

How Charlottesville can Win with Open Data

A Research Paper in STS 4600

Presented to the Faculty of the School of Engineering and Applied Sciences

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Engineering

Author

Edwin Sompayrac

March 6, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signature **Edwin Sompayrac** **Date** 3/6/2020

Approved _____ **Date** _____

Sean Ferguson, Department of Engineering and Society

Understanding Open Data Within Charlottesville

When Michael Signer was elected as the Mayor of Charlottesville in 2016, he decided to hold an open “office-hours” for the people of Charlottesville to come meet him and talk about what they would like to change about the city. (Mandell, 2017) From the office hours, there was an overwhelming amount of support for the creation of an open data portal. Charlottesville already had GIS (Geographic Information System) capability, and a lot of data was already “open”, it just wasn’t easily accessible. (Jason Ness, October 31, 2019) A few teams of engineers and city staffers came together and set up Charlottesville’s first open data portal a few weeks later.

The development of the portal was one of the larger steps taken by the city of Charlottesville to develop into a “smart city”. Wellington Webb, former mayor of Denver, a current smart city, defined a smart city as “The use of Smart Computing technologies to make the critical infrastructure components and services of a city — which include city administration, education, healthcare, public safety, real estate, transportation, and utilities — more intelligent, interconnected, and efficient.” (Washburn, 2010) Open data in Charlottesville or even around the country is not a necessarily new topic, however.

Open data and information is hardly a new concept. In 1967, the Freedom of Information Act (FOIA) was passed which provided the public the right to request access to records from any federal agency. This excludes certain information about national security, public safety, and certain personal information. However, for the most part, if the record exists in a federal agency, any member of the public can request to access the information, and can be granted access to the information.

As digital technologies converge with the physical infrastructure in smart cities, there become more facets for data collection within the city. The number of IoT (Internet of Things) devices is expected to more than double by the end of 2020, meaning there are billions of devices around the world that have the capability to record all sorts of data. Open data initiatives have been taken in different municipalities, locally and globally, in an effort to create more transparency in the relationship between government and its citizens. There is little debate about the potential benefits of open data in a city; it increases trust between governments/corporations and people, it promotes collaboration between government officials and citizens, and increases the accountability of government. (Evans, 2012) While the benefits and goals of open data are clearly defined by numerous organizations and governments, open data initiatives have often fallen flat over the last 15 years, especially in European localities. Charlottesville created its first open data portal in 2016 with the hopes of being more transparent, and providing data that would allow its citizens to use the data to potentially benefit the city. While there have been a few attempts to use Charlottesville's open data portal effectively, most of the efforts prove inconsequential. As Charlottesville works with UVA's (University of Virginia) Link Lab ("UVA's interdisciplinary cyber-physical systems research center") teams and Smart Cville to develop into a "Smart City", the open data portal is at the cornerstone of this goal. (UVA, 2019) Using the portal effectively will help Charlottesville take advantage of technological efficiencies, become more sustainable, and boost economic and social development.

In order to create a successful open data network, there are a multitude of actors that need to work together. While Charlottesville has had an open data portal for four years now, the city is missing key actors that hinder the benefits of the portal. In order to use the data most effectively, UVA's new School of Data Science should be instrumental in developing Charlottesville's open

data environment, and should be fundamental in the success of open data initiatives for Charlottesville moving forward. This paper will talk about the problems different municipalities have had when dealing with open data initiatives, why certain areas have failed with open data, where localities have succeeded with open data and why, and how to make Charlottesville win with open data.

Previous Literature

As was said earlier, the benefits to open data and open government are undeniable. Open public data can improve relations between citizens and governments by creating an attitude of trust and accountability. (Kucera, 2014) This increased transparency can also boost the reputation of the public body releasing the data. Open data also has the ability to stimulate innovation and promote economic growth by developing new products and services, improving current processes, and using the information to better inform investors and companies. One of the main benefits is the ability to tap into the heads of the public, and under the right circumstances, groups of public individuals can make decisions and read trends better than policy makers or other government employees. (Janssen, 2012)

When an open data resource is not designed effectively, it cannot be used effectively. Arguably, the most important actor in the open data network is the portal itself. Open data portals are basic web tools that include a multitude of data published in datasets. Based on the data set, some of the information is curated, some of it is raw, and some of it is a little unclear as to what data is being seen. One of the problems with these open data portals is that there are so many different types of portals that exist. There is no one standardized format to open data portals which makes it difficult when comparing the portals of different communities. Another issue in the design of these open data portals is that each data set looks the exact same from the outside.

