Civic Hackers' User Experiences and Expectations of Seattle's Open Municipal Data Program

Meg Young
Information School
University of Washington
megyoung@uw.edu

An Yan
Information School
University of Washington
yanan15@uw.edu

Abstract

This study examines the challenges and the expectations that civic hackers bring to the use of open government data, building on Gurstein's theory of barriers to effective use. Civic hackers are hobbyists, freelancers, and professionals who use open government data for products and social good applications. Drawing on individual interviews and a focus group with fifteen total civic hackers in Seattle, Washington, we synthesize findings on their experiences using open government data, including their expectations for the kinds of data formats, metadata, API functionality, and datasets that should be provided on the city's open data portal. Respondents report challenges using the data, including low data availability, outdated datasets, limited API functions, proprietary formats, lack of metadata, and untidy datasets. These acted as barriers to their effective use of open data. Respondents expect higher quality data and more usable data portal functionality, in part because of their professional experience in the technology sector. In our discussion, we examine the organizational structure of the open data program, and the constraints it poses for the achievement of respondent expectations. Our analysis points to a demand for an additional, third party civic institution (like a local newspaper) to host cleaned data for wider use.

Keywords—open government data; civic hackers; data use; barriers; effective use, usability

1. Introduction

Open government data is available to the public freely without restrictions [1]. Much work on open data to date highlights its transformative potential [2]. It alters access to publicly produced data, enabling new analysis, creating new forms of transparency and accountability, encouraging social participation, stimulating innovation, and generating public value [3]. As a result, open government data has received considerable attention worldwide. The European Union Public issued by President Barack Obama in 2009. Also Sector Information directive was released in 2003, followed by a Memorandum on Transparency and Open Government in 2009, a non-profit organization was formed called Code for America, "the technology world's equivalent of the Peace Corps" [4]. Volunteers and Fellows at Code for America "help the government work more like the Internet," for example, by promoting and building usable, intuitive interfaces on public-facing applications [5]. One of its programs, Code for America Brigades, facilitates local volunteers to meet regularly as a group. In this study, we examine the user experiences of members of Seattle's local Code for America Brigade, once called Code for Seattle, now known as Open Seattle.

In 2009, The City of Seattle started one of the earliest open data programs in the United States; it was an early adopter of Socrata's open data platform[1]. In 2015, support for the program within the city intensified. The city promoted an Open Data Manager and created a new position, called the Civic Technology Advocate, responsible for connecting the city to its open data users. In February 2016, Mayor Ed Murray signed an Open Data Policy, requiring that Seattle's data becomes 'open by preference.' The policy represents an executive order for the city to release more and higher quality open data, and establishes procedures to ensure data quality and data privacy. The policy also creates 'Open Data Champion' positions within each of Seattle's 33 departments, whose role will be to help publish their department's data assets. At the time of this writing

---

[1] Socrata is a private company that provides open data portal services for governments. It is a leading provider of open government data portal products and services to federal, state, and municipal entities in the US, and is expanding its customer base overseas.

HICSS

there were 2,704 'objects' published on data.Seattle.gov[2], drawn from 797 datasets and maps.

were posted on its open data portal, data.seattle.gov, which is hosted and managed by Socrata. The City of Seattle pays Socrata an annual fee to run this service, with optional add-ons for data visualization services like the Open Budget application.

In the midst of the momentum behind opening Seattle's data, excitement in the local civic hacker community has grown. Our respondents self-identify as civic hackers, which we define as data enthusiasts (hobbyists, freelancers, and professionals) who use open government data for independent projects (most often social good oriented, but also including freelance projects and commercial products). This community shares the hacker sensibility of cleverly circumventing obstacles. In Seattle, many civic hackers attend a local group, Open Seattle (http://openseattle.org). In the last year, Open Seattle has grown from fewer than 20 regular attendees to as many as 40 regular attendees. It has a Slack[3] group with 125 registered users and a Meetup[4] group of over 2,000 subscribers. This increasing participation reflects a growing excitement for the potential to channel Seattle's local tech talent into projects for the public good. Meetings provide an opportunity for volunteers to share knowledge and collaborate with each other on proposed projects.

In 2014-2015, the group met weekly, but in 2016 it began to meet once a month, with more formally organized programming. A typical meeting starts with introductions, moves to short lectures from professionals working on relevant projects, and ends with unstructured collaboration time. In addition to monthly meetings, Open Seattle civic hackers are likely to see each other at one of several relevant local events, held periodically throughout the year. For example, a recent 'hackathon[5]' was held to promote technological innovation for fisheries-related challenges; it was called the 'Fishackathon.'

