Open Data Intermediaries: Motivations, Barriers and Facilitators to Engagement

GRAHAM DOVE, New York University, USA
JACK SHANLEY, New York University, USA
CAMILLIA MATUK, New York University, USA
ODED NOV, New York University, USA

Open data programs have become increasingly established at national and local levels of government. While the degree of success these programs have had in achieving their objectives remains open to question, one factor that has been identified as important to any success is the role of open data intermediaries, individuals and organizations that help others to make use of open data. In this paper we investigate how people become engaged with open data, what their motivations are, and the barriers and facilitators program participants perceive with regard to using open data effectively. We interview participants from a variety of backgrounds with differing levels of experience and engagement with open data. Participants include students learning how to train others in open data techniques and tools; people who attend open data events and use open data for commercial or social benefit; and representatives from local government, municipal agencies and a civic tech non-profit. We identify pathways to successfully developing and nurturing a community of open data intermediaries, and make five recommendations for organizations planning and managing open data programs.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Open Data, open data intermediaries, civic tech

1 INTRODUCTION

Policies promoting public access to data generated by government agencies and services have been adopted at national [27, 50, 68], State [18, 28], and municipal [5, 52] levels. These *open data* policies aim to make government data routinely available without the need for specific requests, and with the intention that data will be reused [24]. Proponents argue that the implementation of these policies can positively impact government transparency and accountability, service efficiency, civic participation, and entrepreneurship [29, 56, 69]. These aspirations also provide the guiding principles and objectives that government organizations typically attach to their open data programs. For example, the U.S. Federal Data Strategy states one of its core principles to be "Ethical Governance", and one of its core practices to be "Building a culture that values data and promotes public use"

[51]. In New York City's vision for open data it states that, "The ultimate success of the Open Data initiative will not be measured in the number of of data sets that are published on the Open Data Portal - it will be in the number of of New Yorkers who use Open Data in their daily lives. And that's not just the tech-savvy New Yorkers - it's all New Yorkers, in all five boroughs" [53]. While it remains somewhat unclear how much impact these open data policies have had on increased transparency and accountability, how much commercial and social value has been created, or the degree to which civic engagement is improved [74, 79], research indicates that the role of *open data intermediaries* is crucial to those successes that are achieved [24, 46, 64]. Broadly speaking, open data intermediaries are those individuals and organizations that help others to access and make use of open data. Some examples of intermediary activities include, but are not limited to, providing training and education, hosting and organizing events, conducting analysis for advocacy, and developing tools and software.

In the US, where this research is based, the big push towards open data initiatives came with the Obama administration's Open Government Initiative in 2009 [33]. Shortly after this announcement, Robinson et al. [59] argued that government should be "providing reusable data", and that private, non-governmental actors would be best placed to "create and reshape the tools individuals use to find and leverage public data". In their concluding remarks, these authors introduce the idea of an open data intermediary, stating that, "[t]hroughout the discussion, we have operated on the premise that citizen interaction with government data requires an intermediary". Building on this, Magalhaes, Rosiera and Strover [46] offer a framework that describes open data intermediaries in terms of "civic startups, open data services, and infomediaries", and outline how products or services that utilize open data for governments, citizens, or third-parties can be organized based on their intersections. However, while intermediaries may increase the accessibility and utility of data, and have the potential to democratize the impacts and use of open data [71], they can often be invisible or absent from official open data narratives. Moreover, open data proponents may be wrongly assuming that just because there is an important role for intermediaries, there will automatically be organizations or individuals ready to step in [24].

In order to better understand how individuals become involved in open data activities, and so have the potential to become open data intermediaries, we report on semi-structured interviews with seventeen open data users whose relationships to the open data ecosystem span a wide spectrum of roles, including those who might already be considered open data intermediaries and those who have the potential to fulfill this role. In particular we investigate their motivations for initial and ongoing engagement with open data, and probe on barriers and facilitators to these engagements. Our stance with regard to who might be considered an open data intermediary is expansive and inclusive. Similarly to Robinson et al. [59] and Magalhaes, Rosiera and Strover [46], this includes government and non-government organizations and individuals, whose involvement includes, but is not limited to, developing tools and products, and providing services, training, and organizing events. We discuss our findings from these interviews in light of previous research and across three main themes: (1) the ways in which open data intermediaries' motives and objectives align with the stated objectives of government open data programs; (2) the need for knowledge of civics and an understanding of local issues of concern, alongside knowledge about data analytics; and (3) the importance of communities of practice to supporting open data intermediaries. Finally, we make five recommendations for supporting pathways to successfully developing and nurturing a community of open data intermediaries, with a focus on organizations planning and managing open data programs.

1.1 Research Context

This research is focused on the experiences of participants whose experiences with open data take place in a major city in the Northeast US. This city has enacted an open data law which mandates the aim of making all 'public' data, i.e. data maintained by municipal agencies, available through a single portal platform. To facilitate the availability of open data, the city has created a team that has responsibility for developing and maintaining the open data program. This team bridges two municipal agencies: the office of data analytics, and department of information technology. This city also has a vibrant civic tech community of volunteers and non-profit organizations that is actively involved in advocating for the further development of open data programs and platforms in the city. This civic tech community also develops tools that utilize open data, and hosts events and training that aim to promote wider open data use. These events and trainings can be independent of or run in conjunction with municipal agencies. For two years we have been working with representatives of the city's office of data analytics and with a prominent civic tech non-profit to investigate informal learning that takes place around engagements with open data in the city.

1.2 Contributions

The research reported in this paper offers two main contributions. First, we identify what motivates individuals to become engaged with open data, together with perceived barriers and facilitators to these engagements. This is a topic that has been the subject of only limited prior research. Second, we consider pathways to successfully develop and nurture a community of open data intermediaries, and make five recommendations for organizations planning and managing open data programs. These include: (1) easing entry and scaffolding progression; (2) adopting a holistic view of 'open data literacy'; (3) integrating teaching and training; (4) supporting community-focused data views; and (5) supporting issue-focused data views. This is important because, communities of open data intermediaries can play a crucial role in developing tools, hosting events and training, and supporting others in using open data to improve lives and increase democratic accountability.

2 RELATED WORK

In this section we discuss prior work in three main areas. First, to motivate our exploration of the barriers and facilitators of effective engagements with open data programs we discuss prior work that has investigated the official objectives of open data programs and whether they are being achieved. Second, we take a focused look at prior research investigating the particular roles and practices of open data intermediaries. Third, we consider prior HCI and CSCW work in areas related to open data and civic technology more widely. To conclude this section we position our own research in relation to our synthesis of reviewed literature.

2.1 Barriers and Facilitators to Achieving Open Data Program Objectives

The degree to which open data programs have successfully achieved objectives, e.g. government transparency and accountability, service efficiency, civic participation, and entrepreneurship [29, 56, 69], has been the focus of much open data research. For example, Peled [55] highlights a potential contradiction between open data transparency goals and government agencies' goals of collaborating through data exchange. Similarly, both Slobodova & Becker [67] and Zuiderwijk, Shinde & Janssen [79] suggest that, despite broader visions, objectives may be narrowed down to economic development and administration efficiency. In addition, program success can be undermined by adopting a conceptually simplistic view of open data, which can lead to myths that offer seductive tales of importance and simplicity [35]. Wilson & Cong [74] investigate who the users of open data are and what impact open data programs have, finding that distinctions between

public and private sectors may be blurred and that the effects of open data initiatives have been mainly experienced within government. Other challenges include an ethical-economic tension between governments, publics, and private sector actors [66], and situations in which the data most useful to entrepreneurs may not best enable citizens' assessments of accountability [76]. Ruijer et al. [61] suggest a need to support richer interactions with users because an absence of shared cognitive frameworks can prevent collaborative learning. Alongside these other challenges, difficulties in using required technology may also slow progress towards open data programs' stated goals [34].

