All for data and data for all?

The open data movement and the influence of the technology industry

Anissa Tanweer

University at Washington

Department of Communication

19 March 2017

Professor Benjamin Mako Hill

INTRODUCTION

The public's right to access data produced by the government is enshrined in the 1966 Freedom of Information Act (FOIA) pertaining to federal data, as well as numerous state and local ordinances that guarantee public access to information at all levels of government. Until recently, the prevailing paradigm for accessing this data has been the FOIA request, in which organizations or individuals petition for the specific documents they wish to access. The government agency responsible for that data produces copies of those specific documents, and the requesting parties pay any administrative costs associated with fulfilling the request, which can be prohibitively expensive. Not surprisingly, although any citizen has the right to request government information, most FOIA requests come from journalists, advocacy organizations, lobby groups, and businesses that can afford to pay.

Although FOIA continues to be a powerful tool, in recent years momentum has shifted toward a newer paradigm. Instead of the public pulling data from clutches of government, the "open data" model involves government pushing data out to the public by proactively making it available online (Lathrop & Ruma, 2010). The origins of the open data movement are often traced back to 2007, when a group of activists calling themselves the Open Government Working Group convened in Sebastopol, CA and penned a manifesto that called on governments to make data openly available by default (Chignard, 2013). The working group, which included prominent figures in the open source and free culture movements such as Tim O'Reilly, Carl Malamud, Lawrence Lessig, and Aaron Swartz (Malamud, 2007), drafted a set of principles that designates government data as "open" when it meets the following eight criteria: it is complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and license-free (Open Government Working Group, 2007). In essence, open data is about making information collected by the government publicly available in a format that is both easy to access and easy to use.

Two years after the Sebastopol meeting, in 2009, the newly inducted Obama administration launched data.gov, the first online portal for open government data in the country. That same year, a group of Chief Information Officers from seven major US cities committed to creating web interfaces for opening their municipal data (Douglas, 2010). As the open data trend continues to spread from within government (McDermott, 2010), organizations of “civic technologists” have sprung up, often working collaboratively with government to make data available and put it to use (McNutt et al., 2016; Thorsby, Stowers, Wolslegel, & Tumbuan, 2015). One notable organization with national reach is Code for America, which places tech-savvy data analysts as employees in local government agencies. Civic technologists have also formed more grassroots community groups, such as local brigades of volunteers organized by Code for American (Goldstein & Dyson, 2013), and independent organizations such OpenSF in San Francisco or BetaNYC in New York City. The open data ecosystem also includes advocacy and philanthropic organizations such as the Sunlight Foundation and Open Knowledge Foundation (McNutt et al., 2016), as well as for-profit companies like Socrata, which services the online open data portals of many American cities (Thorsby et al., 2015).

RELATED WORKS

Literature on open data ranges from normative essays and prescriptive studies (e.g Goldstein & Dyson, 2013; Lathrop & Ruma, 2010; Meijer & Bolivar, 2015; O’Reilly, 2011) to critiques about open data failing to live up to its promise or promoting a neo-liberal economic agenda (e.g. Bates, 2012; Gurstein, 2011; Janssen, 2012). Descriptive scholarship has proposed models for understanding the evolution of open data (e.g. Kalampokis, Tambouris, & Tarabanis, 2011; Sieber & Johnson, 2015), while other studies have set out to understand who is participating in the movement (e.g. O'Connor, 2015) and how people make sense of their involvement (e.g. M. Janssen, Charalabidis, & Zuiderwijk, 2012). A number of evaluative works propose criteria for assessing the success of open data efforts (Dawes, 2010; Vetrò et al., 2016), and a few empirical studies have attempted to identify the conditions and features that predict various components of the open data movement (McNutt et al., 2016; Thorsby et al., 2015).

As Janssen (2012) has noted, dominant discourses have largely drifted from the rights-based argument made at Sebastopol to a more economic rationale. Open data is often justified not only as a transparency measure to help hold democratic governmental bodies accountable, but as a resource to be exploited for economic gain and social innovation. As The Data Foundation puts it in their publication, "The State of the Union of Open Data, 2016":

> Open government data is a powerful resource that our society has only just begun to harness. Access to public-sector information as open data can of course tell us more about what government is doing – how money is spent, how programs are performing, whether progress is being made to address persistent societal issues – but this is just the smallest fraction of the immeasurable economic value of open data…. Within government, open data greatly reduces the costs of sharing and using information….In the private sector, open data can help investors better understand risk and opportunity and provide communities with information to advocate for change and improve their lives.
> (Gill, Hollister, & Hughes, 2016)