Seattle is a particularly interesting case study for research on open data quality and use. The longevity of

its open data program and recent momentum is in part due to its role as the home to a wealth of tech expertise. Its civic hacker community draws from a pool of tech talent, including independent developers, or those who work with Amazon, Microsoft, and Tableau. Other technology companies with local offices include Facebook, Hulu, Zillow, and Google. Seattle is home to the headquarters of Socrata, which is (as mentioned earlier) the leading provider of open data hosting services and API functionality to federal, state, and municipal governments. Open Seattle monthly meetings are often hosted in Socrata's local headquarters; meetings are also hosted in a social enterprise co-working space downtown.

While increasing numbers of datasets are being opened, facilitating access to open data is a separate question from supporting its meaningful use. Gurstein advances the term 'effective use' to delineate between the opportunity to use a technology, and the realization of its transformative potential [6], [7]. He defines it as the capacity to successfully use ICTs (information and communication technologies) to accomplish a particular goal. Originally advanced to critique the notion of the 'digital divide,' the gap between high- and low-resource people in their access to ICTs [6], Gurstein also found that access to open data was also conflated with its use [7]. The theory of effective use proposes multiple factors that determine or impede use of a technology (or data) aside from users' access to it: (1) internet, (2) technology access, (3) user technology skills, (4) usable formats, (5) user knowledge and sense making, (6) community resources, and (7) formal governance structures that support the intended use. The absence of these important factors can create barriers for those hoping to make effective use of open data.

Here, we explore the notion of effective use from a user perspective, and evaluate which of these factors are most salient to the open data user experience. While Gurstein's theory provided the basis for this study; the research questions are framed to be open and exploratory. In the tradition of user experience research, we also focus on 'pain points'-- aspects of the open data program that can be improved [8], [9]. First we ask, what expectations do members of Open Seattle civic hacking group bring to the use of open government data? Second, what challenges do members of the Open Seattle civic hacking group face in using Seattle's open government data? In our findings, we surface the ways in which expectations and challenges are entangled. The discussion interprets our results given users' professional identities as engineers. It foregrounds the relationship between the City of Seattle and its partner Socrata, which provides data hosting and API services. That analysis points to a

---

[2] The bulk of these are filtered views derived from the base datasets and maps. The City government's official count of datasets and maps available is 797.

[3] Slack is a messaging application for teams, most often used in workplace settings. See http://slack.com. Open Seattle uses Slack as a communication tool, for example, to share tips and advice.

[4] Meetup is a platform for organizing meetings for local community groups. See http://meetup.com. Open Seattle uses Meetup to coordinate event announcements.

[5] A hackathon is an event, lasting between several hours to multiple days that brings together programmers (or other engineers) to innovate approaches towards a common goal. While it has roots in free and open software and corporate causes, it has increasingly been adapted for the civic hacking space.

local interest in a third party repository for cleaned data, potentially to be housed in a civic institution like a local newspaper.

It is important to note that civic hackers are not the only users of open government data programs. The portal is used by researchers, non-profits, and policy advocacy groups. Local governments themselves may use open data portals as a means of frictionless access to other levels and departments. Small and large enterprises also commercialize data they access as part of their core business models, for example, Zillow constructs detailed estimates of the real estate market, which it commericalizes. Helbig et al. [10] call this wider stakeholder community an 'information polity.' As such, civic hackers' expectations and needs are not the only ones that governments respond to in their efforts to open data. That said, it is important to understand civic hackers as a usergroup. They are able to use and manipulate open goverment data in more advanced ways than most other local residents. Because they engage with open data over the long term, as a group, they develop expertise that can be leveraged by newcomers. In Seattle, civic hackers are a strong and visible constituency in the open data program.

2. Previous work

Previous work notes the quality of open data as a primary challenge to data use. Janssen et al. [11] explore barriers to use, finding multiple challenges related to information quality, including problems with accuracy, completeness, and clarity. Erete et al [12] note that data collection, cleaning, management, interpretation, and dissemination of open government data is time and resource intensive for non-profit data users. Martin et al. [13] find that there is no single metadata standard for open data; data users have to sort through multiple vocabularies. Metadata is also incomplete in meaningful ways, for example, datasets may lack documentation on their provenance and initial use case [13]. Gonzales-Zapata & Heeks [14] note the relative lack of open data research that addresses the topic from a technical perspective.

A second major category of challenges for the use of open government data is contingent on users' skills. The ability to use government data is limited to a small subset of the population [1], [6]. Although the cost of acquiring information has gone down, the cost of interpreting and using government data is still high [15].