In response, Zuiderwijk & Janssen [78] offer a framework for comparing open data programs, and suggest that improvements can be made through collaboration and a culture in which publicizing open data is incorporated into daily work. Johnson [38] also investigates evaluation metrics for open data programs, suggesting that value should be broadly defined and that institutional contributions can provide important self-evaluation measures. Hivon & Titah [32] highlight 'hands-on activities, greater responsibility, better communication and improved relations between citizens and the open data portal development team' as key to participation in open data programs; while Zuiderwijk & de Reuver [77] suggest that policymakers can overcome barriers to achieving open data objectives by prioritizing efforts on functionality and support, and inclusiveness, as barriers in these areas are most obstructive. Other approaches to overcoming these challenges include adopting an ecosystems lens that uses systems dynamics to model programs and anticipate repercussions from initiatives [31], or to assess existing conditions and consider policies, strategies, and relationships that address barriers [17].

Taken together, this body of research points to the challenges of neatly ascribing and evaluating objectives in a context in which a broad range of actors may have an equally broad range of potentially conflicting goals. It also indicates how members of different communities of practice and interest are involved in ongoing negotiations around what the aims and objectives of government open data programs should be, and how those programs should best be implemented in order to achieve these goals. This strongly suggests that there is no one size fits all solution, and that the process requires a pool of engaged individuals and organizations, both inside and outside of government, willing to participate. This pool of participants in open data and civic tech activities is fed by, but also feeds into, the community of open data intermediaries who develop tools and products, provide services and training, and organize open data events and meetups. We build on this research to investigate ways this community and its potential members are motivated, and to better understand the barriers and facilitators to maintaining these motivations.

2.2 Roles and Practices of Open Data Intermediaries

The role that intermediaries play in providing under-served communities access to data and in assisting with analysis was initially identified in the the 1990s by Sawicki and Craig [63]. Following President Obama's 2009 announcement of the Open Government Initiative [33], Robinson et al. [59] suggested that intermediaries might play a similar role in facilitating wider access to open government data, placing their vision of an 'ecosystem of grassroots, unplanned solutions to online civic needs' within an 'American tradition of entrepreneurial self-reliance'. Magalhaes, Roseira & Strover [46] build on this insight to offer a framework that describes different types of open data intermediaries, based around the intersections between 'civic startups', 'open data services', and 'infomediaries'. Johnson & Greene [39] look more deeply at 'infomediaries', i.e. third parties who accesses government open data to create value-added products, highlighting five classes (government, private sector, NGO, academic, and media), they suggest that activity is largely concentrated within government and private sector types. Frank & Waddell [24] discuss how intermediaries, such as civil society organizations and media, might act as a link bridging government and people.

Their four-level framework builds up from 'support', through 'consultation/protest' and 'deliberation', to 'collaboration', highlighting the important role that trusted intermediaries play in providing practical support and increasing confidence that contributors' efforts are put to good use. Schrock & Shaffer [64] discuss the 'data ideologies' of open data intermediaries, who they argue have an important political role and describe as being 'extra-institutional actors that obtain, use, and translate data for the public'. They also compare perceptions of data and perspectives on government-intermediary collaboration, and show how professional experience influences the tasks intermediaries are likely to undertake, which in turn influences how open data might be translated into alternative products. Although open data intermediaries may be able to increase the accessibility and utility of data [71], Da Silva Craveiro & Albano [16] suggest that alongside the often discussed technical resources there is also a need for 'information conceptualization resources' to fully support the roles that open data intermediaries play in budget transparency. In addition, they highlight how intermediaries need to form partnerships in order to act effectively. Gebka, Crusoe & Arhlin [25] present a set of design principles that aim to extend the range of open data reuse by supporting intermediaries in building open data products based on capturing users' 'activity-based information needs'.

This prior research highlights the importance of intermediaries as a key element in the ecosystems surrounding government open data programs, shows how their roles may be understood and characterized, and provides suggestions for how support might be provided for particular activities. However, an understanding of how individuals and organizations become involved in open data, and more importantly what the barriers and facilitators to the continued engagement necessary to act as an intermediary, has been largely missing from this prior work. This is important because without such consideration the presence of organizations and individuals to fulfil the role may continue to be taken for granted [24], undermining the chances of successfully achieving open data program objectives. We build on and extend prior research by inquiring into these factors through interviews with a broad range of stakeholders involved in using government open data.

2.3 HCI and CSCW Inquiry into Civic Engagement with Open Data and Related Technologies

CSCW and HCI research has considered civic open data, and civic tech more broadly, through a design lens and through inquiry into the ways that technologies are used to collaborate around issues of democracy; typically with reference to the work of community activists. In this section we provide an overview to key streams of this literature to help situate our research.

Choi & Tausczik [11] characterize collaborations between people undertaking the work of open data intermediaries from a CSCW perspective, finding these interactions to be 'interdisciplinary, small-scale, with low turnover, and synchronous communication' and highlighting how they are typically exploratory without sophisticated analyses. Pusaar et al. [57] criticize these efforts as not being sufficiently inclusive of local community members. Addressing this criticism, they describe the process of co-designing, developing and deploying a system for making government open data usable and accessible for a local community in the north east of England, with the explicit aim of supporting civic advocacy and referencing this through Gurstein's term "effective use" [30]. However, researchers seem to play the role of intermediaries in this project, and it also appears likely that their interactions are actually very similar to those described by Choi & Tausik. Furthermore, at the time of writing, the Data:In Place platform that resulted from this work is no longer active. This draws attention to the challenge of maintaining engagements beyond the timescales of research funding, a point which has also been highlighted by Gooch et al. [26] and Johnson, Al-Shahrabi & Vines [36], and involves relations that are discussed in more detail by Le Dantec & Fox [43]. Other similar projects in which researchers adopt the role of open

data intermediaries include, BudgetMap a tool for issue-driven interactions with government budget programs [40], and Accountable a tool for local government and non-profit organization financial data [48]. Boehner & DiSalvo [7] found that civic leaders approach open data from three perspectives, highlighting its use in solving problems, creating opportunities, and facilitating exploratory curiosity. Similarly, Erete et al. [23] investigate non-profit organizations use of open data, highlighting the role open data play in developing narratives that build a case for support and suggesting a need to facilitate communication and support relationships between organizations and expert data analysts. Loukissas and Ntabathia [45] challenge the attribution of openness to data, preferring instead to consider the openness of 'data settings', which are 'contexts in which things of public significance can be presented as evidence'. This situated approach considers openness in three forms: 'accessibility' (material conditions addressed through configuration of the data setting), 'inclusivity' (social relations addressed through convening the data setting), and 'indeterminacy' (rhetorical statements understood through the claim-making that takes place in the data setting). This work acknowledges the centrality of the authors as researchers and designers, and therefore open data intermediaries, and descriptions of their collaborations again seem to closely match those characterized by Choi & Tausczik [11], while also addressing some of the concerns raised by Pusaar et al. [57]. However, as the authors also acknowledge, each configuration and convening of a data setting results in a situation that is both more and less open than its alternatives might be, therefore demanding a high level of reflexivity in the open data intermediaries that bring it into being.