In a recent survey of participants in open data efforts conducted by Socrata, the beneficial impacts of open data are placed into the following categories: economic development, operational efficiency, quality of life, and public safety ("2016 Socrata Open Data Benchmark Study," n.d.). It is widely believed that in order to reap these rewards, there must be "vigorous third party activity to help citizens interact and add value to [open data]" (Robinson, Yu, Zeller, & Felten, 2008). Such non-governmental third party activity is often attributed to both civil society and hi-tech firms, as evidenced by The Data Foundation's observation that "the U.S.

open data movement has no leader, but is invigorated by the nation's government, nonprofit, and tech-industry sectors" (Gill et al., 2016). However, there is reason to believe that non-profit or civil society involvement in open data is not actually independent of the technology industry, but rather inextricably imbricated with it. In places that are recognized as being at the forefront of the evolution of open data, civic engagement on its own is rarely cited as a driver of open data or civic technology; rather a thriving technology company is portrayed as the secret sauce in a recipe that also includes civic engagement. For example, in one of the first seven cities to create an online open data portal, the City of Seattle's Civic Technology Advocate has said, "we have everything we need for an impactful civic technology scene – a robust technology industry, an entrepreneurial attitude, and civic consciousness" (Boyd, 2016). And in a survey of open data community organizations in San Francisco and Raleigh, O'Connor found that the majority of participants in San Francisco identified as either an analyst or a programmer, and that the plurality of participants in Raleigh were employed in the computer and mathematics industry (O'Connor, 2015).

The few empirical studies of that have examined factors predicting adoption and development of open data programs in U.S. cities have not examined the influence of the technology industry, in spite of acknowledging its importance. McNutt et al. suggest that early adopters of civic technology are likely to be cities with "a decent resource base," "good management practices," and are located in "communities with a substantial technology sector" (McNutt et al., 2016, p. 161). Yet when testing for factors associated with civic technology, of which they consider open data to be a fundamental component, they include only population (which they consider to be a proxy for resources), and the receipt of governance awards (which they consider to represent good management practices), without accounting for the presence of a technology sector. Similarly, Thorsby et al. “hypothesize that university cities or others with high tech industries would be more likely to have more features in their open data portals,” because when developing a coding schema for their content analysis of open data portals, they noticed that “cities located near to high-tech areas, even if they were small (Palo Alto), seemed to show high numbers of features and sophistication” (Thorsby et al., 2015, p. 5). But when testing this hypothesis as part of their “inferential analysis of urban portals and the factors that influence their adoption” (Thorsby et al., 2015, p. 4) **,** they only measured the level of educational attainment in the city as their independent variable of interest, according to the percentage of people with bachelor’s degrees.

Both studies found few factors associated with their respective measures of open data. Thorsby et al. found that only population size was a significant predictor of open data portal features, while McNutt et al. found that only population and management orientation were significant in their study. Neither tested for the role of the technology industry, however, in spite of noting its assumed importance. Therefore, this study seeks to fill a gap in extant literature by testing the relationship between the technology industry and open data.

STUDY DESIGN

This study is motivated by an interest in whether or not the technology industry is a key factor in the availability and use of open data. This hunch is based on both anecdotal accounts from the discourse of open data that portray the presence of the technology industry as an enabler (e.g. Boyd, 2016; Gill et al., 2016), as well as previous scholarly literature that gestures to the important role of this industry without empirically investigating that claim (e.g. McNutt et al., 2016; Thorsby et al., 2015). More specifically, the objective of the analysis presented here is to determine whether a high concentration of the technology industry in a particular geographic region is associated with a more favorable environment for open data in the cities of that region.