Other previously identified concerns relate to availability and access. Data are often provided in proprietary formats, making them cost-prohibitive for some to use [11], [13]. The proportion of datasets available by API is generally low, making datasets difficult to use in apps and access efficiently [13]. Data users creating apps and services are concerned about whether open data will be supported over time [13]. Janssen et al. [11] report that users sense a lack of responsiveness by the government to their needs and input. While most work on open government data programs attends narrowly to the data portal, we draw on a definition of open data that includes responses to public records requests as a common means by which data becomes open [16].

Much existing work on open government data use asserts that more data will lead to more democratic engagement. Townsend asserts that civic hackers are part of a seismic shift in city governance—a 'new civics' that create change "one site, one app, one click at a time" [2]. However, little evidence reflects that the promise of open data is being fulfilled [3], [7], [15], [17] Much empirical work on open data focuses on challenges to opening *more* data, but there is insufficient focus on the additional factors that determine the efficacy of an open data program [15] or outline what criteria for success entail. The need for continuing analysis in this arena is intensified by the rapidly shifting technological capability of cities opening data, as well as the unevenness between cities (and even between departments within one city) in these efforts.

Much existing work focuses on open data programs at state, national and international scales. Bertot et al. offer insights into issues small communities face when building data infrastructures, but do not include civic hackers in the group of stakeholders consulted [18]. Civic hackers in particular merit further analysis, due to the fact that they are organized, meet regularly, and develop group expertise and shared resources to circumvent challenges in using open data. For example, our respondents drew on datasets on the portal, as well as those found online or requested directly from government. By looking at civic hackers' user experiences, we see that open government data use is not limited to the assets provided on the portal, but also includes those found online or requested directly from government officials. This broader empirical framing of what constitutes open data makes some aspects of the user experience more visible.

We present the results of an inquiry into civic hackers' expectations for the City of Seattle open data program and their challenges using it. Our work confirms many previous findings on barriers to open data use, albeit in a more mature open data program and portal than those featured in most previous work. Our discussion reflects on the ways that civic hackers' professional identities inform their expectations for the

platform, and the partnership between the City and its partner in providing Seattle's open data program.

3. Methods

Data collection for this study is based on a focus group of 8 people, 7 semi-structured interviews with members and organizers of Open Seattle, and 1 interview with Seattle's Open Data Manager. Focus groups allow for participants to interact and react to each others' input [19]. As a method, focus groups present a risk of respondent bias and group-think; our research design took measures to minimize these risks by piloting the format, soliciting perspectives aside from those already discussed, and asking participants to write down their opinions on paper prior to discussion [20]. The focus group with Open Seattle had 8 members and lasted 90 minutes, and was led by one moderator who asked participants about their challenges for the open data program. We also conducted 7 individual interviews. All participants are members of Open Seattle civic hacker group. At the time this was written, all respondents were male.

Our research questions are:
- *RQ1*: What expectations do members of Open Seattle civic hacking group bring to the use of open government data?
- *RQ2*: What challenges do members of the Open Seattle civic hacking group face in using Seattle's open government data?

We operationalize our definition of expectations to mean only those expectations that apply directly to the execution of the open data program, as opposed to, for example, expectations of open data's transformative potential in society.

Data analysis took the form of iterative inductive qualitative coding based on grounded theory approaches [21], [22]. Each of the two authors coded the contents of the focus groups and interviews separately, and iteratively distilled the codes into a coding manual (See Table 1) which was finalized and used to re-code the data.

In the discussion, these findings are brought to bear on their larger context in the city of Seattle. Background and contextual information used in the discussion is derived from 12 non-consecutive months of ethnographic fieldwork conducted by one of the authors with City of Seattle employees, the Open Seattle civic hacker group, and relevant stakeholder groups. In the course of the fieldwork, the author attended trainings and meetings, conducted additional informal interviews with local open data figures, and examined textual artifacts, like state and local law and policy, including the Open Data Policy and Open Data Champion Playbook. This more study was initially conceived to follow up on a respondent's observation in the ethnographic fieldwork, who reported feeling a mismatch between civic hackers' expectations for data access and the kinds of outcomes the open data program could feasibly acheive.

Table 1. Coding manual emergent from iterative data analysis

Parent nodes		Child nodes
	Data portal usability	Data availability	Lack of important datasets
			Reluctance to release
			Response delay
			Discontinued feeds
		Data updates	Insufficient update notifications
		API functionality	Limited API function
			No analysis function on the server side
Challenges		Format	Data released in proprietary format
			Inconsistent formats
	Dataset quality	Metadata	No data provenance
			Lack of metadata
		Data updates	Outdated datasets
			Data not automatically updated
		Data tidiness	Mistakes in datasets
			Bad data practices
			No naming conventions
			Data not cleaned prior to release
	Data portal usability	Data availability	Data should be available at citizen request
		Data updates	Data updates notifications should be detailed and accessible
		API functionality	API should be able to do joins in the cloud
			Analysis should be able to occur on the server side
Expectations		Format	Data should be released in open formats
	Dataset quality	Metadata	Detailed metadata should be provided
			Information on the provenance of the dataset should be provided
	Citizen Engagement	Dataset request	Citizens should have input on avail. datasets
	Business models	Division of tasks	Some roles are better served outside the city

4. Findings

Based on 7 interviews and the focus group of 8 members with civic hackers, we found that challenges and expectations cohered around two areas: 1) data portal usability and 2) dataset quality.