Moving beyond interactions with open data, CSCW and HCI research has also considered themes associated with the use of technology in civic engagement efforts more broadly, typically through the lens of digital civics [54]. In this context participation may span multiple sites at scale [2], and aim to support collective community engagement [3, 41] for effective civic discourse and deliberative consultation [36, 37]. To address such challenges, Asad et al [3] offer a playbook highlighting a plurality of community engagement that includes residents, municipal officials, and service providers; while Weise, Coulton & Chiasson [72] suggest that analysis of patterns of institutional practice can support changing configurations of civic engagement. Dickinson et al. [19] highlight the role (often unfulfilled) that civic tech might play in building connections between local assets, and strengthening relationships between local government and under-served communities. They contrast this with what they identify as the more typical use of civic tech, i.e. in transactional attempts to increase efficiency. Similarly, Asad & Le Dantec [1] describe activists' use of civic tech to draw attention to issues of concern, translate messaging for different audiences, and facilitate participation in action; suggesting that technology can offer a politically significant resource and means of practicing democracy. Reflecting on the process, outcomes, and challenges of deploying community situated deliberative consultation platforms, Johnson et al. [37] highlight potential gaps between rapid engagement and slower civic decision-making action, and warn that digital technology may reaffirm preexisting issues around trust. These challenges are reiterated by Johnson, Al-Shahrabi & Vines [36], who also highlight the role of social capital in convening a breadth of participation, note that digital collaboration generates a wealth of data that can be difficult to make sense of, and again draw attention to challenges associated with passing on ownership of projects initiated within research contexts. Other considerations include relationships that may initially be distrustful, and so require attending to past experiences and future expectations [12, 13]; the need for data to be situated and to reflect the social relations that lead to data production [42]; the role that small-scale experiments can play in fostering "caring democracy" in the context of community use of civic tech [49]; and tensions between community desire for innovation and governments strict compliance with regulation [22]. Following a systematic literature review, Saldivar et al. [62] highlight a disconnect between civic tech research and practice, calling on researchers to pay more

attention to the experiences of those organizations actively engaged in developing and using civic tech.

2.4 Summary and opportunities

This review of prior research suggests that interactions around open data, and civic tech more broadly, involve a wide range of stakeholders with varying and sometimes competing needs and motivations. Also that open data provide a shared point of reference or platform supporting collaboration for civically engaged groups, including activists and those involved in local government. Studies of open data program effectiveness have highlighted the important roles that intermediaries play in developing tools and supporting others' use of open data, and studies investigating these roles have suggested areas where additional support is needed. CSCW and HCI research also suggests key areas in which mainstream understandings of things like 'openness' can be challenged, and offers lessons in design for civic engagement and participatory democracy. However, the interventions that are developed for research projects typically struggle with longitudinal maintenance beyond the lifespan of research funding or individual PhD studies, and there are few examples where academic projects have been transferred into self-sustaining community ownership. This is not meant as a criticism of researcher motives and objectives, as in our experience similar projects started outside academia are also likely to struggle with sustainability. However, it does highlight just how important it is to better understand what motivates people to become engaged with open data, and what barriers and facilitators they perceive with regard to both initial and continued engagement. It is through these engagements that people come to occupy the role of open data intermediary, and maintaining a pool of engaged intermediaries is key to longitudinal success that supports communities in demanding better and more accountable governance. We build on the prior work reported here by conducting an inquiry into what these motivations, barriers and facilitators might be, a topic that has been the subject of limited research to date. We then use this new understanding to identify pathways to success, and make recommendations for organizations planning and managing open data programs.

3 METHOD

In this section we first describe how our data were collected through remote interview and focus group discussion with a wide range of people engaging with open data. We then provide details about our qualitative analysis method. This research was conducted under the approval of our university's Institutional Review Board (IRB-FY2020-4510).

3.1 Data Collection

Data were collected through individual video interviews and an online focus group. We recruited a total of seventeen participants (7 female and 10 male) who are engaged with open data through a variety of different roles. The purpose of recruiting people who have engaged with open data in this variety of ways was to get the different perspectives of people who are starting their open data journey, and currently represent potential future open data intermediaries, as well as those who are more experienced and established in the role. All participants have used the city's open data portal, and attended open data trainings, meetups, and similar events. Participants included: a) three people who use open data in different forms of public service work and have attended open data training or events, whose open data intermediary standing we characterize as *using services*; b) four students training to use and teach open data tools as intern fellows at a civic tech non-profit, and one student who has attended other open data trainings, whose intermediary standing we characterize as *in training*; c) three people that we characterize as *early stage*, who participate in open data and civic tech community events, and use open data in education, entrepreneurship,

Table 1. Overview of research participants.

ID	Role Using Open Data	Intermediary Standing	Interview	Focus Group
P1	Student Civic Tech Intern	In training	Y	N
P2	Public Health Professional	Using services	N	Y
P3	Civic Tech Professional	Active and experienced	Y	N
P4	Entrepreneur	Early stage	Y	N
P5	Analyst, Educator and Trainer	Early stage	Y	N
P6	Student Civic Tech Intern	In training	Y	N
P7	Educator	Early stage	Y	N
P8	Civic Tech Professional	Active and experienced	Y	Y
P9	Local Government Administrator	Using services	N	Y
P10	Student Civic Tech Intern	In training	Y	N
P11	Open Data Program Administrator	Active and experienced	N	Y
P12	Community Board Member	Using services	Y	N
P13	Student	In training	Y	N
P14	Civic Tech Professional	Active and experienced	Y	N
P15	Civic Tech Community Volunteer	Active	Y	N
P16	Student Civic Tech Intern	In training	Y	Y
P17	Open Data Program Manager	Active and experienced	Y	Y

and for community action; d) one person who is an active community volunteer providing data analytics skills to different projects, and whom we characterize as *active*; e) three members of staff at a civic tech non-profit that provides open data training and analysis services, who also run open data meetups and events, and who we characterize as *active and experienced*; and f) two members of staff at the municipal agency responsible for open data, whom we similarly characterize as *active and experienced*. For participant details see Table 1.

Participants were recruited through the extended networks of research project members, and via a civic tech Slack channel. Fourteen participants were interviewed individually online via Zoom by the first author. Semi-structured interviews lasted between 20 minutes and 60 minutes (mean = 34 minutes). We probed participants on their background and motivations for engagement with open data, their experiences using open data and at open data events (including training events), and their thoughts on barriers and facilitators for starting and continued engagement with open data. A full interview protocol is included as an appendix. We collected additional data from an online focus group discussion (also via Zoom) held as part of an open data event that takes place each year in the city where this research is based. This focus group was an open discussion around the topic of participation in open data facilitated by the first author. It lasted for ninety minutes with six participants, three of whom were also individual interview subjects. The event was open to all and advertised through the umbrella event of which it was a part. An informed consent was obtained from all participants, and those interviewed received a $30 gift certificate as compensation.

Table 2. Summary of findings

Theme	Sub-themes
Motivations for engagement with open government data	(1) As a principle of better government; (2) As a resource for public service; (3) As a resource for education; (4) As a resource for community activism; (5) As a resource for entrepreneurship.
Barriers to engagement with open government data	(1) Range of skills required; (2) Data anxiety; (3) Exclusionary language and challenges to communication; (4) Poor quality or unavailable data; (5) Lack of awareness; (6) Unclear objectives.
Facilitators to engagement with open government data	(1) Inspirational examples and role models; (2) A strong community of practice; (3) Accessible events and training; (4) Easy to use tools and platforms; (5) Data that are personally or domain relevant (6) Civic interest.

3.2 Analysis

Initial transcription of focus group and interview recordings was generated automatically through Zoom. Two of the authors viewed each video individually multiple times, and corrected the transcripts as necessary. Analysis of this data was guided by Seidman's guidelines for approaching interview data as qualitative research [65], and Braun and Clarke's approach to thematic analysis [9, 10, 47]. In each of these approaches, the key activity is drawing connections across sources through themes or threads, rather than attempting to articulate a unifying theory, which is more typically the aim with grounded theory approaches. Analysis started with an initial familiarizing stage, in which interview recordings were viewed and reviewed, and transcripts read and reread. Key passages of interest were highlighted and initial codes applied, using an open coding scheme. During this analysis we paid particular attention to identifying participants' motivations for engaging with open data, and to perceived barriers and facilitators to this engagement. Connections were then drawn between coded interview data in order to generate themes. These were further refined through an iterative process of interpretation and agreement reaching. This analysis was led by the first and second authors with the third and fourth authors providing validity checks for codes and contributing to reaching agreement. Our analysis was then presented back to interview participants for member checking [15], and adjusted accordingly.