The concept of an environment for open data is operationalized along two dimensions related to the U.S. Open Data Census (USODC): 1) inclusion in the USODC, and 2) score assigned by the USODC. The U.S. Open Data Census is a project of the Open Knowledge Foundation, the Sunlight Foundation, and Code for America (McNutt et al., 2016) that seeks to track which datasets are available in American cities, and the degree to which they fulfill various criteria of "openness" ("U.S. Open Data Census Frequently Asked Questions," n.d.). It is worth that noting that in spite of its name, the U.S. Open Data Census is NOT a census, in that it is not a systematic and comprehensive account of all the open data sets in every American city. Rather, individuals voluntarily go to the USODC website and enter information about what kinds of datasets are available in a particular city, based on standardized information solicited through by the U.S. Open Data Census. Because the USODC is not an exhaustive source of information, for the purpose of this study, it is *not* considered to represent a comprehensive account of how all cities in the US are actually performing on open data compared to one another. What the USODC *does* represent is a) an indication that some people care enough about their city's open data reputation to visit the website and systematically enter this detailed information, and b) a reasonable, though possibly imprecise, indication of the degree to which open data is available in those cities. As such, I treat the USODC not as an authoritative measure of performance, but rather as an indication of how favorable the open data environment is in a given city, based on both the presence of open data and the presence of people who are engaged in actively advertising or promoting that presence. In both cases, the unit of analysis is individual cities that are contained within US Census Bureau-designated metropolitan statistical areas

(MSA's). Cities that appear in the USODC but are not part of MSA's have been excluded from quantitative analysis because the dependent variables are not available for those cities.

Dependent variables

Inclusion in the USODC. A list of metropolitan statistical areas was taken from the U.S. Census Bureau and compared against the list of cities appearing in the U.S. Open Data Census. A dichotomous variable was assigned to each metropolitan statistical area indicating whether or not any cities from those MSA's were represented in the USODC.

USODC Score. The USODC assigns an overall open data score and ranking to each of the cities it tracks. The scores are assigned according to a system of weighted criteria for open data. Each data set receives 5 points if the data exists, 5 points if it is publicly available, 5 points if it is available online, 10 points if it is current and timely, 15 points if it is available for free, 15 points if it is machine readable, and 30 points if it is openly licensed ("U.S. Open Data Census Frequently Asked Questions," n.d.). These points are added so that each dataset can earn a maximum of 85 points, and the points for each data set are then added together to yield a cumulative score for each city. There appears to be no upper limit to these cumulative scores, and at the time of data collection, USODC scores ranged from zero to 1,700. Summary statistics for the ODC Scores can be found in Table 0.

```
  odc_score
 Min.   :   0.0
 1st Qu.: 142.5
 Median : 547.5
 Mean   : 624.8
 3rd Qu.:1041.2
 Max.   :1700.0
```

Table 0: Summary statistics of ODC scores for cities in ODC that are part of U.S. Census designated metropolitan statistical areas. n = 66

Independent variable

Concentration of the information sector. Laypersons and scholars alike often refer to the role of the "technology industry" when discussing open data. However, the US federal government does not recognize a single sector that specifically and exclusively covers the entire range of activities that people colloquially understand to encompass this sector, which often includes commercial activities such as the manufacturing of computer hardware (e.g. Intel), online retail (e.g. Amazon), maintenance of social media platforms (e.g. Facebook), and provision of web services (e.g. Google). However, according to the Bureau of Labor Statistics (BLS) NAICS designations of super-sectors, the "Information Sector" encompasses many of the commercial activities that people commonly associate with the tech industry. BLS describes the information sector in the following way:

> The Information sector comprises establishments engaged in the following processes: (a) producing and distributing information and cultural products, (b) providing the means to transmit or distribute these products as well as data or communications, and (c) processing data.
>
> The main components of this sector are the publishing industries, including software publishing, and both traditional publishing and publishing exclusively on the Internet; the motion picture and sound recording industries; the broadcasting industries, including traditional broadcasting and those broadcasting exclusively over the Internet; the telecommunications industries; Web search portals, data processing industries, and the information services industries.
>
> The Information sector groups three types of establishments: (1) those engaged in producing and distributing information and cultural products; (2) those that provide the means to transmit or distribute these products as well as data or communications; and (3) those that process data.
>
> ("Bureau of Labor Statistics industries at a glance: Information," n.d.)

As one can see, the information sector also includes activities that are not typically associated with the tech industry, such as movie and music production. But given the otherwise close alignment between the way people colloquially bound the tech industry and the way BLS defines a large part of the information sector, the concentration of the technology industry is operationalized in this paper as the concentration of the information sector in a given city. BLS measures the concentration of 14 super sectors, including information, with a variable called "location quotient" (LQ). LQ's represent the share of employment attributed to a certain industry in a given geographical area, compared to the national average for that sector's share of total employment in the United States. An LQ of 1 indicates that a given sector's share of employment in a given geographical area is equal to its share of employment in the nation overall, while an LQ of less than 1 indicates that a sector's share of employment in a given geographical area is smaller than its share of employment in the nation overall. Likewise, an LQ greater than 1 indicates that a sector's share of employment is larger than its share of employment in the nation overall ("Bureau of Labor Statistics QCEW Location Quotient Details," n.d.). For this study, LQ data for all US metropolitan statistical areas during the year 2015 was collected from the BLS website. LQ scores for the information sector ranged from .37 to 2.69. Summary statistics of the independent variables are listed in Tables 1 and 2.