4.1. Data portal usability

The data portal is the online website, data.Seattle.gov, where open datasets are hosted and available to the public.

4.1.1. Data availability
Challenges accessing important datasets.

Participants state that the data they are interested in using is often not available on the city's open data portal. Sometimes, commonly needed data is not hosted there, such as a data asset containing the boundaries of local zip codes. The city does not always release datasets when asked (informally, as opposed to a public records request), nor do they often respond in a timely fashion. Respondents report that it was easier and faster to get the datasets that they needed in the context of a hackathon: "A good example when it went right, was when I was at the Parks Hackathon and I was saying, 'Look, I desperately need such and such data.' They went and got it out of their database and put it on the portal" (interview 2).

Expectations for input on which data is released.
Civic hackers report that they want more input into which datasets become available, and expect that datasets be released in a timely fashion at their request. Under Washington State freedom of information law the Public Records Act (RCW 42.56), members of the public may request most public records for disclosure. However, respondents were more likely to email informal requests to city employees than to formally issue a public disclosure request. One respondent spent months searching for a particular dataset on homelessness, by contacting various municipal and county departments. Another person was interested in detailed 911 response data. As we note in the discussion, the legal context of Washington law creates a strong expectation that datasets become available upon request.

4.1.2. Keeping data updated
Challenges with outdated data.
Civic hackers note that there are out-of-date datasets hosted on the portal. For example, there is a dataset of the locations of public restrooms downtown, however, each of the restrooms listed has since been closed. The dataset has not been updated or taken down. Civic hackers also report that in some datasets, errors in a dataset persist over time. Even when a dataset is well-maintained, civic hackers report that it is difficult to know when a dataset is updated. This is challenging for those who want a macro perspective of which datasets are changing.

Expectations for data update notifications.
Civic hackers express that the portal should provide an easy way (e.g. an RSS feed) to have detailed information about dataset updates. An RSS feed currently reports which datasets have been updated by name, but it does not provide important information about the magnitude of the update, such as the number of rows changed. Interviewees note that they would

benefit from a push update notification—in which the user could subscribe for a particular dataset of interest to their application.

4.1.3. API functionality
Challenges using the current API.
While many respondents report positive experiences using the current API (Application Programming Interface) developed by Socrata; some users also note areas for improvement. The current APIs cannot support SQL data operations like join, which is a function that allows the user to unite multiple datasets on common attributes. At this time, the API only supports operations on a single dataset, like aggregation and filtering. The Socrata API does not provide a lot of support for mapping applications. All respondents note a desire for more API functionality than is currently in place, but at least one person said that the new API is a marked improvement over the previous iteration.

Expectations for expanded API capability.
Civic hackers express a strong need for a more powerful API that fulfills fast data retrieval and allows for a variety of database operations such as join. Beyond the API, respondents note a desire for more server-side data analysis functions. One interviewee says that it would be faster to run joins 'in the cloud' instead of processing multiple datasets on the client side, i.e., on the slower processor of the user's own computer.

4.2. Dataset Quality

In the midst of Seattle's open data expansion, respondents note issues related to dataset quality. Note that in June 2016, the City began a new stage of its program by training Open Data Champions in each department to capture rich metadata at the time that a dataset is posted to the portal. According to the Open Data Manager, this measure will address dataset quality issues going forward. These changes will not be reflected in our results, due to the newness of the changes and the preponderance of datasets released prior to the policy change.

4.2.1. Data formats
Challenges using proprietary and non-machine readable formats.
All data downloadable from the city's data portal is machine readable. However, civic hackers sometimes needed and used data from other sources, released in non-machine readable formats that had to be hand-scraped. For example, geospatial data were scraped from PNGs, JPEGs, and PDFs. PDFs are widely

considered an unacceptable format for open data by advocates. This issue is only salient for data that is requested directly from the city when it is not available from the data portal.

Proprietary data formats require civic hackers to convert data prior to use[6]. In the current release of the portal, some datasets are provided in multiple formats, and others are not. For geospatial datasets that have recently been added to Seattle's open data portal, data is already downloadable in multiple data formats, including CSV and GeoJSON. However, this feature requires that city data owners opt-in on datasets that have already been posted, meaning that it is not yet implemented across the board.

