039	−0.214
Population 2013	0.887***	0.042	0.133	0.532***	0.442**
Percent bachelor's degree	0.125	0.275	0.169	0.127	0.179
F	16.453***	1.392	1.81	2.765*	2.585*
Adjusted R squared	0.798	0.07	0.132	0.249	0.229

* ≤0.05.
** ≤0.01.
*** ≤0.000.

Studies in this area are just beginning, and other descriptive questions still need to be addressed. Often, the emphasis of these portals and the hackathons utilizing portal data is driven by the interests of private sector technology firms using the data to create applications for customers. However, with so few of the portals tracking this usage and the subsequent applications developed, the extent of even this obvious usage is still cloudy. What kinds of apps are developed, using what data? How many? Which are effective and which are not? Other uses, either by citizens or by the cities themselves, have yet to be fully elucidated. To what extent are other city agencies using the data themselves to improve their services? To what extent are citizens using the data, and for what purposes? These questions and many others should form the beginnings of a new research agenda in this emerging area of digital government.

More research is needed to develop a more nuanced critical perspective on open data and a deeper understanding of its policy implications. Now that empirical results are available on the types of features and content in open data portals, scholars should be using this information to determine the usefulness of the models presented by Kalampokis, Tambouris and Tarabanis (2011) and Sieber and Johnson (2015). Specifically, scholars could do better documenting and analyzing the classification of information, providing a perspective on the quantity and location of the information, and describing its various contexts. More clear classifications would better inform any future critiques of open data, and would create a better predictive capacity for downstream consequences - intended or otherwise. Future questions should also relate to the administrative mechanisms, policies and/or legal principles that enable and drive open data initiatives, the data consumers use, and the demand for various types of data. Longer-term investigations could describe the degree of innovation that will develop from open government data initiatives.

Finally, we would like to recommend a more consistent presentation by government agencies in their data portals. It would be helpful to citizens, businesses and researchers alike if there were a standard set of categories developed for the types of datasets as well as a consistent type of presentation. The development of ethics and data provision standards would also be helpful to the development of the field, as would a call for more focus on results from these efforts.

This article has contributed to the knowledge of open data by providing an important empirical baseline of the content and features of American city data portals as well as analysis of some of the fundamental patterns among cities in their provision of open data. We hope this will be followed by additional research on the questions highlighted here.

References

Alexopoulos, C., Spiliotopoulou, L., & Charalabidis, Y. (2013). Open Data Movement in Greece: A Case Study on Open Government Data Sources. *ACM*. http://dx.doi.org/10.1145/2491845.2491857 (978-1-4503-1969-0/13/09).

Bates, J. (2012). "This is what modern deregulation looks like": Co-optation and contestation in the shaping of the UK's open government data initiative. *The Journal of Community Informatics*, *8*(2).

Berners-Lee, T. (2010). *Is Your Linked Open Data 5 Star? Retrieved on October 16, 2015, from*http://www.w3.org/DesignIssues/LinkedData.html

Chignard, S. (2013). A brief history of open data. *Paris Tech Review* December 14, 2013. Retrieved from http://www.paristechreview.com/2013/03/29/brief-history-open-data/ (February 28, 2015)

Davis, B., Baxandall, P., & Pierannunzi, R. (2013). *Transparency in city spending: Rating the availability of online government data in American's largest cities.* Public Interest Research Group (PIRG) Education Fund (Retrieved February 18, 2015, from) http://www.uspirg.org/reports/usp/transparency-city-spending

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

Douglas, M. (2010). G7: CIOs from seven big-cities work together to develop open-source IT solutions. *Government Technology* Retrieved from http://www.govtgech.com/G7-Big-City-CIOs-Work-to-Develop-Open-Source-IT-Solutions.html (on February 15, 2015)

Eaves, D. (2009). The three laws of open government data. Retrieved 2013 from https://eaves.ca/2009/09/30/three-law-of-open-government-data/

Ganapati, S., & Reddick, C. G. (2012). Open e-government in US state governments: Survey evidence from Chief Information Officers. *Government Information Quarterly*, *29*(2), 115–122.

Goldstein, B., & Dyson, L. (2013). *Beyond transparency: open data and the future of civic innovation.* San Francisco: Code for American Press.

Gurstein, M. (2013). Should "open government data" be a product or a service (and why does it matter?). *Gurstein's Community Informatics* (Retrieved May 11, 2015, from) https://gurstein.wordpress.com/2013/02/03/is-open-government-data-a-product-or-a-service-and-why-does-it-matter/

Helbig, N., Cresswell, A. M., Burke, G. B., & Luna-Reyes, L. (2012). *The dynamics of opening government data: A white paper.* Center for Technology in Government.

International Development Research Centre, World Wide Web Foundation, & Berkman Center at Harvard University (2012, April 16). Fostering a critical development perspective on open government data. (Retrieved May 11, 2015, from) http://public.webfoundation.org/2012/04/ODRS/ODR-Brasilia-Meeting-Report.pdf

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, *27*(2010), 371–376.

Janssen, K. (2012). Open government data: Right to information 2.0 or its rollback version? *ICRI Research Paper*(8).