4 FINDINGS

We present findings across three main themes. First we provide details on what motivated participants to first use or become engaged with open data, or to later deepen that engagement. We then present the barriers to this engagement that participants identified. Third and last we look into facilitators of engagement. Table 2 provides an overview of our findings.

4.1 Motivations for engagement with open government data

In this section we identify participants motivations for becoming and remaining engaged with open data and the community that supports it. These range from a belief that municipal government data belongs to city residents and should therefor always be made available and accessible, through the resource open data offer those in public service and education, to the opportunities open data support with regard to community activism and entrepreneurship.

4.1.1 Open data as a principle of better government. A common theme among participants was that making government data open and accessible to a wider public is a beneficial principle that will lead to better government. Reasons for this included: (1) a desire to make government more efficient; (2) because analysis can support better informed decision making, and (3) because it can help a greater number of people choose the direction that government takes. The tensions that arise from questions of what should qualify as open data, and the competing principles of privacy and security that are often important to attempting to resolve these questions, were also a motivator for remaining engaged with open data as an issue of concern. Related to this was a broader interest in the role that technology can play in doing local good. However, P8 summed this theme up well when saying, *"It's very much a values based thing. I believe that things should be open, that people should be sharing, that cities are the product of the citizens that live there and we should be able to have access to their information"*.

4.1.2 Open data as a resource for public service. For those working in public service the motivation for engagement with open data was typically how it can support better work practices. This might be: (1) to help understand particular local needs that support budget recommendations or policy decisions; (2) to help inform community health needs and public health dissemination; or (3) as a way to support city-wide understanding across municipal agencies that reduces the friction of official requests. P12 described it in this way, *"One important component is to try to distinguish fact from anecdote, and opinion from verifiable information, and so open data can be a source of that. For example, simple things like exactly how many school children are there in our district? and how many of them go to public versus private school? and of them how many go to the school near them or some other school?"*

4.1.3 Open data as a resource for education. Several participants spoke of open data as being an important resource for education, and as this being a grounding or motivation for engagements. In this group it was fairly common for participants to discuss their first contact with open data coming as students practicing analysis (e.g. P6, P8, P13 & P15). Other participants discussed using open data in teaching, both at K-12 level and university, and for research. In addition, three of the four undergraduates interning with the civic tech organization were introduced to open data through participation in this program. The importance of open data as a resource for education was explained by P7 who told us, *"The driving motivation for me to get into data science was pursuing my own interests and part of it was also to use it in my teaching. In schools it's been a really fun way to just get students thinking about real world scenarios of data science and statistics"*.

4.1.4 Open data as a resource for community activism. Another motivation for engagements with open data is the role they can play in support of community activism. Here participants spoke of sharing the skills they have gained and using technical knowledge to help unpick what is available in order to support people who want to be more informed about the city they live in. In addition, P3 in particular stressed the role of activism in highlighting where there might be a need to advocate for making data open saying *"Displacement issues were really boiling to a head and people being priced out of their apartments that they lived in for years and seeking information about rent stabilization as a protective device to hold on to their apartments, and not being able to find good up to date information on which apartments and buildings were still under rent stabilization restrictions"*.

4.1.5 Open data as a resource for entrepreneurship. The final motivation for engagement with open data was highlighted by P4 who described setting up a company that aims to provide actionable data analyses for advertisers, focused on the pharmaceutical sector, and investigating how municipal open data might play a role. P4 told us how starting out by looking at national medical data was an inspiration for looking at what else might be available more locally, *"I had recently discovered*

federal data sets. Basically CMS data, which is the Center for Medicare and Medicaid Services, I think. And I was blown away at how much data is available from the government in lots of different ways, and so I was just very curious about what [city] is doing, from an open data perspective".

4.2 Barriers to engagement with open government data

In this section we highlight the barriers that participants identified with regard to engaging with and effectively using open data. These include: (1) the range of skills required to use open data effectively; (2) the anxiety that can be associated with attempts to use open data; (3) the ways language around open data can be difficult and associated challenges to communication; (4) the unavailability and poor quality of data sources; (5) lack of awareness about open data resources; and (6) poorly understood or evaluated open data objectives.

4.2.1 The range of skills required to use open data effectively. The perceived range of skills required to make effective use of open data was a potential barrier to engagement that was raised by all bar one of our participants. The skills participants highlighted included technical proficiencies associated with using open data and data analytics tools, and an understanding of civics and how local government works. Participants often mentioned either not coming from a computer science or data analytics background, or conversely not having a background in the social sciences or politics. In these cases they would typically follow up by saying how this missing background was a necessary precursor to being able to effectively carry out analyses of open data. For those who felt they lacked technical skills the tools for conducting analysis and gaining insight were challenging, while those who felt they lacked civic understanding highlighted the challenge of knowing what questions to ask. In addition, participants also highlighted the need to be able to frame a question that would guide analysis. Here they were typically talking about the challenges associated with understanding how to translate concern for an issue into a question that might be answered using the available data, including how to set constraints and think about possible relationships between data that might come from different sources. This was often mentioned in contrast to the idea that you might just explore the data to find something of interest.

4.2.2 Data anxiety. Negative cognitive and emotional experience are a second major barrier to engaging with open data. Primarily this was discussed as feeling out of depth, intimidated or overwhelmed. Participants noted this *data anxiety* with regard to open data training and civic tech events, but also with regard to being confronted with a large repository of open data sources. It was typical for participants to discuss wanting a simpler route in or a clearer pathway. A second and likely related barrier that was identified through our interviews is the possibility that people will have had previous negative experiences often associated with math classes at school, which they need to overcome before engaging with open data.

4.2.3 Exclusionary language and challenges to communication. The language and jargon that is often associated with open data likely exacerbates other potential barriers and can also be a barrier to engagement in itself. Participants highlighted how jargon that can originate from both data analytics and government domains can seem like buzzwords that reduce accessibility because there is an assumption that the audience is equally knowledgeable. As P14 explained, *"You have to be willing to learn a new language. You have to learn the language of government, and you have to learn the language of math together. And so that within itself is a pretty steep hill"*. An associated barrier was the challenge of communication. This might be challenges posed by finding the language to share the results of exploratory analysis, or the challenge of translating complexity that underlies large municipal data sets into supporting documentation and help.

4.2.4   Data that are poor quality or unavailable. Whether or not the data that might help to answer a particular question is even available was highlighted as another important barrier to effectively engaging with and using open data, particularly with regard to supporting community activism. This was seen as a potentially major hurdle that can put people off exploring this route to helping people. An associated barrier participants highlighted was the challenge of data quality and usefulness. Examples included: (1) metadata categories that are misaligned with the particular question being posed; (2) challenges with collecting and therefor making available useful demographic or other potentially personal data; and (3) frustrations that available municipal data can be good at describing the *what* of a situation but less helpful explaining the *why*. In addition to this, there was a perception that while open data can be interesting and support a detailed dive into particular questions, this degree of interesting detail can be a distraction from answering the real question that initiates engagement. Open data tools often support investigation of a particular data set, while the questions people want to answer may relate to data that cross different sources and use different terminology to relate to the same entities. As P12 put it, there is a danger that *"Those things that are easy to research become the subject or the practice of the research, not because they relate to what it is I'm trying to prove but because they're there. It's the old joke about 'why are you looking over here when you lost it over there? Well over here the lights are better'."*.

4.2.5   Lack of awareness about available resources. Participants who are not engaged with open data on a regular or day-to-day basis through their work often highlighted a lack of awareness, particularly about open data events and training, as a barrier to deeper engagement. These participants highlighted the necessary process of discovery as being a particular challenge among competing demands on time. This was discussed with regard to: (1) what government open data is; (2) what data are available; (3) what resources there are to support people; and (4) events and training designed to increase engagement.