Control variables

Population. While considering the information sector as a dependent variable, this analysis controls for population based on previous findings that found population to be an important predictor of open data outcomes (McNutt, Thorsby). Population is operationalized as the US Census Bureau's population estimates for MSA's during the year 2015.

Wealth. Because wealth is a global variable that is likely to affect the resources available to a given city for developing an open data approach, I also control for wealth in this analysis. Wealth is operationalized as the US Census Bureau's median home value estimates for MSA's during the year 2015. Summary statistics of the control variables are listed in Table 1 and 2.

```
   info_lq        home_value        pop_est
 Min.   :0.370   Min.   :118700   Min.   :  114181
 1st Qu.:0.665   1st Qu.:161400   1st Qu.:  473688
 Median :0.820   Median :219700   Median : 1271142
 Mean   :1.009   Mean   :249215   Mean   : 2412752
 3rd Qu.:1.225   3rd Qu.:286650   3rd Qu.: 2894778
 Max.   :2.690   Max.   :718400   Max.   :20182305
```

Table 1: Summary statistics of variables for cities in ODC that are part of U.S. Census designated metropolitan statistical areas. n=59

```
> summary(msa.all.false[,c(3,4,5)])
    info_lq         home_value         pop_est
 Min.   :0.1300   Min.   : 73800   Min.   :  54521
 1st Qu.:0.4200   1st Qu.:127800   1st Qu.: 126850
 Median :0.5900   Median :158900   Median : 193421
 Mean   :0.6144   Mean   :177792   Mean   : 322899
 3rd Qu.:0.7475   3rd Qu.:203500   3rd Qu.: 393789
 Max.   :1.6100   Max.   :668300   Max.   :4489159
```

Table 2: Summary statistics of variables for metropolitan statistical areas that do not appear in the ODC. n=258

Hypotheses

H1: For all cities that are part of US Census Bureau-designated metropolitan statistical areas, concentration of the information sector is associated an increased likelihood of appearing in the ODC, while controlling for population size and median home value. H1 = βinfo>0

H2: For cities that are included in the US Open Data Census, concentration of the information sector are positively associated with ODC score, while controlling for population size and median home value. H3 = βinfo > 0.

Analytic Methods

For H1, a bivariate logistical regression model with simultaneous inclusion of all variables was fitted to the “ODC inclusion” variable, with information sector concentration as the independent variable, and population and median home value as control variables.

For H2, an OLS regression model with simultaneous inclusion of all variables was fitted to the "ODC score" variable, with information sector concentration as the independent variable, and population and median home value as control variables.

Results

The logistic regression model in H1 yielded a null deviance of 304.67, residual deviance of 176.57, and a McFadden pseudo-variance measure of .420. The coefficient for the information sector was 1.660, significant at the .05 level (p = .005). The coefficient for home values was .621 and was not statistically significant. The coefficient for population was 1.494, and was significant at the .001 level (p = .000). Full results are listed in Table 3.

The OLS regression model in H2 yielded an R-squared of .1802 and an adjusted R-squared of .1406. The F-Test (F=4.544) was significant (p = .006). Coefficients were 186.51 for the information sector, 83.99 for median home value, and 114.36 for population. The information sector had the largest beta weight, but none of these coefficients were statistically significant at the .1 level. Full results are listed in Table 4.

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -29.7768     6.4481  -4.618 3.88e-06 ***
info_lq           1.6598     0.5854   2.835  0.00458 **
log(home_value)   0.6205     0.5084   1.221  0.22223
log(pop_est)      1.4937     0.2254   6.626 3.45e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 304.67  on 316  degrees of freedom
Residual deviance: 176.57  on 313  degrees of freedom
AIC: 184.57

Number of Fisher Scoring iterations: 6


         llh      llhNull           G2     McFadden         r2ML         r2CU
-88.2845766 -152.3335802  128.0980072    0.4204523    0.3324191    0.5383076
```

Table 3. Results of binary regression model testing Hypothesis 1.

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2225.58    2024.12  -1.100   0.2758
info_lq           186.51     151.01   1.235   0.2215
log(home_value)    83.99     177.14   0.474   0.6371
log(pop_est)      114.36      58.12   1.968   0.0536 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 487.5 on 62 degrees of freedom
Multiple R-squared:  0.1802,    Adjusted R-squared:  0.1406
F-statistic: 4.544 on 3 and 62 DF,  p-value: 0.006079
```

Table 4. Results of OLS linear regression model testing Hypothesis 2.