For datasets that have not been opted-in, geospatial datasets on the portal are released in proprietary formats like Esri shapefiles, which are natively read in a program called ArcGIS, an expensive piece of software. Shapefiles can be converted into open source formats, but these format conversions require tools that users must search for themselves and download from the open source community. "Sometimes it's hard to get that information in the right format. I've got to learn the use of good conversion tools to change around these different formats but that's-- still now I'll even run across some weird strange format. What was the one I ran across the other day? I forget. It was like a Garmin data base, all these weird data base formats so you've got to find some tool because you definitely don't have time to write your own to convert between" (interview 5). Some formats are better supported in terms of community tools and documentation than others.

Thus, while proprietary data formats do not foreclose the possibility of converting data into open source formats, it is difficult for novice users to accomplish. "I've talked to people that have ran into this issue at hack-a-thons. They're new, they're not familiar with all these mapping technologies, they're still getting their feet wet, maybe they're programmers but they don't really work with open data or civic technology so they run into the issue and they're like "Oh I've got this Esri shape file or this other Esri polygon feature and I don't know what to do with it." (interview 5). Some tools, like ogr2ogr, have a steep learning curve and must be run from the command line—this too limits their accessibility to novices.

Expectations for standardized data formats.

---

[6] We note that challenges related to data format conversion are being addressed on a dataset by dataset basis, because Socrata offers cities the option to make a dataset downloadable in Shapefile, GeoJSON, and XML. Older datasets on Seattle's data portal do not offer these options.

Participants express a strong desire for machine readable formats, and relates instances where data was made available in formats that were pre-aggregated. This issue was most prevalent when asking for datasets directly from the city. "You have to be really explicit about what you're even trying to get in the first place. Getting back a .pdf for example is really not valuable. Even further than that, getting back percentages or any kind of data that's been simplified, if that makes sense. You need the raw data set. Otherwise you can't really ask it new questions" (interview 6).

Among machine readable formats, civic hackers were more ambivalent with respect to standard formats. One respondent said, "I don't think it's really a problem to… have different formats. ...If [software companies] do have a new format they should ensure that they give support to the community to use and convert that format." (interview 5). However, the same respondent said that he has used three different format conversion command-line tools in his latest project.

4.2.2. Metadata
Challenges with a lack of metadata.

All participants report the issue of a lack of metadata associated with datasets. Sometimes civic hackers find it difficult to understand the attributes and the contexts in which a dataset was generated, and reach out to members of the city to try to re-contextualize a particular dataset-- a process that can take months.

"I'm going to get a set of data records, but what I don't get is the date that this data record was created. I don't necessarily get the metadata that tells me who collected this data. Maybe there was a reason that it was collected. What was it collected from? What were even the tools and mechanisms in which the data was collected? What are the limits on that? That's not usually attached to the dataset." (interview 6).

One respondent believes that metadata will improve as the city's open data program matures. "They work so hard to get the data that they don't necessarily have time to get the metadata right too. It's still down to the maturity of the program to fix that, and I think Seattle is getting there, it just takes time" (interview 7). As of June 2016, the City has implemented a new policy to attend to poor metadata quality going forward.

Expectations for increased documentation.

Civic hackers expect the city to provide rich and well-documented metadata on each dataset so that they can understand its context. "The difficulty going forward—what would be nice—is metadata. That's a catch all phrase for where did it come from, why did it appear, where was it made, who had their hands on it, whence was it derived from. In other words, it isn't just

a file with some XML, it's a historical vignette of the context around that dataset that made it what it is. You cannot understand the dataset without that" (focus group).

Metadata provides essential information on data provenance, that is, how each dataset was generated, what changes were made to it, and what was the underlying database structure behind its present representation. Rich metadata contributes to to the understandability, trustworthiness and usability of data. When they are not able to get access to this information, both the usability and trustworthiness of the dataset are undermined.

4.2.3. Tidiness of data
Challenges with cleaning data.

Participants note that data on the portal is sometimes not 'clean;' they refer to missing values, duplications, unreadable characters and inconsistent column naming occur in some of the data on the portal. One respondent asserts that city employees need stronger data and technology literacy by employees within the city.

Expectations for tidy data.

Civic hackers expect the city will devote more efforts to produce tidy datasets, such as recruiting technical consultants and spending more time. They also hope the city will incorporate their advice and contributions into improving open data quality. Often, civic hackers drastically improve the quality of a dataset, but there is no mechanism for them to contribute the improved data back to the city.