4.2.6   Unclear objectives and evaluation. The final barrier we identified from interviews was a potential lack of clarity about the range of objectives for government open data programs, and then difficulties in measuring program effectiveness that may in part result from this. For example P4 noted that the question of whether entrepreneurial use of open data is not merely accepted but actually encouraged was unclear. Because of this it also remained uncertain whether data would remain available and accessible. As P4 put it, *"Part of me kind of feels like not stealing the data but it's oh I'm trying to re-purpose this data for my own personal gain, when it was really intended for some other purpose"*. For P17 the challenge is to identify metrics that adequately reflect objectives when those objectives may be a relatively loose set of aims. As a result it can be hard to identify and promote the success stories that act as facilitators to engagement.

4.3   Facilitators to engagement with open government data

In this section we present the facilitators to engagement with open data that participants discussed. These include: (1) being introduced to inspirational examples and role models; (2) support from an open data community of practice; (3) effective open data training and events; (4) open data tools and platforms; (5) engaging people with data that are personally relevant; and (6) building on existing interest in civics and government.

4.3.1   Inspirational examples and role models. The important role of inspirational examples of open data in use was one facilitator of engagement that was highlighted both by those who work with open data on a day-to-day basis and those new to open data or whose engagement is more sporadic. Participants spoke of attending events in which examples of software or analysis were shared and discussed, covering topics such as housing rights and accessible public transportation, as being

evidence of what can be achieved using open data. Another aspect highlighted was the way in which art based approaches can make data more accessible and approachable, and how journalists can raise the profile of open data by acknowledging its role as a source for news stories. Alongside inspirational analyses and cool apps or tools, role models can also play a key role in helping engage people to find a place in open data. An example of this was offered by P10 who told us, *"I felt I was learning things late and she said no I did not learn until my doctorate. She made the environment so welcoming and so open. I never felt like I was dumb for asking a question"*.

4.3.2 A strong community of practice. Those participants who have become very engaged with open data all discussed how the community of practice that has grown around civic tech and open data was an important facilitator to deepening their engagement. They considered this to be a community of like minded people that are available to share advice and provide technical assistance and local knowledge when asked. The community also extends around particular topics to include local neighborhood groups and domain experts, and people from inside government agencies and academics, and so can become a forum for connecting the pieces and building trust that facilitates longer-term projects and relationships. P15 spoke from this perspective saying, *"Most of the people that I've worked with don't necessarily have the skill set of working with data, they're mostly community initiative driven things. They know data is important and I think then having someone who can help them look further into it and find data is definitely part of that civic community"*.

4.3.3 Accessible events and training. Open data events and training were typically considered a good way to bring people into working with open data. Participants often suggested that even those events where they may not necessarily have understood everything were inspirational and highlighted the presence of a wider community of interested people. What was often highlighted as being important was the creation of a safe space where people feel able to ask questions without being intimated. Perhaps because of this, one suggestion for increasing the effectiveness of these events that participants made was to more clearly identify and organize them by skill level required, e.g. beginner or intermediate technical skills and basic domain knowledge.

4.3.4 Easy to use tools and platforms. The role that tools and platforms play in facilitating peoples' engagement with open data was also highlighted. Here participants discussed how dedicated tools can make interacting with open data and gaining insight simpler than having to download a data set and analyze it using R or Excel. Visualizations and in particular maps were considered an important aspect of this. Participants suggested that one way to improve the effectiveness of these open data tools would be to increase consistency across data sources, e.g. by promoting common naming conventions.

4.3.5 Data that are personally or domain relevant. Making sure the subject domain of data sources used in examples or training is relevant to the audience was discussed by participants as being a key facilitator of initial engagements with open data. For those conducting training this was highlighted as a key factor in planning, while those who had attended training sessions suggested that this was an important factor in making them relevant to a real world inquiry. As an example of the importance of relevant data to getting people interested P17 told us, *"Every time we're talking about open data with anyone, the question is always well what does it look like for my neighborhood. And once you're able to show somebody at that level that's when they're like, oh no there's a mistake in this data. Like, this can't be right because I know this person across the street, they put through a request for this. Or the tree is actually over here not over there. People get really interested when it's about them"*.

4.3.6 Civic interest. The final facilitator for open data engagement participants discussed was simply having an existing interest civics, government and politics. This was typically from the perspective of using someone's existing civic interest as a hook to show them how open data might help make government more democratic or efficient or effective. Engagement with civic life more generally, e.g. by attending community board or council meetings, was also seen as a step towards overcoming a lack of civic knowledge.

5 DISCUSSION

In this section we discuss our findings in light of prior research. First we discuss where the individual motivations of participants we interviewed align with and differ from the beneficial motivations typically ascribed to open data programs at the level of institution or government. We then discuss the implications arising from our findings that interest in civics and local issues of concern are of similar importance to data analytics skills in acting as a barrier or facilitator to engagement with open data. Third, we discuss the role that communities of practice play in supporting potential open data intermediaries. Finally, we make five recommendations for organizations planning and managing open data programs that aim to support pathways to successfully developing and nurturing a community of open data intermediaries.

5.1 Alignment Between Open Data Intermediaries Motivations and Those of Municipal Government

The motivations for government open data programs are typically discussed from the perspective of benefits at a systems or institutions level [35]. Perceived benefits include: (1) greater transparency and more government accountability; (2) improvements to public service efficiency; (3) increased civic engagement and public participation in the process of government; and (4) economic benefits and entrepreneurship [29, 56, 69]. Similar sets of motivations also lie behind systems designed by HCI and CSCW researchers, e.g. [40, 48, 57]. Each of these themes are reflected in the individual reasons that participants in our study gave for their initial and continued engagement with open data. Our findings also show that they often remain motivations because they have not yet been fully realized. For example, the process of negotiating access to data on subjects that may be considered more politically or personally sensitive was discussed both as a goal that motivated continuing involvement and as a barrier to effectively making use of open data. In other areas, these institutional and systems level motivations were not clearly signaled through the programs participants had engaged with. For example, uncertainty around official support for individuals making financial gains from open data entrepreneurship seemed to potentially undermine initial motivations for engagement. The relationship between these interrelated system level, institutional, and individual motivations and how they are implemented in open data programs is complicated, but requires unpicking through further research if we are to understand how longitudinal engagement is best supported for open data communities to flourish.

5.1.1 Open Data Benefits to Education. One area that emerged from our research as being a strong motivator but which is mentioned less often in the literature around open data, and has not been the subject of prior HCI or CSCW research, is the benefit offered to education. Our findings suggest that open data is a resource that supports research activities, and also that open data provides opportunities for students at undergraduate and graduate levels to practice data analytics. Perhaps more importantly, our findings also suggest that open data offers potential benefits to teaching school-age math and by implication to make math relevant to those returning to education. Similar points have also been raised by Atenas, Havemann & Priego [4] who suggest that open datasets should be understood as a form of educational material that offers opportunities for educators to

empower students to engage critically and collaboratively. In a practical example, Wilkersin & Laina [73] have shown how working with open data provides middle school children with opportunities for reasoning that invoked their contextual knowledge of where data are generated when sharing stories about that data with each other. More theoretically, Coughlan [14] suggests a framework based on: (1) identifying shared goals between educators and open data actors; (2) creating shared artifacts and communications; and (3) developing open data literacies, to support educators' use of open data and facilitate these relationships. Future research should further investigate the role that education plays in the motivations of open data intermediaries and work to identify opportunities for synergistic activities.