Threats to Validity

The primary threat to validity in this paper is that the dataset used for measuring the dependent variable – the U.S. Open Data Set – is not comprehensive or systematically sampled.

Conclusion

The results of the binary regression model that tested Hypothesis 1 suggest that, as expected based on previous literature, population is a significant predictor of a favorable open data environment. More importantly for the purpose of this study, however, is that even when controlling for the relationship between population and inclusion in the U.S. Open Data Census, the concentration of the technology sector also is a significant predictor of a favorable open data environment, as measured by presence in the USODC. In fact, when we exponentiate the coefficient for the information sector in the estimated regression model, we see that the odds of a city being included in the USODC increase five-fold for every unit of increase in information sector concentration.

Converting these odds probability ratios can provide an even better sense of the difference the technology sector makes to open data. If we take an average American city in a metropolitan statistical area with an information sector LQ of 1 (indicating that the concentration of the information sector is the same in that area as it is in the rest of the country at large), a median home value of $219,700 (which is the median of median home values for MSA's in this dataset), and a population of 1,271,142 (which is the median population for MSA's in this dataset), that city's likelihood of appearing in the U.S. Open Data Census is approximately .375. Holding that home value and population constant, but increasing the information sector LQ to 2 (a one-unit increase), that city has a likelihood of appearing in the U.S. Open Data Census that is approximately 0.897. In other words, it goes from having a 38 percent chance of being included in the USODC as a typical city, to an almost 90 percent chance of being included in the USODC as a city with a concentration of the information sector that is twice as high as the concentration of that sector in the country at large.

The findings from the test of H2 – that the information sector is not a significant predictor of ODC scores – is in part explained by the strong effect found in testing H1. If cities with a high concentration of the technology industry are more likely to appear in the USODC, then there may not be enough variance in information LQ among cities in the USODC to establish a strong relationship between the information sector and ODC score.

All told, the results of this study can be interpreted as strong, if preliminary, evidence that the technology sector is, in fact, a key factor in the development of open data in the U.S. This is worthy of interrogating for a couple reasons. First, if the potential of open data is as great as its proponents claim, then we should be concerned about whether or not those benefits are being evenly distributed or concentrated geographically in technology hot spots. Because tech hubs tend to be located in urban areas on the east and west coasts, the unbalanced development of open data can potentially exacerbate a deepening economic and cultural divide in the U.S. that has recently been characterized in the popular press as "'Real Americans' vs. 'Coastal Elites'" (Masciotra, 2016). Secondly, if open data becomes central to defining problems and developing solutions, as advocates claim it can and should, then we should be concerned about whose voices are being amplified or ignored in that process. Workers who make up the technology industry are disproportionately white and male compared to the rest of the country (Trop & Jones, 2015), and the data literacy skills demanded by both the technology sector and the open data movement are "unequally distributed according to gender, age, geography and race" (Gregg, 2015, p. 188). We are potentially looking at an open data movement that is dependent upon skills that are out of reach for many Americans, and closely linked with the technology sector in which those skills are abundant. In such a situation, the diversity of perspectives that inform the way open data is leveraged for social and economic benefit is limited, ultimately serving to further disenfranchise the marginalized and further empower the privileged.

Works Cited

2016 Socrata Open Data Benchmark Study. (n.d.). Retrieved from https://21bqi49zoy82acfw93qlpqm1-wpengine.netdna-ssl.com/wp-content/uploads/2016_open_data_benchmark_report.pdf

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), 1–13. Retrieved from http://ci-journal.net/index.php/ciej/article/view/845

Boyd, K. (2016, February 2). Interview with Candace Faber, Seattle's civic technology advocate. *Open Seattle*. Retrieved from http://openseattle.org/2016/02/02/candace-faber-interview

Bureau of Labor Statistics industries at a glance: Information. (n.d.).

Bureau of Labor Statistics QCEW Location Quotient Details. (n.d.).