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open government data: A stage model. In M. J. e. al. (Ed.), *10th IFIP WG 8.5 international conference on electronic government, EGOV 2011. 6846.* (pp. 235–246). LNCS.

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.

Kitchin, R. (2013). Four Critiques of Open Data Initiatives. *The Programmable City* Retrieved from http://www.nuim.ie/progcity/2013/11/four-critiques-of-open-data-initiatives/ (on February 20, 2015)

Lathrop, D., & Ruma, L. (2010). Open government: Collaboration. *Transparency and Participation.* Sebastopol, California: O'Reilly Media.

Lee, G., & Kwak, Y. H. (2011). Open government implementation model: A stage model for achieving increased public engagement. *The proceedings of the 12th annual international conference on digital government research.*

Machado, A. L., & De Oliveira, J. M. P. (2011). DIGO: An open data architecture for e-government. *2011 15th IEEE international enterprise distributed object computing conference workshops.* http://dx.doi.org/10.1109/EDOCW.2011.34 (Retrieved from).

McDermott, P. (2010). Building open government. *Government Information Quarterly*, *27*(2010), 401–413.

Moon, J., & Norris, D. (2005). Does managerial orientation matter? The adoption of reinventing government and E-government at the municipal level. *Info Systems*, *2005*(15), 43–60.

Mossberger, K., & Tai, K. (2014). Panel presentation - #OpenData #BigData: Data, big and small, in the public affairs curriculum. *Presented at the annual conference of the network of schools of public policy, affairs, and administration. Albuquerque, New Mexico, November 4–6, 2014.*

Nam, T. (2012). Citizens' attitudes toward open government and government 2.0. *International Review of Administrative Sciences*, *78*(2), 346–368.

Open Knowledge Foundation (2015). U.S. city open data census. (Retrieved February 18, 2015, from) http://us-city.census.okfn.org

O'Reilly, T. (2005). *What is web 2.0: Design patterns and business models for the next generation of software.* O'Reilly Media, Inc (Retrieved May 11, 2015, from) http://oreilly.com/web2/archive/what-is-web-20.html

Petychakis, M., Vasileiou, O., Georgis, C., Mouszakitis, S., & Psarras, J. (2014). A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 34–47 (May 2014).

Risen, J. (2013, October 17). *Snowden says he took no files to Russia.* New York Times (Retrieved May 11, 2015 from:) http://www.nytimes.com/2013/10/18/world/snowden-says-he-took-no-secret-files-to-russia.html?_r=1

Sandoval-Almazan, R., & Gil-Garcia, J. R. (2012). Are government Internet portals evolving towards more interaction, participation, and collaboration? Revisiting the rhetoric of e-government among municipalities. *Government Information Quarterly*, *29*, S72–S81.

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(2015), 308–315.

Sunlight Foundation (2015). Impacts of open data. (Retrieved February 18, 2015, from) http://assets.sunlightfoundation.com.s3.amazonaws.com/documents/policy/impacts-of-open-data.pdf

Susha, I., Zuiderwijk, A., Janssen, M., & Gronlund, A. (2015). Benchmarks for evaluating the progress of open data adoption: Usage, limitations, and lessons learned. *Social Science Computer Review*, *33*(5), 613–630.

Tolbert, C. J., & Mossberger, K. (2006). The effects of E-government on trust and confidence in government. *Public Administration Review*, *66*(3), 354–369 (May–June 2006).

Veljkovic, N., Bogdanovic-Dinic, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, *31*(2014), 278–290.

Web Foundation (2016a). Open data barometer 2014 technical survey handbook. (Retrieved 2016 from) http://www.opendatabarometer.org/assets/survey/OpenDataBarometer2014-ResearchHandbook.pdf

Web Foundation (2016b). Open data barometer questions. (Retrieved 2016 from) http://www.opendatabarometer.org/assets/survey/WebIndex2014-ODB-Sections-ResearchHandbook.pdf

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

Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new Utopia New York.* W.W. Norton and Company, U.S. General Services Administration. Data.gov https://www.data.gov/.

Jeffrey Thorsby received his M.P.A from San Francisco State University in 2015. He has worked on neighborhood resilience and on data visualizations for the San Francisco Ethics Commission ("Spending to Influence: Campaign Finance and Lobbying in San Francisco in 2014").

Genie N.L. Stowers is Professor of Public Administration at San Francisco State University. Her current research interests are online service delivery in cities and data visualization in the public sector. She has published widely on egovernment and has just published Data Visualization for Public Managers. She teaches budgeting, information technology and statistics.

Kristen Wolslegel received her M.P.A from San Francisco State University in 2015. She was the Study Project Manager for Causes Count: The Economic Power of California's Nonprofit Sector, on the impacts of nonprofits in California, for CalNonprofits and worked on data visualizations for the San Francisco Ethics Commission ("Spending to Influence: Campaign Finance and Lobbying in San Francisco in 2014").

Ellie Tumbuan received her M.P.A from San Francisco State University in 2015. She is an independent consultant, a Fellow at the New Leaders Council in Oakland, CA and Development Director at Open Oakland.