5. Discussion

Our research affirmed findings from related work on challenges that open data users face. We introduce other considerations pertaining to data updates, API functionality, and data availability. In our discussion, we contextualize these findings within larger structural factors in the open data program, for example, discussing whether a particular aspect of the program was provided by the city or by Socrata. In this discussion we begin by highlighting the structure of the provision of the open data program. We then situate the expectations of civic hackers as data professionals and members of the open source community. We then discuss the constraints on meeting civic hackers' expectations, and offer an alternative that emerged from our interviews—that a third party civic institution hosts a repository for cleaned civic data.

5.1. Government provision of data assets

Much of the respondents' feedback is related to decisions and processes around how data is collected, processed, presented and released; these recommendations must be implemented on the municipal government level (as opposed to the city's private partners who manage the data portal). For example, some respondents desired more transparency about why some datasets are opened and others are not, which would require more documentation around decisions to open—or not open-- requested data. Others requested more information about how a particular dataset was generated from an underlying database, which would be included as part of robust metadata.

One respondent highlighted his ambivalence about whether the city is the appropriate actor to provide all necessary data services, or if they should instead be provided by other sources. In his case, the respondent had difficulty finding local boundaries, for example, zip codes, school districts, council districts, and parks. He was building an application that could provide all of the boundaries that enclose a particular point (latitude and longitude); to build this application, he and a team of others had to locate data in a surprising array of locations, as well as manually reconstruct data from PDFs. He considered this a foundational dataset to support other applications, and wondered why the city did not provide it. However, he was unsure of whether the city would be up to the task, saying:

"Presumably this would be something that data.seattle.gov does. I have mixed feelings about that. It's hard to decide which services are best done by the city and by the data portal specifically, and which ones are better outside of that for various getting-things-done reasons. I guess the best example [of why] is that we don't care if we have permission to use the school district boundaries. We don't care. If we expected the city to release those boundaries as part of their service, they would have to care. I don't know what [internal] processes would slow down that resource" (interview 1).

In this case, the respondent's expectations for the city to provide a particular resource are tempered by his familiarity with municipal government liability and responsibility. Namely, that it is unlikely that government can have the same flexibility and experimental attitude that civic hackers have when aggregating and releasing data from disparate sources. In general, we found that Seattle does not release datasets owned by a different government entity (like the county), even if the city regularly uses it. The respondent went on to say that a local newspaper or Open Seattle itself may be a better place to host such a

service, in order to better leverage the flexibility of non-government actors (see Section 5.4).

5.2. Private provision of data portal

The City of Seattle works with a private company, Socrata, to host and manage its open data portal. Findings in our study with respect to API functionality and user experience would fall to Socrata to implement. Socrata is a local company, and regularly hosts Open Seattle at the Socrata headquarters and sponsors pizza for the group. Socrata evangelists and developers often join in on these meetings, through which the company has developed social relationships with local civic hackers. One interviewee notes one developer evangelist's responsiveness to his requests, and recognized that these social ties gave him an increased level of usability support. The Open Data Manager cited local social ties as a factor in the company's responsiveness to feature requests.

5.3. Situated perspectives of civic hackers

During the course of our interviews, we encountered that Open Seattle civic hackers' particular, situated perspective as professional developers. As an organizer of Open Seattle says, "Almost overwhelmingly the people that show up who are interested in working on projects are people that are transitioning jobs, or just finished something like a code school, and are looking for fun projects that they can say they've done [prior to a job search]. At first I just ignored it, but then I was like, 'Wait, this is too common to ignore.'" (interview 1). Civic hacker expectations for the open data program are shaped by their perspective as professional developers. All respondents reported that they were professional developers or engineers. This aspect highlights an aspect to our findings that may be out of reach for the scope and resources of the open data program, insofar as usability standards of private sector professionals may not apply.

Civic hackers also largely self-identified as members of the open source community. This may seem apparent to some because of the collective interest in open data, and open government [23]; however, it affected user expectations for aspects of the open data program to be run like an open source project. For example, instead of interfacing with Socrata for data portal needs, many respondents stated that they prefer that aspects of the program be run more like an open source project. "My interactions with Socrata have been great but I feel like there could be a more open platform for engagement. Perhaps like and open source code base--and I know they do have

that, Socrata does have some open source tools-- but there doesn't seem to be a whole lot of community engagement around it. [I would prefer] more of a distributed approach to improving data.seattle.gov instead of having everyone go through Socrata's community developer relations person. There is sort of like this one gate keeper you've got to work through. I mean for the business sense it makes sense, I can see why they're doing that but I can still how it could be better." Interestingly, respondents did not mention the open source platform solution, CKAN, developed by the Open Knowledge Foundation, an open-source solution developed by the Open Knowledge Foundation. While CKAN has 150 documented instances of use as of the time of this writing[7], civic hackers did not mention it in our interviews and discussions. Seattle's Open Data Manager said that this solution was not available at the time that Seattle began its open data program in 2009. He cited the local presence of Socrata as one factor in the city's decision to contract with the company; others included its affordability, high usability, and early and ongoing presence in the market space, saying, "If Socrata hadn't been there, I don't think we would have been able to get the data portal started when we did" (interview 16).