Chignard, S. (2013). A brief history of Open Data. *ParisTech Review*, 5. Retrieved from http://www.paristechreview.com/2013/03/29/brief-history-open-data/

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, *27*, 377–383. http://doi.org/10.1016/j.giq.2010.07.001

Douglas, M. (2010). G7: CIOs From Seven Big-Cities Work Together to Develop Open-Source IT Solutions. *Government Technology*, *2*. Retrieved from http://www.govtech.com/e-government/G7-Big-City-CIOs-Work-to-Develop-Open-Source-IT-Solutions.html

Gill, A., Hollister, H., & Hughes, A. (2016). *The state of the union of open data, 2016*. Retrieved from http://www.datafoundation.org/

Goldstein, B., & Dyson, L. (Eds.). (2013). *Beyond Transparency: Open Data and the Future of Civic Innovation*. *Beyond transparency: Open data and future of civic innovation*. http://doi.org/10.1017/CBO9781107415324.004

Gregg, M. (2015). Hack for good: Speculative labour, app development and the burden of austerity [Published version]. *The Fibreculture Journal*, (25), 185–202. http://doi.org/10.15307/fcj.25.186.2015

Gurstein, M. B. (2011). (2011). Open data: Empowering the empowered or effective data use for everyone? | Gurstein | First Monday. *First Monday*, *16*(2), 1–8. http://doi.org/10.1177/0170840601223003

Janssen, K. (2012). Open government data: Right to information 2.0 or its rollback version?, (September), 20.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. http://doi.org/10.1080/10580530.2012.716740

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open Government Data: A Stage Model. *LNCS*, *6846*, 235–246. Retrieved from http://s3.amazonaws.com/academia.edu.documents/43429342/978-3-642-22878-0_20.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1489624310&Signature=KJoiWzx2K7h7UY41ujVEq5lSsag%3D&response-content-disposition=inline%3Bfilename%3DOpen_government_data_A_stage

Lathrop, D., & Ruma, L. (2010). *Open government: Collaboration, transparency, and participation in practice*. (D. Lathrop & L. Ruma, Eds.). Sebastopol, CA: O'Reilly.

Malamud, C. (2007). Open government working group meeting in Sebastopol, CA.

Masciotra, D. (2016, November 20). "Real Americans" vs. "coastal elites": What right-wing sneers at city dwellers really mean. *Slate*. Retrieved from http://www.salon.com/2016/11/20/real-americans-vs-coastal-elites-what-right-wing-sneers-at-city-dwellers-really-mean/

McDermott, P. (2010). Building open government. *Government Information Quarterly*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0740624X10000663

McNutt, J. G., Justice, J. B., Melitski, J. M., Ahn, M. J., Siddiqui, S. R., Carter, D. T., & Kline, A. D. (2016). The diffusion of civic technology and open government in the United States. *Information Polity*, *21*(2), 153–170. http://doi.org/10.3233/IP-160385

Meijer, A., & Bolivar, M. P. R. (2015). Governing the smart city: a review of the literature on smart urban governance. *International Review of Administrative Sciences*. http://doi.org/10.1177/0020852314564308

O'Connor, J. (2015). *Towards a profile of open government data users*. University of North Carolina Chapel Hill.

O'Reilly, T. (2011). Government as a platform. *Innovations*, *6*(1), 13–40. Retrieved from http://www.mitpressjournals.org/doi/pdfplus/10.1162/INOV_a_00056

Open Government Working Group. (2007). Open Government Data Principles. Retrieved from https://public.resource.org/8_principles.html

Robinson, D., Yu, H., Zeller, W., & Felten, E. W. (2008). Government Data and the Invisible Hand. *Yale Journal of Law and Technology in Fall*, *11*. Retrieved from http://ssrn.com/abstract=1138083

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*, 308–315. Retrieved from https://www.researchgate.net/profile/Peter_Johnson64/publication/278161820_Civic_open_data_at_a_crossroads_Dominant_models_and_current_challenges/links/57e03b8408aece48e9e1f1bd.pdf

Thorsby, J., Stowers, G. N. L., Wolslegel, K., & Tumbuan, E. (2015). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*. http://doi.org/10.1016/j.giq.2016.07.001

Trop, J., & Jones, S. (2015, July 30). See how the big tech companies compare on employee diversity. *Fortune*. Retrieved from http://fortune.com/2015/07/30/tech-companies-diveristy/

U.S. Open Data Census Frequently Asked Questions. (n.d.).

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, *33*(2), 325–337. http://doi.org/10.1016/j.giq.2016.02.001