In the case of Charlottesville and the majority of other data portals around the world, the datasets are in the style of a CSV (comma separated values) file, or a basic excel file. This gives all of the data in the portal the same agency even though there are more useful files. The dataset with 20,000 rows and data that could be used effectively in a variety of ways looks the same as the 3-row dataset that just has the addresses of the landfills which could be found in a simple Google search. Because of the way the portals are set up, these two data sets are given the same amount of respect and influence. Cesar Hidalgo, professor at MIT, described current open data sites as a supermarket where every item was in the exact same box, whether it be fruit, canned beans, ice cream or shampoo. This makes navigating portals very difficult, and nearly impossible to find the data sets that contain the most useful information.

On January 21, 2009, President Barack Obama signed the “Memorandum on Transparency and Open Government” which asked all federal government agencies to make their data open to the public to establish trust and transparency within the government. Over 272,000 data sets were provided (Lee, Almirall, Wareham, 2020) and only a handful of citizens had looked any of them over. Another key actor in the open data framework are people that want to use the data. Too often, governments or corporations made their data open to the public believing it would foster open data use and that would be the end of it. Wendy Thomas, national lead for Metadata and data flow mapping, called this “the external wait for the non-existing user”. This is the “Field of Dreams effect” and in terms of open data, if you build it, that doesn’t necessarily mean they will come. The Field of Dreams effect is one of the more common scenarios in open data. The main idea is that open data has no value in and of itself; the value only comes when the data is used.

There are even more barriers to open data than just ineffective portals and a lack of people manipulating the data. There is an unclear tradeoff between government transparency and public privacy. In terms of data, many governments are unaware of what crosses the line for its citizens. Another barrier includes the time it takes to analyze data and the lack of incentives for the users. Many portals require registration before users can download data and others require users to pay to access the data. The data itself doesn't always show clear trends and many users are unaware of what to look for within the data. Even if a user has the technical skills to analyze a complex data set, it will take a lot of time to achieve and even more time to implement some sort of change or innovation that was discovered from the data. (Janssen, 2012) Open data is relatively new and has never been a standard for any organization. The benefits of a successful open data network are known but have rarely been achieved. Far too often have the barriers hindered the optimistic approach of establishing an open data system.

Interview, Research, and Data Collected

The State of Charlottesville and Open Data

In an interview with Jason Ness, a business development manager for the city of Charlottesville and an instrumental member of the creation of the open data portal, he explained some of the goals of the portal, and why it was created in the first place. When asked about the data that is available to the public, Ness stated that none of the data in the portal is very “personal” at all, and that every piece of information available on the portal was already FOIA-able. One example of this is in the public safety data, under a crime dataset, a member of the public is able to see the first and last name of a criminal, the crime they were charged, when, and where the crime took place. This may seem like a lot of personal information, but all of this

information is available to public through the Charlottesville court system. The open data portal is merely just one place where all of the information can be easily found. When asked what the goal of the open portal for Charlottesville was, Ness responded by saying data scientists and members of the community can use the data to make Charlottesville a better place through data analysis. He gave two specific examples of when this was the case:

The first is an example with the free Wi-Fi on the Downtown Mall. There are multiple routers on the mall and when a new device gets connected to it, the router can see the name of the device and what time it was connected. As people walk around the mall, different routers pick up the same devices at different times, being able to “track” where and when people are moving. Data scientists are using this data to see how weather patterns on certain days affect traffic on the downtown mall as well as how events affect mall traffic. This can develop patterns of human traffic to help the businesses on the mall with a proper amount of staffing or to help people find open places to park.

The second example was that environmentalists are currently using the utility data to see which neighborhoods are using the most water and gas utilities and which neighborhoods weren't. By analyzing the data over the course of a year, there is information being uncovered about more efficient ways to wire the utility infrastructure of the city, so that neighborhoods are able to get what they need more effectively. This could end up saving the city and the citizens thousands of dollars a year in wasted utilities.