5.4. Two-way data pipeline

Just as the city is cautious with the data it makes available, it is also often unwilling to re-upload datasets that civic hackers have already cleaned. The city has a responsibility for the content and accuracy of the data that it releases; each dataset undergoes privacy and quality assessments prior to release. The awareness about the lack of two-way data provision between the government and citizens appeared in every interview and the focus group; "So down the line, hopefully, there will be a…two-way street for the citizen's input… [to] the governmental data" (focus group).

Adopting a crowdsourcing approach could potentially address several aspects of data quality. Respondents often discussed cleaning and aggregating data in ways that added value to the city's original data files. One respondent used the example of city boundaries, which he had assembled from a variety of sources into a usable boundary API (featuring for example school boundaries, political boundaries, park boundaries). However, as an expectation for city government data programs, the idea of a 'two-way street' is perhaps not realistic in light of the legal reality of the government as an authoritative source of

---

[7] "CKAN instances around the world" <http://ckan.org/instances/>.

information, and city governments' acute concerns about ethical and liability risks in privacy-invasive, inaccurate, or misinterpreted information [16].
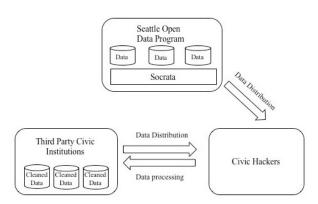

Figure 1. A diagram of a third party civic institution as a host for cleaned data

Given that the city is not able to accept cleaned data from citizens, he also considered whether other civic actors, like a local non-profit newspaper, could host such data services and do a better job of creating a two-way data pipeline. Thus, the idea outlined in Section 5.1 is a promising solution to this widely reported need; namely, that third party civic institutions may be best positioned to respond to an end-user need for a repository of cleaned data. This finding is a primary contribution of this research, as we believe it helps civic hackers not to duplicate their efforts to clean datasets, and creates a more agile alternative that meets their expectations with respect to format, and open-source engagement. Other respondents noted that any third party re-hosting data would need to create processes to support transparency in changes to datasets. This would be necessary to avoid serious risks, for example, the risk of propagating a dataset that had been tampered with.

6. Limitations and Future Research

This study is closely bound to a particular context in Seattle. The City of Seattle has a large and growing community of open data users who meet monthly, often at Socrata's own headquarters. The City of Seattle also has an unusual commitment to open data, and has created positions to formalize their efforts—the Open Data Manager is responsible for overseeing the management of the Seattle Open Data Program, and the Civic Technology Advocate who acts as a liaison between open data users and the city in order to anticipate and facilitate their needs. The city's Open Data Policy, promulgated by a Mayoral executive order, instructs City departments to

release more and higher quality open data. These factors have contributed to growing interest in open data within the City of Seattle, as well as increasing participation within the Open Seattle group. Although a small number of cities may exceed the total number of datasets Seattle has opened, most other cities do not share these qualities.

Local laws are a further factor that may play a role in respondents' expectations. Washington State has a strong, pro-transparency freedom of information law called the Public Records Act (PRA). The PRA creates a mandate for government agencies to release information upon request—compounding the expectation that as much data would be released by local agencies as possible. The PRA context is likely to have affected respondents' expectations to have access to any dataset they deem necessary. Some respondents in the study had actively filed public records requests in the past.

This study represents a preliminary look at the community of local open data users. It represents an open-ended effort to solicit as many challenges and expectations as possible. Further work is needed to assess these challenges in detail and to examine the resources needed to meet the aforementioned expectations.

7. Conclusion

This study highlights civic hackers as an important, under-researched stakeholder group. Our findings produced mostly technical feedback, cohering around two aspects; data portal usability and data quality. First, civic hackers perceived challenges with accessing important datasets, keeping data updated, and using current APIs of the data portal. Correspondingly, they expressed an expectation that their voices would be heard about which data should be released, data update notifications, and expanded API capability. Second, open data is considered to be often under-curated. Proprietary formats, lack of metadata, and untidy data were report as the main barriers when using the open data. Civic hackers appealed for standardized data formats, increased documentation and tidy data. Addressing these challenges requires changes to both private sector portal functionality, and the public sector provision of data and documentation. This confirms what has been previously reflected in the literature about usability problems of open government data; it is interesting to note that these problems persist even as more sophisticated data visualization tools and APIs are becoming available on data portals.