5.2 An Understanding of Data Analytics and Civics are Complementary Needs

When we looked at what participants perceived to be barriers and facilitators to successfully using and becoming more deeply engaged with open data, we found that an understanding of and skills in data analytics were only one part of the story. While these are often discussed as a major barrier to successful open government data programs [74, 77], we found that an understanding of how local government works, and also a preexisting interest in civics and local issues of concern were also considered important by participants. It was not that the need for data analytics skills and training were not considered necessary for accessing and using open data, but rather that they were not considered sufficient. Our findings suggest that actionable engagement with open data typically requires an understanding of local issues and local government to help frame questions to ask of data. For example, as P12 explained to us, *"it's how one bit of data relates to another bit of data or to a policy or to a program that makes data useful as a tool"*. This often requires understanding of a bigger policy picture alongside more specific knowledge, such as the different terms that municipal departments and agencies use to refer to the same or similar entities that in turn become the labels describing individual data sets. Similar challenges are hinted at when Boehner & DiSalvo [7] say, "[d]ata could be open through more access but closed through lack of context", and by Asad and Le Dantec [1] when they argue that "[i]t is not enough to design a system for activists that reveals information, but it must also be assembled and contextualized". Our findings also indicate that actionable engagements with open data require an understanding of how to interact with government and municipal agencies in order to use the answers found. This finding is similar to the suggestion by Puussaar et al. [57] that "it takes social capital to transform this data into activities for local benefit". Place-based approaches, e.g. [45, 57] have been proposed as one way to address similar challenges by grounding open data engagements in contexts where concerns can be more narrowly focused, so that data can be concretized, and be more easily grasped and meaningful. However, map-based approaches are implemented in a number of open data portals (e.g. [52]), and were familiar to the participants in our study. We believe that this raises larger issues about the accessibility and effective use of open data, which have not been addressed by prior research, and which point towards a need for more holistic understandings of *'open data literacy'* (see Sections 5.2.2 and 5.4.2)

5.2.1 Connecting With an Open Data Community. The need to combine multiple different skills for successful engagements with open data aligns with our finding that connecting with a wider open data community can be a strong facilitator for continued engagement. A number of participants discussed how this community of interest, practice and action consistently helps to bring people with technical skills and experience together with people who have questions that might be answered through data analysis, and then with people who may have the advocacy skills and civic knowledge to use analytic results in support of community action. This is particularly important because the potential to connect people to personally relevant topics is often suggested as being a

key facilitator with regard to data literacy, e.g. [44, 60], which our research suggests is a barrier to meaningful engagement with open data. Our findings here seem to be in agreement with those of Boehner & DiSalvo [7] when they discuss the value of civic hackathons as being largely in the exchanges between participants.

5.2.2 Understanding Open Data Literacy. We believe that this also has important implications for what it might mean to be considered data literate in this context. In prior literature understandings of what it means to be data literate have understandably been based in acquiring core skills from math and statics, e.g. [70]. Where an understanding of 'real-world' issues, local concerns or civics have been invoked it has typically been with regard to providing a context or grounding for learners who may struggle with more abstract concepts [60, 75]. While we largely agree with these previous characterizations, our findings also suggest that when we think in particular about what we might call *open data literacy* such civic knowledge should also be considered a core aspect. Loukissas & Ntabathia [45] point towards similar concerns of open data in use when they suggest that, "something only becomes data when it is taken up as evidence in support of a claim". We also find connections to the work of Boehner & DiSalvo [7] who pose the related question, "what does civic literacy mean in a world saturated with public data, when the public is the data?". The intersection of these two concerns, the impact of data on our understanding of civic literacy and of civics on our understanding of data literacy, represents a key area for future research, and we believe that thinking about these relationships will help to broaden our understanding of data literacy, in a similar way to how it has previously been expanded to include art-based practices [6, 20, 21].

5.3 Open Data Intermediaries and Communities of Practice and Action

Our findings strongly suggest that connecting with a civic tech community of practice, and with communities of action that coalesce around local issues, are both strong facilitators for engagement with open data. These communities can offer local and issue specific knowledge, technical and scientific assistance, and knowledge and experience with civic or political advocacy. This finding highlights the way in which engagements with open data are often collaborative and how the community serves a social function beyond individual engagements. The downside of this can be seen when participants highlight how the use of language and poor communication, in particular in-group jargon, can be a barrier to engagement. Choi & Tausczic [11] characterize collaborations around open data as being small-scale, and highlight their use of simple analysis methods and visualization. They suggest that in this way there are strong connections to maker movements and open science. While we do not disagree with this interpretation, we believe that it misses the important role of civics, politics, and community advocacy over local issues in building these collaborations. We believe that when successfully convened, these communities of open data practice and action are an important factor in the infrastructures that Asad & Le Dantec [1] highlight as being crucial to supporting contexts of activism. Also that it is these communities that can often work out how to leverage the social capital that Johnson, Al-Shahrabi & Vines highlight as important to work in digital civics [36]. However, because our research is located in a single municipality further research should be undertaken to discover how these communities of practice and action manifest in relation to open data elsewhere.

5.4 Pathways to Success and Recommendations for Open Data Programs

We now consider steps towards developing pathways to successfully nurturing a community of open data intermediaries. To support these pathways we make five recommendations for organizations planning and managing open data programs.

5.4.1 Ease entry and scaffold progression. To support wider participation and engagement with municipal open data programs we suggest following Mitch Resnick's advice for learning technologies and creating entry points with "low floors, high ceilings, and wide walls" [58]. This is because our findings suggest that newcomers may initially feel overwhelmed by data and intimidated by terminology and jargon, and so the aim should be to support easy initial access (low floors) and support diverse perspectives and paths (wide walls). Following these principles, programs should then try to help scaffold the transition from newcomer to experienced and skilled open data intermediary who uses open data creatively (high ceilings), e.g. by creating a consistent framework for describing events that reflects their target audience and requirements for technical and civic know-how, and provides opportunities to develop and extend skills and knowledge. Loukissass & Ntabathia [45] offer additional advice about how the configuration of settings and convening of concerned publics impacts the ways in which people are able to approach open data, and point to the 'interdependence of data with their social and material settings'. While it is relatively straight forward to see how this understanding can be translated into the context of workshops, events, and trainings, it is less clear how these important considerations can be incorporated into online open data portals or community spaces. Because of this, further research is needed to better understand how, or indeed if, such a translation is possible.

5.4.2 Adopt a holistic view of 'open data literacy'. Our research strongly suggests that engagement with government open data programs requires an extended view of data literacy that incorporates elements of civic literacy, and implies a participation in what Boyte refers to as 'everyday politics' [8], i.e. the ability to address issues people face in their day-to-day lives. This should include skills in framing and asking questions, where possible answering them with reference to data, and then acting on those answers politically. As Boehner & DiSalvo [7] highlight, there is a complex interaction emerging between civics and public data that has implications for both, and which is not yet well understood. Supporting the development of open data literacy will require open data programs to consider support and training that moves beyond technical assistance towards opening municipal processes and practices up to questioning. This extended view of data literacy should draw further inspiration from art-based practices, e.g. as described in [6, 20, 21, 45]. As an additional benefit, this more holistic view may also help provide a framework for developing traditional data literacy and technical data science skills in a way that can overcome prior negative experiences, e.g. around school math, that were discussed by participants in our study.

5.4.3 Integrate teaching and training. Our interviews with students interning at the civic tech non-profit suggest that opportunities to teach the open data tools and skills they gained experience with were a key motivator and facilitator to continued engagement. To support this we recommend developing train-the-trainer and peer-teaching models that can capitalize on the skills, knowledge, and goodwill available within open data and civic tech communities. Such programs should aim to provide role models, as learning from role models was considered to strongly support learning and engagement, and create entry points into the open data and civic tech communities considered helpful in supporting extended engagement.

5.4.4 Support community-focused data views. One way to help organize open data resources so that they appear more accessible and less overwhelming to people is to increase the involvement of community representatives in the design of open data portals and decisions about the release and maintenance of particular data sets. In this way programs would hope to increase users' ability to relate open data to their everyday lives, which was considered an important motivator and facilitator for engagement. Including community representatives in these decisions and discussions can increase program relevancy and enable people to directly focus on data that reflect issues

they are interested in. While place-based approaches, such as [45, 57], offer one approach to thinking about community, community membership is not necessarily linked to geographically defined attributes suitable for mapping. They can also be a complex mix of many other factors, e.g. interests, abilities or demographics. Similarly, issues of concern may cross boundaries of scale and granularities of data representation (e.g. zip code, census track, county, city, State, etc.), particularly in large cities with multiple layers of local government. HCI, CSCW, and participatory design (PD) research into how different communities can be included in the design of sociotechnical systems for civic engagement is an ongoing and active area of inquiry, e.g. [3, 26, 41, 43], which future studies of design for community-focused open data can both inform and be informed by.