While Ness gave these examples of potential initiatives undertaken by the Charlottesville community when it comes to open data, neither have been fully investigated and conclusions to both examples are incomplete. There simply are actors missing that is preventing Charlottesville from reaching its full open data potential.

New York City's Success

There are places, however, where open data is actually working, despite the hinderance. In 2012, New York City launched its “Open Data for All” initiative what was designed to give the collective public access to government data. This included data from city agencies as well as city apps that use APIs. Some of the main benefits from this initiative were that crime data was released and citizens and visitors could view an interactive map of the city that showed where crime was located. There were even filters on the map so you could search by time, location, and a few other categories. Another thing this allowed New Yorkers to see was where their taxpayer dollars were being allocated and utilized.

By 2016, more than 5 million hits were recorded on the portal, and New York continues to roll out new and improved data sets, included a couple hundred automated data sets. Citizens also benefit from the data portal during elections, as New York is continuing ways to increase the communication between political officials and citizens using their open data resources. In a city as large as New York, this previous communication has often been a challenge.

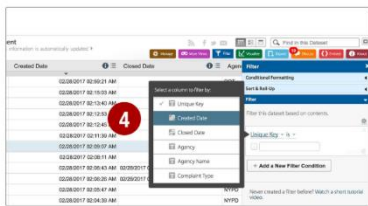


Fig. 1.0: A screenshot from New York's open data portal that shows step by step how to navigate through the portal. (City of New York, n.d.)

One aspect that makes the New York portal so successful is the amount of resources that the city has provided the citizens. Going onto the portal for the first time, a user will immediately see a link to join the open data mailing list to receive updates about the city's data and new open data policies, training events, and upcoming events. Upon scrolling down the website, there are large, self-describing links that make the site easy to navigate. For instance, the first link explains what open data is exactly and how to use the open data portal. The site actually

goes through and teaches users how to navigate through the portal to filter, visualize, and manage all of the data. It then walks you through an example data set, giving step by step instructions on how to make a map of 311 service requests. The website also gives tips and tricks for navigating specific and large data sets, and even has instructional videos that anyone with an internet connection can watch and learn from. New York also has an open data team that is willing to answer and questions, comments or concerns left for them.

New York City also has two universities that developed data science programs: Columbia and NYU (in the Fall of 2014 and 2013 respectively). These data science programs are training students and future leaders on the computer science skills necessary to navigate and make use of large data sets given, while also learning the technical skills to read data, find trends, and then take advantage of the resources available to maximize the benefits from open data.

New York's portal is successful because not only does it present data in an open way, but the portal is self-explanatory, able to be filtered and visualized, and there is a training program *within* the portal itself that allow novices to delve into the world of open data. Not only that, but because of the academic initiatives taken by universities in the city, there are people who *want* to get involved with data and use it to the cities advantage. New York has also continued to pour resources into the portal and *continually update* the portal with new data sets and updating old ones. All of the actors are in their places.

United Kingdom's Setbacks

In 2006, The Guardian newspaper's Technology section began a campaign titled "Free Our Data", calling for the release of data by government authorities to be free and available to the public. (The Guardian, 2006) In 2010, the UK removed its barrier to public data and unveiled

its new Open Government License which allows for public reuse of government data. Tim Berners-Lee, inventor of the World Wide Web, said about the license “It’s great to see a simple and straightforward license for people to re-use government data in any way they want. It will enable inventive people to build innovative new applications and websites which help people in their everyday lives.” (Cabinet Office, 2010) Since 2010, The UK has been consistently named a leader in open data initiatives by the European Data Portal.

Some of the successes from the UK’s open data initiatives are DEFRA (Department for Environment, Food, and Rural Affairs) and its subordinates have built data analytic services that provide the public with analysis of its own data sets including sewer water and air quality, releasing over 8,000 datasets a year. Also, in 2011, Citymapper was created in London. Citymapper is a public transit/map app used in a lot of major cities. In London, the app was used to predict the economic value of areas within London by tracking the public transport being used within the different districts. All of the data used to create the app was reused from UK open data portals (Rizzi, 2015).