We found that the expectations that civic hackers bring to open data are related in part to their role as professional developers and members of the open source community. We also found a widely held hope that the city government would be able to integrate cleaned and improved data assets back into its data catalog. As we note in the discussion, the city government's interest in data privacy, accuracy, risk, and liability make it more difficult for them to accept responsibility for data assets that they have not generated. Therefore, we argue that this expectation is unrealistic. A primary outcome of our study is to point to the promise of a third-party civic institution, like a local newspaper, which could host cleaned datasets for wider use, and create processes to support transparency in such changes to data. Such an institution would remediate many of the expectations and needs expressed by civic hackers, for example, problems with data cleaning, or non-machine readable formats. It would enable them to build on data already cleaned and compiled by their fellow local residents.

8. References

[1]   A. Zuiderwijk and M. Janssen, "Open data policies, their implementation and impact: A framework for comparison," *Government Information Quarterly*, vol. 31, no. 1, pp. 17–29, 2014.

[2]   A. M. Townsend, *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company, 2013.

[3]   R. Kitchin, "Four critiques of open data initiatives," *LSE: The impact blog*, vol. 27, 2013.

[4]   V. Wadhwa, "Code for America: An Elegant Solution for Government IT Problems," *The Washington Post*, 2011.

[5]   "Code for America," *Wikipedia*. Wikimedia Foundation.

[6]   M. Gurstein, "Effective use: A community informatics strategy beyond the digital divide," *First Monday*, vol. 8, no. 12, 2003.

[7]   M. B. Gurstein, "Open data: Empowering the empowered or effective data use for everyone?," *First Monday*, vol. 16, no. 2, Jan. 2011.

[8]   J. Nielsen, "Heuristic evaluation," *Usability inspection methods*, vol. 17, no. 1, pp. 25–62, 1994.

[9]   D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with Big Data Analytics," *interactions*, vol. 19, no. 3, pp. 50–59, May 2012.

[10]   N. Helbig, A. M. Cresswell, G. B. Burke, and L. Luna-Reyes, "The dynamics of opening government data," *Center for Technology in Government.[Online]. Available: http://www. ctg. albany. edu/publications/reports/opendata*, 2012.

[11]   M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government," *Information Systems Management*, vol. 29, no. 4, pp. 258–268, Sep. 2012.

[12]   S. L. Erete, E. Ryou, G. Smith, K. M. Fassett, and S. Duda, "Storytelling with Data: Examining the Use of Data by Non-Profit Organizations," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 1271–1281.

[13]   S. Martin, M. Foulonneau, S. Turki, and M. Ihadjadene, "Risk analysis to overcome barriers to open data," *Electronic Journal of e-Government*, vol. 11, no. 1, pp. 348–359, 2013.

[14]   F. Gonzalez-Zapata and R. Heeks, "The multiple meanings of open government data: Understanding different stakeholders and their perspectives," *Government Information Quarterly*, vol. 32, no. 4, pp. 441–452, 2015.

[15]   T. McClean, "Not with a Bang but a Whimper: The Politics of Accountability and Open Data in the UK," in *APSA 2011 Annual Meeting Paper*, 2011.

[16]   J. Whittington, R. Calo, M. Simon, J. Woo, M. Young, and P. Schmiedeskamp, "Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government," *Berkeley Technology Law Journal*, vol. 30, no. 3, May 2016.

[17]   A. Wiig, "The empty rhetoric of the smart city: from digital inclusion to economic promotion in Philadelphia," *Urban Geography*, pp. 1–19, 2015.

[18]   J. C. Bertot, B. S. Butler, and D. M. Travis, "Local Big Data: The Role of Libraries in Building Community Data Infrastructures," in *Proceedings of the 15th Annual International Conference on Digital Government Research*, New York, NY, USA, 2014, pp. 17–23.

[19]   D. L. Morgan, *Focus groups as qualitative research*, vol. 16. Sage publications, 1996.

[20]   J. Kitzinger, "Qualitative research. Introducing focus groups.," *BMJ: British medical journal*, vol. 311, no. 7000, p. 299, 1995.

[21]   J. M. Corbin and A. Strauss, "Grounded theory research: Procedures, canons, and evaluative criteria," *Qualitative sociology*, vol. 13, no. 1, pp. 3–21, 1990.

[22]   B. G. Glaser and A. L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Transaction Publishers, 2009.

[23]   H. Yu and D. G. Robinson, "The New Ambiguity of 'Open Government,'" Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2012489, Feb. 2012.