5.4.5 Support issue-focused data views. To help address the challenges that participants face with regard to framing questions and aligning data that intersect with these questions but come from different sources, we recommend developing tools that support issue-focused views of open data. While there examples of open data portals that currently frame queries as being issue-focused, e.g. [5], they typically use the issue-focused approach as an entry point to individual data sets. Our findings suggest that more is needed, and that an important next step is to be able to link data from different sources using issues of concern as a common link. This is likely to require open data program administrators to adopt and enforce consistent meta data dictionaries so that entities are named consistently, and to think about open data in ways that cut across municipal silos. Similarly to developing community-focused views into data this will be facilitated by including the people who create, curate, manage, administer, and use open data in processes for deciding how they are released, organized, and accessed. and in designing tools and platforms to support their effective use.

5.5 Limitations

There are two main limitations to this study that should be acknowledged. First the work reported here involved engaging with a limited number of participants, and its focus was on the open data ecosystem in a single US city. This can have significant implications in terms of the study's generalizability. Municipalities and the open data ecosystems they support differ in a myriad of ways and have developed at very different paces, from the engagement of local government and agencies through to the pool of civic tech, data analytics and community advocates engaged in the process. Future HCI and CSCW research should investigate these issues across different contexts and in different locations. Second, while this study focused on people who have been engaged with open data, the research would benefit from the perspective of community members who have not. In particular this would help us better understand the barriers people might face and the potential facilitators that could help in their mitigation.

6 CONCLUSION

In this paper we report on the results of semi-structured interviews with 17 people whose engagements with open data fulfill different roles associated with acting as or becoming open data intermediaries. These include: students learning how to train others in open data techniques and tools; people who attend and host open data events and use open data for commercial or social benefit; and representatives from local government, municipal agencies, and a civic tech non-profit. Our analysis identifies their motivations for initial and ongoing engagements with open data, and the barriers and facilitators they perceive with regard to that engagement. Based on this data, and building on prior research identifying the important role that intermediaries can play in developing and supporting open data programs, we identify pathways to successfully developing and nurturing civic tech and open data communities, and make five recommendations for organizations planning

and managing open data programs. These include: easing initial entry and scaffolding progression to expertise; adopting a holistic view of 'open data literacy' that includes civics; integrating teaching and learning in a train-the-trainer model; and supporting community-focused and issue-focused views into data. Better understanding of and designing for each of these will help support open data programs in realizing the promises they make with regard to improved government practice and civic engagement.

ACKNOWLEDGMENTS

REFERENCES

[1] Mariam Asad and Christopher A Le Dantec. 2015. Illegitimate civic participation: supporting community activists on the ground. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1694–1703.

[2] Mariam Asad and Christopher A Le Dantec. 2017. Tap the" make this public" button: A design-based inquiry into issue advocacy and digital civics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6304–6316.

[3] Mariam Asad, Christopher A Le Dantec, Becky Nielsen, and Kate Diedrick. 2017. Creating a sociotechnical API: Designing city-scale community engagement. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2295–2306.

[4] Javiera Atenas, Leo Havemann, and Ernesto Priego. 2015. Open data as open educational resources: Towards transversal skills and global citizenship. *Open praxis* 7, 4 (2015), 377–389.

[5] Greater London Authority. [n.d.]. London Data Store. Retrieved May 21, 2021 from https://data.london.gov.uk

[6] Rahul Bhargava, Ricardo Kadouaki, Emily Bhargava, Guilherme Castro, and Catherine D'Ignazio. 2016. Data murals: Using the arts to build data literacy. *The Journal of Community Informatics* 12, 3 (2016).

[7] Kirsten Boehner and Carl DiSalvo. 2016. Data, design and civics: An exploratory study of civic tech. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2970–2981.

[8] Harry C Boyte. 2010. *Everyday politics*. University of Pennsylvania Press.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.

[11] Joohee Choi and Yla Tausczik. 2017. Characteristics of collaboration in the emerging practice of open data analysis. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 835–846.

[12] Eric Corbett and Christopher A Le Dantec. 2018. Going the distance: Trust work for citizen participation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[13] Eric Corbett and Christopher A Le Dantec. 2018. The problem of community engagement: Disentangling the practices of municipal government. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[14] Tim Coughlan. 2020. The use of open data as a material for learning. *Educational Technology Research and Development* 68, 1 (2020), 383–411.

[15] John W Creswell and Dana L Miller. 2000. Determining validity in qualitative inquiry. *Theory into practice* 39, 3 (2000), 124–130.

[16] Gisele da Silva Craveiro and Claudio Albano. 2017. Open data intermediaries: coproduction in budget transparency. *Transforming Government: People, Process and Policy* (2017).

[17] Sharon S Dawes, Lyudmila Vidiasova, and Olga Parkhimovich. 2016. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly* 33, 1 (2016), 15–27.

[18] Maryland Department of Information Technology. [n.d.]. Maryland's Open Data Portal. Retrieved May 21, 2021 from https://opendata.maryland.gov

[19] Jessa Dickinson, Mark Díaz, Christopher A Le Dantec, and Sheena Erete. 2019. "The cavalry ain't coming in to save us" Supporting Capacities and Relationships through Civic Tech. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[20] Catherine D'Ignazio. 2017. Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal* 23, 1 (2017), 6–18.

[21] Catherine D'Ignazio and Rahul Bhargava. 2016. DataBasic: Design principles, tools and activities for data literacy learners. *The Journal of Community Informatics* 12, 3 (2016).

[22] Andy Dow, Rob Comber, and John Vines. 2018. Between grassroots and the hierarchy: Lessons learned from the design of a public services directory. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[23] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. 2016. 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*. 1273–1283.

[24] Mark Frank and Phil Waddell. 2014. Solving the Democratic Deficit: The Role of Open Data and Intermediaries. In *Conference for E-Democracy and Open Governement*. 487.

[25] Elisabeth Gebka, Jonathan Crusoe, and Karin Ahlin. 2021. Methods to Capture User Information Needs: Design Principles for Open Data Intermediaries and Data Providers: Design Principles for Open Data Intermediaries and Data Providers. *Data Science Journal* 20, 37 (2021), 1–19.

[26] Daniel Gooch, Matthew Barker, Lorraine Hudson, Ryan Kelly, Gerd Kortuem, Janet Van Der Linden, Marian Petre, Rebecca Brown, Anna Klis-Davies, Hannah Forbes, et al. 2018. Amplifying quiet voices: Challenges and opportunities for participatory design at an urban scale. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 1 (2018), 1–34.

[27] UK Government Digital Service. [n.d.]. Find open data. Retrieved May 21, 2021 from http://data.gov.uk/

[28] California Government Operations Agency. [n.d.]. California Open Data Portal. Retrieved May 21, 2021 from https://data.ca.gov

[29] Joel Gurin. 2014. Open Governments, Open Data. *The SAIS Review of International Affairs* 34, 1 (2014), 71–82.

[30] Michael B Gurstein. 2011. Open data: Empowering the empowered or effective data use for everyone? *First Monday* (2011).

[31] Teresa M Harrison, Theresa A Pardo, and Meghan Cook. 2012. Creating open government ecosystems: A research and development agenda. *Future Internet* 4, 4 (2012), 900–928.

[32] Julien Hivon and Ryad Titah. 2017. Conceptualizing citizen participation in open data use at the city level. *Transforming Government: People, Process and Policy* (2017).

[33] The White House. [n.d.]. Open Government Initiative. Retrieved April 25, 2021 from https://obamawhitehouse.archives.gov/open

[34] Hannu Jaakkola, Timo Mäkinen, and Anna Eteläaho. 2014. Open data: opportunities and challenges. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*. 25–39.