While the UK has been a leader in open data for well over a decade, in 2018 the Open Data Barometer declared the UK has stalled in its open data initiatives. (World Wide Web Foundation, 2018). Although it is still a leader worldwide in open data infrastructure and use, the Open Data Barometer report says that to continue improving, the UK needs to focus on open data governance instead of just isolated projects. The report also suggests literal engagement with citizens in order to work with data sets the people most need and want. Though there are many successes, open data is still not a standard with only 20% of the datasets being truly open and having an impact score of only 50%. The UK is worldwide open data leader, but still the desired impact is lacking. (World Wide Web Foundation, 2018)

STS Framework and Methodology

In all of these examples, there are actors lacking within the open data network. The most common problem was a deficiency in public interaction with open data. Governments and corporations have their data open to create transparency and opportunities to act within social, political, economic, and climate realms by allowing the public to openly and freely interact with, download, and analyze data sets. Once the data is published, someone needs to do something with it in order for some action or change to occur, and often times that is lacking. New York has been successful in its initiatives because of the resources they have given the public to tackle open data. New York is actively collaborating with its citizens to make open data initiatives possible by working data in with university programs, utilizing a whole government sector toward open data, and training available to anyone with an internet connection. Charlottesville and the UK have struggled to find the people that not only care enough to analyze datasets, but also possess the skills needed to make use of large amounts of open data. New York also has led the way in amount of data published, while Charlottesville and the United Kingdom have settled for only a fraction of updated data. Open data only becomes useful when all of the necessary actors are able to perform their roles to create and sustain the network.

What Can Charlottesville Improve?

While UVA is developing their Data Science curriculum, Charlottesville at large is relatively unchanged by the portal. The portal has been out for four years now, and still no one really knows about it. The data is presented in bland Excel files with little to no information regarding the sets. There is no way to filter or visualize the data unless the user is proficient in Excel and the data isn't continually updated. The latest economic data on the Charlottesville portal is from April 2018, nearly two years ago. There is no team of data agents, there is no new

relevant data, and there is hardly anyone proficient enough in the Charlottesville area who cares enough to make sense of the data. The data itself is useless.

For Charlottesville to be successful with open data, UVA needs to be working even closer with the Charlottesville government than ever before. Students and professors within the Data Science School need to be addressing the deficiencies within the open data portal as it currently is, and do projects and research to not only provide more data and better updated data, but also mold Charlottesville into a good and great city. The curriculum will not only provide Charlottesville with individuals that are technically skilled to analyze and work with data, but will provide enthusiastic young adults that will want to make a difference within their community. UVA has increased its role within the Charlottesville community by positively impacting education, the arts, economic stability, and social justice. With UVA's substantial resources, endowment, and academic reach worldwide, it would undoubtedly stimulate the open data initiatives and success within Charlottesville boosting citizen trust and working to increase economic, political, and social prosperity.

Conclusion

Open data success is the most direct and promising route to becoming a smart city for Charlottesville. While the creation of the open data portal was a good start, an absence of awareness and citizen initiative and knowledge ultimately let the portal down. The portal is also lacking in its design, ease of navigation, and age of some of the data sets. For Charlottesville to continue innovating, the new data science program at UVA should expect to be the backbone of Charlottesville's open data, working directly with the city to optimize the portal and achieve open data goals. The benefits of open data cannot be quantified but are boundless if Charlottesville is encouraged to win with open data.

Bibliography

Cabinet Office. (2010, September 30). Government removes barrier to public data use. Retrieved March 2, 2020, from <https://www.gov.uk/government/news/government-removes-barrier-to-public-data-use>

Carolan, L. (2016). Open data, transparency and accountability: Topic guide. Birmingham, UK: GSDRC, University of Birmingham. Retrieved from https://assets.publishing.service.gov.uk/media/5857fdb40f0b60e4a0000d6/OpenDataTA_GSDRC.pdf

: City of New York, N. Y. C. O. D. (n.d.). NYC Open Data. Retrieved February 19, 2020, from <https://opendata.cityofnewyork.us/>

Evans, A. M., & Campos, A. (2012, June 13). Open Government Initiatives: Challenges of Citizen Participation. Retrieved February 5, 2020, from <https://onlinelibrary.wiley.com/doi/full/10.1002/pam.21651>

(FOIA), F. of I. A. (n.d.). FOIA.gov (Freedom of Information Act) Frequently Asked Questions (FAQ). Retrieved from <https://www.foia.gov/faq.html>.

Hidalgo, C. A. (2016, May 2). What's Wrong with Open-Data Sites--and How We Can Fix Them. Retrieved February 11, 2020, from <https://blogs.scientificamerican.com/guest-blog/what-s-wrong-with-open-data-sites-and-how-we-can-fix-them/>

IoT Has Quietly and Quickly Changed Our Lives. (2019, February 1). Retrieved from <https://www.ncta.com/whats-new/iot-has-quietly-and-quickly-changed-our-lives>.

- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012, October 5). Benefits, Adoption Barriers and Myths of Open Data and Open Government. Retrieved February 2, 2020, from <https://www.tandfonline.com/doi/abs/10.1080/10580530.2012.716740>
- Joo, J. Y.-M., & Lee Kuan Yew School of Public Policy. (n.d.). Smart Cities: A New Age of Digital Insecurity. Retrieved October 10, 2019, from <https://iiss.tandfonline.com/doi/full/10.1080/00396338.2018.1448577#.XZ9gXEZKg2w>.
- Kucera, J., & Chlapek, D. (2014, November 1). Benefits and Risks of Open Government Data. Retrieved February 20, 2020, from <http://www.si-journal.org/index.php/JSI>
- Lee, M., Almirall, E., & Wareham, J. (2020). Digital Library: Communications of the ACM. Retrieved February 13, 2020, from <https://dl.acm.org/doi/fullHtml/10.1145/2756542>
- Mandell, J. R., Woods, C. R., Robinson, E. undefined, & Hays, E. undefined. (2017, August 27). City's open data portal goes online with hopes of creative use by public • Charlottesville Tomorrow. Retrieved from <https://www.cvilletomorrow.org/articles/charlottesville-launches-website-for-open-data>.
- Millennials Throw Caution to the Wind Regarding Internet Security. (n.d.). Retrieved from <https://securethoughts.com/younger-generations-throw-caution-wind-internet-security/>.
- Rizzi, D. (2015). Open Data Maturity in Europe 2015. Insights into the European State of Play, 7–41. Retrieved from <https://www.europeandataportal.eu/>
- Team. (2017, March 2). Three Places Where Open Government Data is Making a Positive Impact. Retrieved February 24, 2020, from <https://www.opendatasoft.com/blog/three-ways-open-government-data-is-making-a-positive-impact>

The Guardian. (n.d.). Free our data | Technology. Retrieved March 4, 2020, from <https://www.theguardian.com/technology/free-our-data>

Thomas, W., & Environment Agency. (2016, December 19). How to fail at Open Data.... and other learnings from the Open Government Partnership Summit 2016. Retrieved February 10, 2020, from <https://defradigital.blog.gov.uk/2016/12/19/how-to-fail-at-open-data-and-other-learnings-from-the-open-government-partnership-summit-2016/>

UVA's Link Lab Teams with Smart Cville to Help Launch Network to Inspire Smart-city Innovation. (2019, April 9). Retrieved March 2, 2020, from <https://engineering.virginia.edu/news/2019/04/uva's-link-lab-teams-smart-cville-help-launch-network-inspire-smart-city-innovation>

Vaughan, F., & University of Birmingham. (2017, October 24). Open Data, Transparency and Accountability: Topic guide. Retrieved October 10, 2019, from <https://opendatacharter.net/open-data-transparency-accountability-topic-guide/>.

Washburn, D., & Sindhu, U. (2010). Helping CIOs Understand “Smart City” Initiatives. Forrester, 2–3. Retrieved from <https://go.forrester.com/>

World Wide Web Foundation. (2018, September 20). A decade into open data, leading governments struggle to make it a reality. Retrieved March 2, 2020, from <https://webfoundation.org/2018/09/a-decade-into-open-data-leading-governments-struggle-to-make-it-a-reality/>