[35] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. Benefits, adoption barriers and myths of open data and open government. *Information systems management* 29, 4 (2012), 258–268.

[36] Ian G Johnson, Dalya Al-Shahrabi, and John Vines. 2020. From Creating Spaces for Civic Discourse to Creating Resources for Action. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[37] Ian G Johnson, John Vines, Nick Taylor, Edward Jenkins, and Justin Marshall. 2016. Reflections on deploying distributed consultation technologies with community organisations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2945–2957.

[38] Peter A Johnson. 2016. Reflecting on the success of open data: How municipal government evaluates their open data programs. *International Journal of E-Planning Research (IJEPR)* 5, 3 (2016), 1–12.

[39] Peter A Johnson and Sarah Greene. 2017. Who are government opendata infomediaries? A preliminary scan and classification of open data users and products. (2017).

[40] Nam Wook Kim, Jonghyuk Jung, Eun-Young Ko, Songyi Han, Chang Won Lee, Juho Kim, and Jihee Kim. 2016. Budgetmap: Engaging taxpayers in the issue-driven classification of a government budget. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1028–1039.

[41] Christopher Le Dantec. 2012. Participation and publics: supporting community engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1351–1360.

[42] Christopher A Le Dantec, Mariam Asad, Aditi Misra, and Kari E Watkins. 2015. Planning with crowdsourced data: rhetoric and representation in transportation planning. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1717–1727.

[43] Christopher A Le Dantec and Sarah Fox. 2015. Strangers at the gate: Gaining access, building rapport, and co-constructing community-based research. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1348–1358.

[44] Victor R Lee, Joel R Drake, and Jeffrey L Thayne. 2016. Appropriating quantified self technologies to support elementary statistical teaching and learning. *IEEE Transactions on Learning Technologies* 9, 4 (2016), 354–365.

[45] Yanni Alexander Loukissas and Jude Mwenda Ntabathia. 2021. Open Data Settings: A Conceptual Framework Explored Through the Map Room Project. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–24.

[46] Gustavo Magalhaes, Catarina Roseira, and Sharon Strover. 2013. Open government data intermediaries: A terminology framework. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*. 330–333.

[47] Moira Maguire and Brid Delahunt. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education* 9, 3 (2017).

[48] Matthew Marshall, David S Kirk, and John Vines. 2016. Accountable: Exploring the inadequacies of transparent financial practice in the non-profit sector. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1620–1631.

[49] Amanda Meng, Carl DiSalvo, and Ellen Zegura. 2019. Collaborative data work towards a caring democracy. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[50] Ministry of Electronics National Informatics Centre (NIC) and Government of India Information Technology. [n.d.]. Open Government Data (OGD) Platform India. Retrieved May 21, 2021 from http://data.gov.in/

[51] Office of Management and Budget. [n.d.]. Federal Data Stretegy: Overview. Retrieved May 21, 2021 from https://strategy.data.gov/overview/

[52] City of New York. [n.d.]. NYC OpenData. Retrieved May 21, 2021 from https://opendata.cityofnewyork.us

[53] City of New York. [n.d.]. Open Data For All. Retrieved May 21, 2021 from https://www1.nyc.gov/assets/home/downloads/pdf/reports/2015/NYC-Open-Data-Plan-2015.pdf

[54] Patrick Olivier and Peter Wright. 2015. Digital civics: taking a local turn. *interactions* 22, 4 (2015), 61–63.

[55] Alon Peled. 2011. When transparency and collaboration collide: The USA open data program. *Journal of the American society for information science and technology* 62, 11 (2011), 2085–2094.

[56] Oleg Petrov, Joel Gurin, and Laura Manley. 2016. Open data for sustainable development. (2016).

[57] Aare Puussaar, Ian G Johnson, Kyle Montague, Philip James, and Peter Wright. 2018. Making open data work for civic advocacy. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

[58] Mitchel Resnick and Ken Robinson. 2017. *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play.* MIT press.

[59] David Robinson, Harlan Yu, William P Zeller, and Edward W Felten. 2008. Government data and the invisible hand. *Yale JL & Tech.* 11 (2008), 159.

[60] Laurie H Rubel, Maren Hall-Wieckert, and Vivian Y Lim. 2016. Teaching mathematics for spatial justice: Beyond a victory narrative. *Harvard Educational Review* 86, 4 (2016), 556–579.

[61] Erna Ruijer, Stephan Grimmelikhuijsen, Jochem Van Den Berg, and Albert Meijer. 2020. Open data work: understanding open data usage from a practice lens. *International Review of Administrative Sciences* 86, 1 (2020), 3–19.

[62] Jorge Saldivar, Cristhian Parra, Marcelo Alcaraz, Rebeca Arteta, and Luca Cernuzzi. 2019. Civic technology for social innovation. *Computer Supported Cooperative Work (CSCW)* 28, 1 (2019), 169–207.

[63] David S Sawicki and William J Craig. 1996. The democratization of data: Bridging the gap for community groups. *Journal of the American Planning Association* 62, 4 (1996), 512–523.

[64] Andrew Schrock and Gwen Shaffer. 2017. Data ideologies of an interested public: A study of grassroots open government data intermediaries. *Big Data & Society* 4, 1 (2017), 2053951717690750.

[65] Irving Seidman. 2006. *Interviewing as qualitative research: A guide for researchers in education and the social sciences.* Teachers college press.

[66] Renee E Sieber and Peter A Johnson. 2015. Civic open data at a crossroads: Dominant models and current challenges. *Government information quarterly* 32, 3 (2015), 308–315.

[67] Olga Slobodova and Sören Becker. 2020. Zooming Into the Ecosystem: Agency and Politics Around Open Data Platforms in Lyon and Berlin. *Frontiers in Sustainable Cities* 2 (2020), 20.

[68] US General Services Administration Technology Transformation Service. [n.d.]. DATA.GOV: The home of the U.S. Government's open data. Retrieved May 21, 2021 from http://data.gov/

[69] Barbara Ubaldi. 2013. Open government data: Towards empirical analysis of open government data initiatives. (2013).

[70] Phil Vahey, Louise Yarnall, Charles Patton, Daniel Zalles, and Karen Swan. 2006. Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In *Annual Meeting of the American Educational Research Association, San Francisco, CA*.

[71] Francois Van Schalkwyk, Michelle Willmers, and Maurice McNaughton. 2016. Viscous open data: The roles of intermediaries in an open data ecosystem. *Information Technology for Development* 22, sup1 (2016), 68–83.

[72] Sebastian Weise, Paul Coulton, and Mike Chiasson. 2017. Designing in between local government and the public–Using institutional analysis in interventions on civic infrastructures. *Computer Supported Cooperative Work (CSCW)* 26, 4 (2017), 927–958.

[73] Michelle Hoda Wilkerson and Vasiliki Laina. 2018. Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM* 50, 7 (2018), 1223–1235.

[74] Bev Wilson and Cong Cong. 2021. Beyond the supply side: Use and impact of municipal open data in the US. *Telematics and Informatics* 58 (2021), 101526.

[75] Annika Wolff, Daniel Gooch, Jose J Cavero Montaner, Umar Rashid, and Gerd Kortuem. 2016. Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics* 12, 3 (2016).

[76] Harlan Yu and David G Robinson. 2011. The new ambiguity of open government. *UCLA L. Rev. Discourse* 59 (2011), 178.

[77] Anneke Zuiderwijk and Mark de Reuver. 2021. Why open government data initiatives fail to achieve their objectives: categorizing and prioritizing barriers through a global survey. *Transforming Government: People, Process and Policy* (2021).

[78] Anneke Zuiderwijk and Marijn Janssen. 2014. Open data policies, their implementation and impact: A framework for comparison. *Government information quarterly* 31, 1 (2014), 17–29.

[79] Anneke Zuiderwijk, Rhythima Shinde, and Marijn Janssen. 2019. Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results? *International Review of Administrative Sciences* 85, 4 (2019), 645–672.